



Università
di Catania



FSE FONDO SOCIALE EUROPEO
SICILIA 2020
PROGRAMMA OPERATIVO



REPUBBLICA ITALIANA



Unione Europea



Fondo Sociale Europeo

Università degli studi di Catania
Dipartimento di Ingegneria Civile e Architettura – DICAR

Dottorato di Ricerca XXXV Ciclo
“Valutazione e mitigazione dei rischi urbani e territoriali”

**Derivazione di soglie di innesco idro-
meteorologiche basate su algoritmi di machine
learning**

A.A. 2021-2022

Tesi di Dottorato

Candidato

Pierpaolo Distefano

Gruppo Tesi

Prof. Ing. Pietro Scandura - Tutor

Prof. Ing. Antonino Cancelliere - Co-tutor

Prof. Ing. David J. Peres

Coordinatore del corso di Dottorato

Prof. Ing. Antonino Cancelliere

*A Emme, che mi è stata vicina
e ha provato a capirmi
nel momento più buio.*

INDICE

INDICE DELLE FIGURE		III
INDICE DELLE TABELLE.....		VIII
Abstract.....		1
CAPITOLO 1	Introduzione	3
1.1	PREMESSE	3
1.2	OBIETTIVI DELLA RICERCA	5
1.3	METODOLOGIA DELLA RICERCA	6
1.4	STRUTTURA DELL'ELABORATO	7
CAPITOLO 2	Frane e soglie pluviometriche di innesco.....	9
2.1	GENERALITÀ	9
2.2	TIPOLOGIE DI FENOMENI FRANOSI.....	9
2.3	SOGLIE PLUVIOMETRICHE DI INNESCO	11
2.4	SOGLIE IDRO-METEOROLOGICHE.....	16
CAPITOLO 3	Reti neurali artificiali	21
3.1	GENERALITÀ	21
3.2	LE RETI NEURALI ARTIFICIALI	21
3.3	CARATTERISTICHE DELLE RETI NEURALI ARTIFICIALI	23
3.4	STRUTTURA DELLE RETI NEURALI ARTIFICIALI	23
3.5	ALGORITMI DI TRAINING	27
3.6	VALUTAZIONE DELLE PERFORMANCE DELLE RETI NEURALI	29

3.7	APPLICAZIONI DELLE RETI NEURALI NEL CAMPO DELLE FRANE.....	31
CAPITOLO 4	Descrizione dei casi di studio e costruzione dei dataset.....	37
4.1	GENERALITÀ.....	37
4.2	CASO STUDIO IN ITALIA.....	37
4.3	CASO STUDIO IN NORVEGIA.....	39
4.4	DATI DI UMIDITÀ DEL SUOLO DERIVANTI DAL PROGETTO DI RIANALISI ERA5-LAND.....	41
4.5	CONSTRUZIONE DEI DATASET.....	45
CAPITOLO 5	Soglie pluviometriche di innesco basate su reti neurali artificiali.....	51
5.1	GENERALITÀ.....	51
5.2	INTRODUZIONE.....	51
5.3	DATI E SOGLIE EMPIRICHE PER LA REGIONE SICILIA.....	53
5.4	APPLICAZIONE DELLE RETI NEURALI ARTIFICIALI.....	56
5.5	SOGLIE PLUVIOMETRICHE DETERMINATE PER LA NORVEGIA.....	61
5.6	CLASSIFICATORE CHE CONSIDERA SOLO DATI DI PRECIPITAZIONE.....	63
5.7	LIMITI DELL'APPROCCIO IMPIEGATO E SVILUPPI FUTURI.....	67
CAPITOLO 6	Soglie idro-meteorologiche multivariate basate su reti neurali artificiali.....	69
6.1	GENERALITÀ.....	69
6.2	DATI DI UMIDITÀ.....	69
6.3	DERIVAZIONE DELLE SOGLIE IDRO-METEOROLOGICHE.....	70

6.4	SOGLIE IDRO-METEOROLOGICHE CON DATI DI UMIDITÀ RELATIVI AD UN UNICO STRATO.....	72
6.5	SOGLIE IDRO-METEOROLOGICHE CON DATI DI UMIDITÀ RELATIVI A PIÙ STRATI.....	82
6.6	DISCUSSIONE DEI RISULTATI	94
CAPITOLO 7	Conclusioni	97
7.1	CONCLUSIONI	97
APPENDICE A	Confronti umidità del suolo.....	101
A.1	CONFRONTO DATI DI UMIDITÀ DEL SUOLO PER CHIARAVALLE CENTRALE.....	101
A.2	CONFRONTO DATI DI UMIDITÀ DEL SUOLO PER FITTERIZZI	104
A.3	CONFRONTO DATI DI UMIDITÀ DEL SUOLO PER MONGRASSANO	107
A.4	CONFRONTO DATI DI UMIDITÀ DEL SUOLO PER SATRIANO	110
A.5	CONFRONTO DATI DI UMIDITÀ DEL SUOLO PER TORANO ..	113
A.6	CONFRONTO DATI DI UMIDITÀ DEL SUOLO PER EIDSVOLL	116
A.7	CONFRONTO DATI DI UMIDITÀ DEL SUOLO HOBØL	119
APPENDICE B	Codici MATLAB	125
	Bibliografia.....	151
	Ringraziamenti.....	161
 INDICE DELLE FIGURE		
	Figura 1.1: Diagramma a blocchi della metodologia proposta.	7
	Figura 2.1: Tipologia di frane (Highland and Bobrowsky, 2008).	10

Figura 2.2: Rappresentazione di una soglia pluviometrica derivata da dati di pioggia osservati in concomitanza con l'innescamento di frane (Peres & Cancelliere, 2021).....	13
Figura 2.3: Distribuzione spaziale degli studi che propongono soglie pluviometriche di innescamento (Segoni et al., 2018)	14
Figura 3.1: Algoritmi di machine learning (Batta, 2020)	21
Figura 3.2: Architettura delle reti neurali: (da sinistra a destra) una rete neurale feedforward a singolo strato totalmente connessa, una rete neurale con uno strato nascosto totalmente connessa e una di rete neurale di tipo ricorsivo (Haykin, 1999).	24
Figura 3.3: Rete feedforward (Zhang et al., 1998).	25
Figura 3.4: Procedura di early stopping (Haykin, 1999).....	30
Figura 3.5: Rete neurale utilizzata da Conforti et al. (2014)	34
Figura 4.1: Inquadramento geografico e DEM della Sicilia e localizzazione dei pluviometri utilizzati in Sicilia	38
Figura 4.2: Disposizione dei pluviometri fittizi nell'area di studio norvegese	40
Figura 4.3: Schematizzazione della generazione dei dati di rianalisi di ERA5-Land (modello ECMWF)	42
Figura 4.4: Diagramma a blocchi del software CTRL-T (Melillo et al., 2018)	47
Figura 4.5: Ubicazione delle frane ricostruite dall'algoritmo CTRL-T per la Sicilia	49
Figura 4.6: Ubicazione delle frane ricostruite dall'algoritmo CTRL-T per i bacini Norvegesi	49
Figura 5.1: Metodologia proposta	54
Figura 5.2: Confronto tra le soglie pluviometriche per la Sicilia nel piano cartesiano	55
Figura 5.3: Confronto tra le soglie pluviometriche per la Sicilia nel piano bi-logaritmico	55
Figura 5.4: Schema concettuale della rete neurale artificiale.	56
Figura 5.5: Confronto tra la soglia che considera eventi innescanti e non innescanti (viola) e quelle che considerano solo eventi innescanti nel piano cartesiano	60

Figura 5.6: Confronto tra la soglia che considera eventi innescanti e non innescanti (viola) e quelle che considerano solo eventi innescanti nel piano bi-logaritmico	60
Figura 5.7: Soglia derivata da CTRL-T per la Norvegia	62
Figura 5.8: Soglia cumulata-durata che tiene conto dei dati innescanti e non innescanti (NO)	62
Figura 5.9: Confronto tra soglie pluviometriche derivate per la Norvegia.	63
Figura 5.10: Boxplot del classificatore che considera solo le caratteristiche della precipitazione (IT)	65
Figura 5.11: Boxplot del classificatore che considera solo le caratteristiche della precipitazione (NO)	66
Figura 6.1: Boxplot del classificatore basato sulle variabili D-E-S per singoli strati (IT)	72
Figura 6.2: Boxplot del classificatore che considera come variabili D-E-S per singoli strati (NO)	74
Figura 6.3: Boxplot del classificatore che considera come dati di input D-I-S singoli strati (IT)	76
Figura 6.4: Boxplot del classificatore che considera come dati di input i valori D-I-S per singoli strati (NO)	78
Figura 6.5: Boxplot del classificatore che considera come dati di input i valori di D-E-I-S per singoli strati (IT)	79
Figura 6.6: Boxplot del classificatore che considera i valori di D-E-I-S per singoli strati (NO)	81
Figura 6.7 Boxplot del classificatore che considera come variabili di input i valori di D-E-S relativi a più strati (IT)	82
Figura 6.8: Boxplot del classificatore che considera i valori di D-E-S relativi a più strati (NO)	84
Figura 6.9: Boxplot del classificatore che considera i dati D-I-S relativi a più strati (IT)	86
Figura 6.10: Boxplot del classificatore che considera i valori di D-I-S relativi a più strati (NO)	88
Figura 6.11: Boxplot del classificatore che considera i valori di D-E-I-S relativi a più strati (IT)	90

Figura 6.12: Boxplot del classificatore che considera valori di D-E-I-S relativi a più strati (NO)	92
Figura A.1: Boxplot dati di umidità del suolo di Chiaravalle Centrale.....	101
Figura A.2: Confronto umidità del suolo tra secondo strato di ERA5-Land e i dati della ISMN alla profondità di 30cm per la stazione di Chiaravalle Centrale	102
Figura A.3: Confronto umidità del suolo tra il terzo strato di ERA5-Land e i dati della ISMN alla profondità di 60cm per la stazione di Chiaravalle Centrale	102
Figura A.4: Confronto umidità del suolo tra il terzo strato di ERA5-Land e i dati della ISMN profondità di 90cm per la stazione di Chiaravalle Centrale	103
Figura A.5: Diagramma di Taylor per la stazione di Chiaravalle Centrale.....	103
Figura A.6: Boxplot dati di umidità del suolo di Fitterizzi	104
Figura A.7: Confronto umidità del suolo tra secondo strato di ERA5-Land e i dati della ISMN alla profondità di 30cm per la stazione di Fitterizzi.	104
Figura A.8: Confronto umidità del suolo tra il terzo strato di ERA5-Land e i dati della ISMN alla profondità di 60cm per la stazione di Fitterizzi	105
Figura A.9: Confronto umidità del suolo tra il terzo strato di ERA5-Land e i dati della ISMN profondità di 90cm per la stazione di Fitterizzi.....	105
Figura A.10: Diagramma di Taylor per la stazione di Fitterizzi	106
Figura A.11: Boxplot dati di umidità del suolo di Mongrassano	107
Figura A.12: Confronto umidità del suolo tra secondo strato di ERA5-Land e i dati della ISMN alla profondità di 30cm per la stazione di Mongrassano	107
Figura A.13: Confronto umidità del suolo tra il terzo strato di ERA5-Land e i dati della ISMN alla profondità di 60cm per la stazione di Mongrassano	108
Figura A.14: Confronto umidità del suolo tra il terzo strato di ERA5-Land e i dati della ISMN profondità di 90cm per la stazione di Mongrassano	108
Figura A. 15: Diagramma di Taylor per la stazione di Mongrassano	109
Figura A.16: Boxplot dati di umidità del suolo di Satriano.....	110
Figura A. 17: Confronto umidità del suolo tra secondo strato di ERA5-Land e i dati della ISMN alla profondità di 30cm per la stazione di Satriano ..	110

Figura A.18: Confronto umidità del suolo tra il terzo strato di ERA5-Land e i dati della ISMN alla profondità di 60cm per la stazione di Satriano .	111
Figura A.19: Confronto umidità del suolo tra il terzo strato di ERA5-Land e i dati della ISMN profondità di 90cm per la stazione di Satriano	111
Figura A.20: Diagramma di Taylor per la stazione di Satriano.....	112
Figura A.21: Boxplot dati di umidità del suolo di Torano	113
Figura A.22: Confronto umidità del suolo tra secondo strato di ERA5-Land e i dati della ISMN alla profondità di 30cm per la stazione di Torano....	113
Figura A. 23: Confronto umidità del suolo tra il terzo strato di ERA5-Land e i dati della ISMN alla profondità di 60cm per la stazione di Torano....	114
Figura A. 24: Confronto umidità del suolo tra il terzo strato di ERA5-Land e i dati della ISMN profondità di 90cm per la stazione di Torano	114
Figura A.25: Diagramma di Taylor per la stazione di Torano	115
Figura A.26: Boxplot dati di umidità del suolo di Eidsvoll.....	116
Figura A. 27: Confronto umidità del suolo tra il secondo strato di ERA5-Land e i dati della stazione di Eidsvoll alla profondità di 10cm.....	116
Figura A.28: Confronto umidità del suolo tra il terzo strato di ERA5-Land e i dati della stazione di Eidsvoll alla profondità di 50 e 100cm	117
Figura A. 29: Confronto umidità del suolo tra il quarto strato di ERA5-Land e i dati della stazione di Eidsvoll alla profondità di 200cm	117
Figura A 30: Boxplot dati di umidità del suolo di Hobøl (zona erbosa).....	119
Figura A.31: Confronto umidità del suolo tra il secondo strato di ERA5e i dati della stazione di Hobøl (zona erbosa) alla profondità di 10,20 e 30 cm	120
Figura A.32: Confronto umidità del suolo tra il terzo strato di ERA5 e i dati della stazione di Hobøl (zona erbosa) alla profondità di 30,40 e 50 cm	120
Figura A.33: Boxplot dati di umidità del suolo di Hobøl (arbusti)	121
Figura A.34: Confronto umidità del suolo tra il secondo strato di ERA5e i dati della stazione di Hobøl (arbusti) alla profondità di 10,20 e 30 cm	122
Figura A.35: Confronto umidità del suolo tra il terzo strato di ERA5 e i dati della stazione di Hobøl (arbusti) alla profondità di 30,40 e 50 cm	122
Figura A.36: Boxplot dati di umidità del suolo di Hobøl (alberi).....	123

Figura A.37: Confronto umidità del suolo tra il secondo strato di ERA5e i dati della stazione di Hobøl (alberi) alla profondità di 30 cm.....124

Figura A.38: Confronto umidità del suolo tra il terzo strato di ERA5 e i dati della stazione di Hobøl (alberi) alla profondità di 30,40, 50 e 80cm ...124

INDICE DELLE TABELLE

Tabella 3.1:Funzioni di trasferimento 26

Tabella 3.2:Indicatori performance delle reti neurali (Zhang et al, 1998)..... 29

Tabella 4.1: Caratteristiche del database ERA5-Land 43

Tabella 4.2:Grandezze del database ERA5-Land utilizzate per le elaborazioni..... 44

Tabella 4.3:Valori utilizzati nell'algoritmo CTRL-T per il caso studio italiano..... 46

Tabella 4.4: Valori utilizzati nell'algoritmo CTRL-T per il caso studio norvegese... 48

Tabella 5.1: Matrice di confusione 57

Tabella 5.2:Risultati dei test con reti neurali, che mostrano il numero ottimale di neuroni nascosti (il cui numero è variabile da 5 a 20), il TSS per l'intero campione, i campioni di allenamento, validazione e test, i valori di TPR e FPR 58

Tabella 5.3: Risultati dei classificatori che considerano solo le caratteristiche della precipitazione (IT)..... 65

Tabella 5.4: Risultati dei classificatori che considerano solo le caratteristiche della precipitazione (NO)..... 66

Tabella 6.1:Risultati del classificatore che utilizza come variabili D-E-S considerando singoli strati (IT)..... 73

Tabella 6.2: Risultati del classificatore che considera come input i dati di D-E- S singoli strati (NO) 75

Tabella 6.3: Risultati del classificatore che considera come input i dati di D-I-S singoli strati (IT)..... 77

Tabella 6.4: Risultati del classificatore che considera come input i dati D-I-S singoli strati (NO) 77

Tabella 6.5: Risultati del classificatore che considera i valori di D-E-I-S per singoli strati (IT)..... 80

Tabella 6.6: Risultati del classificatore che considera i dati D-E-I-S per singoli strati (NO)..... 80

Tabella 6.7: Risultati del classificatore basato sulle variabili di D-E-S relativi a più strati (IT)..... 83

Tabella 6.8: Risultati del classificatore basato sulle variabili D-E-S relativi a più strati (NO)..... 85

Tabella 6.9: Risultati del classificatore che considera i dati D-I-S relativi a più strati (IT) 87

Tabella 6.10: Risultati del classificatore che considera i valori di D-I-S relativi a più strati (NO)..... 89

Tabella 6.11: Risultati del classificatore che considera i valori di D-E-I-S relativi a più strati (IT) 91

Tabella 6.12: Risultati del classificatore che considera i valori di D-E-I-S relativi a più strati (NO)..... 93

ABSTRACT

Il recente report delle Nazioni Unite sui disastri naturali presentato nel 2020 ha evidenziato come le calamità osservate considerando due finestre temporali, 1980/1999 e 2000/2019, siano globalmente aumentate. Analogamente sono aumentate le popolazioni colpite, le vittime e i fondi da destinare per fronteggiare queste emergenze. Tra le calamità prese in considerazione sono considerate anche le frane, che aumentano da 254 a 376.

Gli impatti delle frane possono essere devastanti quando l'innesco è prossimo ai centri abitati e spesso, sfortunatamente, oltre ai danni a beni materiali e infrastrutture di vario genere si riscontrano anche vittime tra la popolazione.

Fra le principali misure di mitigazione di tipo non strutturale del rischio di frana, vi sono i sistemi di preallerta. Fra le componenti fondamentali di questi sistemi vi sono le soglie di innesco. Nel presente lavoro di tesi ci si è posti l'obiettivo di migliorare la comprensione dei fenomeni di innesco, proponendo approcci innovativi per superare alcuni limiti legati alla determinazione delle soglie pluviometriche di innesco. Per fare ciò si propone di: (1) superare l'utilizzo di relazioni predeterminate per la soglia, in genere di tipo potenza, (2) analizzare la possibilità dell'utilizzo di variabili non legate alla precipitazione.

Per raggiungere il primo obiettivo si propone un approccio basato sull'impiego di reti neurali artificiali, uno degli algoritmi di intelligenza artificiale già utilizzato in numerosi campi, per la capacità di apprendere dai dati, di generalizzazione e di riconoscimento di pattern. Utilizzando questo approccio si vuole verificare se è possibile mettere a punto un classificatore, che in funzione dei dati in input possa fornire una relazione di innesco, senza effettuare nessuna ipotesi a priori sulla forma dell'equazione della soglia.

La metodologia prevede, in un primo momento, di determinare per due aree di studio, una italiana e una norvegese (dove è stato svolto il periodo di ricerca all'estero), le soglie pluviometriche cumulata-durata. Per entrambe le soglie empiriche saranno valutate le performance tramite indicatori statistici di performance. In un secondo momento, si utilizzeranno le capacità di pattern recognition delle reti neurali artificiali per riconoscere, partendo dal medesimo dataset utilizzato per determinare le soglie cumulata-durata, le precipitazioni che

possono innescare movimenti franosi. I risultati ottenuti sembrano indicare che a parità di informazioni di input le performance di un classificatore basato su reti neurali risultano migliori.

Per quanto riguarda il secondo limite, cioè l'uso di informazioni legate alle caratteristiche della precipitazione (durata della precipitazione, cumulata e/o intensità di pioggia), si costruirà un classificatore neurale che permetterà di valutare le performance di nuove soglie a n variabili. Tra le variabili utilizzate si terrà conto anche dei dati di umidità del suolo (derivati dai dati di Rianalisi ERA5-Land e disponibili per entrambe le aree di studio), valutando quale tra i dati di input permetta di ottenere la migliori performance in termini di riconoscimento delle precipitazioni innescanti. I risultati ottenuti per quanto riguarda le soglie idrometeorologiche mostrano che introdurre dati di umidità del suolo permette di migliorare in modo sostanziale le performance del classificatore, sebbene all'aumentare dei parametri di input in concomitanza a un miglioramento delle performance si ottiene una maggiore variabilità legata all'inizializzazione delle reti neurali.

Il sistema di riconoscimento di eventi innescanti e non innescanti allenato e tarato sui dati di un'area di studio potrebbe costituire la parte di un sistema di preallerta, responsabile del riconoscimento delle condizioni di innesco. Tale componente poi deve essere integrata all'interno del sistema che permette la comunicazione con l'esterno. La procedura messa a punto, ben si presta ad essere automatizzata fermo restando il controllo di un operatore nelle fasi critiche.

Tra gli svantaggi di questo approccio, deve essere menzionata la mancanza di una equazione operativa che permetta di valutare in tempo reale l'evoluzione dei fenomeni meteorologici. La soluzione potrebbe essere quella di fornire agli enti proposti all'allertamento un software, che permetta un grado di libertà tale da poter far coesistere le previsioni degli algoritmi di machine learning con le decisioni prese sull'esperienza di un operatore.

CAPITOLO 1 INTRODUZIONE

1.1 PREMESSE

I fenomeni di dissesto idrogeologici costituiscono una problematica a rischio molto elevato che in Italia ha provocato ingenti danni e numerose vittime nel recente passato. Ogni anno, infatti, la penisola è colpita da eventi meteorologici intensi con gravi effetti al suolo, riportando costantemente all'attualità il problema. Il recente rapporto sul dissesto idrogeologico in Italia (Trigila et al., 2021) riporta un quadro aggiornato sulla pericolosità da frane e alluvioni. Il rapporto rivela come 7423 comuni (93.9% del totale) risulta essere a rischio frane, alluvioni e/o erosione costiera, il 18.4% del territorio nazionale è classificato a maggiore pericolosità per frane e alluvioni. Sono 1.3 milioni gli abitanti esposti a rischio frana e 6.8 milioni gli abitanti esposti a rischio alluvioni.

Tra i fenomeni naturali più impattanti, catastrofici e ricorrenti devono sicuramente essere menzionate le frane, con particolare attenzione a quelle innescate da eventi pluviometrici. Sebbene queste siano le più frequenti esistono altre cause che possono provocare il movimento franoso, sia di origine naturale (e.g. terremoti) che di natura antropica (scavi minerari, costruzioni, movimento terra ecc.). Sebbene le frane innescate da queste cause non siano oggetto della presente attività di ricerca, esse rappresentano sicuramente eventi potenzialmente dannosi per la popolazione e le infrastrutture. Secondo lo studio effettuato da Froude and Petley (2018) su scala globale, nell'arco temporale 2004-2016 (incluso), sulla base dei soli report disponibili in lingua inglese, sono state registrate 4862 frane mortali

con un totale di 55997 deceduti. Uno studio simile condotto da Haque et al. (2019) su un totale di 128 nazioni in una scala temporale di 20 anni, che spazia dal gennaio 1995 al dicembre del 2014, ha conteggiato un totale di 3876 frane che hanno provocato circa 11690 feriti e 163660 perdite umane.

I dati riportati sebbene non siano direttamente confrontabili, avendo una diversa scala spazio-temporale e non attingendo dagli stessi database di frane, possono dare una idea dell'importanza dell'argomento trattato.

Il rischio, nel caso in questione quello idrogeologico, indicato con R , può essere valutato con una formula a tre fattori in funzione dei parametri pericolosità P , esposizione Es e, infine, vulnerabilità V . Noto che il rischio non può essere annullato, è possibile sviluppare e intraprendere delle misure di mitigazione con l'obiettivo di ridurre al minimo il numero di vittime. Tali misure possono essere di tipo strutturale o non strutturale. Le prime prevedono la costruzione di infrastrutture di difesa (muri di sostegno, reti paramassi, etc), strategia non sempre adottabile sia a causa dell'impatto economico che tali misure richiedono, rappresentando una importante voce per il bilancio pubblico, sia per l'impatto ambientale di tali opere. Tra le misure non strutturali, quelle considerate in questo lavoro, possono essere citate le attività di divulgazione e diffusione di informazioni riguardanti il dissesto del territorio e i sistemi di preallerta. La divulgazione, effettuata da esperti e/o enti preposti, ha lo scopo di sensibilizzare la popolazione alla tematica della percezione del rischio, su quelle che sono le buone pratiche e i comportamenti da tenere in caso di eventi di dissesto, come l'iniziativa "Io non rischio" divulgata dal Dipartimento di Protezione Civile ([Homepage - Io non rischio \(protezionecivile.it\)](http://protezionecivile.it)). Altra strategia di mitigazione riguarda la gestione del rischio e la creazione di una pianificazione per il monitoraggio e il controllo, sia nei periodi ordinari che in quelli di crisi, conoscendo quelle che sono le criticità all'interno del territorio, ad esempio a scala comunale. Altri sistemi di tipo non strutturale riguardano i sistemi di preallerta che hanno lo scopo di generare in tempi adeguati messaggi di allerta per fronteggiare le minacce di eventi calamitosi (Intrieri et al., 2012; Calvello et al., 2014; Pumo et al., 2016; Piciullo et al., 2018). Generalmente una combinazione delle due strategie di mitigazione rappresenta la scelta migliore, sia dal punto di vista economico che sociale (Foti et al., 2013).

Le relazioni che legano grandezze meteorologiche con l'innescò delle frane, chiamate anche soglie di innesco, costituiscono una delle parti fondamentali dei

sistemi di preallerta. Quelle maggiormente usate sono ottenute generalmente tramite modelli empirici. Le soglie forniscono informazioni su eventi innescanti avvenuti in passato correlando, con una legge di tipo potenza, le caratteristiche della precipitazione come la durata e l'intensità di precipitazione (Caine, 1980; Guzzetti et al., 2007) o la precipitazione cumulata (Peruccacci et al., 2012) con l'innescamento del movimento franoso. Considerando che dopo più di 50 anni di letteratura scientifica il problema della determinazione delle soglie e il loro impiego nei sistemi di preallerta sia ancora aperto, e a tutt'oggi oggetto di ricerca, può dare una idea della complessità del fenomeno oggetto del presente studio.

1.2 OBIETTIVI DELLA RICERCA

Obiettivo del presente lavoro di tesi è quello di mettere a punto misure non strutturali al fine di migliorare la mitigazione del rischio dovuto alle frane indotte da precipitazione.

Nello specifico ci si è focalizzati sulle seguenti tematiche:

1. Determinazione delle soglie pluviometriche di innescamento dei movimenti franosi utilizzando metodi empirici applicati per le due aree di studio;
2. Addestramento di algoritmi di machine learning al fine di testare la capacità di un classificatore di riconoscere eventi innescanti e non innescanti;
3. Valutare la possibilità di utilizzare i dati di umidità del suolo ricavati dal database di rianalisi ERA5-Land;
4. Introduzione all'interno del classificatore di nuove variabili collegate all'innescamento della frana (intensità di picco, umidità del suolo, ecc.);
5. Valutare le performance dei classificatori neurali al variare delle combinazioni dei dati di input.

Al fine di raggiungere gli obiettivi sopra indicati è stato necessario raccogliere e organizzare dati e informazioni riguardanti le precipitazioni e l'innescamento dei movimenti franosi, all'interno delle due aree di studio esaminate in database utilizzabili anche per future ricerche.

1.3 METODOLOGIA DELLA RICERCA

Già a partire dalla metà degli anni 70' numerosi studi (Campbell, 1975; Caine, 1980; Aleotti, 2004; Guzzetti et al., 2007; Peruccacci et al., 2012; 2017; Gariano et al., 2015) hanno tentato di correlare le caratteristiche delle precipitazioni all'innescamento della frana (soglie I-D), per comprendere quale fosse la condizione capace di mobilitare il movimento franoso. Altri tipi di soglie pluviometriche empiriche includono variabili relative alle precipitazioni antecedenti la mobilitazione della frana come l'umidità del terreno (ad esempio il contenuto volumetrico d'acqua) e vengono definite soglie idro-meteorologiche (Bogaard & Greco, 2018; Mirus et al., 2018a; 2018b; Conrad et al., 2021). Queste ultime legano le caratteristiche della precipitazione (in genere la precipitazione cumulata) con le caratteristiche correlabili all'innescamento come le condizioni antecedenti del terreno o i valori di umidità (v. Paragrafo 2.4).

La metodologia proposta prevede di utilizzare algoritmi di intelligenza artificiale, nello specifico le reti neurali artificiali, per proporre un nuovo approccio nell'identificazione dei parametri più idonei da tenere in considerazione nella determinazione delle soglie pluviometriche.

La prima applicazione della metodologia è schematizzata nel diagramma a blocchi di Figura 1.1 e riguarda la regione Sicilia. Disponendo di un database di eventi franosi, dopo aver costruito un database di precipitazioni innescanti e non innescanti, per l'area di studio, è stata implementata una funzione di precursori di frana come quella espressa dalla funzione (1):

$$f_I(D; E; I; I_p) \quad (1)$$

dove D ; E ; I ; I_p sono rispettivamente i valori di durata, precipitazione cumulata, intensità media e intensità di picco sia delle condizioni pluviometriche innescanti che di quelle non innescanti (Peres and Cancelliere, 2021). Le performance sono state valutate costruendo la matrice di confusione e successivamente determinando il True Skill Statistic.

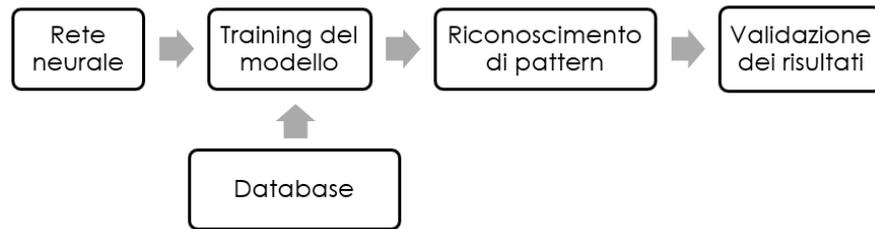


Figura 1.1: Diagramma a blocchi della metodologia proposta.

Una seconda applicazione, più evoluta della prima, ha visto l'utilizzo della medesima metodologia ma una diversa funzione di precursori di frana (Eq. 2). In questo secondo caso sono stati inseriti i dati di umidità del suolo rilevati nel progetto ERA5-Land (Muñoz-Sabater et al., 2021), dopo aver confrontato i dati di rianalisi con quelli misurati da stazioni di monitoraggio.

$$f_{II}(D; E; I; S) \quad (2)$$

Nella relazione (2) S indica l'umidità del suolo, nello specifico il contenuto volumetrico d'acqua relativo ai 4 strati disponibili in cui è stata discretizzata la colonna di terreno. Le applicazioni della seconda metodologia riguardano sia il territorio della regione Sicilia che un'area di studio ubicata a est della città di Bergen, in Norvegia. In Norvegia, infatti, è stata svolta parte dell'attività del dottorato presso il Norwegian Geotechnical Institute (NGI) di Oslo, dove è stata ideata inizialmente la fase che prevede l'utilizzo dei dati di umidità del suolo.

1.4 STRUTTURA DELL'ELABORATO

La tesi è suddivisa in sette capitoli (compreso il presente capitolo introduttivo) e due appendici.

Nel Capitolo 2 viene presentata una panoramica della letteratura esistente sulle caratteristiche delle soglie pluviometriche, i metodi utilizzati per la loro determinazione, vantaggi e svantaggi dei vari metodi.

Nel Capitolo 3 vengono presentate le generalità degli algoritmi di machine learning e le applicazioni nel campo del rischio idrogeologico

Al Capitolo 4 vengono presentate le aree di studio, i dati disponibili e le informazioni consultate per la creazione dei database di frane e precipitazione, nonché i dati di umidità utilizzati.

Il Capitolo 5 riassume una parte della metodologia proposta e i risultati ottenuti per i casi studio analizzati, l'Italia e la Norvegia. Il lavoro presentato in questo capitolo è stato oggetto di pubblicazione (Distefano et al., 2022).

Nel Capitolo 6 viene presentato un approccio che ha come obiettivo l'utilizzo dei dati di umidità del suolo per valutare le performance di un predittore a n variabili considerando le due diverse aree di studio.

Il Capitolo sette riporta le conclusioni del lavoro presentato, i limiti degli approcci proposti e spunti per lavori futuri.

La tesi è completata da due appendici. Nel primo sono riportati i confronti tra l'umidità del suolo registrate da stazioni di misura e i dati di umidità del progetto di rianalisi ERA5-Land mentre nel secondo sono riportati i codici di calcolo sviluppati per la maggior parte delle elaborazioni svolte nel presente lavoro.

CAPITOLO 2 FRANE E SOGLIE PLUVIOMETRICHE DI INNESCO

2.1 GENERALITÀ

Nel presente capitolo verranno presentate le caratteristiche principali delle frane e la nomenclatura comunemente utilizzata per descrivere tali fenomeni. Sarà presentata anche una panoramica sulla letteratura esistente riguardante le soglie pluviometriche di innesco nonché le metodologie impiegate per la loro determinazione mettendo in luce punti di forza e debolezza degli approcci esposti.

2.2 TIPOLOGIE DI FENOMENI FRANOSI

Le frane sono fenomeni che si verificano per cause naturali o a seguito di interventi antropici che possono avere ripercussioni sul territorio, sulle infrastrutture, su beni pubblici e privati. Il termine frana è nel gergo comune indicato per descrivere il movimento di roccia e/o masse di terreno che può manifestarsi quando esiste un dislivello (Highland and Bobrowsky, 2008). Nello studio tecnico e scientifico di fenomeni così complessi, appare chiaro che una singola parola non può descrivere esaustivamente tutti i movimenti possibili né la tipologia di materiale movimentato. La nomenclatura più diffusa e utilizzata è quella inizialmente proposta da Varnes (1978) e successivamente modificata da Cruden and Varnes (1996).

La classificazione prevede una distinzione sulla base del tipo di movimento e una riferita alla tipologia di materiale coinvolto. Considerando la tipologia di movimento possiamo distinguere fenomeni di caduta (fall), ribaltamento (topple), scivolamento (slide) o scorrimento (flow), mentre il materiale mobilitato può essere distinto in roccia, suolo o una combinazione dei due. Il materiale mobilitato può quindi essere classificato come terra se composto prevalentemente da granulometria fine o sabbiosa, mentre si parla di detrito se composto da granulometria grossolana. La Figura 2.1 illustra le tipologie di frane più comuni che possono essere osservate in natura, sebbene sia possibile che una frana cambi la propria tipologia da un tipo a un altro tra il momento di innesco e la fase di arresto. Questo, ad esempio, accade nelle colate di detrito che si manifestano come fase successiva di un'altra tipologia

di frana innescata a seguito di un aumento della pressione interstiziale come conseguenza della saturazione del pendio.

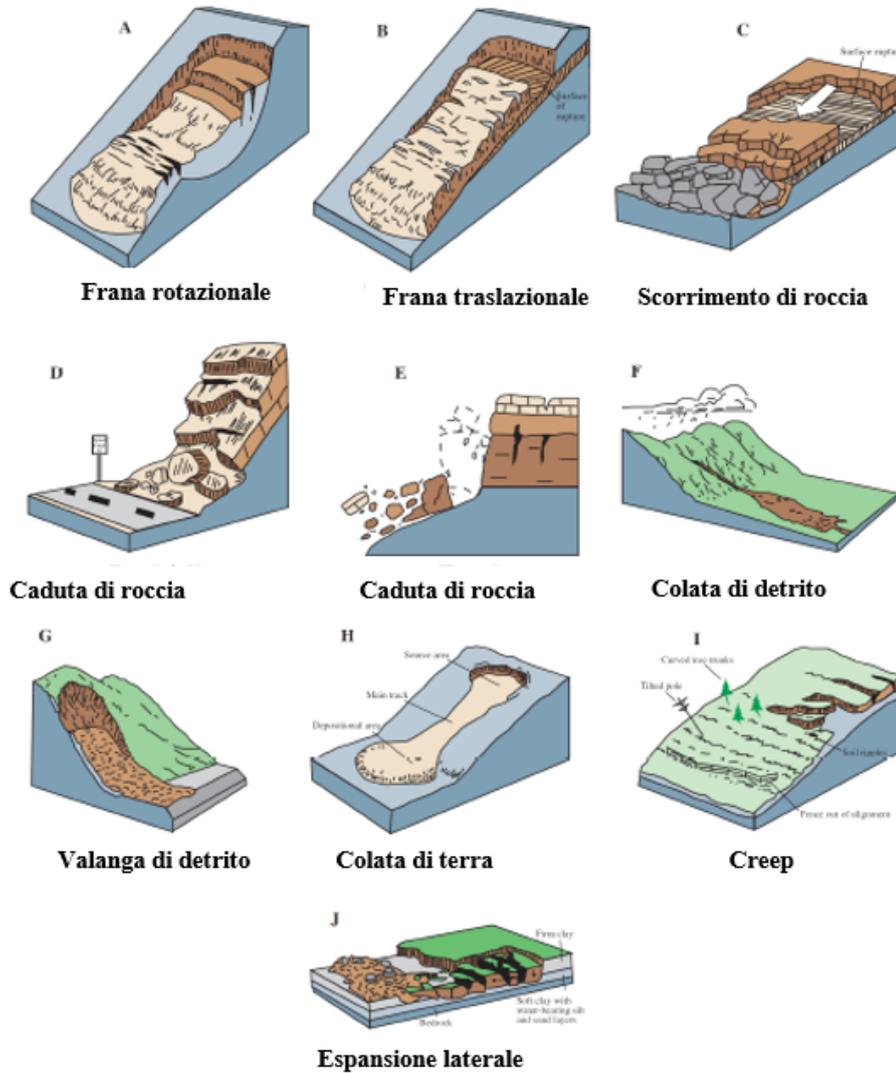


Figura 2.1: Tipologia di frane (Highland and Bobrowsky, 2008).

Numerosi sono i fattori che rendono un'area predisposta a movimenti franosi, tra questi quelli più comunemente presi in considerazione sono la quota, le

caratteristiche dei pendii (pendenza, orientamento, concavità), le caratteristiche litologiche e geologiche dell'area, l'uso del suolo (Pereira et al., 2012; Wang et al., 2015; Yu et al., 2022) nonché l'umidità del suolo (Lin et al., 2020) e l'altezza della falda rispetto al piano campagna (Chacon et al., 2006).

Tra i fattori che invece possono mobilitare il movimento franoso è doveroso citare gli eventi sismici, le precipitazioni, le cause antropiche come le operazioni di movimento terra (Wang et al., 2015; Calvello & Pecoraro, 2018) o la concomitanza di più cause.

Il presente lavoro ha come obiettivo quello di determinare le soglie pluviometriche di innesco, per cui nel seguito verranno prese in considerazione solo frane pluvio-indotte. A seconda della stagione, del regime pluviometrico, della geomorfologia e delle condizioni del pendio e dai meccanismi di innesco (Aleotti, 2004) della zona in esame si evince che le frane possono essere essenzialmente di due tipi, frane superficiali sono per lo più innescate da precipitazioni intense (Campbell, 1975; Peres & Cancelliere, 2014) e frane profonde sono per lo più innescate da precipitazioni prolungate (Martelloni et al., 2012).

2.3 SOGLIE PLUVIOMETRICHE DI INNESCO

Le soglie pluviometriche di innesco identificano i valori critici di precipitazione al superamento dei quali ci si attende il verificarsi di un evento franoso (Guzzetti et al., 2007). Queste costituiscono una parte fondamentale degli early warning system che hanno lo scopo di avvertire la popolazione di una possibile minaccia allertando gli interessati e quindi minimizzando l'esposizione della popolazione (Intrieri et al., 2012; Calvello et al., 2014; Pumo et al., 2016; Piciullo et al., 2018).

La determinazione delle soglie pluviometriche può avvenire utilizzando due possibili approcci, uno fisicamente basato e uno empirico.

Attraverso i modelli fisicamente basati (Baum et al., 2010; Lepore et al., 2013; Alvioli and Baum, 2016) si cerca di trovare una correlazione tra le misure di pioggia a scala locale o regionale e le caratteristiche idrauliche e geotecniche del terreno (Antinoro et al., 2017). I processi legati all'innesco sono descritti dal "*Fattore di Sicurezza (FS)*", definito come rapporto tra le forze stabilizzanti e le forze instabilizzanti. Nel caso in cui $FS > 1$ il terreno è in condizioni di equilibrio statico, per $FS = 1$ è in condizioni di equilibrio limite, mentre per $FS < 1$ è in condizioni di

instabilità. Nel caso di frane innescate da pioggia, il fattore di sicurezza dipende esplicitamente dalla pressione neutra. Una parte fondamentale di questo tipo di modelli è basata sui processi di infiltrazione (Richards, 1931) quindi richiedono il calcolo delle pressioni neutre a partire dai dati di pioggia. Per queste tipologie di modelli sono richiesti dati relativi al sito ed è anche necessario introdurre ipotesi semplificative, ad esempio l'ipotesi di pendio infinito (Peres and Cancelliere, 2016). A causa della loro sito-specificità le soglie di questa tipologia possono solo essere utilizzate a scala di bacino, ma non si prestano a essere implementate all'interno dei sistemi di preallerta (Segoni et al., 2009; Mercogliano et al., 2013), sebbene alcuni studi provino a superare tali difficoltà (Rossi et al., 2013).

Le soglie empiriche, che costituiscono il punto focale di questa ricerca, sono quelle che maggiormente si prestano all'essere implementate nei sistemi di preallerta (Keefer et al., 1987). Queste soglie sono determinate correlando le caratteristiche delle piogge, all'innesco delle frane (Campbell, 1975; Caine, 1980; Aleotti, 2004; Guzzetti et al., 2007; Peruccacci et al., 2012; Peruccacci et al., 2017; Gariano et al., 2015). In alcuni casi, tali soglie sono state implementate all'interno di sistemi di preallerta. Le grandezze che vengono solitamente correlate sono principalmente la durata della pioggia, indicata in genere con D , e la precipitazione cumulata, indicata con la lettera E o l'intensità di pioggia in genere indicata con I . Le leggi che legano le grandezze sopra citate sono legate da una funzione matematica di tipo esponenziale come mostrato dalle Eq. (3) e (4):

$$I = \alpha D^{-\beta} \quad (3)$$

$$E = \alpha D^\gamma \quad (4)$$

dove I , E , D rappresentano rispettivamente l'intensità della precipitazione espressa in $[\text{mm}\cdot\text{h}^{-1}]$, la pioggia cumulata espressa in $[\text{mm}]$ e la durata in $[\text{h}]$ registrati in concomitanza dei fenomeni di innesco. I parametri α e β rappresentano rispettivamente, il parametro di scala e il parametro di forma della soglia. Esistendo una relazione che lega intensità e durata secondo la relazione (5)

$$I = E/D \quad (5)$$

i parametri β e γ sono legati dall'equazione (6).

$$\gamma = -\beta + 1 \quad (6)$$

Riportando in un piano cartesiano la durata della precipitazione e i valori di intensità (o precipitazione cumulata) è possibile tracciare la curva in genere di tipo potenza. che involuppa i punti e divide idealmente il piano in due zone come

mostrato in Figura 2.2 (Peres & Cancelliere, 2021). Al di sotto della soglia le precipitazioni non dovrebbero avere caratteristiche tali da innescare movimenti franosi, i punti sopra la soglia possono invece innescare eventi di dissesto idrogeologico.

I primi lavori in questo campo risalgono tra metà degli anni 70' e gli anni 80' con i lavori pionieristici di Campbell (1975) e Caine (1980), in particolare il primo ipotizzava già che le piogge antecedenti potessero avere un peso non trascurabile nell'innescare delle frane, mentre il secondo determinò una soglia sulla base di 73 eventi innescanti e ipotizzava di poter utilizzare i risultati ottenuti all'interno di un sistema di preallerta

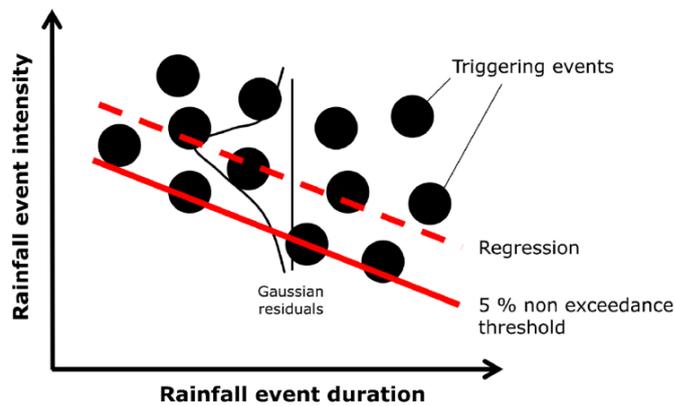


Figura 2.2: Rappresentazione di una soglia pluviometrica derivata da dati di pioggia osservati in concomitanza con l'innescare di frane (Peres & Cancelliere, 2021)

A distanza di quasi 50 anni il dibattito e la ricerca sulla determinazione delle soglie pluviometriche e delle possibili applicazioni è più che mai attivo, infatti, tenendo conto dei dati forniti da Guzzetti et al. (2007, 2008) e da Segoni et al. (2018) tra gli anni 70' e il 2007 sono state proposte 124 diverse soglie pluviometriche, mentre nell'arco temporale 2006-2018 più di cento studi hanno definito 115 nuove soglie. Tuttavia, come è possibile osservare nella mappa in Figura 2.3 si nota come gli studi non siano omogenei e che ampie aree non siano state oggetto di studio. Alcuni tentativi di fornire soglie a scala globale sono descritti in Caine, (1980) e Guzzetti et al., (2007). Studi su macroaree come quelle continentali sono descritti in Jibson, (1989), Chien-Yuan et al. (2005), Guzzetti et al. (2008), Dahal and Hasegawa (2008).

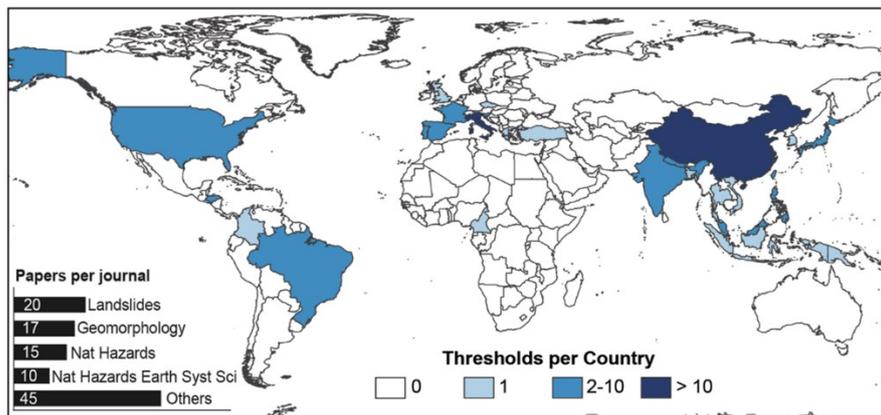


Figura 2.3: Distribuzione spaziale degli studi che propongono soglie pluviometriche di innesco (Segoni et al., 2018)

Molti più studi (Brunetti et al., 2010; Peruccacci et al., 2012; Gariano et al., 2015, 2020) si focalizzano su scale di dettaglio maggiore proponendo soglie nazionali o regionali. Indagando su aree estese (già a livello nazionale) risulta essere pressoché impossibile tenere conto del regime climatico che influisce sulle caratteristiche delle precipitazioni. Gli studi riguardanti le soglie di innesco, tendono ad uniformare i dati di pioggia, per poter meglio rappresentare e descrivere l'innesco delle frane, ottenendo serie temporali più lunghe che meglio si prestano a descrivere la precipitazione, in termini probabilistici. Questo può essere causa di errore, a titolo di esempio, nelle soglie calcolate per le regioni tropicali (Saito et al., 2010; Li et al., 2011) e sub-tropicali si deve ricorrere a due soglie differenti a seconda che si sia nella stagione secca o in quella umida.

Sebbene le soglie empiriche richiedano un numero di informazioni minori (e comunque diverse rispetto a quelle necessarie per ricavare le soglie con modelli fisicamente basati), numerose sono le criticità che si presentano quando si prova a calcolare o a validare un nuovo set di soglie di innesco. Le principali cause di incertezza e/o errore sono di seguito riportate (Peres & Cancelliere, 2014; Bogaard & Greco, 2018):

- Disponibilità e qualità dei dati di pioggia, da intendersi come quantità (serie storiche di estensione adeguata), completezza e qualità dei dati e distribuzione spaziale degli strumenti di misura sul territorio al fine di

rendere la modellazione quanto più prossima e veritiera (Nikolopoulos et al., 2015; Marra, 2019);

- Disponibilità dei dati di frana, in genere ricostruiti tramite i rapporti delle forze dell'ordine e/o dei vigili del fuoco, tramite le informazioni ricavate dai quotidiani, siti web o articoli scientifici. Sono in genere organizzati in database. Non sempre si conosce il momento di innesco, il che può portare alla determinazione di soglie non ben calibrate (Peres et al., 2018).

Ultimo ma non meno importante dei problemi nella determinazione delle soglie pluviometriche di innesco riguarda la validazione dei risultati ottenuti. La capacità di saper distinguere in modo netto precipitazioni innescanti da quelle non innescanti è un passo necessario, specie per le soglie implementate in sistemi di preallerta. Secondo la revisione sullo stato dell'arte pubblicata da Segoni et al. (2018) nell'arco temporale 2008-2016 considerando più di 110 soglie, si possono estrapolare alcuni importanti dati. Secondo l'analisi proposta, nel 40% delle ricerche vagliate non è presente nessuna tipologia di validazione dei risultati, nel 27% dei casi la validazione viene effettuata con gli stessi dati utilizzati per la calibrazione dei modelli e solo nel restante 33% dei casi viene utilizzata una corretta procedura di validazione. La fase di validazione rappresenta uno di quei fattori (come, ad esempio, l'utilizzo di procedure obiettive e riproducibili) di estrema importanza, soprattutto se le soglie sono implementate in un sistema di early warning. Un mancato allarme, infatti, può esporre al rischio la popolazione mentre un numero troppo elevato di falsi allarmi può indurre nella popolazione una mancanza di fiducia del sistema e l'innescarsi della "crying wolf syndrome" (Breznitz 1984). La validazione viene effettuata generalmente popolando una "matrice di confusione" o "matrice di contingenza", una tabella a doppia entrata che mette in relazione le previsioni del sistema con quello che è accaduto nella realtà. Dalle istanze che popolano la matrice di contingenza, attraverso l'utilizzo di "skill score", ovvero parametri numerici adimensionali che riassumono le caratteristiche del classificatore (Accadia et al., 2003) come il True Skill Statistics (Peirce, 1884), o tramite l'analisi Receiver Operating Characteristic (ROC) (Fawcett, 2006) si può determinare la capacità di fare previsioni corrette del sistema.

Nuove tipologie di soglie sono continuamente in fase di studio come nel caso del lavoro presentato da Rosi et al. (2021) dove viene mostrata una soglia 3D. Oltre alle

grandezze intensità e durata viene aggiunta una nuova variabile sull'asse z la piovosità media areale (MeAR), definita come la piovosità media cumulata registrata da tutti i pluviometri di una data area di allerta in un determinato periodo di tempo (sono state considerate finestre temporali comprese tra 7 e 30 giorni prima di un determinato evento di frana). Questa nuova grandezza rappresenta una condizione di saturazione del terreno. L'introduzione del nuovo parametro ha mostrato un miglioramento delle condizioni di innesco dovuto alla riduzione dei falsi allarmi (eventi non innescanti riconosciuti come innescanti).

2.4 SOGLIE IDRO-METEOROLOGICHE

Altre tipologie di soglie, note in letteratura come soglie di innesco idro-meteorologiche (Bogaard & Greco, 2018), prevedono di sostituire la durata dell'evento, in genere riportato sull'asse delle ascisse, con alcuni indicatori di tipo idrologico come le piogge antecedenti, il grado di saturazione del suolo, il livello della falda. Le grandezze menzionate che possono costituire alcune tra le "cause" che predispongono al collasso, vengono correlate con grandezze come l'intensità o la precipitazione cumulata, che rappresentano il fattore che provoca l'innescamento.

Tentativi di mettere in correlazione le caratteristiche delle precipitazioni e valori di umidità del suolo in concomitanza con l'innescamento dei movimenti franosi iniziano nei primi anni del 2000, prima che il termine "soglie idro-meteorologiche" venisse coniato.

Glade et al. (2000) indagano sulle condizioni di innesco, per definire le soglie di innesco, in tre regioni soggette a frane in Nuova Zelanda. Le soglie regionali sono determinate utilizzando il "Antecedent Daily Rainfall Model." (Crozier and Eyles, 1980). Le condizioni di innesco della frana sono rappresentate da una combinazione di pioggia antecedente e caratteristiche delle precipitazioni del giorno dell'evento.

Nello studio di Godt et al. (2006) le soglie di intensità-durata sono state combinate con un modello di bilancio idrico. Gli autori dispongono di un database di frane di 25 anni e informazioni di precipitazione a scala oraria, provenienti da 17 pluviometri. Utilizzando un albero decisionale, a tre livelli di attenzione (Null, Watch, Warning), sono state definite le condizioni che possono portare all'innescamento di frane nell'area di Seattle. Dal confronto tra i risultati dell'albero decisionale e le

date di frane emerge come per le 58 precipitazioni associate allo stato di “No frana (Null)” solo 3 sono state classificate erroneamente. Per 28 delle precipitazioni ricadenti nella soglia “Attenzione (Watch)”, esistono prove di frane superficiali per il 42.9% dei casi, mentre per le 13 tempeste a cui è stata assegnato il livello “Pericolo (Warning)” si è verificata una frana nel 61.5% dei casi.

Ponziani et al. (2012) utilizzano il database AVI (Aree Vulnerabili Italiane), per creare tre diverse soglie cumulata-grado di saturazione considerando tre diverse finestre temporali (24, 36 e 48 ore). L'area di studio è la regione Umbria. Lo studio mostra una correlazione lineare tra le soglie di precipitazione e le condizioni iniziali di umidità del suolo con coefficienti di correlazione fino a 0.60 (nel caso di cumulata a 48 ore), confermando il ruolo chiave che l'umidità del suolo ha sull'innesco di frane.

Nello studio condotto da Mirus et al. (2018a) per la zona della linea ferroviaria di Seattle viene adottato un sistema di early warning che tiene conto di una soglia pioggia-saturazione. Le informazioni riguardanti lo stato del suolo sono ricavate da strumenti di monitoraggio che forniscono il contenuto volumetrico d'acqua disposti in prossimità dei pendii presenti nell'area di studio. Le valutazioni sulle prestazioni delle soglie vengono valutate tramite l'analisi ROC sulla base di un database di frane avvenute nel periodo 2015-2017.

In uno studio successivo Mirus et al. (2018b) hanno confrontato diverse soglie idro meteorologiche giornaliere bilineari costruite considerando diversi precursori di frana. Le soglie legano la precipitazione cumulata, rispettivamente del giorno precedente (P1), dei tre giorni (P3), dei cinque (P5) e dieci giorni (P10) precedenti al collasso con le condizioni di saturazione del terreno del giorno antecedente all'innesco (S1) e le condizioni di saturazione antecedente a nove giorni (S9) e dieci giorni (S10). Le aree considerate per lo studio sono due, la linea ferroviaria di Seattle e la città di Portland. Le performance delle soglie presentate sono state valutate tramite analisi ROC. Lo studio evidenzia come ogni tipologia soglia, in funzione dei parametri scelti per caratterizzarla sia in grado di prevedere alcune frane, ad esempio quelle all'inizio della stagione piovosa, ma non riesce a prevedere quelle dove la saturazione del terreno è responsabile dell'innesco. Per tale scopo gli autori consigliano di utilizzare un sistema a più soglie che insieme riescano a prevedere più tipologie di eventi in funzione di quelle che possono essere le sito-specificità dell'area di studio.

Conrad et al. (2021) hanno sviluppato un algoritmo in linguaggio “Python” che permette all’utente di determinare soglie idro-meteorologiche per l’area di studio. Utilizzando serie di dati riguardanti diverse variabili come informazioni idrologiche, condizioni di umidità del terreno, pressioni interstiziali e informazioni sull’innesco dei movimenti franosi possono essere determinate diverse tipologie di soglie le cui performance possono essere valutate considerando diversi indicatori deducibili dall’analisi ROC.

Alcune applicazioni di soglie idro-meteorologiche sono in fase di test anche in Europa come nel caso della regione della Catalogna (Palau et al., 2021) in cui alle informazioni radar sono state associate informazioni derivanti da modelli che simulano il contenuto volumetrico di acqua del terreno. Il sistema di allertamento della regione Emilia-Romagna (Martelloni et al., 2012; Lagomarsino et al., 2013), basato inizialmente solo su soglie pluviometriche intensità-durata, è stato aggiornato integrando anche valori di umidità del suolo come descritto in Segoni et al. (2018b). Tramite una procedura di back analysis sono state confrontate le due versioni del sistema di early warning. Il confronto ha mostrato che l’introduzione dei dati di umidità del suolo comporta una diminuzione dei falsi allarmi rispettivamente dell’8%, del 31% e 48% per i livelli di allerta uno (allertamento più lieve), di livello due e di livello tre. Il numero di mancati allarmi è stato ridotto del 22% ed è aumentata la capacità di riconoscere eventi innescanti.

Altri studi provano a determinare soglie idro-meteorologiche che tengano conto delle condizioni di umidità del suolo come nel caso utilizzando la crescente disponibilità di dati di telerilevamento in termini di qualità dei prodotti considerando la crescente risoluzione spaziale e temporale. Gli studi presentati da Marino et al. (2020) e da Uwihirwe et al. (2022) dimostrano come l’utilizzo dei dati provenienti da telerilevamento possa migliorare le capacità di capire le meccaniche legate all’innesco.

Per la Sicilia Palazzolo et al. (2022) hanno proposto un sistema di soglie idrometeorologiche considerando l’intensità media delle precipitazioni e dati di umidità del suolo del progetto di rianalisi ERA5-Land considerando sia i singoli strati in cui è discretizzato il terreno, che una combinazione di tutti gli strati tramite la “Principal Component Analysis (PCA)”. L’utilizzo di dei dati di umidità del suolo ha mostrato un apprezzabile miglioramento delle performance quando sono stati considerati i valori di umidità relativi agli strati superiori. Il valore del TSS della

soglia idrometeorologica risulta pari a 0.71, di molto superiore a quello ottenuto con l'approccio tradizionale intensità-durata ($TSS_0 = 0.50$).

CAPITOLO 3 RETI NEURALI ARTIFICIALI

3.1 GENERALITÀ

Nel presente capitolo sono presentati gli algoritmi di machine learning più noti e utilizzati. Più in dettaglio saranno esaminati gli elementi che compongono le reti neurali artificiali e i concetti alla base della modellazione con questi specifici algoritmi. Saranno passati in rassegna gli studi in cui le ANN vengono maggiormente utilizzate, ovvero la realizzazione di mappe di suscettibilità e la previsione degli spostamenti di frane lente e/o profonde.

3.2 LE RETI NEURALI ARTIFICIALI

Tramite gli algoritmi di machine learning un calcolatore, a cui viene assegnato un compito tramite un programma, può raggiungere l'obiettivo migliorando iterativamente le proprie prestazioni man mano che acquisisce sempre più esperienza nell'esecuzione di queste attività. Quindi basandosi sui dati disponibili la macchina è capace di effettuare previsioni e/o prendere decisioni (Ray, 2019). Numerosi sono gli algoritmi di machine learning messi a punto negli anni, come riportato in Figura 3.1 (Batta, 2020).

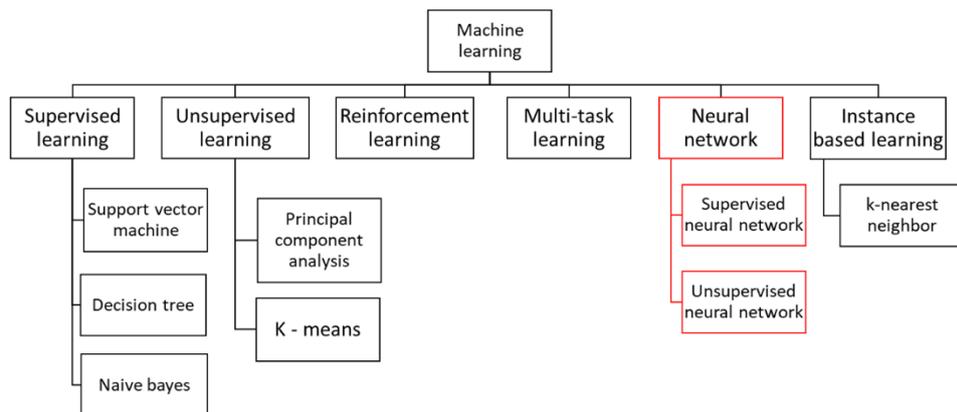


Figura 3.1: Algoritmi di machine learning (Batta, 2020)

Tra gli algoritmi più comuni possono essere citati:

- Naive Bayes (NB): l'algoritmo si basa sul teorema di Bayes (John & Langley, 2013). Può essere utilizzato sia per categorizzare istanze binarie che multi-classe. Utile per classificare efficacemente le istanze affette da rumore (noise nei dati) e costruire un modello di previsione robusto (Sarker, 2019). Il vantaggio chiave è che, rispetto ad approcci più sofisticati, necessita di un minore numero di epoche di addestramento per stimare i parametri necessari (Pedregosa et al., 2011).
- K-nearest neighbors (KNN): è un "apprendimento basato sull'istanza" o apprendimento non generalizzante. Non si concentra sulla costruzione di un modello generale, invece, memorizza tutte le istanze corrispondenti a dati di allenamento in uno spazio n-dimensionale (Aha et al., 1991). Il KNN utilizza e classifica i nuovi dati in base a misure di similarità come, ad esempio, la distanza euclidea (Pedregosa et al., 2011). La classificazione di un nuovo dato è calcolata basandosi sul "voto" della maggioranza dei "k" vicini di ogni punto. L'algoritmo risulta abbastanza robusto ma le performance dipendono dalla qualità dei dati di partenza. Il problema più grande con il KNN è scegliere il numero ottimale dei vicini da considerare. Può essere utilizzato sia per la classificazione che per la regressione.
- Random Forest (RF): (Breiman, 2001) questo metodo utilizza il "parallel ensembling" o "apprendimento d'insieme" creando diversi classificatori chiamati alberi decisionali. Gli alberi lavorano in parallelo su diversi sotto-campioni di dati. Il risultato finale dipende dal voto della maggioranza o dalle medie degli alberi decisionali. In tale modo si riduce al minimo il problema dell'overfitting e aumenta l'accuratezza e il controllo della previsione (Pedregosa et al., 2011). Il modello di apprendimento RF con più alberi decisionali è in genere più accurato di un modello basato su un singolo albero decisionale (Sarker et al., 2019). È adattabile sia a problemi di classificazione che di regressione e si adatta bene sia ai valori categoriali che a quelli continui.

3.3 CARATTERISTICHE DELLE RETI NEURALI ARTIFICIALI

Tra le possibili scelte la nostra attenzione si è focalizzata sulle reti neurali artificiali.

L'idea di un sistema che potesse riuscire a processare i dati in maniera analoga al cervello umano risale alla prima metà degli anni 40' con il lavoro di McCulloch & Pitts (1943), in cui a seguito di un input le informazioni fossero processate da un sistema interconnesso simile al sistema nervoso. Lo sviluppo delle scienze informatiche e un incremento esponenziale della potenza di calcolo hanno permesso, a partire degli anni 80', l'affermarsi di questi strumenti. Questa tipologia di sistema computazionale risulta essere molto apprezzata per le capacità di auto-apprendimento, classificazione e il riconoscimento di schemi. Ciò ha permesso una grande diffusione su una moltitudine di campi che vanno dalla medicina, allo studio finanziario ed economico, in campo produttivo e industriale, nonché nel campo della ricerca (Widrow et al., 1994).

Tra i punti di forza delle reti neurali possiamo sicuramente annoverare la capacità di fornire previsioni sulla base di un set o campione di dati (ammesso che questo sia sufficientemente esteso) anche quando non si conoscono esplicitamente le leggi che regolano il processo in questione, grazie ai sopra menzionati sistemi di riconoscimento di pattern e auto-apprendimento. Risultano inoltre capaci di migliori approssimazioni rispetto ai modelli statistici e hanno la caratteristica di essere non lineari, riuscendo a descrivere meglio molti dei processi sia naturali (e.g. serie temporali) che socioeconomico-produttivi (Granger and Terasvirta, 1993).

3.4 STRUTTURA DELLE RETI NEURALI ARTIFICIALI

Le reti neurali sono composte da un numero di unità computative elementari chiamate neuroni o nodi collegate tra esse con connessioni pesate chiamate sinapsi, in analogia con il sistema cerebrale umano. Ogni nodo riceve un segnale di input o uno stimolo esterno che elabora secondo una funzione di attivazione, che successivamente si propaga all'interno della rete. La disposizione di uno o più strati di neuroni, il numero di connessioni e la possibilità che il segnale possa essere propagato in una o in due direzioni, danno origine a diverse architetture delle ANN che influenzano il comportamento, la capacità di approssimare e generalizzare del modello stesso.

Il primo aspetto è la definizione dell'architettura della rete, che influisce sulle performance dell'algoritmo così come il numero di hidden layer e il numero di neuroni in ogni singolo strato.

Tra le possibili architetture di rete neurale possibili abbiamo (Figura 3.2):

- Reti neurali feedforward a singolo strato (single-layer feedforward networks), composte da uno strato di input (input layer) e uno strato di output (output layer). Essendo lo strato di output l'unico ad essere composto da neuroni (detti anche computational nodes) la rete è detta essere a singolo strato: lo strato di input non prevede calcoli ma solo l'immissione di segnali nella rete, lo strato di output è composto da neuroni e quindi è in esso che si hanno le trasformazioni dei segnali in ingresso;
- Reti neurali feedforward multi-strato (multi-layer feedforward networks), oltre agli strati di input e di output presenti nelle reti a singolo strato, sono presenti uno o più strati nascosti composti da neuroni;
- Reti neurali di tipo ricorsivo (recurrent); in questo tipo di reti si ha almeno un feedback loop: l'informazione in uscita da uno strato è utilizzata come input in uno strato precedente per pervenire (ri-trasformato) allo stesso strato, secondo un processo ricorsivo.

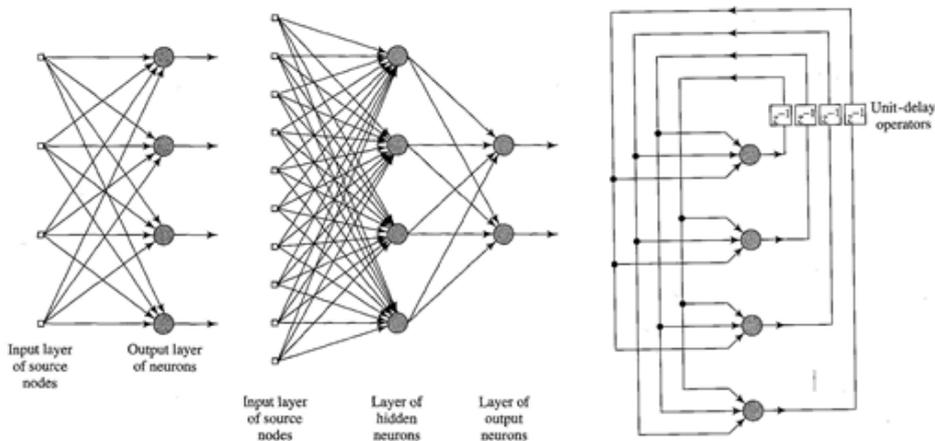


Figura 3.2: Architettura delle reti neurali: (da sinistra a destra) una rete neurale feedforward a singolo strato totalmente connessa, una rete neurale con uno strato nascosto totalmente connessa e una di rete neurale di tipo ricorsivo (Haykin, 1999).

La seconda tipologia è probabilmente la più utilizzata ed è caratterizzata dal fatto che il flusso di informazioni si propaga in una sola direzione (Zhang et al., 1998). Un esempio di questo tipo di ANN è riportato in Figura 3.3. Dalla figura è possibile vedere la disposizione di diversi strati di neuroni, in particolare si hanno:

1. Uno strato di neuroni di input;
2. Uno o più strati di neuroni nascosti che elaborano i dati;
3. Uno strato di neuroni di output;
4. La presenza di connessioni pesate che permettono alla rete di apprendere.

Il numero di neuroni presente in ogni strato dipende dalla tipologia di problema che si vuole risolvere. Per gli strati di input e output esiste una “regola empirica”, in genere il numero di neuroni nel primo strato coincide con il numero di variabili che la rete neurale dovrà processare, in maniera analoga, il numero di neuroni nello strato di output sarà pari al numero di variabili fornite dalla rete.

Determinare il numero di neuroni negli strati nascosti e il numero di hidden layer risulta essere in genere molto più complesso, perché questo strato è responsabile del riconoscimento dei pattern e dell’analisi non lineare tra dati in entrata e uscita.

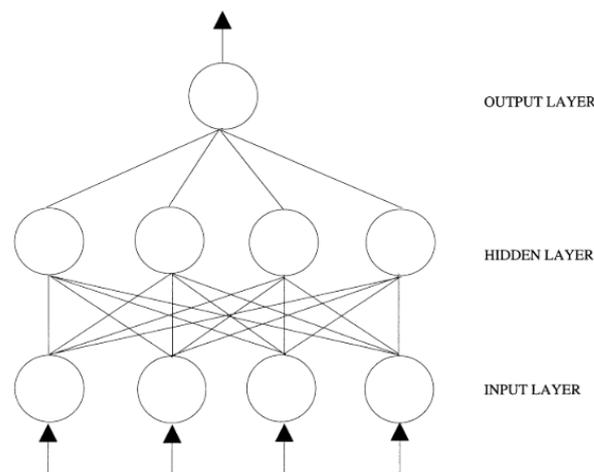


Figura 3.3: Rete feedforward (Zhang et al., 1998).

Una rete con un singolo strato è capace di processare e approssimare ogni tipo di funzione, anche complessa (Hornik et al., 1989), motivo per cui questa tipologia

risulta essere la più diffusa. Reti con più di due strati hanno invece mostrato una maggiore efficienza nella fase di training della rete neurale e capacità previsionale migliore.

Per quanto riguarda invece il numero di neuroni negli hidden layer è stato osservato che, in genere, reti con pochi neuroni sono preferibili in quanto hanno una migliore capacità di generalizzare dando così minori problemi di overfitting. Tuttavia, reti che non hanno abbastanza nodi nascosti potrebbero non essere in grado di imparare dai dati. Grazie alla potenza di calcolo attualmente disponibile, la possibilità di testare più tipologie di rete e diverse architetture procedendo “a tentativi”, non rappresenta un limite, anzi permette il confronto tra le diverse performance ottenibili.

Il passaggio di informazioni tra i vari layer è affidato a quella che prende il nome di “funzione di attivazione” o “funzione di trasferimento”, funzioni matematiche dotate di determinate proprietà matematiche (e.g. funzione limitata, monotona crescente e differenziabile) che introducono nella rete neurale un grado di non linearità. Le funzioni più utilizzate sono mostrate in Tabella 3.1.

Tabella 3.1: Funzioni di trasferimento

Funzione	Formula
Funzione sigmoide o logistica	$f(x) = (1 - \exp(-x))^{-1}$
Funzione tangente iperbolica	$f(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)}$
Funzione lineare	$f(x) = x$

Tra le funzioni citate quella più utilizzata è sicuramente la funzione sigmoide. In determinati casi alcune funzioni risultano essere più indicate di altre. Come suggerito da Klimasauskas (1991), ad esempio, la funzione sigmoide risulta essere la più adatta per problemi di classificazione quando l'apprendimento è basato su grandezze medie o su trend mediati, mentre la funzione tangente iperbolica risulta essere quella più indicata nei problemi di previsione, o quando in genere si è interessati a studiare i comportamenti in cui si cerca la deviazione da parametri medi.

3.5 ALGORITMI DI TRAINING

Le capacità di apprendimento delle reti neurali derivano da tecniche di training mediante cui si sviluppano le abilità di classificare e il riconoscimento di pattern. Le procedure più utilizzate sono essenzialmente due: l'apprendimento non supervisionato (unsupervised learning) e l'apprendimento supervisionato (supervised learning), utilizzato nelle elaborazioni e maggiormente attenzionato.

Nell'apprendimento supervisionato il training della rete neurale viene effettuato sulla base di un campione (training sample) mostrando alla rete neurale degli esempi (input e corrispondenti "target"). Tramite questa modalità di apprendimento (utilizzato per lo più in reti feedforward a uno o più strati), che può essere paragonata alla spiegazione effettuata da un insegnante, gli strati intermedi della rete hanno lo scopo di calibrare i pesi delle connessioni per poter emulare i risultati degli esempi forniti.

Tra gli algoritmi di training più utilizzati per le reti neurali artificiali il più utilizzato è il "error back-propagation" proposto da Rumelhart et al. (1986). Dato un campione di training la procedura seguita dall'algoritmo può essere così schematizzata (Haykin, 1999):

1. Inizializzazione (initialization): si assegnano dei valori ai pesi della rete;
2. Presentazione degli esempi di allenamento (presentations of training examples): la rete apprende dagli esempi. Per ciascun esempio del training set si eseguono quindi le sequenze di calcoli dei punti 3 e 4 successivi. Ciascuna presentazione alla rete di tutto il training set, con l'esecuzione di tutti i relativi cicli di calcoli dell'algoritmo di allenamento, costituisce un'epoca;
3. Calcoli in avanti (forward computation): sia $(\mathbf{x}(n), \mathbf{d}(n))$ un esempio del training set; il vettore-input $\mathbf{x}(n)$ è inserito nella rete a partire dallo strato di input e la risposta desiderata $\mathbf{d}(n)$ è presentata allo strato di output. Si calcolano i potenziali di attivazione e i segnali di output della rete procedendo strato-per-strato, da quello di input a quello di output, calcolando l'errore;
4. Calcoli indietro (backward computation). Si calcolano i gradienti locali della rete e si correggono quindi i pesi della rete, nel generico j -esimo strato.

Questo è il passaggio caratteristico dell'algoritmo back-propagation in cui l'errore viene "propagato all'indietro";

5. Iterazione (iteration): terminata la presentazione dell'intero insieme di esempi $\{(\mathbf{x}(n), \mathbf{d}(n))\}_{n=1, \dots, N}$, si procede, fino al soddisfacimento di un criterio di arresto, presentando nuovamente alla rete tale set di esempi. Si ha cioè una procedura iterativa di correzione dei pesi, che procede epoca per epoca.

Per riassumere, nella prima fase i dati presentati nello strato di input sono processati dagli strati intermedi. L'elaborazione prosegue all'interno della rete fino allo strato di output, in cui si ha il confronto tra i risultati ottenuti e quelli voluti attraverso una metrica di prestazione. Nella seconda fase partendo dall'errore osservato il segnale viene inviato in direzione contraria verso gli strati intermedi, ed i pesi della rete vengono modificati i pesi delle connessioni per migliorare ulteriormente i risultati (Sathya and Abraham, 2013).

Con il tempo diverse varianti di questo algoritmo sono stati proposti come il "modified back-propagation" (Liang et al., 1994), "conjugate-gradient back-propagation" (Charalambous, 1992) e quello utilizzato per le elaborazioni lo "scaled conjugate-gradient back-propagation" caratterizzato da velocità superiori rispetto agli altri algoritmi (Møller, 1993).

Altre tipologie di reti neurali utilizzano algoritmi di apprendimento non supervisionato per identificare relazioni input-output non note a priori (non vengono forniti esempi alla rete) (Kohonen et al., 2005). La procedura più comune di apprendimento non supervisionato è probabilmente quella delle "Self-Organizing neural networks" che prevede una mappatura "Self-Organizing Maps - SOM" degli input in ingresso. La procedura può essere riassunta in tre step. Nel primo step i neuroni sono in competizione, il neurone il cui peso sinaptico (in genere assegnato in maniera randomica) è simile al dato in ingresso, sulla base di una metrica euclidea, vincerà la competizione e costituirà il centro della mappa. Il secondo step prende il nome di fase cooperativa; in essa viene decretato il vincitore della competizione che costituisce il centro della rete topologica e vengono creati dei legami laterali con altri neuroni. La terza fase è quella di adattamento in cui il neurone vincitore e quelli limitrofi calibrano i pesi sinaptici sulla base dei dati in ingresso (Haykin, 1999).

3.6 VALUTAZIONE DELLE PERFORMANCE DELLE RETI NEURALI

La valutazione delle performance di una rete neurale può essere espressa in termini di efficienza computazionale attraverso molteplici parametri, come ad esempio il tempo di training o il tempo necessario alla modellazione. Tuttavia, il dato sicuramente più importante e rappresentativo è l'accuratezza delle previsioni effettuate dalla rete stessa, definita tramite una metrica di performance che esprime in modo sintetico le differenze tra il valore desiderato (target) e il valore predetto dall'algoritmo (output).

Indicando con e_t il singolo errore di previsione, con y_t il valore attuale (valore atteso) e con N il numero dei termini di errore, si possono definire una serie di indicatori, mostrati in Tabella 3.2, che permettono di valutare le performance.

Tabella 3.2: Indicatori performance delle reti neurali (Zhang et al., 1998).

Indicatore	Formula
Mean Absolute Deviation	$MAD = \frac{\sum e_t }{N}$
Sum of Squared Error	$SSE = \sum (e_t)^2$
Mean Squared Error	$MSE = \frac{\sum (e_t)^2}{N}$
Root Mean Squared Error	$RMSE = \sqrt{MSE}$
Mean Absolute Percentage Error	$MAD = \frac{1}{N} \sum \left \frac{e_t}{y_t} \right \cdot 100$

Quelli presentati sono gli indicatori più utilizzati, sebbene nella letteratura di settore possono esserne trovati molti altri. Bisogna inoltre precisare che non esiste un singolo indicatore che possa riassumere le proprietà della rete ma che ognuno di questi ha vantaggi e limitazioni (Makridakis et al., 1983).

Considerate le numerose e interessanti caratteristiche delle reti neurali, risulta giustificato il loro crescente utilizzo in un numero sempre maggiore di campi di applicazioni; tuttavia, esistono alcuni aspetti che possono renderne difficoltosa l'applicazione o portare a risultati inesatti o scorretti.

Importante problematica che si deve affrontare in fase di apprendimento (dopo aver scelto l'algoritmo più adatto al caso) è la necessità di evitare l'overfitting, ovvero garantire una sufficiente capacità di generalizzazione alla ANN. Arrestare il

training con troppo anticipo può pregiudicare le capacità di riconoscere le caratteristiche del dataset, in maniera opposta non arrestarlo per tempo potrebbe far apprendere errori immancabilmente presenti nel training sample per poi provocare problemi di overfitting (Weigend et al., 1992). Spingendo il più possibile il periodo di allenamento della rete (anche noto come numero di epoche), la rete neurale esplica le sue potenzialità interpolative in maniera graduale: risulta quindi conveniente sfruttare la potenza di calcolo della rete e aumentare il numero di epoche di allenamento. Il secondo fattore è che essendo il training set un campione, esso presenta delle fluttuazioni casuali o rumore (noise), specifiche di quel campione, che tendono a rendere maggiormente complesso l'andamento dell'output in funzione dell'input. Aumentando il numero di epoche, aumenta la possibilità che ciò che la rete neurale stia imparando sia il proprio rumore e non la possibile relazione tra input e target. Uno fra i metodi per garantire buone capacità di generalizzazione è l'early stopping.

La procedura di early stopping prevede di fermare periodicamente l'allenamento della rete (sull'estimation subset), e poi testare la stessa sul campione di validazione. Scegliendo come parametro di controllo, ad esempio, l'MSE (o un altro indice di prestazione, a seconda del problema da risolvere), si costruiscono così, per punti, le due curve dell'MSE relative alla calibrazione e alla validazione (Haykin, 1999). In Figura 3.4 è possibile osservare come il numero di epoche ottimale corrisponda è quello corrispondente al minimo errore in fase di validazione.

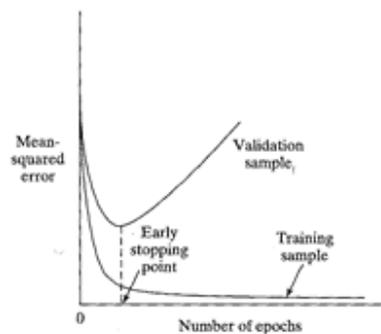


Figura 3.4: Procedura di early stopping (Haykin, 1999)

3.7 APPLICAZIONI DELLE RETI NEURALI NEL CAMPO DELLE FRANE

L'utilizzo delle reti neurali o più in generale l'adozione di strumenti che sfruttino l'intelligenza artificiale risulta essere diffuso in modo particolare nella realizzazione di mappe di suscettibilità di frana e nel monitoraggio di frane sospese o caratterizzate da movimenti lenti, mentre l'applicazione ai fini del preannuncio è piuttosto limitata.

Storicamente la zonizzazione e la mappatura di zone con predisposizione a franare viene effettuata attribuendo dei pesi ad alcuni parametri. Le grandezze più comunemente oggetto di studio sono ad esempio le caratteristiche geomorfologiche, la pendenza e l'orientamento dei versanti, l'uso del suolo ma anche la distanza e/o l'oscillazione della falda dal piano campagna, le precipitazioni o la pericolosità sismica associata all'area di studio nonché la possibile interazione dei versanti con fabbricati o infrastrutture.

Secondo la classificazione proposta da Reichenbach et al. (2018) la suscettibilità a franare può essere determinata seguendo 5 approcci (di seguito elencati):

1. Mappatura geomorfologica;
2. Analisi dei database di frana;
3. Approcci euristici o basati su indici;
4. Modelli statisticamente basati
5. Metodi basati sui processi fisici;

Per i primi 4 punti le analisi si basano sul giudizio di un esperto. Il primo passo consiste nel suddividere l'area di studio, in genere utilizzando una maglia più o meno estesa. Successivamente si assegna un peso per ognuno degli indicatori o selezionati e si effettua la somma dei valori pesati (Carrara, 1983). Le procedure che determinano la suscettibilità a franare non forniscono, in genere, una grandezza fisica che quantifica la propensione a franare, ma alle varie aree vengono assegnati generalmente dei giudizi qualitativi in ordine gerarchico (e.g. suscettibilità molto bassa, bassa, modesta, elevata, molto elevata).

Il punto 5 considera una modellazione delle frane tramite un approccio fisicamente basato semplificato che analizza le condizioni di stabilità/instabilità utilizzando modelli di equilibrio limite e l'ipotesi di pendio infinito (Rigon et al., 2006; Alvioli and Baum, 2016).

L'approccio presentato ha dei limiti tra cui:

- Gli indici utilizzati e i pesi assegnati sono in genere frutto dell'esperienza e del giudizio soggettivo di chi effettua lo studio;
- È complesso confrontare queste tipologie di mappe, per la varietà di grandezze considerate in ogni studio e perché sito-specifiche;
- Le ricerche e le informazioni necessarie per ottenere un buon prodotto non possono prescindere dall'acquisizione di informazioni reperibili solo tramite indagini in situ e sopralluoghi.

A partire dagli anni 90' un numero crescente di autori hanno iniziato a servirsi delle reti neurali per la creazione di mappe di suscettibilità, in cui il giudizio dell'esperto è stato sostituito nell'assegnazione dei pesi dall'utilizzo di algoritmi di machine learning.

Ermini et al. (2005) hanno creato una mappa di suscettibilità per il sottobacino del Riomaggiore (che occupa la parte settentrionale del bacino del Reno) nella catena degli Appennini tra la Toscana e l'Emilia-Romagna. Per lo studio sono state utilizzate due diverse tipologie di reti neurali. La prima è una Multilayer Perceptions (MLP) con apprendimento supervisionato, la seconda è una Probabilistic Neural Networks (PNN), una rete a tre strati (input, strato radiale responsabile dell'apprendimento e output) ad apprendimento non supervisionato. I cinque parametri scelti per il training sono la litologia, l'uso del suolo, l'inclinazione del versante (espressa in gradi, da zero a novanta) la curvatura (quattro classi distinte in concavo, leggermente convesso, convesso, molto convesso) e il numero di celle drenate a monte (calcolate con strumenti GIS utilizzando algoritmo D8). Lo studio mostra come entrambe le reti presentate riescano a cogliere le relazioni tra i parametri utilizzati e la suscettibilità a franare dell'area di studio. I risultati ottenuti, in entrambi i casi sono molto buoni, mostrano che la rete MLP risulta avere performance superiori rispetto alla concorrente.

Melchiorre et al. (2008) hanno messo a punto una procedura di zonizzazione nel comune di Val Brembilla (Bergamo) ponendo particolare attenzione su come distinguere durante la fase di training della rete neurale le unità spaziali stabili da quelle che invece presentano una suscettibilità a franare. Secondo gli autori, infatti, mentre sono chiaramente definiti i punti che presentano la propensione a franare, non è così semplice definire i punti che possono essere considerati come stabili. Nello studio presentato vengono confrontate due procedure per distinguere i punti

instabili da quelli stabili. La prima tecnica il "Random Sampling", quella più comunemente utilizzata, considera come punti stabili quelli dal dataset dei punti non etichettati, la seconda tecnica è il "Cluster Sampling (CS)". Il CS è un algoritmo di riconoscimento non supervisionato più evoluto di quello presentato in precedenza, in cui i dati vengono raggruppati sulla base di proprietà comuni partendo da una inizializzazione random. La procedura si affina per step successivi basandosi sulla distanza dei centroidi, estratti in maniera randomica, considerando una metrica la euclidea. I parametri considerati per determinare la suscettibilità nell'area di studio sono sei, l'erosione da agenti atmosferici, l'uso del suolo, la pendenza, l'orientamento dei versanti, la permeabilità e il numero di celle drenate a monte. I due campioni vengono quindi processati da una rete neurale MLP costituita da sei neuroni di input un singolo hidden layer con quattordici neuroni e un singolo neurone di output. Sebbene entrambe le procedure abbiano portato a buoni risultati, le mappe ottenute utilizzando il CS forniscono risultati migliori in quanto Random Sampling, potrebbe considerare stabili unità territoriali che non lo sono (ad esempio a causa di notizie mancanti) compromettendo le capacità di distinguere le caratteristiche dei punti e il training della rete neurale.

Conforti et al. (2014) presenta uno studio di suscettibilità nel bacino idrografico del fiume Tubolo, nella parte settentrionale della regione Calabria, utilizzando una rete neurale. Gli indicatori utilizzati sono dieci, litologia, distanza dalle superfici di rottura, rete di drenaggio, uso del suolo, inclinazione dei versanti (slope angle), orientamento dei versanti, elevazione, curvatura dei versanti (tre classi con una suddivisione in concavo, piatto, convesso), il topographic wetness index (TWI) un indice che mette in relazione l'effetto della topologia con gli effetti di ruscellamento superficiale e lo stream power index (SPI) che misura la capacità di erosione del ruscellamento superficiale. La rete creata è mostrata in Figura 3.5, si tratta di un MLP a 3 layer addestrata con un training supervisionato e caratterizzata da dieci neuroni di input, 21 neuroni nell'hidden layer e un neurone di output. Nello studio riportato, il numero di neuroni dello strato intermedio è stato ricavato con una regola empirica.

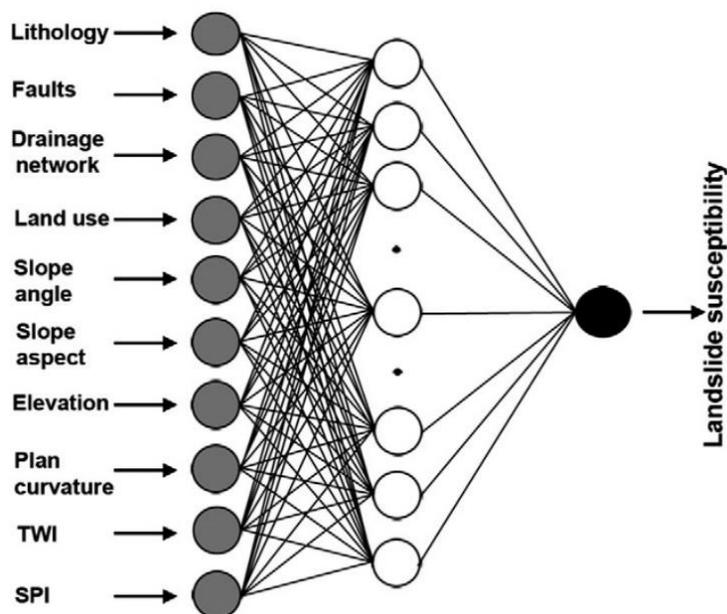


Figura 3.5: Rete neurale utilizzata da Conforti et al. (2014)

I risultati ottenuti in termini di previsione e robustezza del modello sono stati valutati tramite analisi ROC ottenendo risultati pari a 0.90. Considerando che il valore massimo, fissato pari ad 1, è attribuito ad un algoritmo che effettua solo previsioni perfette, questo studio conferma la validità dell'uso delle reti neurali nello sviluppo di mappe di suscettibilità.

Come accennato in precedenza, le reti neurali possono essere utilizzate anche per prevedere i movimenti franosi di versanti opportunamente strumentati. Infatti oltre ad un rischio diretto per le popolazioni, i movimenti lenti del terreno possono causare disagi come la creazione di superfici di rottura e/o subsidenza lungo gli assi viari o arrecare danni alle strutture presenti in zona. Partendo dai dati registrati da sensori (inclinometri, estensimetri, stazioni GPS, pluviometri, pozzi e/o piezometri di monitoraggio) e individuate quali sono le variabili più adatte a descrivere il problema, e quindi da monitorare, si cerca di prevedere gli sviluppi futuri.

Krkač et al., (2017) per prevedere i movimenti di una frana profonda a ridosso della città di Zagabria in Croazia, hanno utilizzato un algoritmo di machine learning chiamato "Random Forest" (Breiman, 2001) che permette di ottenere previsioni su scala mensile. Questo metodo computazionale permette di creare un gran numero di alberi decisionali indipendenti inizializzati in maniera randomica, che non

presentano sovrapposizioni tra le varie componenti. I parametri investigati sono stati principalmente la precipitazione, da cui sono stimati anche i dati di umidità del terreno, l'evapotraspirazione, le condizioni antecedenti di pioggia in varie condizioni climatiche, il livello della falda. Sono stanti anche considerati i movimenti e le velocità registrate in due anni di osservazioni. Il modello è stato valutato confrontando le previsioni con gli spostamenti. I risultati ottenuti permettono di ottenere buone previsioni sugli spostamenti in una finestra temporale di 30 giorni rappresentando una importante fonte di informazione per la valutazione del rischio e la gestione delle crisi.

Numerosi studi sul monitoraggio e lo spostamento di frane lente sono stati condotti in Cina nell'area della "Diga delle Tre Gole" sfruttando diversi algoritmi di machine learning (Cao et al., 2016; Miao et al., 2018), un'area con una elevata predisposizione a franare, a causa di un ambiente geomorfologico complesso e a precipitazioni molto intense, dove sono state censite più di 4200 frane (Yin et al., 2010). Per tale motivazione numerosi versanti sono strumentati al fine di condurre studi approfonditi. I parametri di input utilizzati per questo genere di applicazioni sono lo spostamento cumulato, il livello idrico del serbatoio o del corso d'acqua, la precipitazione, il livello della falda e gli spostamenti storici. In tutti i casi analizzati gli algoritmi hanno fornito previsioni accurate.

CAPITOLO 4 DESCRIZIONE DEI CASI DI STUDIO E COSTRUZIONE DEI DATASET

4.1 GENERALITÀ

Nel presente capitolo verranno descritte le aree di studio oggetto della ricerca presentata, cioè la regione Sicilia per l'Italia e l'area a est di Bergen per la Norvegia. L'individuazione dell'area di studio norvegese è stata dettata dall'esigenza di avere una grande mole di dati di frana e una estensione paragonabile a quella del caso studio italiano. La possibilità di paragonare due aree caratterizzate da un differente regime climatico permette di valutare la robustezza del sistema messo a punto. Saranno inoltre presentati i dati utilizzati per la costruzione dei database che sono utilizzati per le elaborazioni e le fonti da cui tali dati derivano.

4.2 CASO STUDIO IN ITALIA

L'area di studio relativa al territorio italiano è regione la Sicilia. La Sicilia ha una estensione pari a 25,711 km² ed è la più grande isola del mar Mediterraneo. Il territorio è principalmente collinoso e montuoso, la quota varia dal livello dal mare ai 3320 metri in corrispondenza del monte Etna, il più alto vulcano attivo d'Europa. Il clima è temperato-caldo, con estati calde e secche, soprattutto sulle coste meridionali, ed eventi di pioggia più frequenti durante i mesi invernali più freddi, nelle zone interne montuose. Le precipitazioni medie si aggirano attorno a 700 - 800 mm/anno, distribuite principalmente durante le stagioni autunnali e invernali. Le precipitazioni più abbondanti si registrano nella parte orientale dell'isola nel versante ionico tra il monte Etna e la catena dei monti Peloritani. Precipitazioni inferiori rispetto alla media si registrano nella parte più a sud dell'isola caratterizzata da quote minori ed esposta ai caldi venti provenienti dall'Africa (Alecci and Rossi, 2007). La mappa in Figura 4.1 mostra l'inquadramento geografico e come variano le quote nell'area di studio. I dati di precipitazione, alla scala oraria, utilizzati nell'analisi svolta provengono da 306 pluviometri, gestiti per il 69% dall'Osservatorio delle Acque (OdA), per il 28.4% dal Servizio Informativo Agrometeorologico Siciliano (SIAS) e per il 2.6% dal Dipartimento Regionale della Protezione Civile (DRPC). In Figura 4.1 è mostrata l'ubicazione dei pluviometri

disponibili per il periodo gennaio 2009–ottobre 2018 (cerchietti bianchi, i 211 strumenti gestiti dall' Oda) e quelli per cui si hanno informazioni solo per il periodo gennaio 2014–ottobre 2018 (cerchietti neri, 8 strumenti del DRPC, 87 del SIAS).

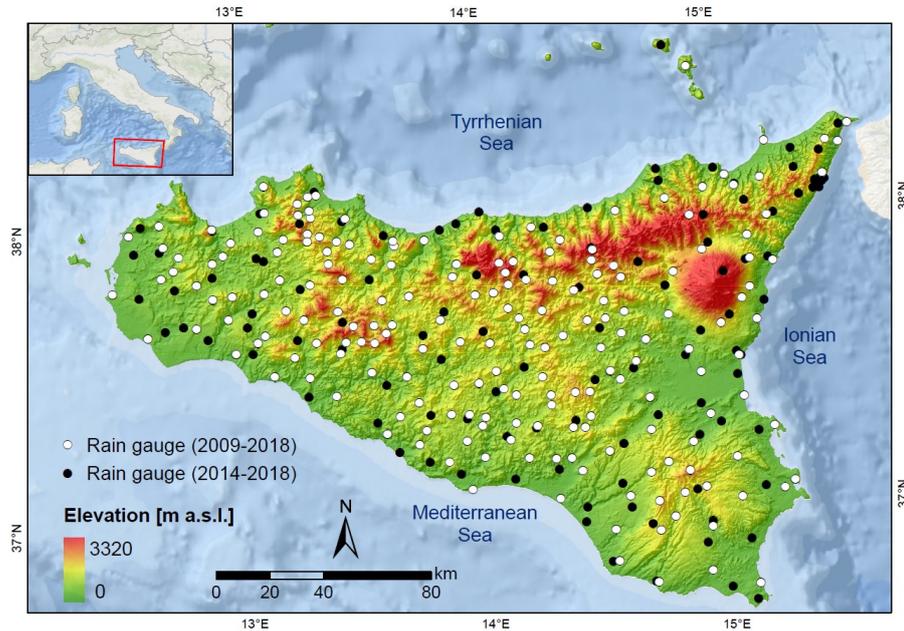


Figura 4.1: Inquadramento geografico e DEM della Sicilia e localizzazione dei pluviometri utilizzati in Sicilia

I dati di frana utilizzati provengono dal database “FraneItalia” realizzato da Calvello & Pecoraro (2018). Le informazioni, nella prima versione del catalogo, coprivano l’arco temporale compreso tra il mese di Gennaio del 2010 e Dicembre del 2017. Il catalogo continua ad essere aggiornato con cadenza annuale ed è consultabile al sito <https://franeitalia.wordpress.com/database/>, dove al momento sono disponibili dati per l’intero 2020. Esso tiene conto non solo delle frane innescate da pioggia ma anche di quelle innescate da cause antropiche e terremoti. Le informazioni per popolare il database provengono dal web; in particolare, la procedura messa a punto prevede sia l’utilizzo di Google Alert che la consultazione del Giornale della protezione civile (www.ilgiornaledellaprotezionecivile.it), nei quali vengono inserite le parole “frana” e “frane”. Ogni qual volta si riscontra una notizia non precedentemente immagazzinata, nel DB viene aggiunto un nuovo campo; in caso contrario si procede all’aggiornamento di notizie esistenti.

L'inventario distingue due tipologie di frane ovvero "frana singola" indicata dall'acronimo SLE (single landslide events) per i campi in cui è riportata una sola frana, e "franosità diffusa", indicata dall'acronimo ALE (areal landslide events), se sono riportati più fenomeni di dissesto. Le conseguenze delle frane sulla popolazione sono descritte da un codice alfanumerico, che distingue: frane che hanno provocato vittime o dispersi (C1), fenomeni di dissesto per cui sono stati registrati feriti o sfollati (C2), frane che non hanno avuto ripercussioni note sulla popolazione (C3). Una codifica simile è utilizzata anche per valutare la qualità delle informazioni sulla localizzazione della frana; nello specifico, con le sigle "Sd1", "Sd2", "Sd3", "Sd4" si indicano rispettivamente: posizione certa, posizione approssimata, comune in cui si è verificato l'evento, provincia in cui si è verificato l'evento. Una classificazione simile è utilizzata anche per identificare l'incertezza temporale dell'ora dell'innescò, sulla base di due soli classi: "Td1" (certo) e Td2 (stimato). Discorso analogo vale per il numero di frane nel caso di eventi diffusi: Nd1 (riportato) e Nd2 (stimato). Per le elaborazioni presentate sono state selezionate solo frane pluvio-indotte o che avessero come materiale mobilitato quelli riconducibili a frane innescate da precipitazione (ad esempio, sono stati eliminati crolli in roccia). Si è così arrivati ad un numero di frane compatibili con gli scopi dell'analisi da effettuare pari a 207.

Le caratteristiche delle frane individuate per la regione Sicilia sono presentate al paragrafo 4.5.

4.3 CASO STUDIO IN NORVEGIA

L'area studio in Norvegia è composta da 5 bacini situati a est di Bergen, sulla costa ovest del paese, ed è caratterizzata da una estensione di circa 18 mila km², con quote comprese tra lo zero e i mille-seicento m s.l.m., come mostrato in Figura 4.2. Le precipitazioni medie annuali per l'area di studio sono comprese tra i 3500-4700 mm (Meyer et al., 2012) considerando anche lo scioglimento nivale. La delimitazione dei bacini idrografici, utilizzati per individuare l'area di studio, è stata effettuata dal Norwegian Water Resources and Energy Directorate (NVE), un ufficio sotto l'amministrazione del "Ministero del Petrolio e dell'Energia". Tale ente, responsabile della gestione delle risorse idriche ed energetiche della Norvegia,

opera anche per ridurre il rischio di danni associati a frane e inondazioni. Attraverso tool GIS è stato possibile inquadrare l'area di studio, raggruppare in un unico poligono i cinque bacini selezionati sulla base del numero di frane disponibili, e valutare l'estensione dell'area di studio.

I dati di precipitazione utilizzati per le elaborazioni provengono dal database "seNorge_2018" (Lussana et al., 2019), che copre un intervallo temporale che spazia dal 1957 al 2020; tuttavia sono stati utilizzati solo i dati di precipitazione compresi nell'intervallo 2000-2020 in quanto le informazioni sulle frane coprono tale periodo (https://thredds.met.no/thredds/catalog/senorge/seNorge_2018/catalog.html, ultimo accesso 28/02/22). I dati di precipitazione, a scala giornaliera, sono disponibili per tutto il territorio norvegese come GRID con maglia di 1km x 1km. L'introduzione di dati alla scala giornaliera introduce una ulteriore incertezza nella determinazione della soglia come dimostrato dallo studio Marra (2019). Al fine di semplificare le analisi, si è ridotto in numero di celle da analizzare scegliendone 175, i quali sono posizionati su una griglia regolare, con una distanza tra ogni singolo pluviometro pari a 5km (Figura 4.2)

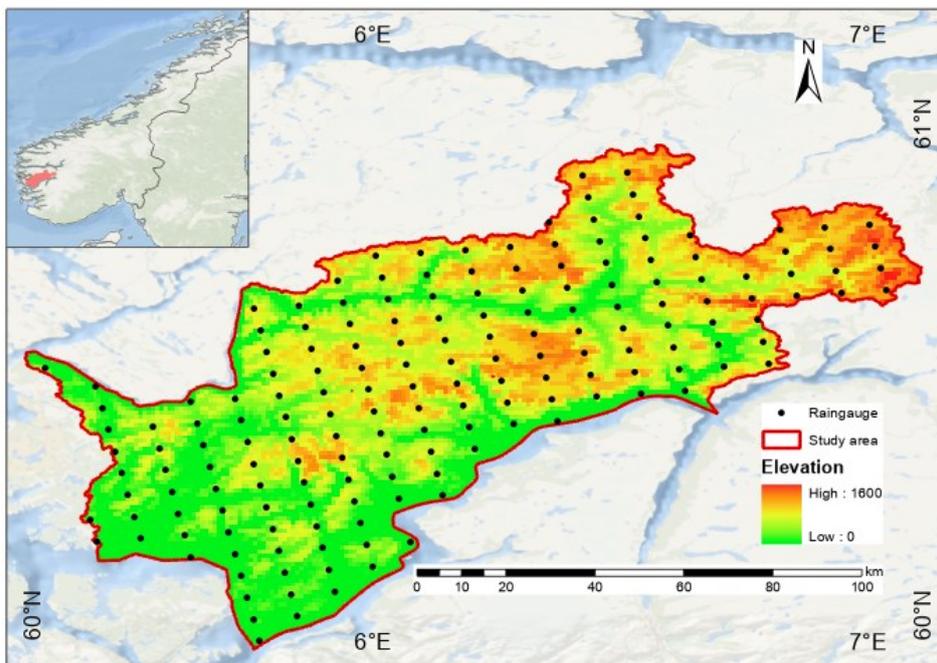


Figura 4.2: Disposizione dei pluviometri fittizi nell'area di studio norvegese

Le informazioni sulle frane nell'area di studio questi provengono dal database nazionale messo a disposizione dal NVE (<https://nedlasting.nve.no/gis/>, ultimo accesso 31/01/2022). Utilizzando gli strati informativi dell'NVE, è stato possibile selezionare dal dB di frane, contenente 73528 istanze, quelle pluvio-indotte, nello specifico: loose material (LM), debris flow (DF) and debris slide (DS). Sono state quindi selezionate le frane avvenute nell'intervallo temporale dei dati di precipitazione (2000÷2020), con incertezza spaziale inferiore ai 5km e temporale inferiore alle 24 ore. Per le elaborazioni il numero di frane utilizzabili con le caratteristiche richieste è sceso a 298, di cui 147 LM, 51 DF e 100 DS.

4.4 DATI DI UMIDITÀ DEL SUOLO DERIVANTI DAL PROGETTO DI RIANALISI ERA5-LAND

L'umidità del suolo è una variabile influenzata in maniera complessa da tutte le componenti del ciclo idrologico (Jung et al., 2010; Crow et al., 2012). Questa informazione può essere utilizzata, ad esempio, per la gestione dei serbatoi, la programmazione dell'irrigazione, la previsione della resa delle colture e per meglio capire le condizioni di innesco (Marino et al., 2020; Mirus et al., 2018a, 2018b) o all'interno dei sistemi di early warning dei movimenti franosi (Calvello et al., 2014; Piciullo et al., 2018).

Tre sono i possibili approcci per determinare misure di umidità del suolo:

1. Osservazioni in situ;
2. Dati derivanti da telerilevamento (Brocca et al., 2016);
3. Progetti di Rianalisi (Palazzolo et al., 2021; Reder and Rianna, 2021).

I dati di umidità del suolo utilizzati in questo lavoro di tesi derivano dal database di rianalisi ERA-5Land (Muñoz-Sabater et al., 2021) forniti dall' European Centre for Medium-Range Weather Forecasts. Le caratteristiche del database ERA5-Land sono mostrate in Tabella 4.1. I dati ERA5 già utilizzati in alcuni studi (Piciullo and Gilbert, 2022; Palazzolo et al., 2021) hanno mostrato che i dati di umidità del suolo possono essere usati come proxy per meglio capire le condizioni che portano all'innesco, ma che presentano alcuni limiti di accuratezza, come rivelato dal confronto con misure in situ (Reder and Rianna, 2021, Wicki et al., 2020; 2021).

I dati ottenuti tramite la tecnica della rianalisi forniscono una descrizione accurata del passato utilizzando le osservazioni atmosferiche provenienti da diverse

fonti e rilevate sulla superficie terrestre. Queste informazioni vengono utilizzate come input nei più avanzati modelli numerici di previsione meteorologica fisicamente basati, per creare un set di dati che copre diversi decenni indietro nel tempo e che è coerente con i dati registrati (Soci et al., 2016). Una schematizzazione del modello è presentata in Figura 4.3.

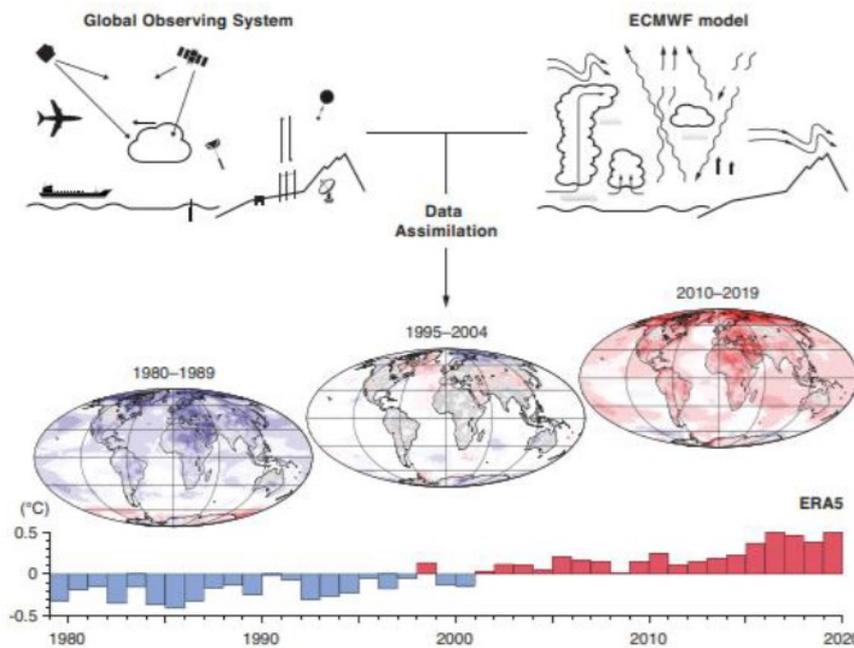


Figura 4.3: Schematizzazione della generazione dei dati di rianalisi di ERA5-Land (modello ECMWF)

Ad oggi, i dati di rianalisi forniscono il quadro più completo attualmente possibile nel tempo, del clima del passato. Sono generalmente disponibili gratuitamente su piattaforme dedicate del web, rilasciati in formato grid, con un ritardo di alcuni mesi dal presente (Pelosi et al., 2020).

I dati di umidità del suolo ottenuti dal dB Era5-Land sono stati confrontati con quelli registrati da stazioni di monitoraggio. Per il caso studio in Norvegia i dati di umidità provengono da 4 differenti stazioni di monitoraggio, una ubicata a Eidsvoll e gestita dal Norwegian Geotechnical Institute (NGI), le altre in corrispondenza del fiume Hobøl (sud-est della Norvegia) in prossimità di differenti tipi di vegetazione, erba, arbusti, e alberi (Krzeminska et al., 2019). Per la regione Sicilia non si dispone di stazioni di monitoraggio dei dati di umidità del suolo, per cui sono stati utilizzati

i dati della regione Calabria, la più prossima sia geograficamente che per clima utilizzando i dati dell' "International Soil Moisture Network (ISMN)" (Dorigo et al., 2011; 2021) (ultimo accesso 22 novembre 2021).

Tabella 4.1: Caratteristiche del database ERA5-Land

Descrizione	Caratteristiche
Tipo di dati	Gridded
Proiezione	Griglia regolare su latitudine e longitudine
Copertura orizzontale	Globale
Risoluzione orizzontale	0.1° x 0.1°; Risoluzione di 9 km.
Copertura verticale	Da 2 metri sul piano campagna a una profondità del suolo di 289 cm
Risoluzione verticale	4 livelli: primo strato 0-7 cm; secondo strato 7-28cm; terzo strato 28-100 cm; quarto strato 100-289cm
Copertura temporale	Da Gennaio 1950 al presente
Risoluzione temporale	Oraria
Formato file	GRIB
Frequenza di aggiornamento	Mensile con un ritardo di circa tre mesi

I dati provenienti dai dati di Rianalisi di ERA5-Land utilizzati per le elaborazioni nel seguito illustrate sono riportati in Tabella 4.2.

I dati di umidità del suolo registrati da strumenti e stazioni di monitoraggio e i quelli del progetto ERA5-Land sono stati confrontati al fine di verificare con dettaglio la loro accuratezza (v. Appendice A). Per la regione Calabria (utilizzata per i confronti come sostituta della regione Sicilia dove non sono disponibili dati misurati di umidità del suolo) è possibile osservare che i dati di umidità misurati e quelli di rianalisi seguono il medesimo trend, sebbene sia possibile osservare un offset medio tra le serie con valori compresi di $0.1 \div 0.2 \text{ m}^3 \cdot \text{m}^{-3}$. Le elaborazioni e i confronti effettuati per la Norvegia mostrano un migliore adattamento dei dati registrati e quelli di rianalisi, tranne che nei periodi invernali, dove il congelamento del suolo può giocare un ruolo fondamentale.

Sebbene, talvolta, le differenze tra i dati misurati e quelli restituiti dal dB ERA5 non siano trascurabili, la possibilità di utilizzare un database a scala globale che permette il confronto tra le due aree di studio, e che utilizza la medesima

metodologia per la determinazione dei valori di umidità del suolo rappresenta la scelta migliore considerando l'estensione delle aree esaminate.

I confronti grafici e i diagrammi di Taylor (Taylor, 2001), indicano che la migliore corrispondenza tra dati in situ e dati di rianalisi è quella relativo al terzo strato di suolo per tutte le area analizzate. I risultati sono riportati in Appendice A.

Tabella 4.2: Grandezze del database ERA5-Land utilizzate per le elaborazioni

Grandezza	Unità di misura	Descrizione
Contenuto volumetrico d'acqua del primo livello	$m^3 m^{-3}$	Volume d'acqua nel primo strato di suolo (0-7 cm) del Sistema Integrato di Previsione ECMWF. La superficie è a 0 cm. Il volume d'acqua del suolo è associato alla struttura del suolo (o classificazione), alla profondità del suolo e al livello della falda freatica sottostante.
Contenuto volumetrico d'acqua del secondo livello	$m^3 m^{-3}$	Volume d'acqua nel secondo strato di suolo 2 (7 -28 cm) del Sistema Integrato di Previsione ECMWF
Contenuto volumetrico d'acqua del terzo livello	$m^3 m^{-3}$	Volume d'acqua nel terzo strato di suolo (28 - 100 cm) del Sistema Integrato di Previsione ECMWF
Contenuto volumetrico d'acqua del quarto livello	$m^3 m^{-3}$	Volume d'acqua nel quarto strato di suolo (100 -289 cm) del Sistema Integrato di Previsione ECMWF

4.5 COSTRUZIONE DEI DATASET

Partendo dai dati presentati nei precedenti paragrafi è stato necessario ricostruire gli eventi pluviometrici innescanti e non innescanti. Al fine di utilizzare criteri replicabili e oggettivi per l'identificazione degli eventi pluviometrici, Melillo et al. (2015, 2016, 2018) hanno proposto una procedura automatizzata che restituisce soglie pluviometriche utilizzando un codice in linguaggio R. Il software messo a punto è identificato (nella sua ultima versione) dall'acronimo CTRL-T (Calculation of Thresholds for Rainfall-induced Landslides-Tool). Le operazioni eseguite dall'algoritmo sono mostrate nello schema a blocchi, riportato in Figura 4.4.

In particolare, l'algoritmo utilizza come dati di input, serie di precipitazioni ad una determinata risoluzione temporale (tipicamente oraria) e informazioni sulle frane nell'area di studio. È prevista l'impostazione di diversi parametri climatici e spaziali. Si può distinguere tra una stagione calda in (cw) in cui due differenti precipitazioni sono separate da un periodo Pw (espresso in ore), e una stagione fredda in un (cc) in cui le precipitazioni consecutive sono separate da un periodo Pc . È possibile impostare la sensibilità del pluviometro (gs) e l'estensione dell'intorno circolare entro cui è possibile considerare un pluviometro prossimo alla frana (Rb). I parametri degli eventi piovosi sono stati calibrati adottando il modello mensile di bilancio idrico del suolo e l'analisi dell'evapotraspirazione (Thornthwaite, 1948; Thornthwaite & Mather, 1957) per il caso studio italiano. Considerando la maggiore piovosità del caso studio norvegese empiricamente sono stati dimezzati i tempi di inter-storm. Un ulteriore modulo, invece, effettua la selezione del pluviometro più rappresentativo per la frana. La distanza massima consentita tra una frana e un pluviometro è limitata dall'area circolare con raggio pari a Rb . Vengono quindi identificate le condizioni di pioggia (MRC) che sono molto probabilmente responsabile dell'innescamento. L'MRC può essere una coppia (D_L , E_L) di durata dell'evento piovoso (D_L) e precipitazione cumulata dell'evento (E_L) o un insieme di due o più coppie. Ad ogni MRC viene assegnato un peso per selezionare il pluviometro rappresentativo e le condizioni di pioggia associate alla frana. Il peso è inversamente proporzionale al quadrato della distanza tra il pluviometro e la frana, e proporzionale al prodotto tra la cumulata e l'intensità media delle precipitazioni. La condizione MRC con il peso maggiore è quella che viene identificata come

innescante e viene promossa a MPRC (condizione di pioggia maggiormente probabile).

L'algoritmo fornisce successivamente soglie pluviometriche di innesco in termini di cumulata di pioggia-durata (E-D) utilizzando il metodo frequentista (Guzzetti et al., 2007; 2008; Brunetti et al., 2010), basato solo sugli eventi innescanti e su una fissata probabilità di non superamento degli stessi. La probabilità è valutata attraverso la tecnica del "bootstrap", una procedura statistica non parametrica che permette di associare un termine di incertezza $\Delta\alpha$ e un secondo $\Delta\gamma$ all'intercetta (α) e al parametro di forma (γ).

La tecnica del "bootstrap" genera k serie di dati composti da un numero m di istanze sul totale di n eventi presenti dal database da cui si vuole estrapolare la soglia di innesco. L'analisi delle k serie sintetiche permette di calcolare i valori medi associati ai parametri includendo anche la deviazione standard σ (Peruccacci et al., 2012). La relazione (4) può quindi essere riscritta come:

$$E = (\alpha \pm \Delta\alpha) \cdot D^{(\gamma \pm \Delta\gamma)} \quad (7)$$

L'algoritmo dopo essere stato lanciato restituisce una lista di tutte le precipitazioni avvenute nell'area di studio, tutte le condizioni responsabili dell'innescamento di movimenti franosi (MRC e MPRC) caratterizzandole in termini di durata, precipitazione cumulata, intensità media e intensità di picco. L'algoritmo è stato utilizzato per determinare le condizioni di innesco dei movimenti franosi per entrambe le aree di studio utilizzando i parametri riportati in Tabella 4.3 per il caso studio italiano e in Tabella 4.4 per il caso studio norvegese.

Tabella 4.3: Valori utilizzati nell'algoritmo CTRL-T per il caso studio italiano

Variabile	Simbolo	Unità di misura	Valore
Sensibilità dello strumento	gs	mm	0.1
Periodo asciutto nella stagione calda	Pw	h	96
Periodi asciutto nella stagione fredda	Pc	h	48
Intorno circolare del buffer del pluviometro	Rb	km	16

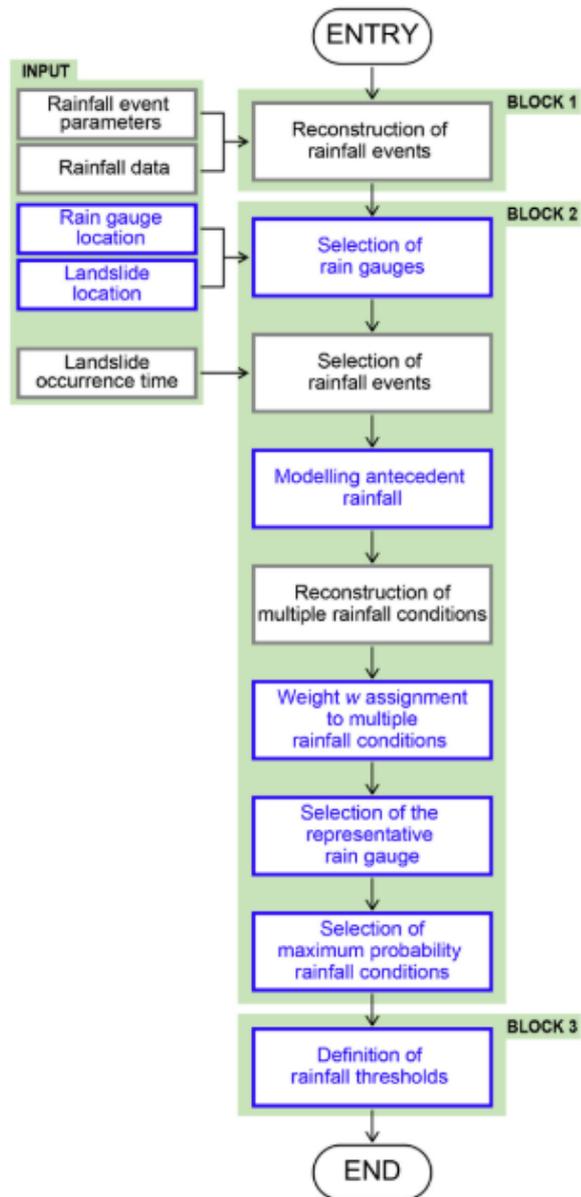


Figura 4.4: Diagramma a blocchi del software CTRL-T (Melillo et al., 2018)

Tabella 4.4: Valori utilizzati nell'algoritmo CTRL-T per il caso studio norvegese

Variabile	Simbolo	Unità di misura	Valore
Sensibilità dello strumento	gs	mm	0.1
Periodo asciutto nella stagione calda	Pw	h	48
Periodi asciutto nella stagione fredda	Pc	h	24
Intorno circolare del buffer del pluviometro	Rb	km	16

L'utilizzo del software ha permesso di ricostruire le condizioni di innesco di 144 delle 207 frane all'interno del territorio siciliano. Per 126 (87.5%) di queste non si dispone di informazioni sulla tipologia di movimento franoso. La tipologia dei rimanenti eventi è distribuita come segue: 10 (6.9%) crolli in eventi franosi areali (fall), 4 (5.6%) colate (flow) e 4 scivolamenti (slide). Per quanto riguarda l'accuratezza temporale dei dati di frana, per 103 eventi è noto solo il giorno di innesco, mentre per i restanti è disponibile un'indicazione più precisa dell'istante di distacco. Nel primo caso l'istante di innesco è stato attribuito alla fine della giornata, nel secondo caso al picco di pioggia nell'intervallo di tempo in cui si è verificato l'innesco. L'ubicazione delle frane, per cui sono state determinate dall'algoritmo le condizioni innescanti in termini di parametri idrologici, è riportata in Figura 4.5.

Per il caso studio in Norvegia l'algoritmo ha permesso di ricostruire 247 precipitazioni innescanti di cui 117 loose material, 48 debris flow e 82 debris slide. Essendo il dato di precipitazione a scala giornaliera e il dato della cumulata è riferito al giorno seguente, l'innesco della frana è stato attribuito all'ora per cui si dispone del valore di cumulata di precipitazione cioè alle 06:00 AM. L'ubicazione delle frane ricostruite è riportata nella mappa presentata in Figura 4.6.

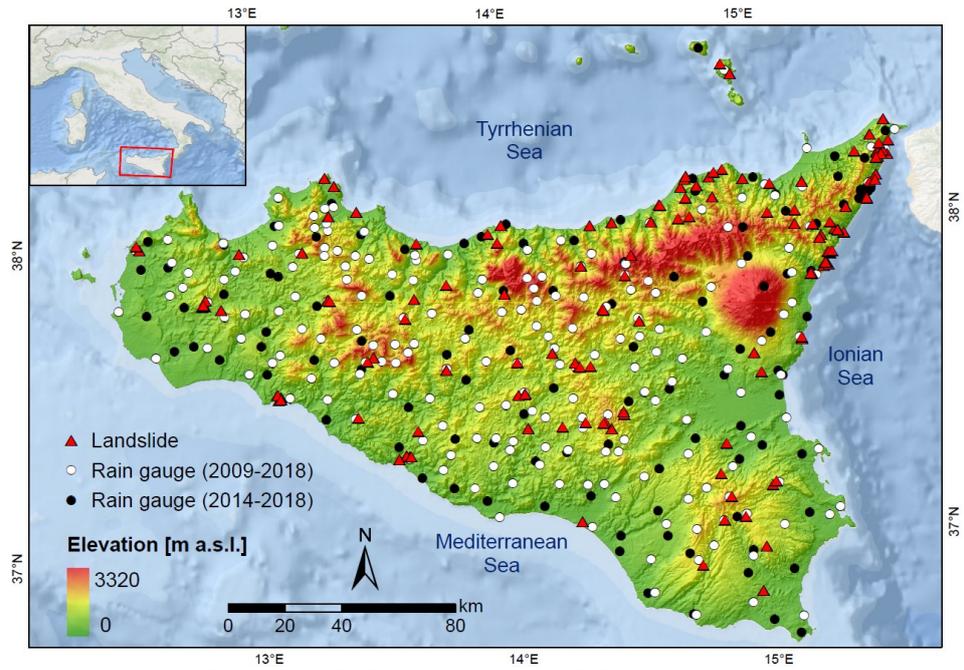


Figura 4.5: Ubicazione delle frane ricostruite dall' algoritmo CTRL-T per la Sicilia

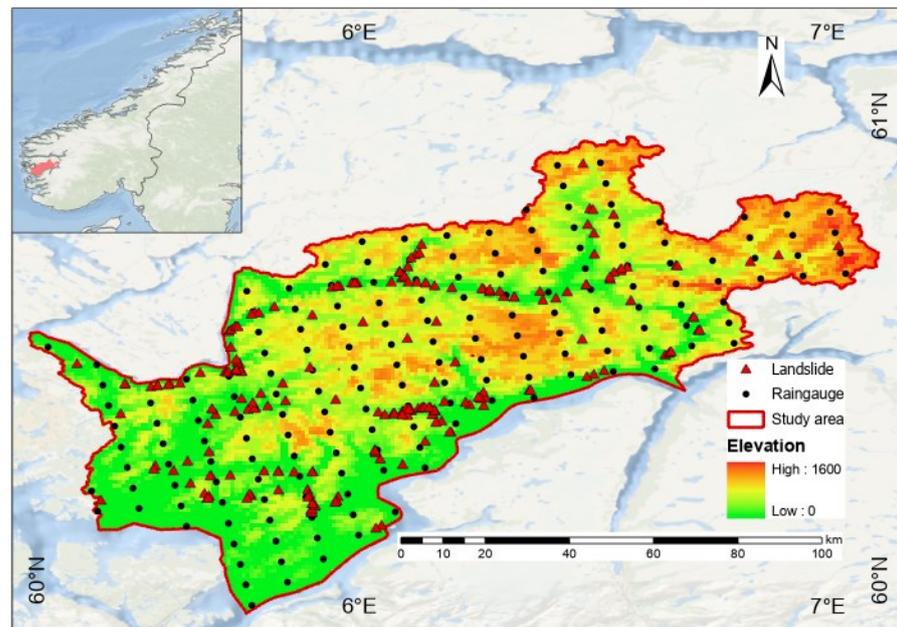


Figura 4.6: Ubicazione delle frane ricostruite dall' algoritmo CTRL-T per i bacini Norvegesi

CAPITOLO 5 SOGLIE PLUVIOMETRICHE DI INNESCO BASATE SU RETI NEURALI ARTIFICIALI

5.1 GENERALITÀ

In questo capitolo verranno presentate le soglie pluviometriche empiriche determinate per la regione Sicilia e il caso studio norvegese. Sulla base dei dati presentati al capitolo precedente; quindi, verranno effettuati dei confronti con altre soglie determinate per le due aree di studio esaminate. Per entrambi i casi verrà presentata una applicazione che mostra le capacità delle reti neurali artificiali di riconoscere eventi innescanti utilizzando diverse configurazioni di input che tengono conto delle caratteristiche della precipitazione come la durata, la precipitazione cumulata, l'intensità media e l'intensità di picco. Tutti i risultati ottenuti per la Sicilia sono stati pubblicati sulla rivista "Natural hazard and Earth System Sciences" come: "Distefano, P., Peres, D. J., Scandura, P., and Cancelliere, A.: Brief communication: Introducing rainfall thresholds for landslide triggering based on artificial neural networks, *Nat. Hazards Earth Syst. Sci.*, 22, 1151–1157, <https://doi.org/10.5194/nhess-22-1151-2022>, 2022."

5.2 INTRODUZIONE

Il territorio siciliano, prevalentemente collinare e montuoso, nel passato è stato colpito da eventi di frana che in alcuni casi hanno provocato vittime tra la popolazione. Uno dei casi più drammatici è la frana che il primo Ottobre 2009 ha investito il comune di Giampileri e le zone limitrofe innescata da circa 220 mm di pioggia caduti in un intervallo di 7 ore (Ardizzone et al., 2012).

Nel tempo diversi autori si sono occupati della determinazione di soglie pluviometriche di innesco per il territorio siciliano. Possono, ad esempio, essere citati gli studi di Gariano et al. (2015) e Melillo et al. (2016) che determinano le soglie descritte rispettivamente dalle relazioni (8) e (9), considerando una probabilità di non superamento del 5%

$$T_{G5\%} : E = 10.4 \cdot D^{0.27} \quad (8)$$

$$T_{M5\%} : E = 5.6 \cdot D^{0.40} \quad (9)$$

Le soglie presentate sono entrambe derivate considerando l'intervallo temporale che copre il periodo Luglio 2002– Dicembre 2012, che è diverso da quello che abbiamo considerato nella nostra analisi (2009-2018). La prima soglia è stata derivata con una versione precedente del codice CTRL-T, che richiedeva la selezione manuale del pluviometro più rappresentativo (Melillo et al., 2015), mentre il secondo studio deriva dall'algoritmo aggiornato, dove questa selezione viene effettuata automaticamente.

In generale, per un dato insieme di predittori, la scelta di una forma di equazione di soglia predeterminata, come quella di tipo potenza, può potenzialmente limitare le prestazioni. Il contenuto informativo delle variabili predittive considerate, potrebbe non essere sfruttato appieno. Ciò vale, a maggior ragione, quando si cercano variabili alternative (a quelle comunemente usate come intensità o durata) o aggiuntive con l'obiettivo di migliorare le prestazioni delle soglie, come le condizioni di pioggia antecedenti (Glade et al., 2000b), i livelli di falda e i dati sull'umidità del suolo (Bogaard e Greco, 2018; Marino et al., 2020). Per il caso delle soglie cumulata–durata o intensità–durata è consuetudine l'uso di una legge di potenza. La sua logica è stata verificata anche sulla base di un approccio combinato stocastico e fisicamente-basato (Peres e Cancelliere, 2014). Al contrario, nel caso di una coppia diversa di variabili o dell'analisi di più di due variabili, non esiste un'analoga forma parametrica prominente dell'equazione di soglia. Ad esempio, come riportato da Conrad et al. (2021), sono state prese in considerazione formule alternative per soglie idrometeorologiche – ovvero basate su precipitazioni e umidità del suolo o capacità di invaso del bacino – comprese funzioni lineari e bilineari, segmenti di linee interpolati senza funzione matematica e operatori logici più complessi. L'uso di forme parametriche predeterminate può infine mettere a repentaglio la fondatezza scientifica dei confronti tra diverse tipologie di soglie (soglie I-D e soglie idro-meteorologiche) utilizzando diversi insiemi di predittori (ad es. soglie di precipitazione rispetto a soglie che utilizzano l'umidità del suolo). Le reti neurali artificiali (ANN), appartenenti a tecniche di intelligenza artificiale, consentono di rimuovere potenzialmente la citata limitazione in quanto sono approssimatori universali, cioè in grado di riprodurre qualsiasi funzione continua (Haykin, 1999).

Le potenzialità delle reti neurali artificiali sono utilizzate in molti studi incentrati sulla mappatura della suscettibilità e sull'instabilità individuale dei pendii (Reichenbach et al., 2018). In altri studi, l'attenzione si concentra sulla previsione

degli spostamenti individuali di frane profonde mediante algoritmi di apprendimento automatico utilizzando dati ricavati in situ (van Natijne et al., 2020).

5.3 DATI E SOGLIE EMPIRICHE PER LA REGIONE SICILIA

Il caso studio è la regione Sicilia una delle venti regioni d'Italia. I dati di precipitazione a scala oraria derivano dagli strumenti dell'Osservatorio delle Acque (Osservatorio delle Acque, OdA), dal SIAS (Servizio Informativo Agrometeorologico Siciliano) e dal Dipartimento Regionale della Protezione Civile (DRPC) mentre i dati di frana provengono dal database "FraneItalia" compilato da Calvello and Pecoraro (2018). L'area di studio è mostrata in Figura 4.5, l'analisi si basa sul periodo da gennaio 2010 - ottobre 2018, durante il quale sono disponibili informazioni sia sulle precipitazioni che sulle frane. Sono stati considerati solo gli eventi che hanno come causa di innesco le voci "pioggia" o "pioggia e altro" in modo da escludere gli smottamenti dovuti a terremoti e ad attività antropiche. Fenomeni come la caduta di massi sono stati rimossi dalle analisi perché non sempre possono essere collegate alle precipitazioni. I dati sulle precipitazioni sono stati controllati al fine di rimuovere i dati sospetti. In particolare, sono stati considerati dati sospetti le precipitazioni orarie maggiori di 250 mm - corrispondenti a circa un terzo della piovosità media annua per la Sicilia, maggiori di circa 2 volte la piovosità massima mai registrata in 1 h. Per le serie ispezionate visivamente quando si riscontrava un errore evidente (ad esempio un malfunzionamento del pluviometro), l'intero evento piovoso che circonda il picco è stato rimosso. Alla luce di quanto sopra, in Figura 5.1 è mostrato un diagramma di flusso che rappresenta la metodologia applicata.

I dati pre-processati sono stati utilizzati come input per l'algoritmo CTRL-T (Melillo et al., 2018) che ha ricostruito le precipitazioni innescanti e non innescanti caratterizzandole in termini di durata, precipitazione cumulata, intensità media e intensità di picco. Le condizioni di pioggia più probabili associate a ciascun evento di frana (nel caso in cui sono disponibili più pluviometri per una data località per una determinata frana) sono calcolate dal software in base alla distanza tra il pluviometro e la posizione della frana nonché le caratteristiche dell'evento pluviometrico ricostruito. In particolare, per una data frana, vengono ricercati tutti i pluviometri entro un intorno circolare, di raggio $R_b = 16$ km, specificato dall'utente.

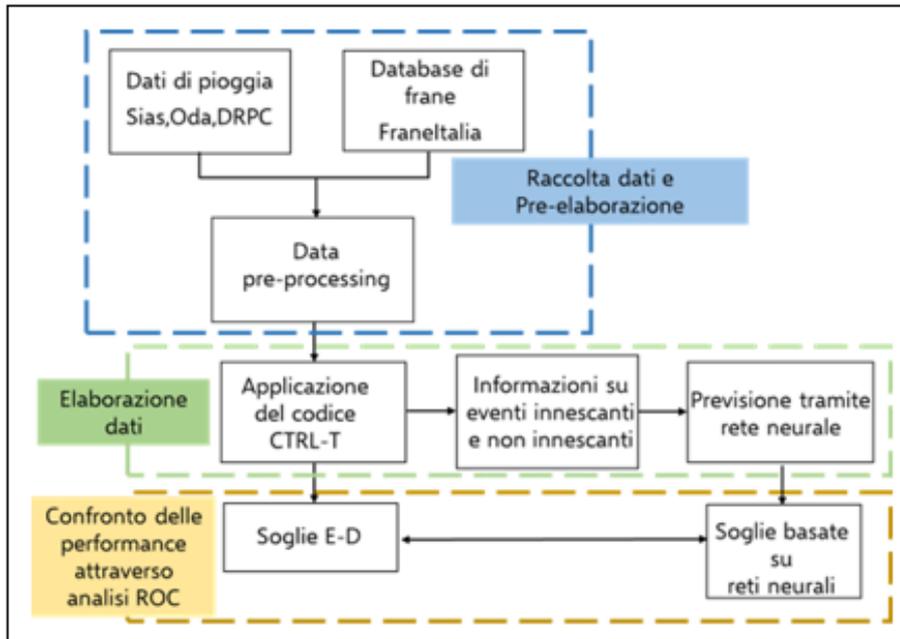


Figura 5.1: Metodologia proposta

Quando all'interno dell'intorno è presente più di un pluviometro, gli eventi pluviometrici di ciascun pluviometro vengono ponderati in base alla distanza pluviometro-frana e caratteristiche dell'evento piovoso (precipitazioni cumulate e durata). Il peso viene utilizzato per stimare la "probabilità" associata a ciascuna condizione di precipitazione potenzialmente attribuibile a ciascun evento franoso. In particolare, la probabilità, nel caso di più condizioni di pioggia, viene calcolata dividendo ciascun peso per la somma. CTRL-T determina quindi le condizioni di innesco delle precipitazioni di ciascuna frana come quelle corrispondenti alla più alta probabilità. L'applicazione dell'algoritmo ha permesso di ricostruire 144 frane su 207 frane disponibili e le relative precipitazioni innescanti e 47398 precipitazioni non innescanti. La distanza media tra pluviometro e frana ricostruita è pari a 5 km. Infine, il codice fornisce soglie E-D della legge di potenza per diversi livelli di frequenza di non superamento degli eventi scatenanti, quella ottenuta considerando una probabilità di non superamento del 5% è riportata di seguito (10).

$$T_{D5\%}: E = 4.9 \cdot D^{0.26} \quad (10)$$

I confronti tra le soglie di Gariano et al. (2015) e Melillo et al. (2016) (Eq. 8 e 9) e quella determinata in Distefano et al. (2022) (Eq 10) sono presentati in Figura 5.2 e

Figura 5.3, dove è possibile osservare che i risultati delle Eq. (9) e (10) sono molto simili invece i valori della soglia determinata dall'Eq (8) risultano essere più elevati.

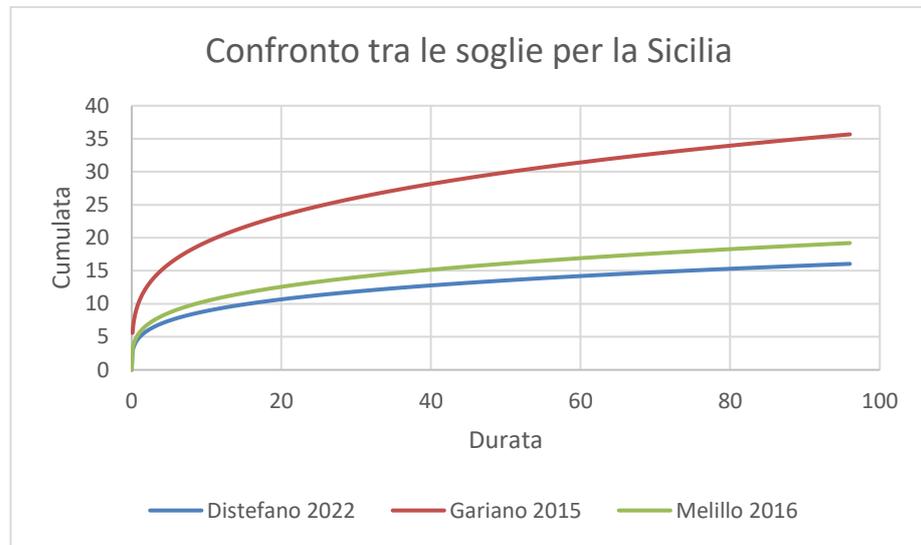


Figura 5.2: Confronto tra le soglie pluviometriche per la Sicilia nel piano cartesiano

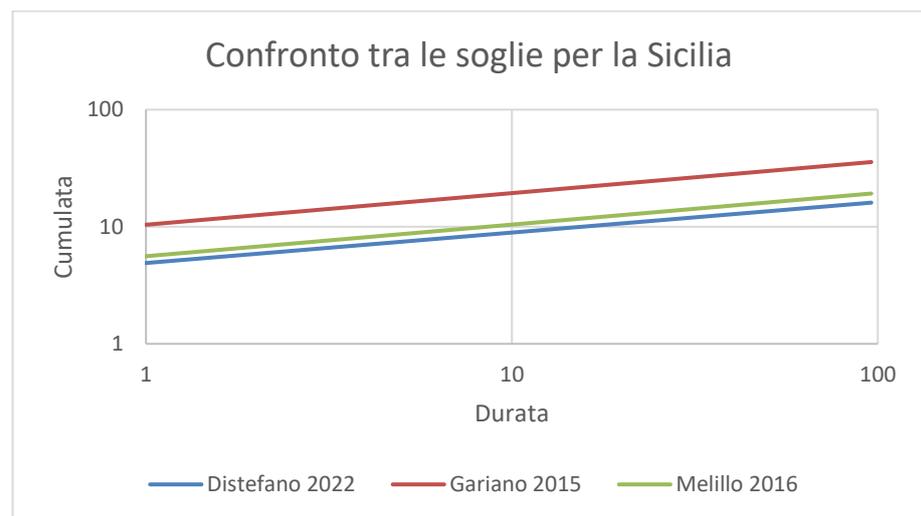


Figura 5.3: Confronto tra le soglie pluviometriche per la Sicilia nel piano bi-logaritmico

5.4 APPLICAZIONE DELLE RETI NEURALI ARTIFICIALI

Le caratteristiche idrologiche degli eventi di precipitazione innescanti e non innescanti restituite dall'algoritmo CTRL-T sono state utilizzate come input all'interno di una rete neurale ideata per il riconoscimento di pattern, in MATLAB®. La rete neurale, caratterizzata da una struttura feed-forward, è composta da tre strati: input, hidden e output (Figura 5.4).

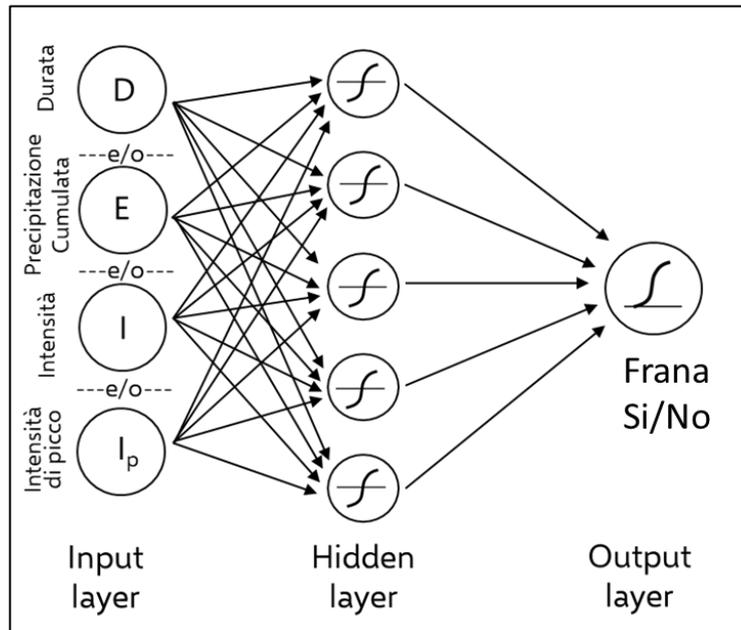


Figura 5.4: Schema concettuale della rete neurale artificiale.

Lo strato di input trasferisce i dati in ingresso al livello nascosto, dove le serie vengono combinate e trasformate attraverso una specifica funzione di attivazione. Sono state considerate due diverse funzioni di attivazione: una funzione tan-sigmoide $f(n)$ (Eq.11) per lo strato nascosto e una log-sigmoide $g(n)$ (Eq.12) per lo strato di output:

$$f(n) = \frac{2}{(1-e^{-2n})} \quad (11)$$

$$g(n) = \frac{1}{(1-e^{-n})} \quad (12)$$

La rete neurale artificiale è stata addestrata utilizzando un l’algoritmo chiamato “scaled conjugate gradient backpropagation”, mentre la funzione utilizzata per il training è chiamato “cross-entropy”.

Indicando con y_i l’output generico della rete neurale (con valori nell’intervallo aperto tra 0 e 1) e con t_i il target, dove $i = 1, 2, \dots, N$, la funzione “cross-entropy” F penalizza pesantemente le previsioni imprecise e assegna sanzioni minime per le previsioni corrette:

$$F = \frac{1}{N} \sum_{i=1}^n [t_i \cdot \log(y_i) + (1 - y_i) \cdot \log(1 - y_i)] \quad (13)$$

La capacità della rete di distinguere gli eventi innescanti da quelli non innescanti viene effettuata utilizzando una matrice di confusione distinguendo i “veri positivi” TP (eventi franosi correttamente classificati) i “veri negativi” TN (eventi non innescanti correttamente classificati), i “falsi positivi” FP (eventi non innescanti classificati come innescanti) e i “falsi negativi” FN (eventi innescanti classificati non innescanti). Utilizzando le statistiche che compongono la matrice di confusione, è possibile definire il “tasso di veri positivi” TPR (Eq. 14) il “tasso di falsi negativi” FPR (Eq. 15) e il *true skill statistic* TSS (Eq. 16) per valutare le prestazioni predittive della rete neurale.

Tabella 5.1: Matrice di confusione

	Evento	
	Previsione	TP
	FP	TN

$$TPR = \frac{TP}{(TP+FN)} \quad (14)$$

$$FPR = \frac{FP}{(TN+FP)} \quad (15)$$

$$TSS = TPR - FPR \quad (16)$$

L’output della rete viene trasformato in codice binario assumendo un valore pari a 1 (la rete prevede una frana) quando l’uscita è maggiore di un valore di soglia e un valore pari a 0 in caso contrario. Quindi identifichiamo il valore soglia menzionato massimizzando il TSS

Le analisi effettuate hanno portato al test di diverse combinazioni di dati in input (durata D , intensità I , precipitazione cumulata E e intensità di picco I_p) e diverse architetture, modificando il numero di neuroni nello strato nascosto. In particolare, sono state studiate le seguenti configurazioni di variabili di input: (1) D ; (2) E ; (3) I ; (4) I_p ; (5) D & E ; (6) D & I ; (7) D & I_p ; (8) E & I_p ; (9) I e I_p ; (10) D , E & I_p .

Le configurazioni di ingresso elencate sono infatti tutte possibili, tranne quelle che combinano sia E che I con la durata D . Ciò è stato fatto perché le due coppie D - I e D - E hanno lo stesso contenuto informativo per costruzione. Questo è confermato dal fatto che le prestazioni delle reti neurali D - I e D - E non differiscono in modo significativo (Tabella 5.2). Possono verificarsi lievi differenze poiché le ANN possono essere sensibili al modo in cui un insieme di variabili, con lo stesso contenuto informativo di un altro set, viene presentato alla rete stessa. Tutti i dati sono stati inseriti prendendo i loro logaritmi naturali. Sono state considerate diverse reti, variando il numero di neuroni nascosti da 5 a 20, al fine di ricercare il valore migliore, cioè quello che produce il TSS più alto.

Tabella 5.2: Risultati dei test con reti neurali, che mostrano il numero ottimale di neuroni nascosti (il cui numero è variabile da 5 a 20), il TSS per l'intero campione, i campioni di allenamento, validazione e test, i valori di TPR e FPR.

Input data	N. neuroni nascosti	TSS max train	TSS max valid	TSS max test	TPR totale	FPR totale	TSS max totale
D	14	0.35	0.13	0.29	0.74	0.44	0.30
E	18	0.45	0.36	0.41	0.86	0.43	0.43
I	9	0.52	0.41	0.29	0.74	0.31	0.44
Ipeak	16	0.37	0.35	0.35	0.81	0.46	0.35
D-E	20	0.61	0.59	0.55	0.82	0.23	0.59
D-I	11	0.60	0.60	0.59	0.80	0.20	0.60
D-Ipeak	7	0.53	0.48	0.47	0.83	0.32	0.51
E-Ipeak	17	0.46	0.46	0.39	0.84	0.40	0.44
I-Ipeak	13	0.61	0.54	0.60	0.82	0.24	0.58
D-E-Ipeak	8	0.68	0.65	0.60	0.81	0.15	0.66

La Tabella 5.2 mostra i risultati ottenuti dalle 160 configurazioni di reti neurali testate (10 diversi configurazioni di dati in input e 16 strutture in cui varia il numero di neuroni nello strato nascosto). L'intero set di dati degli eventi piovosi è stato suddiviso in un set di dati di addestramento, convalida e test, selezionati casualmente da quello complessivo, nelle proporzioni del 70 % - 15 % - 15 %. Il set di dati di addestramento è costituito dai dati utilizzati per adattare il modello, mentre la convalida fornisce una valutazione imparziale di un adattamento del modello al set di dati durante addestramento. Infine, il set di dati del test fornisce una valutazione imparziale di un adattamento del modello finale. Questa suddivisione ha consentito l'applicazione del criterio del "early stopping" per prevenire l'overfitting. Secondo questo criterio, l'allenamento della rete neurale viene interrotto quando i valori della funzione di prestazione calcolati sul set di dati di validazione iniziano a peggiorare.

Al fine di garantire la rappresentatività dei dati assegnati in modo casuale ai campioni di addestramento, convalida e test, i risultati in cui il TSS nella fase di test o nel campione di convalida sono maggiori del TSS nella fase di addestramento vengono rimossi dall'analisi. Una volta che la rete è stata addestrata considerando i tre subset di dati e l'early stopping, vengono "bloccate" le metriche della matrice di confusione (es. TSS), che possono essere calcolate sull'intero campione e le prestazioni corrispondenti possono essere considerate generalizzabili. In questo modo è possibile confrontare l'approccio proposto con quello tradizionale, definendo le performance che in Tabella 5.2 sono etichettate come "totale".

A differenza delle soglie presentate precedentemente, in cui si considerano solo degli eventi innescanti l'approccio proposto, che prevede di utilizzare le reti neurali, permette di tenere conto sia di eventi innescanti che di eventi non innescanti (Peres and Cancelliere, 2021). Sebbene non confrontabili con le precedenti, in quanto tiene conto di eventi innescanti e non innescanti, è stata ricavata una soglia con legge di tipo potenza corrispondente al TSS massimo considerando i dati di output del software CTRL-T, tramite lo strumento MATLAB® "global optimisation toolbox" ottenendo il seguente risultato (Eq. 17):

$$T_{I_NoI_5\%}: E = 2.40 \cdot D^{0.68} \quad (17)$$

La soglia determinata ha un TSS=TSS₀ =0.50, ottenuto da un TPR=0.76 e un FPR=0.26. La soglia ha un'intercetta più bassa ma una pendenza maggiore, quindi,

dopo una durata di circa 5 h, è superiore a quella riportata dall'Eq. (10). La *Figura 5.5* e *Figura 5.6* riportano il confronto grafico tra le soglie.

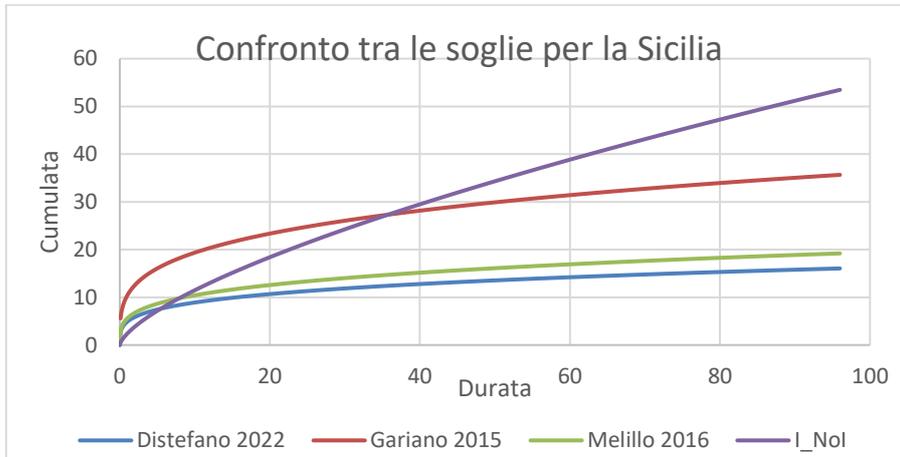


Figura 5.5: Confronto tra la soglia che considera eventi innescenti e non innescenti (viola) e quelle che considerano solo eventi innescenti nel piano cartesiano

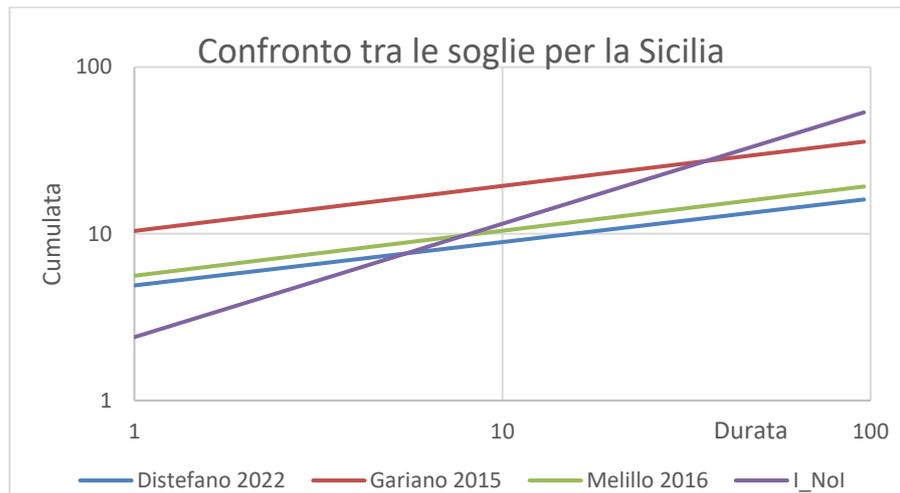


Figura 5.6: Confronto tra la soglia che considera eventi innescenti e non innescenti (viola) e quelle che considerano solo eventi innescenti nel piano bi-logaritmico

Come si può vedere dalla Tabella 5.2, utilizzando una sola variabile come input all'interno della rete neurale, le prestazioni sono significativamente inferiori a quelle ottenute dall'utilizzo della soglia di tipo potenza dell'Eq. (17): tuttavia, per la

variabile con il più alto contenuto informativo, l'intensità media delle precipitazioni I , si osserva che il $TSS=0,44$ è abbastanza vicino a $TSS_0 = 0,50$. Quando si utilizzano le variabili di input in coppia, le prestazioni aumentano significativamente. In particolare, nel caso delle coppie $D-I$ e $D-E$ – ovvero le stesse variabili utilizzate per la legge di potenza, il $TSS=0,59$ (0,60) è significativamente superiore a TSS_0 . Ciò è ottenuto sia da un aumento del TPR sia da una diminuzione dell'FPR. Il fatto che a parità di dati di input la rete neurale fornisca prestazioni significativamente migliori rispetto alla legge di tipo potenza dimostra che l'utilizzo di una forma parametrica predeterminata per l'equazione di soglia non consente di sfruttare a pieno il contenuto informativo delle variabili di input più completo, mentre la flessibilità delle ANN consente di ottenere una migliore classificazione. In altre parole, uno degli inconvenienti di una legge di potenza è che la stessa equazione è generalmente presupposta valida per tutte le durate, mentre le ANN sono più flessibili. Infine, aggiungendo una terza variabile (input nella rete $D-E-Ip$), si ottiene un ulteriore miglioramento ($TSS=0,66$), dovuto principalmente ad una diminuzione del FPR. Questo risultato dimostra come le reti neurali possono essere un aiuto nella ricerca di variabili aggiuntive in grado di fornire una previsione dinamica e più affidabile delle condizioni di attivazione di una frana. In particolare, in questo caso, è stato dimostrato che l'intensità di picco può avere un importante contenuto informativo, aspetto forse non sufficientemente indagato in letteratura, anche se alcuni ricercatori hanno riscontrato che l'aggiunta di una terza variabile è un possibile modo per derivare soglie che si adattano meglio a casi di studio complessi (Rosi et al., 2021).

5.5 SOGLIE PLUVIOMETRICHE DETERMINATE PER LA NORVEGIA

Per l'area studio norvegese è stata eseguita la stessa procedura proposta per la regione Sicilia, costruendo il database di precipitazioni innescanti e non, utilizzando il software CTRL-T. L'algoritmo ha ricostruito 247 precipitazioni innescanti di cui 117 loose material, 48 debris flow, 82 debris slide. Essendo il dato di precipitazione quotidiano, l'innescamento della frana è stato riferito al dato di precipitazione del giorno successivo a cui è associata la cumulata di precipitazione del giorno della frana. L'ubicazione delle frane ricostruite è riportata nella mappa presentata in Figura 4.6, la soglia ottenuta descritta dall'Eq. (18) è mostrata in Figura 5.7.

$$T_{Nor\ 5\%}: E = 5 \cdot D^{0.44} \quad (18)$$

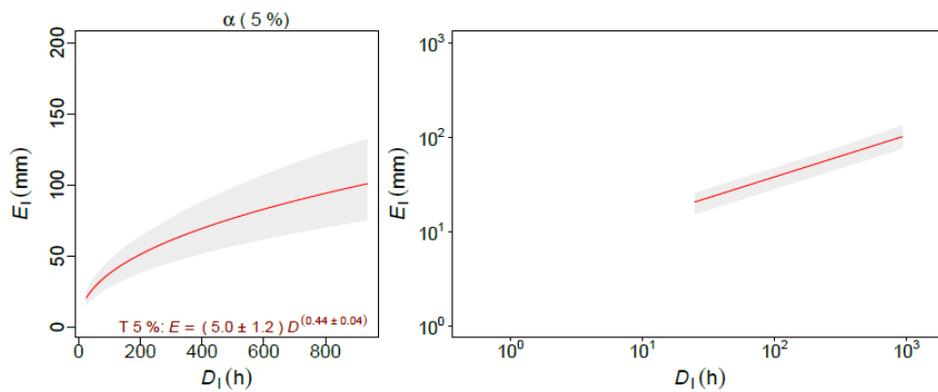


Figura 5.7: Soglia derivata da CTRL-T per la Norvegia

Come effettuato per il caso studio della Sicilia è stata ricavata una soglia con legge di tipo potenza che tiene conto di eventi innescanti e non innescanti tale da massimizzare il TSS considerando i dati di output del software CTRL-T. Tramite lo strumento MATLAB® "global optimisation toolbox" si è ottenuto il seguente risultato (Eq. 19) e mostrata in Figura 5.8:

$$T_{Nor\ I_NoI\ 5\%}: E = 9.05 \cdot D^{0.33} \quad (19)$$

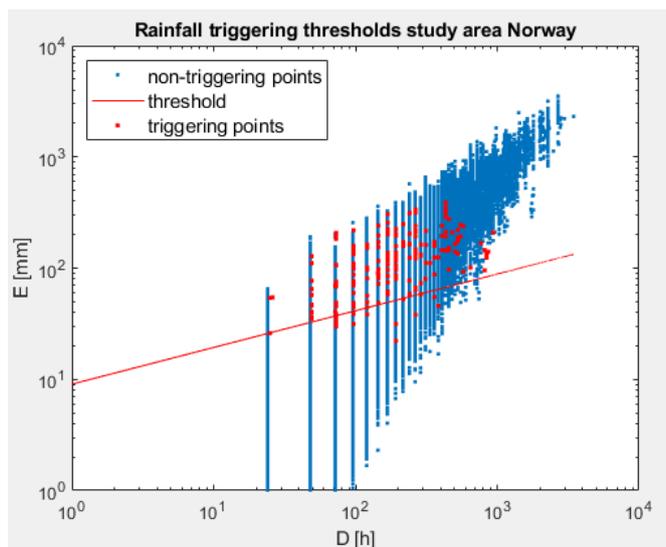


Figura 5.8: Soglia cumulata-durata che tiene conto dei dati innescanti e non innescanti (NO)

Di seguito è riportato il confronto delle soglie determinate con quelle proposte da Meyer et al. (2012) (Figura 5.9) caratterizzata dalla seguente equazione, con la durata D è espressa in giorni:

$$T_{Meyer\ 5\%}: E = 23.3 \cdot D^{0.76} \quad (20)$$

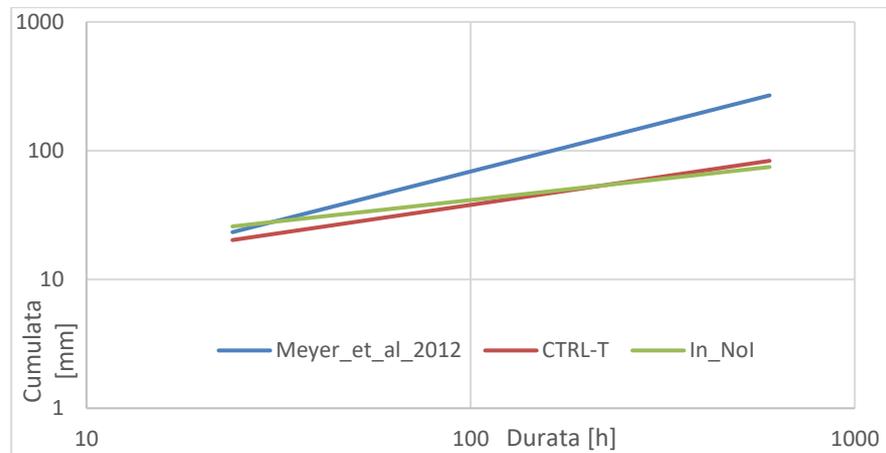


Figura 5.9: Confronto tra soglie pluviometriche derivate per la Norvegia.

Confrontando le soglie, è possibile osservare che quella realizzata da Meyer et al. (2012) assume valori paragonabili a quelle mostrate in precedenza, per valori compresi tra le 24 e le 48 ore. Per durate maggiori esiste un discostamento netto dalle soglie espresse dalle relazioni (18) e (19). Queste ultime risultano essere paragonabili sebbene quella determinata dall'algoritmo CTRL-T (18) supera quella che considera eventi innescenti e non (19) per precipitazioni di durata superiori a 10 giorni.

5.6 CLASSIFICATORE CHE CONSIDERA SOLO DATI DI PRECIPITAZIONE

Nel seguito verranno mostrati e discussi i risultati ottenuti dalle elaborazioni effettuate mostrando le differenze tra le due aree di studio.

In questo caso a differenza della prima applicazione, presentata precedentemente, per valutare l'influenza che la suddivisione randomica del database comporta sui campioni di training, validazione e test e sui valori in output, le reti sono state addestrate 30 volte. In questo modo non è stato ottenuto un singolo valore, come mostrato in Tabella 5.2, bensì un intervallo, visualizzato tramite box and whiskers plot.

Per poter valutare in maniera oggettiva le differenze di performance tra le due aree di studio non ci si riferisce più ai valori di intensità di picco, ma ai valori di intensità media dell'evento, essendo il dato di precipitazione norvegese aggregato alla scala oraria.

Nelle Figura 5.10 e Figura 5.11 sono riportati i box plot relativi alle analisi che considerano esclusivamente i dati di precipitazione per le due aree di studio, in Italia e Norvegia rispettivamente. I valori medi in termini di TSS, FPR e TPR che per i valori di TSS per le fasi di training, validazione e test mostrati rispettivamente in Tabella 5.3 e in Tabella 5.4 per l'Italia e la Norvegia. Rispetto alle tabelle mostrate precedentemente sono state inserite due nuove colonne, il numero di neuroni medio per cui si ottiene il TSS medio, nonché il range Δ di variazione di performance determinato come differenza tra valore di TSS massimo e quello minimo.

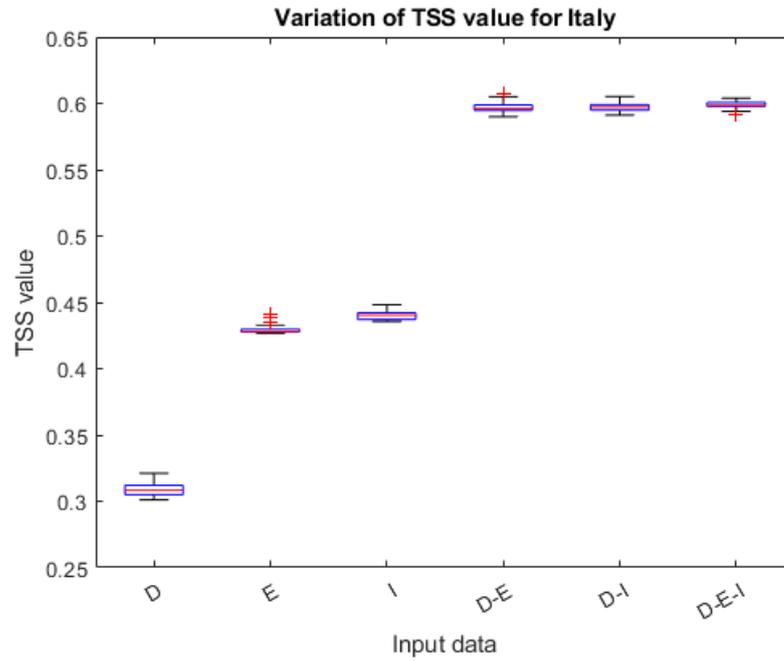


Figura 5.10: Boxplot del classificatore che considera solo le caratteristiche della precipitazione (IT)

Tabella 5.3: Risultati dei classificatori che considerano solo le caratteristiche della precipitazione (IT)

Input (IT)	Num. Med. Neuro.	TSS tot med.	TSS med. train	TSS med. valid	TSS med. test	TPR med.	FPR med.	Δ
D	15.83	0.31	0.35	0.29	0.28	0.75	0.44	0.02
E	16.70	0.43	0.45	0.41	0.40	0.87	0.44	0.01
I	16.67	0.44	0.48	0.41	0.42	0.71	0.27	0.01
DE	13.20	0.60	0.63	0.55	0.57	0.80	0.20	0.02
DI	11.67	0.60	0.63	0.56	0.55	0.79	0.20	0.01
DEI	13.10	0.60	0.63	0.57	0.56	0.80	0.20	0.01

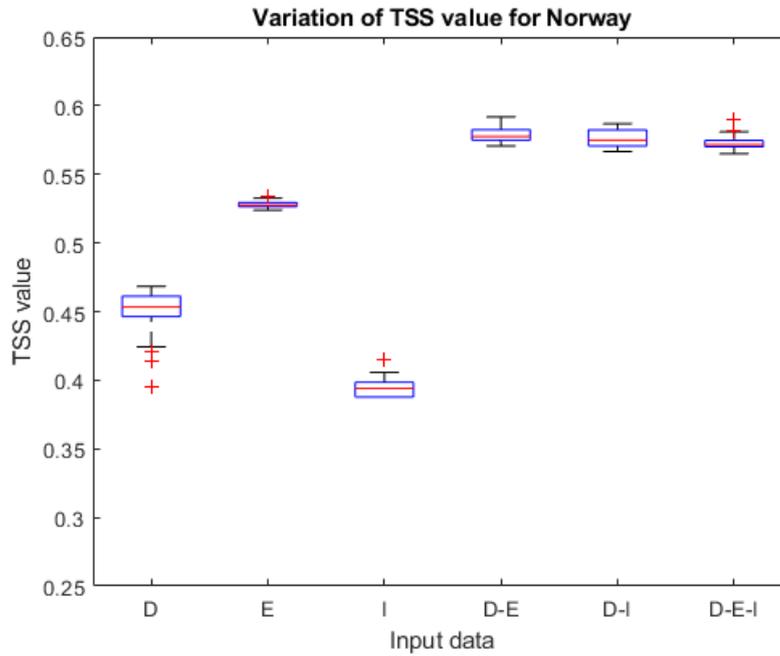


Figura 5.11: Boxplot del classificatore che considera solo le caratteristiche della precipitazione (NO)

Tabella 5.4: Risultati dei classificatori che considerano solo le caratteristiche della precipitazione (NO)

Input (NO)	Num. Neuron.	TSS tot med.	TSS med. train	TSS med. valid	TSS med. test	TPR med.	FPR med.	Δ
D	16.27	0.45	0.47	0.42	0.41	0.84	0.39	0.07
E	16.83	0.53	0.55	0.53	0.53	0.87	0.34	0.01
I	13.93	0.39	0.42	0.37	0.39	0.82	0.42	0.03
DE	16.33	0.58	0.60	0.57	0.57	0.91	0.33	0.02
DI	16.13	0.58	0.59	0.57	0.57	0.93	0.36	0.02
DEI	15.87	0.57	0.59	0.56	0.57	0.92	0.34	0.02

Considerando solo le caratteristiche della precipitazione è possibile osservare che per la regione Sicilia i risultati in termini di TSS considerando singole variabili risultano essere inferiori a quelli ottenuti considerando le leggi di tipo potenza.

Considerando le coppie di dati in input, $D-E$ e $D-I$ invece i risultati in termini di prestazioni sono comparabili. A differenza di quanto visto in Distefano et al. (2022) utilizzare l'intensità media piuttosto che quella di picco non migliora le performance.

Per l'area a est di Bergen (Figura 5.11 e Tabella 5.4) è possibile osservare che, eccezione fatta per la precipitazione cumulata, considerando singole variabili in ingresso si ottiene un valore di TSS inferiore a 0.50. Per la durata dell'evento D , è possibile osservare la variabilità più elevata di tutti gli output a causa dei tre valori outlier. I valori di performance considerando le configurazioni con due variabili in ingresso ($D-E$ e $D-I$) ottenuti risultano essere paragonabili a quelli visti per la Sicilia. Risulta essere controintuitiva la diminuzione di performance, seppur minima, che si ottiene introducendo il valore di intensità media, passando da un valore di TSS=0.58 per i casi $D-E$ e $D-I$ ad un valore di TSS =0.57 per la terna $D-E-I$.

5.7 LIMITI DELL'APPROCCIO IMPIEGATO E SVILUPPI FUTURI

L'identificazione delle soglie di precipitazione che indicano le condizioni di innesco delle frane è un passaggio fondamentale per l'attuazione di sistemi di early warning affidabili. Comunemente, le soglie sono vincolate a una forma parametrica predeterminata, che è generalmente una legge di potenza che collega l'evento piovoso, la durata D e l'intensità media I o precipitazione cumulata E . Dai dati in nostro possesso, è stato possibile riscontrare che la scelta di una predeterminata legge per la soglia pluviometrica di innesco può potenzialmente limitare le prestazioni del modello empirico. Le reti neurali artificiali possano costituire un prezioso strumento per superare alcune limitazioni. L'analisi effettuata, con riferimento a entrambi i casi di studio, ha mostrato che una soglia di legge di potenza $E-D$ ha un valore di TSS=TSS₀ =0.50.

D'altra parte, il classificatore basato su reti neurali, utilizzando la stessa coppia di variabili di input, ha prodotto un TSS=0.60 significativamente maggiore.

Da non sottovalutare è anche la possibilità di introdurre altre tipologie di informazioni sotto forma di altre variabili che possano migliorare le prestazioni predittive. Ad esempio, è stato dimostrato che l'inclusione dell'intensità di picco delle precipitazioni come variabile aggiuntiva, nel caso di studio italiano, può

portare a un miglioramento delle prestazioni. È importante che durante l'addestramento delle reti neurali, le capacità di generalizzazione siano garantite, ad esempio tramite la tecnica di "early stopping". L'overfitting non è un problema per l'approccio tradizionale basato sulla legge di potenza – o qualsiasi altra equazione parametrica – poiché in generale il numero di parametri liberi è molto basso (due per una legge di potenza). Questo può essere uno svantaggio nell'utilizzo delle reti neurali anche se permette di considerare sia gli eventi innescanti che quelli non innescanti, fondamentale per ottenere soglie con caratteristiche statistiche accettabili (Peres e Cancelliere, 2021). Un altro possibile svantaggio delle reti neurali rispetto alle soglie di forma predeterminata è rappresentato anche dal fatto che può essere difficile esprimere il classificatore delle reti neurali come una semplice equazione. Ciò potrebbe limitare l'implementazione pratica di soglie di attivazione basate su reti neurali, che potrebbero essere percepite come impraticabili dai professionisti. Tuttavia, questo limite può essere superato, ad esempio, fornendo un software di facile utilizzo all'utente finale.

CAPITOLO 6 SOGLIE IDRO-METEOROLOGICHE MULTIVARIATE BASATE SU RETI NEURALI ARTIFICIALI

6.1 GENERALITÀ

Nel presente capitolo saranno presentati i modelli di soglia di tipo idro-meteorologico, ovvero che includono dati di umidità del suolo. Dopo aver descritto come i dati di umidità del suolo sono integrati all'interno del database di eventi innescanti e non innescanti, verranno presentati i risultati e i confronti per i due casi di studio, in Italia e in Norvegia.

6.2 DATI DI UMIDITÀ

I dati relativi al contenuto volumetrico di acqua del database ERA5-Land (Muñoz-Sabater et al., 2021), forniti dall'European Centre for Medium-Range Weather Forecasts mostrati in Tabella 4.1 sono stati integrati nei database delle due aree di studio. I dati ERA5-Land sono disponibili per entrambe le aree di studio, determinati con la stessa metodologia per gli intervalli temporali considerati. Tutte le motivazioni sopra elencate giustificano l'utilizzo di tale dataset. Per entrambe le aree di studio e per i periodi di indagine sono stati scaricati i dati di umidità del suolo tramite gli strumenti messi a disposizione dall'ECMWF ([ERA5-Land hourly data from 1950 to present \(copernicus.eu\)](https://era5-climate.copernicus.eu/)).

L'attribuzione del valore di umidità del suolo, effettuata in maniera automatica, tramite un codice MATLAB® accuratamente messo a punto per lo scopo, avviene sulla base dell'appartenenza o meno di un pluviometro o di una frana all'interno di una delle celle di ERA5. Più nello specifico si è stabilito che:

1. I valori di umidità del suolo associate alle frane ricostruite sono quelli che corrispondono alla cella ERA5 in cui ricadono le coordinate del movimento franoso.
2. I valori di umidità del suolo delle precipitazioni non innescanti sono quelli ricavati dalla cella ERA5 in cui è posizionato il pluviometro (per il caso studio italiano) o il pluviometro fittizio, ovvero i punti del raster individuati come pluviometri, per il caso studio norvegese

Essendo note il giorno di inizio (per il caso studio norvegese) e/o l'ora (per il caso studio italiano) della precipitazione, e nota o ipotizzata l'ora di innesco della

frana è stato possibile determinare due diversi dati di umidità che possono essere usati come proxy per le indagini. Nel primo caso si ha il valore di contenuto volumetrico d'acqua in corrispondenza dell'evento di precipitazione che può essere visto come una condizione iniziale che può portare (o meno) all'innescò, e il valore finale di umidità del suolo per cui è avvenuta, o no, la mobilitazione della frana che può rappresentare la "capacità di immagazzinamento massima" del terreno.

Per la tipologia di dati disponibili, dati di precipitazione aggregati alla scala giornaliera per il caso studio norvegese, nelle elaborazioni che seguono non si farà riferimento all'intensità di picco ma all'intensità dell'evento, determinata come rapporto tra la precipitazione cumulata e la durata totale dell'evento espresso in ore. I valori di umidità del suolo per il database norvegese sono ricavati in corrispondenza dello stesso orario in cui viene fornito il valore di precipitazione cumulata giornaliero (06:00 AM).

6.3 DERIVAZIONE DELLE SOGLIE IDRO-METEOROLOGICHE

Sulla base dei dati presentati nei paragrafi precedenti e utilizzando gli algoritmi di riconoscimento di pattern basati su reti neurali descritti nel Paragrafo 5.4 è stato messo a punto un classificatore che consente di tenere in considerazione oltre che le caratteristiche delle precipitazioni anche i valori di umidità del suolo.

A differenza della prima applicazione, presentata precedentemente, in questo caso per valutare anche l'influenza che la suddivisione randomica del database ha sui subset di training, validazione e test e sui valori di output delle reti neurali, queste sono state addestrate 30 volte. In questo modo non è stato ottenuto un singolo valore, come mostrato in Tabella 5.2, bensì un intervallo, visualizzato tramite box and whiskers plot.

Maggiore è l'ampiezza dei diagrammi a scatola e baffi, maggiore è l'incertezza della previsione legata all'effetto che l'inizializzazione random ha sulla rete neurale.

In totale sono state analizzate, per entrambe le aree di studio, ben 81 combinazioni di input (considerando sempre 16 differenti reti con numero di neuroni nello strato nascosto variabili tra 5 e 20), per un totale di $(2 \cdot 30 \cdot 16 \cdot 81)$ 77760 reti neurali addestrate, così distinte:

- 12 reti considerano Durata - Cumulata - Umidità del suolo considerando un singolo layer (istante iniziale, finale, iniziale e finale);

- 12 reti considerano Durata – Intensità – Umidità del suolo considerando un singolo layer (istante iniziale, finale, iniziale e finale);
- 12 reti considerano Durata – Cumulata – Intensità – Umidità del suolo considerando un singolo layer (istante iniziale, finale, iniziale e finale);
- 15 reti considerano Durata – Cumulata – Umidità del suolo usando combinazioni di più layer (istante iniziale, finale, iniziale e finale);
- 15 reti considerano Durata – Intensità – Umidità del suolo usando combinazioni di più layer (istante iniziale, finale, iniziale e finale);
- 15 reti considerano Durata – Cumulata – Intensità – Umidità del suolo usando combinazioni di più layer (istante iniziale, finale, iniziale e finale).

6.4 SOGLIE IDRO-METEOROLOGICHE CON DATI DI UMIDITÀ RELATIVI AD UN UNICO STRATO

In questo paragrafo verranno mostrati e discussi i risultati ottenuti dal classificatore neurale. Si considerando come dati di input le variabili idrologiche e i valori di umidità del suolo dei singoli strati in cui è stato discretizzato il terreno (S_i indica l'*i-esimo* layer) nel dB ERA5-Land. I valori di umidità possono riferirsi al valore dell'umidità all'istante iniziale (S_i), finale (S'_i) o la combinazione iniziale e finale ($S_i S'_i$). Con numero medio di neuroni (Num. med. Neuron, vedi Tabella 6.1 e successive) si intende mostrare per quale configurazione di rete si ottengono le migliori performance della rete. Considerando come input i valori di **D-E-S** i risultati ottenuti per il caso studio italiano sono riportati in Figura 6.1 e Tabella 6.1, mentre quelli relativi al caso di studio norvegese sono riportati in Figura 6.2 e Tabella 6.2.

I valori di TSS sono compresi tra $-1 \div 1$. Nelle immagini seguenti si propone un ingrandimento nell'intervallo $0.5 \div 0.85$, dove sono concentrati i risultati ottenuti per facilitare la lettura del grafico.

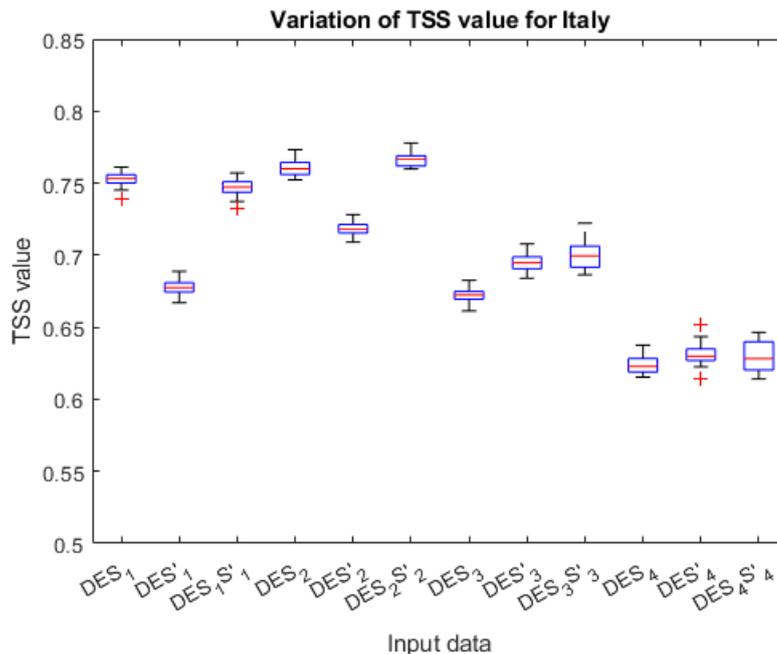


Figura 6.1: Boxplot del classificatore basato sulle variabili D-E-S per singoli strati (IT)

Tabella 6.1: Risultati del classificatore che utilizza come variabili D-E-S considerando singoli strati (IT)

Input (IT)	Num. med. Neuron.	TSS tot med.	TSS med. train	TSS med. valid	TSS med. test	TPR med.	FPR med.	Δ
DES ₁	14.90	0.75	0.78	0.74	0.72	0.88	0.12	0.02
DES' ₁	15.27	0.68	0.71	0.64	0.63	0.85	0.17	0.02
DES ₁ S' ₁	15.83	0.75	0.78	0.73	0.72	0.88	0.13	0.02
DES ₂	13.10	0.76	0.78	0.75	0.74	0.89	0.13	0.02
DES' ₂	13.70	0.72	0.75	0.70	0.68	0.88	0.16	0.02
DES ₂ S' ₂	12.63	0.77	0.79	0.74	0.73	0.89	0.12	0.02
DES ₃	14.77	0.67	0.70	0.66	0.66	0.85	0.17	0.02
DES' ₃	16.27	0.69	0.72	0.70	0.67	0.85	0.15	0.02
DES ₃ S' ₃	14.97	0.70	0.73	0.69	0.67	0.88	0.18	0.04
DES ₄	16.67	0.62	0.65	0.61	0.59	0.84	0.21	0.02
DES' ₄	15.50	0.63	0.66	0.60	0.60	0.83	0.20	0.04
DES ₄ S' ₄	15.17	0.63	0.67	0.61	0.60	0.82	0.19	0.03

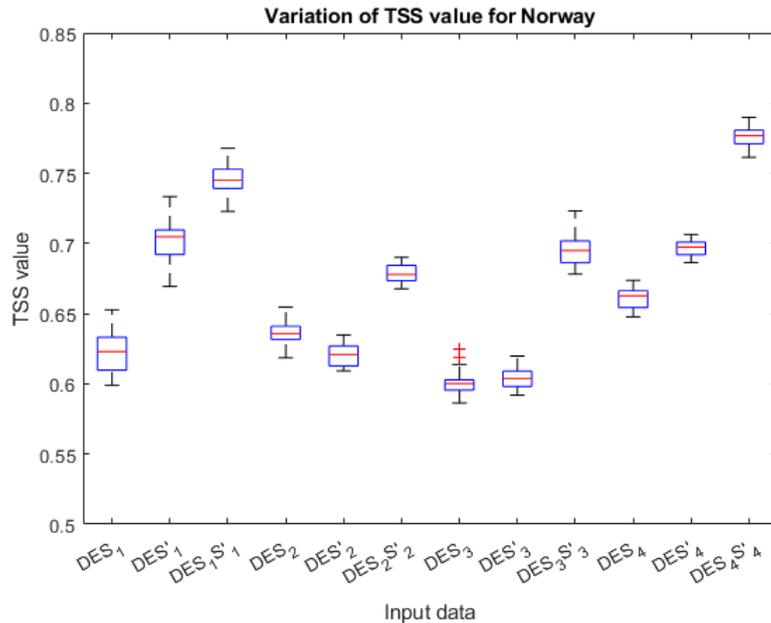


Figura 6.2: Boxplot del classificatore che considera come variabili D-E-S per singoli strati (NO)

Con l'introduzione dei dati di umidità del suolo in entrambi i casi di studio si osserva un miglioramento delle performance del classificatore considerando che in entrambi i casi il TSS medio non scende sotto 0.62.

Per l'Italia è possibile osservare come i valori di performance migliori si ottengono utilizzando i valori dei primi strati di umidità del suolo, in ordine crescente DES_1 , DES_2 , $DES_2S'_2$, sebbene il valore di TSS massimo si ottenga considerando 4 variabili in ingresso a differenza di DES_1 , e DES_2 ottenuti con solo 3 variabili.

Per la Norvegia le combinazioni che danno i valori di TSS migliori sono in ordine crescente $DES_1S'_1$ e $DES_4S'_4$. In entrambi i casi bisogna ricorrere a un classificatore che utilizza 4 variabili. Il risultato per cui il quarto layer di suolo sia quello con il TSS (con valori prossimi a 0.78) risulta poco intuitivo in quanto rappresentativo di profondità comprese tra 100 e 289 centimetri per cui poco correlabile con la tipologia di frane superficiali.

Tabella 6.2: Risultati del classificatore che considera come input i dati di D-E- S singoli strati (NO)

Input (NO)	Num. med. Neuron.	TSS tot med.	TSS med. train	TSS med. valid	TSS med. test	TPR med.	FPR med.	Δ
DES ₁	15.07	0.62	0.64	0.61	0.61	0.87	0.25	0.05
DES' ₁	16.63	0.70	0.72	0.69	0.68	0.89	0.19	0.06
DES ₁ S' ₁	15.73	0.75	0.77	0.72	0.72	0.87	0.13	0.04
DES ₂	16.50	0.64	0.66	0.62	0.61	0.85	0.21	0.04
DES' ₂	15.80	0.62	0.65	0.60	0.60	0.83	0.20	0.03
DES ₂ S' ₂	15.50	0.68	0.70	0.67	0.65	0.85	0.17	0.02
DES ₃	15.97	0.60	0.62	0.59	0.59	0.88	0.28	0.04
DES' ₃	16.40	0.60	0.62	0.59	0.59	0.89	0.29	0.03
DES ₃ S' ₃	16.37	0.70	0.72	0.69	0.68	0.91	0.21	0.05
DES ₄	15.90	0.66	0.68	0.64	0.64	0.89	0.23	0.03
DES' ₄	15.17	0.70	0.71	0.69	0.68	0.93	0.23	0.02
DES ₄ S' ₄	15.80	0.78	0.79	0.78	0.76	0.94	0.17	0.03

Verranno ora mostrati e discussi i risultati ottenuti dal classificatore neurale che considera come variabili in ingresso i valori di **durata**, **intensità** e i **valori di umidità del suolo dei singoli strati** in cui è stato discretizzato il terreno. I dati per il caso studio italiano sono riportati in Figura 6.3 e Tabella 6.3, mentre quelli relativi al caso di studio norvegese sono riportati in Figura 6.4 e Tabella 6.4.

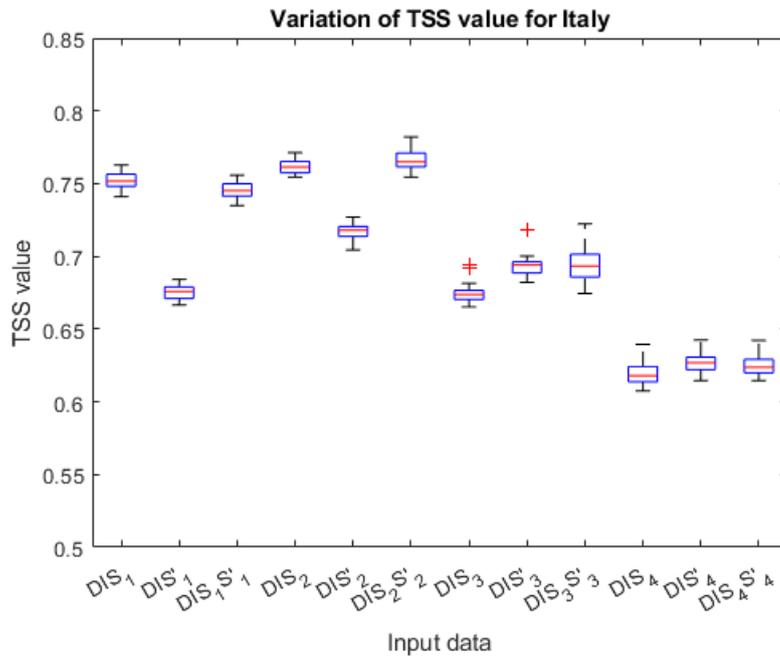


Figura 6.3: Boxplot del classificatore che considera come dati di input D-I-S singoli strati (IT)

Tabella 6.3: Risultati del classificatore che considera come input i dati di D-I-S singoli strati (IT)

Input (IT)	Num. med. Neuron.	TSS tot med.	TSS med. train	TSS med. valid	TSS med. test	TPR med.	FPR med.	Δ
DIS ₁	15.23	0.75	0.78	0.73	0.73	0.87	0.12	0.02
DIS' ₁	14.60	0.68	0.70	0.65	0.65	0.84	0.16	0.02
DIS ₁ S' ₁	15.60	0.75	0.78	0.73	0.72	0.88	0.14	0.02
DIS ₂	13.57	0.76	0.79	0.73	0.73	0.88	0.12	0.02
DIS' ₂	14.07	0.72	0.75	0.70	0.68	0.87	0.16	0.02
DIS ₂ S' ₂	12.40	0.77	0.79	0.72	0.75	0.89	0.12	0.03
DIS ₃	13.73	0.67	0.71	0.65	0.64	0.83	0.16	0.03
DIS' ₃	14.83	0.69	0.73	0.67	0.68	0.85	0.15	0.04
DIS ₃ S' ₃	15.17	0.69	0.73	0.67	0.67	0.87	0.17	0.05
DIS ₄	15.27	0.62	0.65	0.59	0.58	0.84	0.22	0.03
DIS' ₄	14.97	0.63	0.66	0.59	0.61	0.83	0.20	0.03
DIS ₄ S' ₄	15.37	0.63	0.66	0.60	0.61	0.82	0.20	0.03

Tabella 6.4: Risultati del classificatore che considera come input i dati D-I-S singoli strati (NO)

Input (NO)	Num. med. Neuron.	TSS tot med.	TSS med. train	TSS med. valid	TSS med. test	TPR med.	FPR med.	Δ
DIS ₁	14.63	0.62	0.65	0.61	0.60	0.85	0.23	0.06
DIS' ₁	13.97	0.69	0.71	0.67	0.68	0.88	0.19	0.05
DIS ₁ S' ₁	15.73	0.74	0.77	0.73	0.73	0.87	0.13	0.05
DIS ₂	14.17	0.63	0.65	0.61	0.62	0.82	0.19	0.02
DIS' ₂	16.27	0.62	0.64	0.60	0.60	0.82	0.20	0.05
DIS ₂ S' ₂	15.43	0.68	0.70	0.66	0.67	0.86	0.18	0.04
DIS ₃	16.53	0.60	0.62	0.58	0.58	0.87	0.27	0.02
DIS' ₃	16.73	0.60	0.62	0.59	0.58	0.86	0.26	0.06
DIS ₃ S' ₃	16.17	0.69	0.71	0.67	0.68	0.91	0.22	0.06
DIS ₄	15.43	0.66	0.68	0.63	0.64	0.88	0.23	0.04
DIS' ₄	15.80	0.69	0.71	0.68	0.68	0.93	0.23	0.03
DIS ₄ S' ₄	15.77	0.78	0.79	0.77	0.77	0.94	0.16	0.03

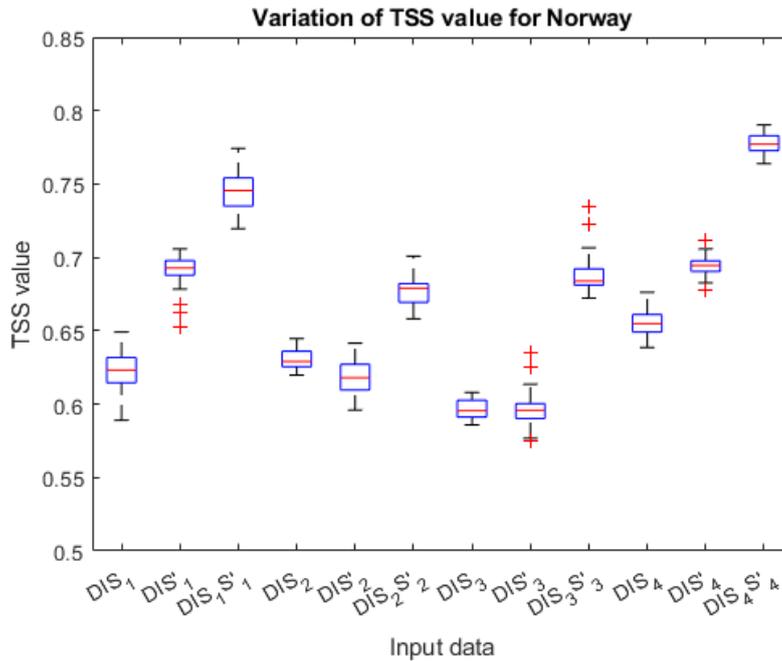


Figura 6.4: Boxplot del classificatore che considera come dati di input i valori D-I-S per singoli strati (NO)

Considerando come variabili di input la durata l'intensità di precipitazione e i singoli strati di umidità ai vari istanti si nota come non ci siano differenze sostanziali dai risultati mostrati al paragrafo precedente. Per la regione Sicilia le migliori performance si ottengono considerando i primi due strati di terreno con TSS medio compreso tra 0.72 e 0.79 (valore massimo ottenuto considerando un classificatore con 4 variabili in input).

Per l'area di studio norvegese gli strati più rappresentativi sono il primo e il quarto, considerando i valori di umidità iniziali e finali con valori di TSS rispettivamente di 0.74 e 0.78. L'introduzione dell'intensità seppure mantenga valori di performance paragonabili a quelli visti precedentemente introduce una maggiore variabilità media delle performance come può essere osservato dai valori maggiori di Δ , e graficamente dall'ampiezza dei diagrammi a scatola e baffi.

Di seguito verranno mostrati e discussi i risultati ottenuti dal classificatore neurale che considera come variabili di input **durata, cumulata, intensità** e i valori di **umidità del suolo dei singoli strati** in cui è stato discretizzato il terreno. I dati per il caso studio italiano sono riportati in Figura 6.5 e Tabella 6.5, mentre quelli relativi al caso di studio norvegese sono riportati in Figura 6.6 e Tabella 6.6.

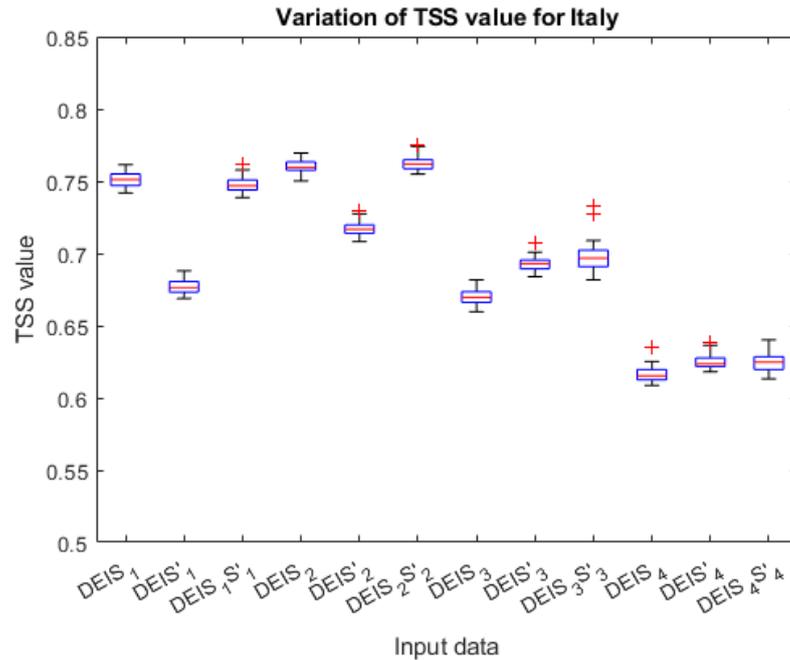


Figura 6.5: Boxplot del classificatore che considera come dati di input i valori di D-E-I-S per singoli strati (IT)

Tabella 6.5: Risultati del classificatore che considera i valori di D-E-I-S per singoli strati (IT)

Input (IT)	Num. med. Neuron.	TSS tot med.	TSS med. train	TSS med. valid	TSS med. test	TPR med.	FPR med.	Δ
DEIS ₁	15.63	0.75	0.78	0.73	0.72	0.88	0.13	0.02
DEIS' ₁	14.83	0.68	0.71	0.63	0.65	0.84	0.17	0.02
DEIS _{1S'} ₁	14.90	0.75	0.77	0.72	0.74	0.88	0.13	0.02
DEIS ₂	12.33	0.76	0.79	0.72	0.73	0.88	0.12	0.02
DEIS' ₂	13.80	0.72	0.74	0.69	0.69	0.88	0.16	0.02
DEIS _{2S'} ₂	13.90	0.76	0.79	0.72	0.73	0.89	0.12	0.02
DEIS ₃	16.17	0.67	0.70	0.65	0.64	0.86	0.19	0.02
DEIS' ₃	13.53	0.69	0.72	0.67	0.68	0.86	0.16	0.02
DEIS _{3S'} ₃	14.50	0.70	0.73	0.68	0.68	0.88	0.18	0.05
DEIS ₄	15.43	0.62	0.64	0.61	0.60	0.84	0.23	0.03
DEIS' ₄	15.73	0.63	0.66	0.61	0.59	0.82	0.19	0.02
DEIS _{4S'} ₄	13.83	0.63	0.66	0.60	0.60	0.82	0.20	0.03

Tabella 6.6: Risultati del classificatore che considera i dati D-E-I-S per singoli strati (NO)

Input (NO)	Num. med. Neuron.	TSS tot med.	TSS med. train	TSS med. valid	TSS med. test	TPR med.	FPR med.	Δ
DEIS ₁	14.40	0.62	0.63	0.61	0.61	0.87	0.26	0.07
DEIS' ₁	15.33	0.69	0.71	0.67	0.67	0.87	0.19	0.07
DEIS _{1S'} ₁	15.33	0.74	0.76	0.71	0.72	0.86	0.13	0.05
DEIS ₂	15.37	0.63	0.66	0.62	0.60	0.84	0.21	0.03
DEIS' ₂	14.60	0.61	0.64	0.60	0.59	0.81	0.20	0.03
DEIS _{2S'} ₂	15.37	0.68	0.70	0.67	0.65	0.85	0.18	0.04
DEIS ₃	15.70	0.60	0.62	0.59	0.57	0.89	0.29	0.03
DEIS' ₃	16.97	0.60	0.62	0.59	0.58	0.87	0.27	0.04
DEIS _{3S'} ₃	16.67	0.69	0.71	0.68	0.68	0.90	0.21	0.05
DEIS ₄	16.10	0.66	0.68	0.63	0.64	0.88	0.22	0.03
DEIS' ₄	13.83	0.69	0.71	0.68	0.69	0.93	0.24	0.02
DEIS _{4S'} ₄	16.60	0.78	0.80	0.77	0.77	0.94	0.16	0.03

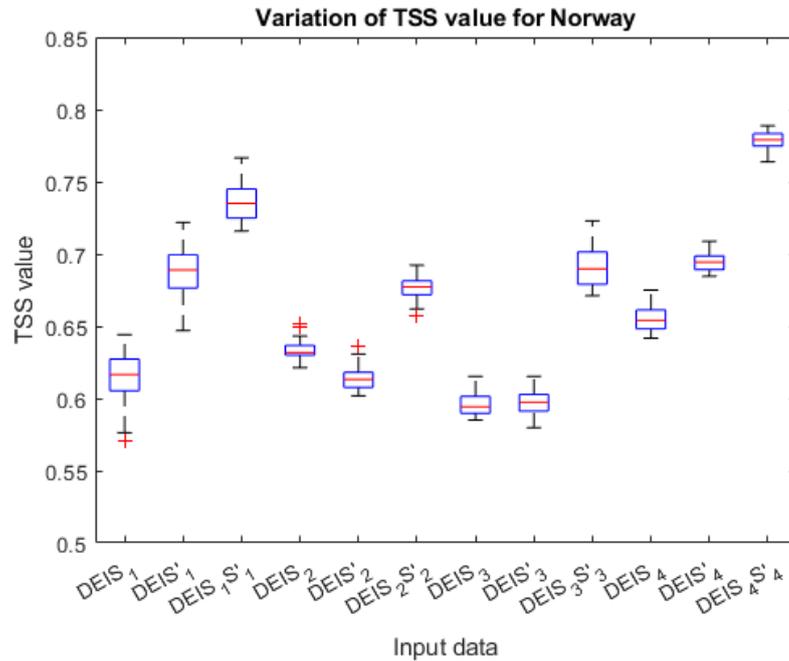


Figura 6.6: Boxplot del classificatore che considera i valori di D-E-I-S per singoli strati (NO)

Come è possibile osservare dai grafici e dai dati tabellati utilizzare sia la cumulata che l'intensità di precipitazione non comporta un miglioramento sostanziale delle performance del predittore neurale. I valori di TSS rimangono confrontabili con quelli visti precedentemente nonostante l'introduzione di una nuova variabile.

I trend visibili dai box plot confermano l'andamento mostrato dalle elaborazioni precedenti in cui si evince che sono più rappresentativi i primi due layer per il caso siciliano mentre il primo e il quarto strato per l'area di studio di Bergen. L'introduzione dei valori di intensità di precipitazione stabilizza le performance per l'area di studio siciliana, ma ha l'effetto opposto per l'area norvegese dove il Δ medio è prossimo a 0.04.

6.5 SOGLIE IDRO-METEOROLOGICHE CON DATI DI UMIDITÀ RELATIVI A PIÙ STRATI

In questo paragrafo verranno mostrati e discussi i risultati ottenuti dal classificatore neurale che considera come dati di input i valori delle variabili idrologiche e i valori di umidità del suolo, intesi come composizioni dei più strati in cui è stato discretizzato il terreno (S_i indica l'*i-esimo* layer). I valori di umidità presi in considerazione riguardano l'istante iniziale (S_i), finale (S'_i) o la combinazione iniziale e finale ($S_i S'_i$). In questo caso considerando come variabili di input la **durata** e la **precipitazione cumulata**, la dicitura *DESall* si riferisce alla combinazione che tiene conto di tutti i singoli strati di umidità e di entrambi i valori temporali (iniziale e finale). I dati per il caso studio italiano sono riportati in Figura 6.7 e Tabella 6.7, mentre quelli relativi al caso di studio norvegese sono riportati in Figura 6.8 e Tabella 6.8. I valori di TSS sono compresi tra $-1 \div 1$. Nelle immagini seguenti si propone un ingrandimento nell'intervallo $0.5 \div 0.85$, dove sono concentrati i risultati ottenuti per facilitare la lettura del grafico.

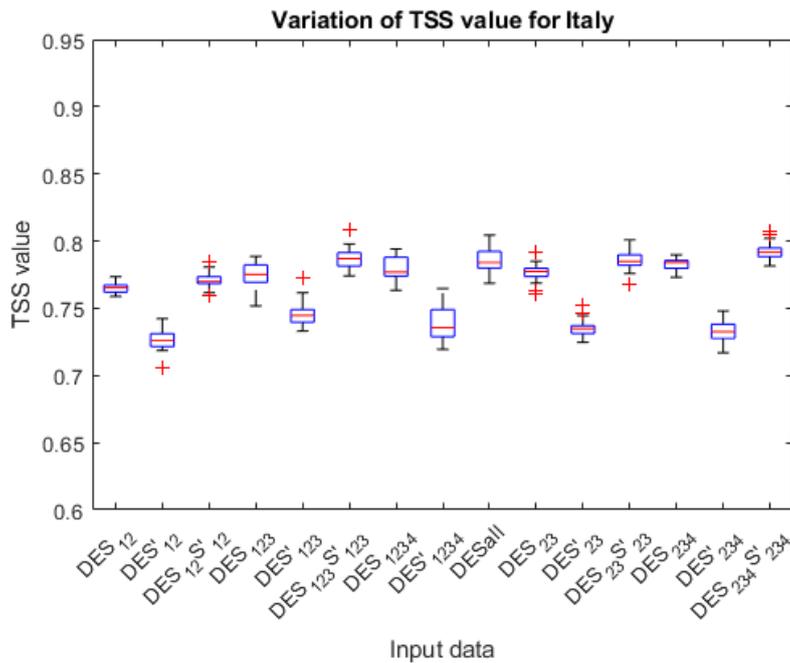


Figura 6.7 Boxplot del classificatore che considera come variabili di input i valori di D-E-S relativi a più strati (IT)

Tabella 6.7: Risultati del classificatore basato sulle variabili di D-E-S relativi a più strati (IT)

Input (IT)	Num. med. Neuron.	TSS tot med. med.	TSS med. train	TSS med. valid	TSS med. test	TPR med.	FPR med.	Δ
DES ₁ S ₂	13.90	0.77	0.79	0.74	0.73	0.88	0.11	0.01
DES' ₁ S' ₂	15.03	0.73	0.75	0.70	0.71	0.88	0.15	0.04
DES ₁₂ S' ₁₂	14.90	0.77	0.80	0.75	0.74	0.88	0.11	0.03
DES ₁₂₃	15.67	0.77	0.80	0.76	0.74	0.89	0.12	0.04
DES' ₁₂₃	14.17	0.75	0.77	0.73	0.73	0.90	0.16	0.04
DES ₁₂₃ S' ₁₂₃	15.97	0.79	0.81	0.77	0.76	0.89	0.10	0.03
DES ₁₂₃₄	14.63	0.78	0.81	0.74	0.75	0.89	0.11	0.03
DES' ₁₂₃₄	15.87	0.74	0.76	0.72	0.72	0.90	0.16	0.05
DES _{all}	15.03	0.79	0.81	0.77	0.76	0.89	0.10	0.04
DES ₂₃	14.97	0.78	0.80	0.75	0.74	0.88	0.10	0.03
DES' ₂₃	14.67	0.74	0.76	0.70	0.72	0.90	0.16	0.03
DES ₂₃ S' ₂₃	14.93	0.79	0.81	0.75	0.76	0.88	0.10	0.03
DES ₂₃₄	14.77	0.78	0.81	0.75	0.75	0.89	0.11	0.02
DES' ₂₃₄	14.17	0.73	0.76	0.72	0.71	0.89	0.16	0.03
DES ₂₃₄ S' ₂₃₄	15.23	0.79	0.82	0.76	0.77	0.89	0.09	0.03

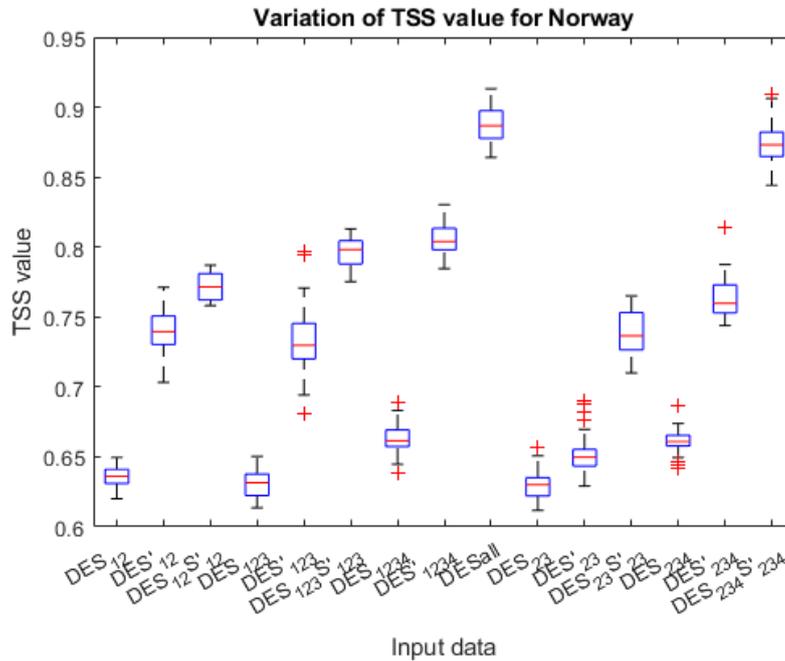


Figura 6.8: Boxplot del classificatore che considera i valori di D-E-S relativi a più strati (NO)

Come è possibile osservare dai grafici presentati per la regione Sicilia considerando diverse combinazioni di strati di suolo si ottengono performance in termini di TSS medio compresi tra 0.7 e 0.8 e la variabilità, escluso un singolo caso, risulta essere molto bassa considerando che si utilizzano minimo 4 variabili in ingresso.

Per la Norvegia invece, i risultati ottenuti sono molto discordanti sia in termini di TSS (compresi tra 0.64 e 0.89) che in termini di range come testimoniano alcuni box plot caratterizzati da un elevato valore di range interquartile. Le combinazioni migliori sono quelle che si ottengono considerando tutti i dati disponibili, in ordine crescente le combinazioni $DES_{234}S'_{234}$ e DES_{all} caratterizzate rispettivamente da 8 e 10 variabili in ingresso al classificatore.

Per il caso studio norvegese è possibile osservare una marcata oscillazione dei valori medi di TSS, mentre il trend per il caso studio italiano è molto più uniforme.

Tabella 6.8: Risultati del classificatore basato sulle variabili D-E-S relativi a più strati (NO)

Input (NO)	Num. med. Neuron.	TSS tot med.	TSS med. train	TSS med. valid	TSS med. test	TPR med.	FPR med.	Δ
DES _{1S₂}	16.23	0.64	0.66	0.62	0.61	0.85	0.21	0.03
DES' _{1S'₂}	16.50	0.74	0.76	0.73	0.72	0.90	0.16	0.07
DES _{12S'₁₂}	16.70	0.77	0.79	0.74	0.75	0.88	0.11	0.03
DES ₁₂₃	15.00	0.63	0.65	0.61	0.61	0.83	0.20	0.04
DES' ₁₂₃	15.10	0.73	0.75	0.72	0.71	0.86	0.12	0.12
DES _{123S'₁₂₃}	16.63	0.80	0.82	0.77	0.78	0.86	0.06	0.04
DES ₁₂₃₄	16.10	0.66	0.69	0.65	0.65	0.87	0.21	0.05
DES' ₁₂₃₄	15.20	0.81	0.83	0.79	0.78	0.89	0.08	0.05
DES _{all}	16.60	0.89	0.90	0.87	0.86	0.94	0.05	0.05
DES ₂₃	13.93	0.63	0.65	0.61	0.61	0.81	0.18	0.04
DES' ₂₃	16.30	0.65	0.68	0.63	0.62	0.84	0.19	0.06
DES _{23S'₂₃}	17.23	0.74	0.76	0.72	0.71	0.88	0.14	0.06
DES ₂₃₄	16.20	0.66	0.68	0.65	0.64	0.87	0.21	0.05
DES' ₂₃₄	16.07	0.76	0.78	0.75	0.75	0.91	0.14	0.07
DES _{234S'₂₃₄}	16.17	0.88	0.89	0.86	0.86	0.95	0.07	0.06

Sono di seguito mostrati e discussi i risultati ottenuti dal classificatore neurale che considera come variabili in ingresso **durata**, **intensità di pioggia** e i **valori di umidità del suolo di composizioni di più strati** in cui è stato discretizzato il terreno. I dati per il caso studio italiano sono riportati in Figura 6.9 e Tabella 6.9, mentre quelli relativi al caso di studio norvegese sono riportati in Figura 6.10 e Tabella 6.10.

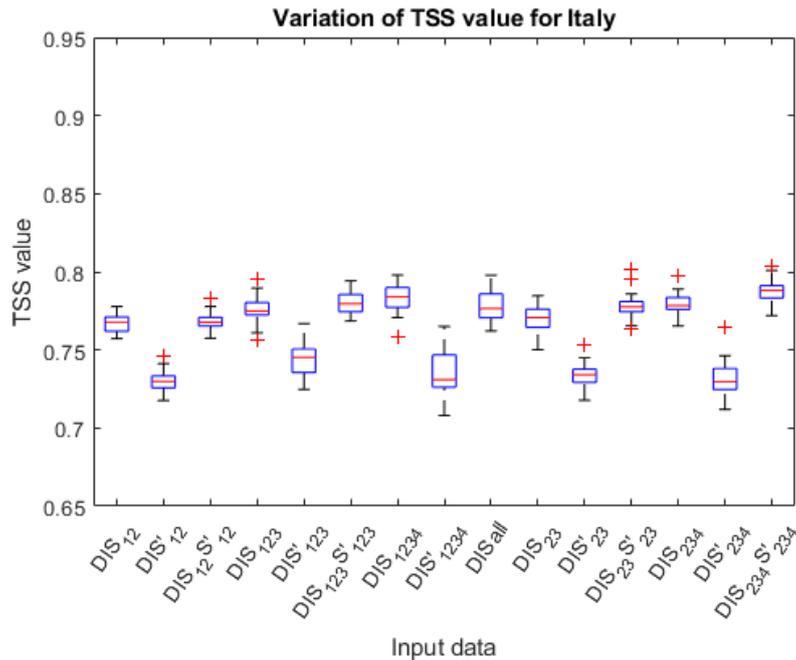


Figura 6.9: Boxplot del classificatore che considera i dati D-I-S relativi a più strati (IT)

Tabella 6.9: Risultati del classificatore che considera i dati D-I-S relativi a più strati (IT)

Input (IT)	Num. med. Neuron.	TSS tot med.	TSS med. train	TSS med. valid	TSS med. test	TPR med.	FPR med.	Δ
DIS ₁ S ₂	13.20	0.77	0.79	0.74	0.74	0.87	0.21	0.03
DIS' ₁ S' ₂	15.43	0.73	0.76	0.71	0.71	0.88	0.16	0.07
DIS ₁₂ S' ₁₂	14.70	0.77	0.80	0.75	0.73	0.88	0.11	0.03
DIS ₁₂₃	14.10	0.78	0.80	0.74	0.75	0.88	0.20	0.04
DIS' ₁₂₃	16.10	0.74	0.77	0.73	0.72	0.91	0.12	0.12
DIS ₁₂₃ S' ₁₂₃	13.17	0.78	0.81	0.75	0.74	0.89	0.06	0.04
DIS ₁₂₃₄	15.53	0.78	0.81	0.76	0.76	0.89	0.21	0.05
DIS' ₁₂₃₄	14.47	0.74	0.76	0.72	0.72	0.90	0.08	0.05
DIS _{all}	14.80	0.78	0.80	0.76	0.75	0.89	0.05	0.05
DIS ₂₃	14.43	0.77	0.79	0.75	0.75	0.89	0.18	0.04
DIS' ₂₃	15.70	0.73	0.76	0.71	0.71	0.90	0.19	0.06
DIS ₂₃ S' ₂₃	12.20	0.78	0.80	0.75	0.74	0.88	0.14	0.06
DIS ₂₃₄	14.60	0.78	0.80	0.76	0.75	0.89	0.21	0.05
DIS' ₂₃₄	15.43	0.73	0.76	0.72	0.71	0.90	0.14	0.07
DIS ₂₃₄ S' ₂₃₄	15.90	0.79	0.82	0.76	0.74	0.89	0.07	0.06

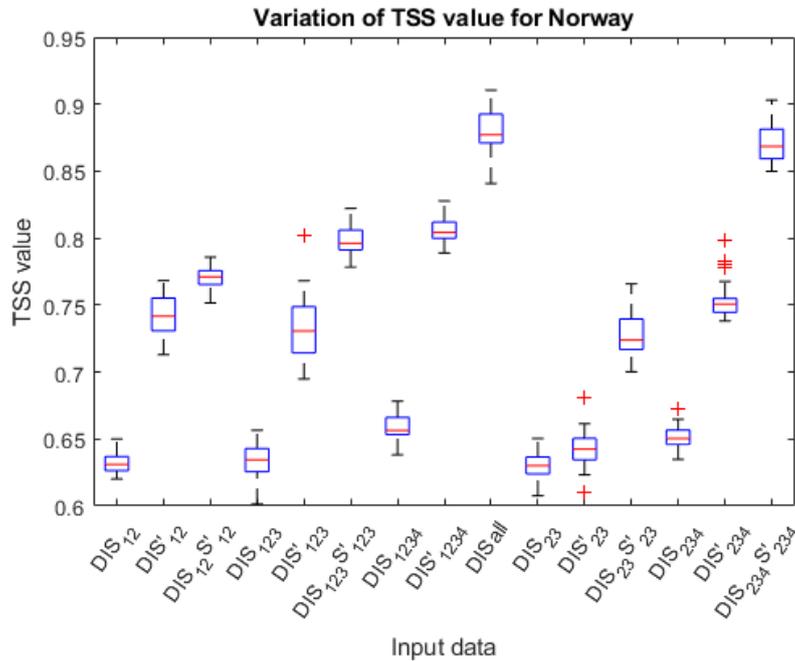


Figura 6.10: Boxplot del classificatore che considera i valori di D-I-S relativi a più strati (NO)

I risultati del classificatore che considera come variabili in ingresso la durata, l'intensità di precipitazione e i valori dei vari strati di umidità del suolo seguono lo stesso trend dell'elaborazioni presentate al paragrafo precedente. Per la regione Sicilia anche in questo caso i valori di TSS medi sono compresi tra 0.7 e 0.8 sebbene il range Δ risulti essere più elevato. Le performance migliori si ottengono considerando le combinazioni DIS_{123} (5 variabili in input) e la combinazione $DIS_{234}S'_{234}$ (8 variabili in ingresso).

Per il caso studio norvegese si nota come in funzione delle variabili in ingresso le performance del predittore variano nel range di TSS compreso tra 0.6 e 0.9. Anche in questo caso le combinazioni $DIS_{234}S'_{234}$ e DIS_{all} caratterizzate da un elevato numero di variabili sono quelle per cui si ottiene il valore di TSS massimo prossimo a 0.9.

Tabella 6.10: Risultati del classificatore che considera i valori di D-I-S relativi a più strati (NO)

Input (NO)	Num.	TSS	TSS	TSS	TSS	TPR med.	FPR med.	Δ
	med. Neuron.	tot med.	med. train	med. valid	med. test			
DIS _{1S₂}	15.27	0.63	0.66	0.61	0.61	0.84	0.20	0.03
DIS' _{1S'₂}	16.20	0.74	0.76	0.73	0.72	0.88	0.14	0.06
DIS _{12S'₁₂}	16.33	0.77	0.79	0.74	0.75	0.86	0.09	0.03
DIS ₁₂₃	13.87	0.63	0.66	0.61	0.61	0.82	0.19	0.06
DIS' ₁₂₃	13.90	0.73	0.76	0.72	0.71	0.85	0.12	0.11
DIS _{123S'₁₂₃}	16.07	0.80	0.82	0.79	0.78	0.87	0.07	0.04
DIS ₁₂₃₄	15.77	0.66	0.68	0.65	0.64	0.86	0.20	0.04
DIS' ₁₂₃₄	14.17	0.81	0.82	0.80	0.79	0.90	0.10	0.04
DIS _{all}	16.50	0.88	0.90	0.86	0.86	0.93	0.05	0.07
DIS ₂₃	15.40	0.63	0.65	0.61	0.61	0.82	0.19	0.04
DIS' ₂₃	14.37	0.64	0.67	0.62	0.62	0.83	0.19	0.07
DIS _{23S'₂₃}	16.10	0.73	0.75	0.71	0.70	0.88	0.16	0.07
DIS ₂₃₄	14.93	0.65	0.67	0.63	0.63	0.86	0.21	0.04
DIS' ₂₃₄	14.63	0.75	0.77	0.74	0.74	0.89	0.14	0.06
DIS _{234S'₂₃₄}	16.40	0.87	0.89	0.86	0.85	0.94	0.07	0.05

Di seguito verranno mostrati e discussi i risultati ottenuti dal classificatore neurale che considera come dati di input i valori di **durata**, **precipitazione cumulata**, **intensità di pioggia** e i **valori di umidità del suolo della composizione di più strati**, in cui è stato discretizzato il terreno nel database ERA5. I dati per il caso studio italiano sono riportati in Figura 6.11 e Tabella 6.11, mentre quelli relativi al caso di studio norvegese sono riportati in Figura 6.12 e Tabella 6.12.

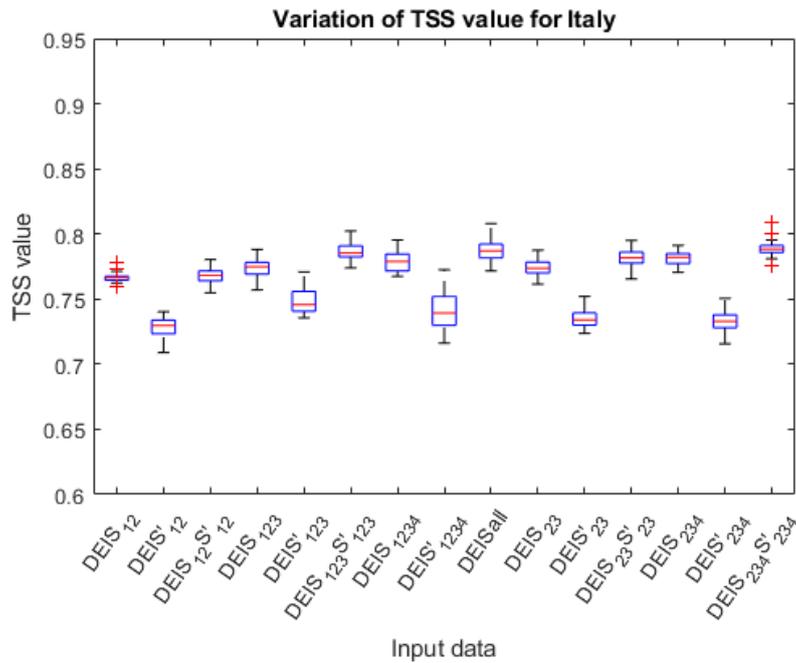


Figura 6.11: Boxplot del classificatore che considera i valori di D-E-I-S relativi a più strati (IT)

Tabella 6.11: Risultati del classificatore che considera i valori di D-E-I-S relativi a più strati (IT)

Input (IT)	Num. med. Neuron.	TSS tot med.	TSS med. train	TSS med. valid	TSS med. test	TPR med.	FPR med.	Δ
DEIS ₁ S ₂	14.17	0.77	0.80	0.74	0.74	0.89	0.12	0.02
DEIS' ₁ S' ₂	15.47	0.73	0.75	0.70	0.71	0.88	0.15	0.03
DEIS ₁₂ S' ₁₂	14.53	0.77	0.80	0.74	0.73	0.87	0.10	0.03
DEIS ₁₂₃	13.97	0.77	0.80	0.75	0.75	0.89	0.11	0.03
DEIS' ₁₂₃	15.30	0.75	0.78	0.73	0.73	0.91	0.16	0.04
DEIS ₁₂₃ S' ₁₂₃	15.63	0.79	0.81	0.76	0.75	0.89	0.10	0.03
DEIS ₁₂₃₄	14.23	0.78	0.81	0.76	0.75	0.89	0.11	0.03
DEIS' ₁₂₃₄	15.53	0.74	0.77	0.72	0.71	0.90	0.16	0.06
DEIS _{all}	15.33	0.79	0.81	0.76	0.76	0.89	0.10	0.04
DEIS ₂₃	14.57	0.77	0.80	0.75	0.74	0.89	0.12	0.03
DEIS' ₂₃	13.93	0.74	0.76	0.71	0.73	0.90	0.17	0.03
DEIS ₂₃ S' ₂₃	14.00	0.78	0.81	0.75	0.75	0.89	0.11	0.03
DEIS ₂₃₄	14.00	0.78	0.81	0.76	0.74	0.88	0.10	0.02
DEIS' ₂₃₄	15.50	0.73	0.76	0.71	0.70	0.90	0.17	0.03
DEIS ₂₃₄ S' ₂₃₄	14.83	0.79	0.82	0.75	0.75	0.88	0.09	0.03

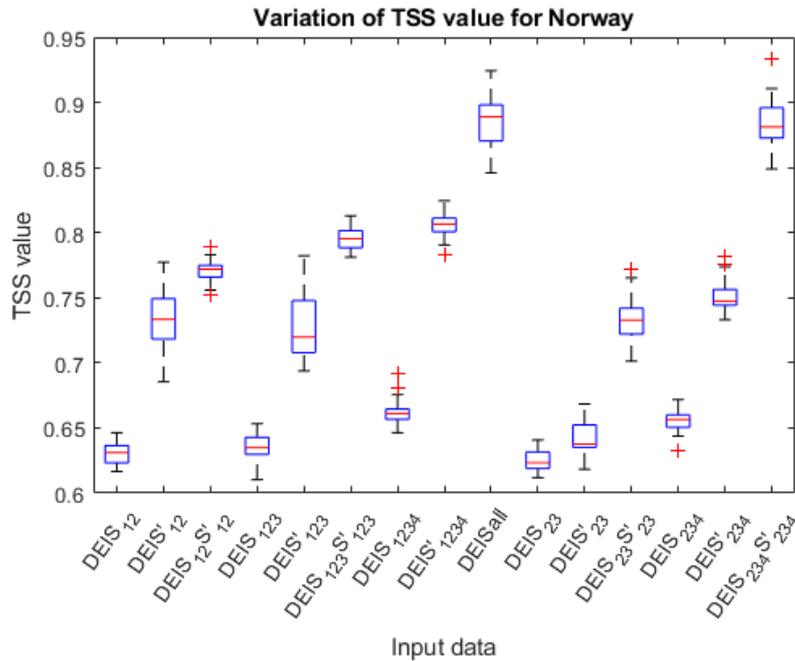


Figura 6.12: Boxplot del classificatore che considera valori di D-E-I-S relativi a più strati (NO)

Le elaborazioni presentate considerano tutte le informazioni relative alle precipitazioni disponibili, durata, cumulata, intensità, e le combinazioni dei valori di umidità del suolo come variabili di input per il classificatore. Per quanto riguarda la regione Sicilia è possibile osservare che l'utilizzo delle tre variabili di precipitazione non fa ottenere al classificatore un incremento di performance che continuano ad attestarsi su valori medi di TSS compresi tra 0.7 e 0.8, tuttavia diminuisce di molto la variabilità Δ che l'inizializzazione random della rete neurale ha sulla capacità previsionale.

Per l'area di studio in Norvegia, invece, la presenza di tre variabili di precipitazione incrementa la capacità del classificatore di riconoscere le condizioni che possono portare all'innescò considerando un TSS medio in alcune configurazioni prossimo a 0.9, tuttavia il valore di Δ medio risulta essere molto elevato e prossimo a 0.053.

Tabella 6.12: Risultati del classificatore che considera i valori di D-E-I-S relativi a più strati (NO)

Input (NO)	Num. med. Neuron.	TSS tot med.	TSS med. train	TSS med. valid	TSS med. test	TPR med.	FPR med.	Δ
DEIS _{1S₂}	15.33	0.63	0.66	0.61	0.62	0.85	0.21	0.03
DEIS' _{1S'₂}	15.47	0.74	0.76	0.73	0.71	0.88	0.14	0.09
DEIS _{12S'₁₂}	15.77	0.77	0.79	0.75	0.76	0.88	0.10	0.04
DEIS ₁₂₃	15.47	0.64	0.66	0.62	0.60	0.83	0.19	0.04
DEIS' ₁₂₃	16.43	0.73	0.75	0.71	0.71	0.85	0.12	0.09
DEIS _{123S'₁₂₃}	14.93	0.80	0.81	0.77	0.78	0.85	0.06	0.03
DEIS ₁₂₃₄	14.63	0.66	0.68	0.65	0.65	0.86	0.20	0.04
DEIS' ₁₂₃₄	14.57	0.81	0.83	0.79	0.78	0.90	0.10	0.04
DEIS _{all}	16.00	0.89	0.90	0.88	0.87	0.94	0.05	0.08
DEIS ₂₃	14.03	0.62	0.65	0.61	0.60	0.81	0.18	0.03
DEIS' ₂₃	16.33	0.64	0.66	0.62	0.62	0.83	0.19	0.05
DEIS _{23S'₂₃}	16.13	0.73	0.75	0.72	0.72	0.88	0.14	0.07
DEIS ₂₃₄	15.33	0.66	0.68	0.65	0.63	0.86	0.21	0.04
DEIS' ₂₃₄	13.60	0.75	0.77	0.74	0.73	0.89	0.14	0.05
DEIS _{234S'₂₃₄}	16.87	0.88	0.90	0.87	0.87	0.95	0.07	0.08

6.6 DISCUSSIONE DEI RISULTATI

Le elaborazioni hanno mostrato come all'alternarsi delle caratteristiche delle precipitazioni utilizzate come input (D, E, I), considerando gli stessi valori di umidità del suolo, è possibile vedere per le singole aree di studio, un trend simile. Risulta invece variabile l'incertezza legata alla suddivisione dell'intero database nei tre subset (training, validazione e test) utilizzati della rete neurale, mostrata graficamente dall'ampiezza dei diagrammi a scatola e baffi e descritto dal parametro Δ .

Sulla base dei risultati ottenuti per l'area di studio della Sicilia è possibile trarre le seguenti conclusioni:

1. Le performance ottenute da tutte le configurazioni che considerano un singolo layer di suolo ($D-E$; $D-I$; $D-E-I$) si attestano su un TSS massimo prossimo a 0.76;
2. Il range Δ medio per le configurazioni viste al punto precedente è compreso tra 0.025 e 0.027, ma il valore medio minimo si ottiene per la configurazione $D-E-I$ in cui l'introduzione dell'intensità di precipitazione abbassa la variabilità media ($\Delta = 0.024$);
3. Per le configurazioni che considerano combinazioni di più strati di umidità il valore massimo del TSS medio si attesta attorno a valori compresi tra 0.79 e 0.80;
4. Considerando le composizioni di più strati le performance migliori si ottengono per la configurazione che considera come dati in input la durata cumulata e le informazioni di umidità del suolo del secondo, terzo e quarto strato considerando i valori all'istante iniziale e finale.
5. Al variare delle combinazioni dei dati di umidità del suolo si registra una limitata variabilità delle prestazioni medie in termini di TSS compresa tra 0.7 e 0.8.

Per quanto riguarda l'area di studio norvegese si può arrivare alle seguenti conclusioni:

1. Considerando i risultati ottenuti per i singoli layer, i valori di TSS migliori si ottengono in corrispondenza del quarto strato di suolo con tutte le configurazioni di dati di precipitazione ($D-E$; $D-I$; $D-E-I$);

2. Le performance in termini di TSS risultano essere simili e si attestano a valori prossimi a $TSS = 0.79$, sebbene il valore di Δ minore si ottenga per la configurazione *D-E*;
3. Considerando la combinazione di più strati di umidità in tutti i casi si è possibile vedere che la composizione che porta al migliore risultato è quella che considera tutte le informazioni disponibili (*Sall*) con valori di TSS compresi tra 0.87 e 0.89.
4. Anche in questo caso il migliore compromesso tra le performance e la variabilità è dato dalla combinazione che considera come valori di input la durata e la precipitazione cumulata
5. Al variare delle combinazioni dei dati di umidità del suolo, considerando il caso in cui sono utilizzati più layer, si osserva una marcata variabilità media del TSS. I valori di TSS medi, infatti, oscillano tra 0.64 e 0.89. La variabilità risulta molto più marcata di quella vista per il caso italiano)

Di particolare rilevanza risultano anche le informazioni legate a combinazioni che considerano due variabili di precipitazione e un singolo valore di umidità del suolo. Utilizzando queste informazioni, in studi futuri potrebbero essere messe a punto soglie 3D, come quelle presentate da Rosi et al. (2021).

Restringendo il campo a questa specifica categoria, considerando quindi solo tre variabili di input, per l'area di studio siciliana è possibile osservare che:

1. Le performance migliori si ottengono considerando il valore di umidità iniziale per il secondo strato di suolo;
2. I valori di performance in termini di TSS delle coppie *D-E* e *D-I* risultano essere simili ($TSS_{tot} = 0.76$) sebbene la configurazione che consideri durata e intensità abbia valori di Δ minori

Per l'aria studio che considera l'unione dei 5 bacini a est di Bergen è possibile osservare che:

1. Considerando la configurazione che utilizza come dati di input la durata e la precipitazione cumulata, le performance migliori si ottengono considerando il valore di umidità del primo strato all'istante finale, anche se il valore di Δ risulta essere elevato ($TSS = 0.7$; $\Delta = 0.06$). L'utilizzo del valore di umidità all'istante finale non può tuttavia essere noto in un sistema di early warning ma è comunque utile per procedure di back-analysis;

2. Considerando come dati di input i valori di intensità e durata le performance migliori si ottengono considerando i valori di umidità del suolo del quarto strato nell'istante finale ($TSS = 0.69$; $\Delta=0.03$). Come detto al punto precedente il valore di umidità all'istante finale può presentare problemi nella determinazione, inoltre il fatto che il quarto strato è quello più rappresentativo è controintuitivo quando si parla di frane superficiali.

CAPITOLO 7 CONCLUSIONI

7.1 CONCLUSIONI

Tra i fenomeni naturali di dissesto idrogeologico le frane innescata da precipitazione sono tra i più impattanti, ricorrenti e talvolta catastrofici. Le strategie per la mitigazione del rischio, per questa tipologia di eventi possono prevedere un approccio di tipo strutturale o uno di tipo non strutturale. La presente attività di ricerca ha preso in considerazione l'approccio non strutturale.

Il lavoro presentato ha avuto come obiettivo quello di proporre strategie innovative per la determinazione di soglie pluviometriche riguardanti l'innescio di movimenti franosi.

Le soglie pluviometriche di innescio sono una delle componenti alla base di un sistema di early warning relativo a fenomeni di movimenti franosi affidabile (caratterizzato da un basso numero di falsi positivi e falsi negativi). La tematica è di grande importanza in quanto ha un impatto diretto su beni materiali e sulla sicurezza delle persone.

L'esame della bibliografia e le elaborazioni effettuate mostrano come, sebbene nel campo delle frane innescate da precipitazione esiste una vasta letteratura scientifica, siano ancora presenti alcune criticità come:

1. Procedure non sempre oggettive e ripetibili che rendono difficile generalizzare i risultati ottenuti;
2. Mancanza di informazioni riguardanti l'innescio di movimenti franosi;
3. Strumenti di misura delle precipitazioni che non garantiscono la totale copertura delle aree di studio e/o mancanza di manutenzione.

In tal senso molto è stato fatto da svariati autori e da molti enti che mettono a disposizione in maniera free e open source dati di precipitazione e database di eventi franosi.

I dati di precipitazione e i database contenenti informazioni sulle frane avvenute nel passato, sono state utilizzate per determinare per le due aree di studio analizzate, la regione Sicilia per l'Italia e l'area a est di Bergen per la Norvegia, soglie empiriche di innescio durata-precipitazione cumulata. Il primo approccio prevede di utilizzare una procedura automatizzata, che determina le soglie di innescio tenendo conto solo delle caratteristiche degli eventi innescanti. Attraverso un approccio più

innovativo sono state determinate, per entrambe le aree di studio, soglie che tengono conto sia di eventi innescanti che di eventi non innescanti utilizzando una procedura che mira a massimizzare il True Skill Statistic, un indicatore generalmente utilizzato per definire le performance di un sistema di previsione.

Attraverso un approccio che utilizza le reti neurali artificiali sono state testate le capacità degli algoritmi di machine learning di riconoscere e distinguere gli eventi innescanti da quelli non innescanti. Dai risultati mostrati nei Capitoli 5 e 6, è possibile osservare che utilizzando gli stessi dati in ingresso le performance del classificatore neurale in termini di TSS risultino superiori a quelle delle comuni soglie intensità-durata determinate con legge di tipo potenza. Si riscontra, infatti, un incremento del TSS che passa da 0.5, delle soglie pluviometriche con legge potenza, a valori prossimi a 0.6 considerando i risultati delle ANN. In particolare, per il caso studio della Sicilia dove si dispone di dati a scala oraria, introducendo l'intensità di picco è possibile osservare un importante incremento del valore di TSS, che passa da 0.5 a 0.66.

Anche per il caso studio norvegese è stato riscontrato un aumento delle performance, infatti, ricorrendo al classificatore neurale il TSS passa da valori prossimi a 0.5 a valori prossimi a 0.58. Contrariamente a quanto visto per la Sicilia l'introduzione di una terza variabile idrologica (intensità media) non migliora le performance del sistema di previsione. Ragionevolmente questo accade per la differenza di regime climatico tra le due aree esaminate.

La possibilità di introdurre più variabili all'interno delle reti neurali è stata sfruttata introducendo, in una seconda fase del lavoro, i dati di umidità del suolo ottenuti dal progetto di rianalisi ERA5-Land. Tale scelta dipende dalla possibilità di avere dati coerenti per entrambe le aree di studio esaminate, alla scala oraria per una ampia finestra temporale (1950-giorni nostri).

I dati di umidità del suolo di ERA5-Land, prima di essere utilizzati per le elaborazioni sono stati confrontati con dati provenienti da stazioni di misura. Il risultato di questi confronti ha confermato che i dati di ERA5 forniscono una buona rappresentazione dell'andamento dell'umidità del suolo; tuttavia, i valori soffrono di una sistematica sovrastima rispetto ai valori rilevati dagli strumenti sia per l'Italia che per la stazione di Eidsvoll in Norvegia. I dati relativi alla stazione di Hobøl invece, per la presenza di più sensori installati a diverse profondità descrivono meglio l'andamento del contenuto volumetrico d'acqua. Per tutte le stazioni studiate è stato

osservato che il miglior riscontro tra dati ERA5-Land e misurazioni in situ si ottiene considerando la discretizzazione del terzo strato di terreno

Introducendo i valori di umidità del suolo, come singoli strati o come combinazione di più strati, sono state effettuate 77760 simulazioni al fine di investigare anche gli effetti dell'inizializzazione random delle reti neurali sul database di partenza.

L'introduzione dei dati di umidità del suolo all'interno del classificatore neurale comporta un notevole miglioramento delle performance dei valori in termini di TSS. Considerando già i dati per singoli strati di umidità del suolo, si ha un incremento dei valori medi del TSS da 0.58 ± 0.59 a valori compresi tra $0.75 \div 0.8$ per entrambe le aree di studio. È stato osservato che i valori migliori di TSS nelle combinazioni con singoli strati si ottengono considerando il primo e il secondo strato nel caso studio italiano e quelli relativi al primo e al quarto strato per il classificatore messo a punto per la Norvegia.

L'introduzione di combinazioni di più strati di umidità insieme alle variabili idrologiche, durata cumulata e intensità media fornisce risultati migliori per il caso studio norvegese con le configurazioni che tengono conto di tutti i dati disponibili. La presenza di più variabili in ingresso, tuttavia, comporta una maggiore variabilità nel range delle performance dovuta all'inizializzazione della rete neurale e alla fase di training che influisce sulle performance complessive delle reti. In genere è stato possibile osservare che le combinazioni in termini di performance previsionali e stabilità di previsione si ottengono considerando i valori di durata e cumulata di pioggia.

L'approccio proposto basato su un classificatore neurale, presenta tre criticità fondamentali. Fra queste la prima è la necessità di disporre di un gran mole di dati. Sebbene sia possibile ricostruire i dati di precipitazioni partendo dalle serie storiche di pioggia e necessario conoscere anche date e posizioni di eventi innescanti che spesso risultano approssimativi, incompleti o poco attendibili. Per tale motivo è stato necessario considerare aree estese, come quelle regionali, in modo poter disporre di un numero congruo di istanze di frana.

In secondo luogo, un classificatore come quello messo a punto non fornisce all'utente una legge matematica, come quella di tipo potenza (anche se, come evidenziato limita il contenuto informativo dei dati), che può essere integrata facilmente in un sistema di preallerta basata su una o più soglie.

Infine, la possibilità di introdurre un numero molto elevato di variabili che può essere visto come la possibilità di spostarsi da un modello empirico ad uno fisicamente basato, apre diversi interrogativi su quali siano le variabili da valutare di caso in caso.

La terna di criticità presentate può costituire un punto di partenza per riflessioni e lavori futuri.

APPENDICE A CONFRONTI UMIDITÀ DEL SUOLO

A.1 CONFRONTO DATI DI UMIDITÀ DEL SUOLO PER CHIARAVALLE CENTRALE

In questo capitolo verranno presentati tutti i confronti effettuati considerando i dati di umidità del suolo provenienti dal database ERA5-Land (nel seguito ERA5) e quelli ottenuti dalla rete di strumenti dell'International Soil Moisture Network. (nel seguito ISMN)

Con la nomenclatura *ERA5-nL*, si intendono i valori di umidità dell'*n-esimo* strato di suolo dei dati di rianalisi di ERA5.

Nei diamanti del diagramma di Taylor il primo numero indica lo strato perso in considerazione (primo, secondo o terzo) mentre il secondo numero indica la profondità dello strato considerato dall' International Soil Moisture Network. (ISMN)

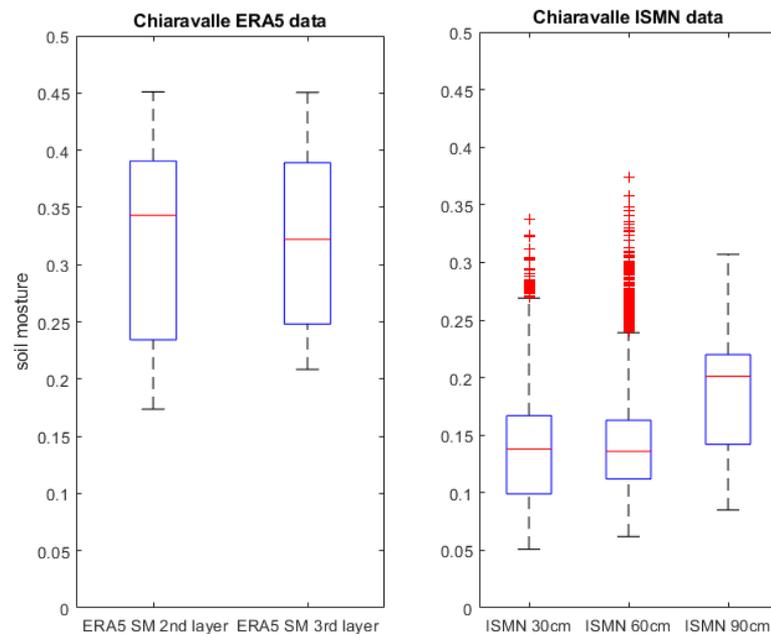


Figura A.1: Boxplot dati di umidità del suolo di Chiaravalle Centrale

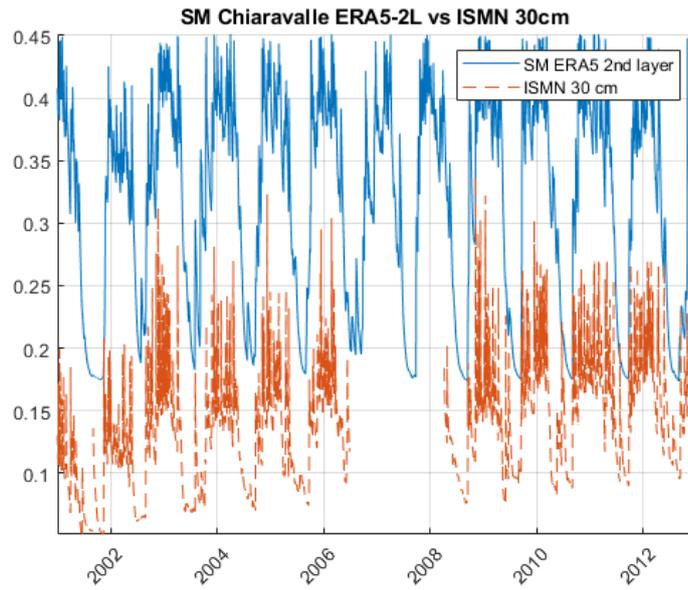


Figura A.2: Confronto umidità del suolo tra secondo strato di ERA5-Land e i dati della ISMN alla profondità di 30cm per la stazione di Chiaravalle Centrale

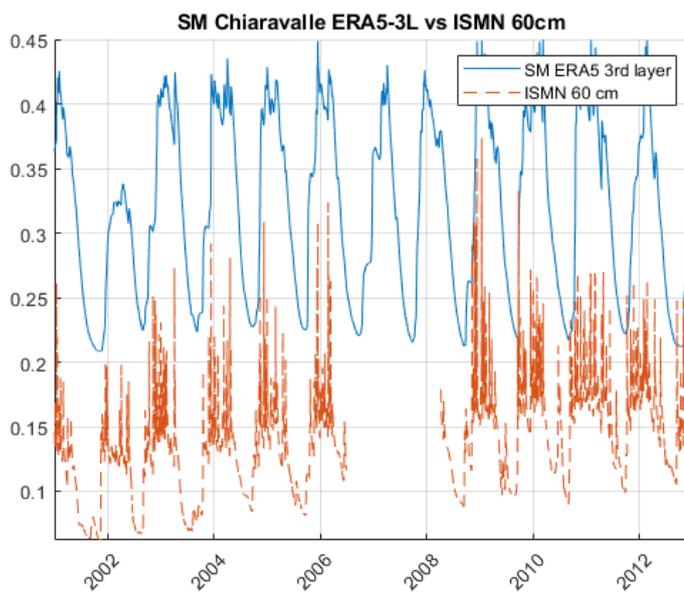


Figura A.3: Confronto umidità del suolo tra il terzo strato di ERA5-Land e i dati della ISMN alla profondità di 60cm per la stazione di Chiaravalle Centrale

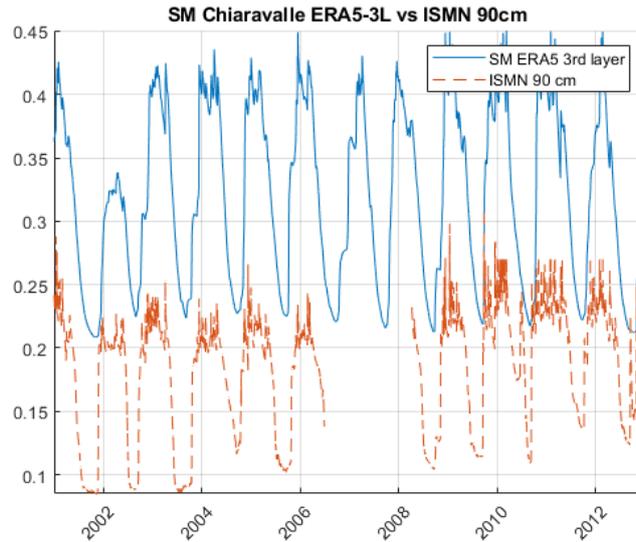


Figura A.4: Confronto umidità del suolo tra il terzo strato di ERA5-Land e i dati della ISMN profondità di 90cm per la stazione di Chiaravalle Centrale

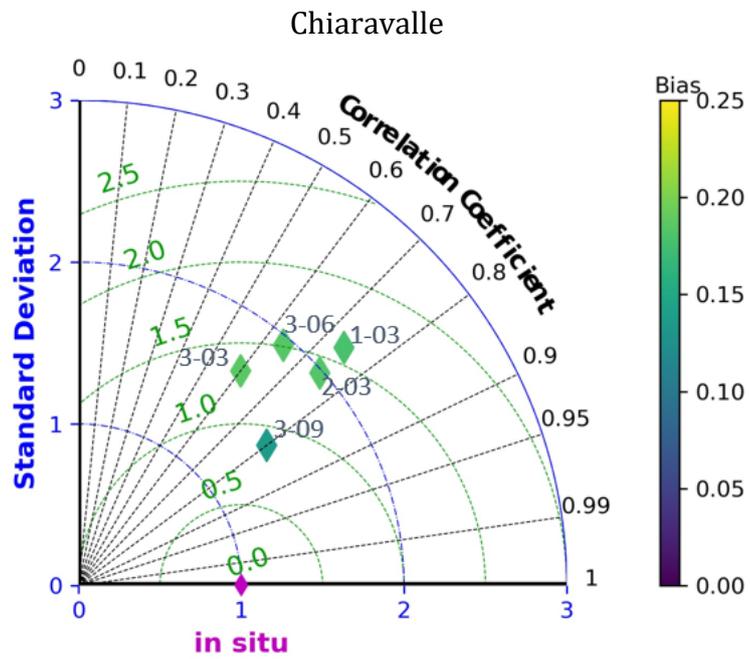


Figura A.5: Diagramma di Taylor per la stazione di Chiaravalle Centrale

A.2 CONFRONTO DATI DI UMIDITÀ DEL SUOLO PER FITTERIZZI

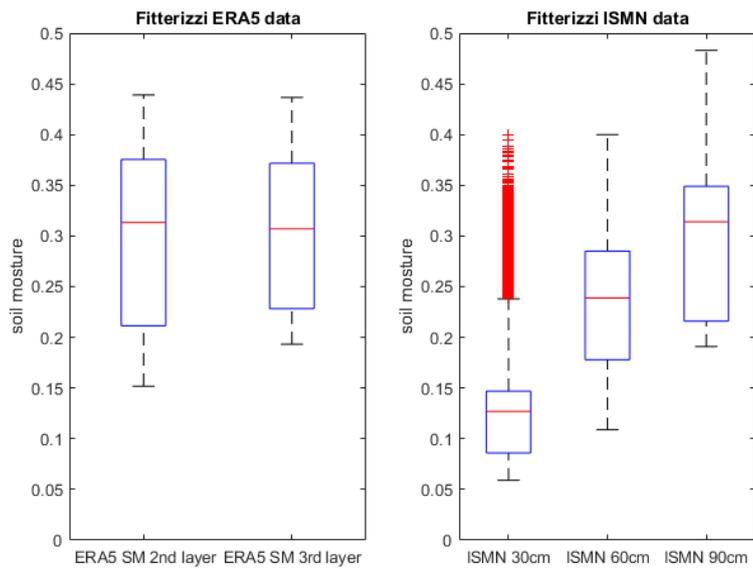


Figura A.6: Boxplot dati di umidità del suolo di Fitterizzi

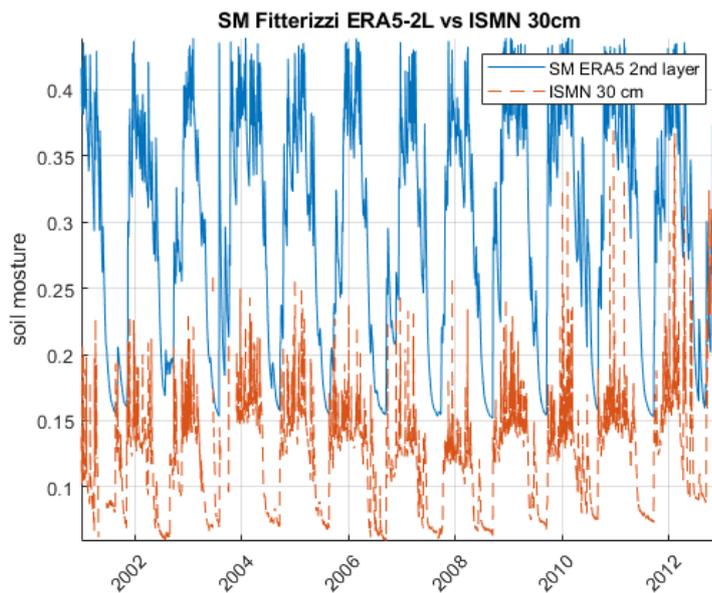


Figura A.7: Confronto umidità del suolo tra secondo strato di ERA5-Land e i dati della ISMN alla profondità di 30cm per la stazione di Fitterizzi

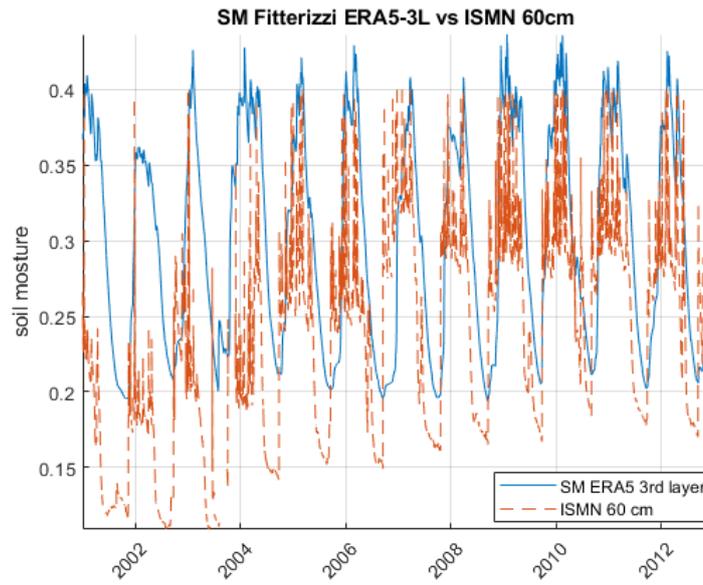


Figura A.8: Confronto umidità del suolo tra il terzo strato di ERA5-Land e i dati della ISMN alla profondità di 60cm per la stazione di Fitterizzi

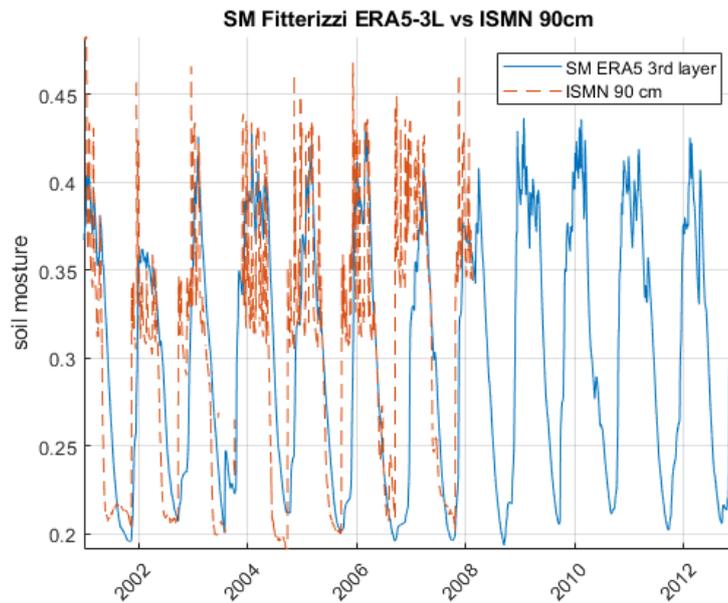


Figura A.9: Confronto umidità del suolo tra il terzo strato di ERA5-Land e i dati della ISMN profondità di 90cm per la stazione di Fitterizzi

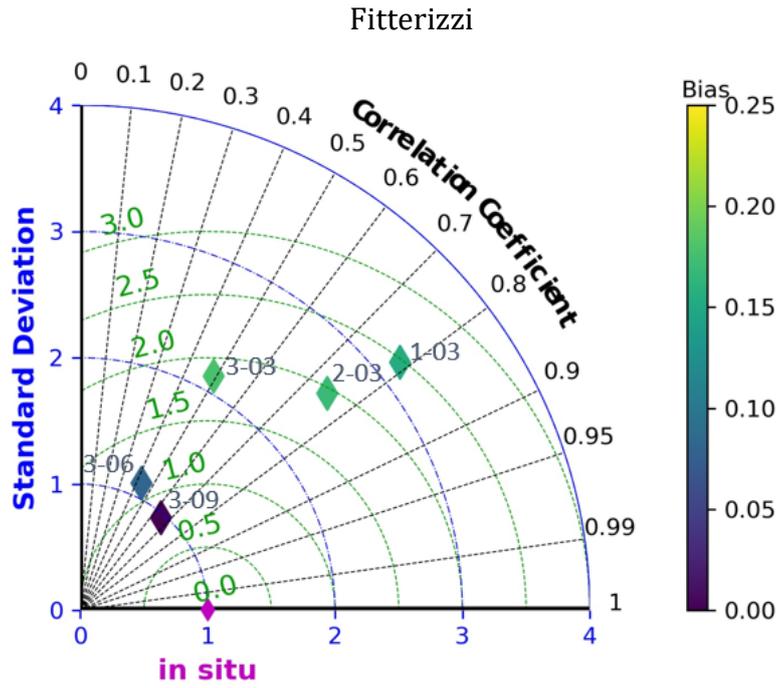


Figura A.10: Diagramma di Taylor per la stazione di Fitterizzi

A.3 CONFRONTO DATI DI UMIDITÀ DEL SUOLO PER MONGRASSANO

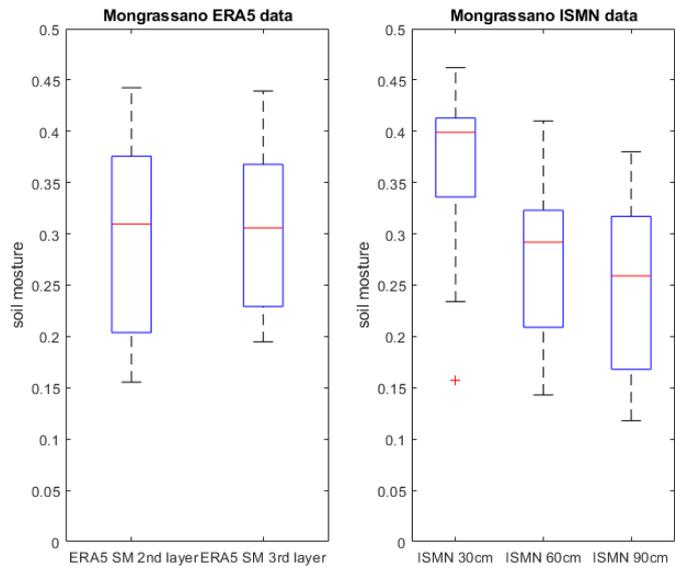


Figura A.11: Boxplot dati di umidità del suolo di Mongrassano

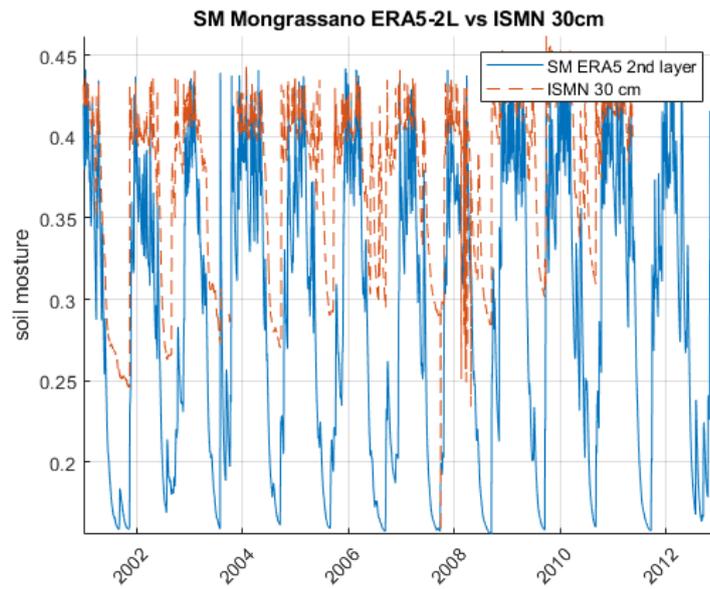


Figura A.12: Confronto umidità del suolo tra secondo strato di ERA5-Land e i dati della ISMN alla profondità di 30cm per la stazione di Mongrassano

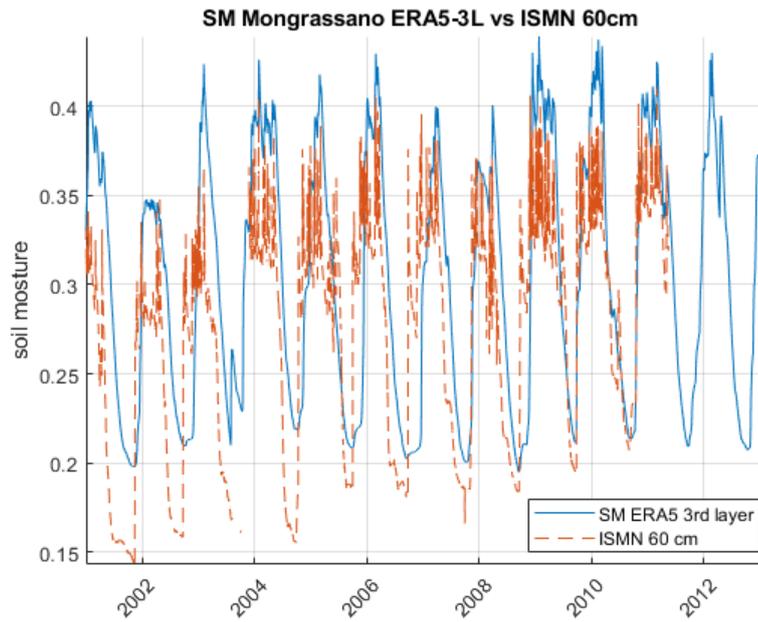


Figura A.13: Confronto umidità del suolo tra il terzo strato di ERA5-Land e i dati della ISMN alla profondità di 60cm per la stazione di Mongrassano

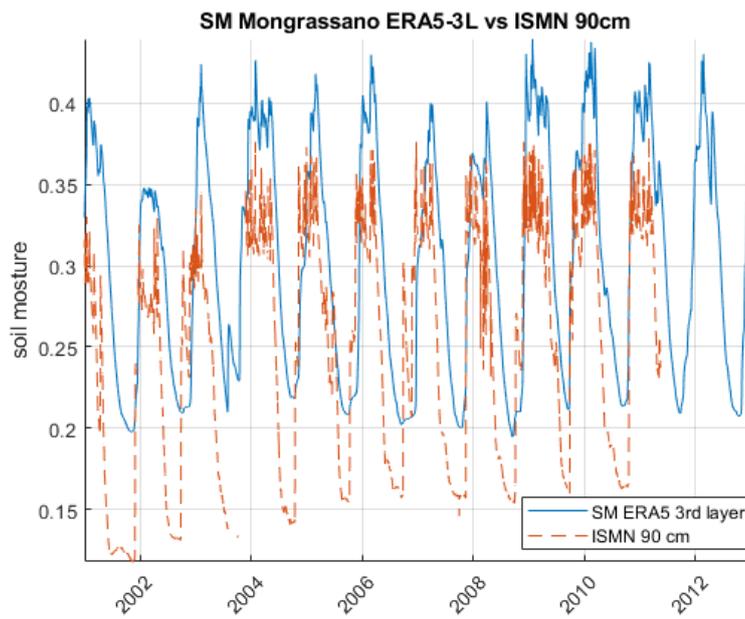


Figura A.14: Confronto umidità del suolo tra il terzo strato di ERA5-Land e i dati della ISMN profondità di 90cm per la stazione di Mongrassano

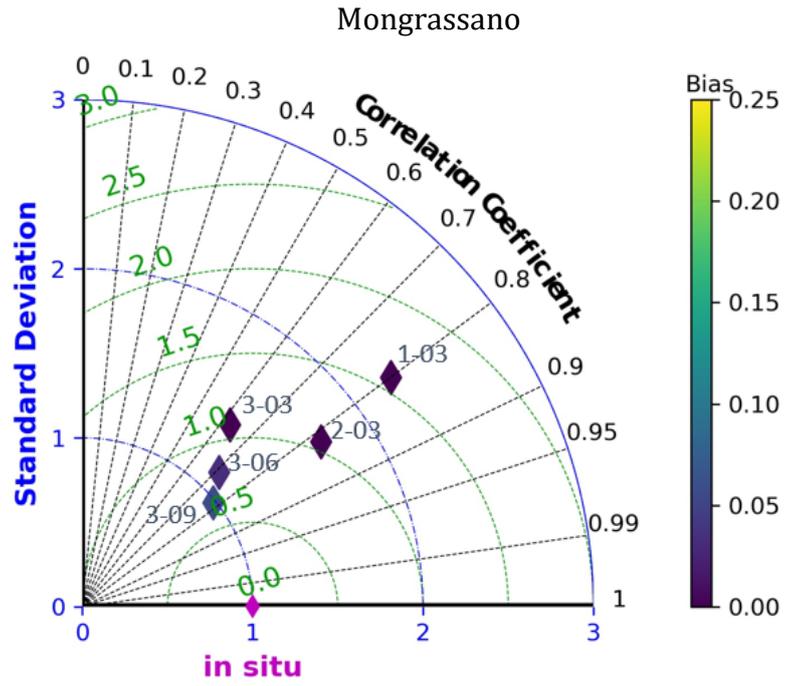


Figura A. 15: Diagramma di Taylor per la stazione di Mongrassano

A.4 CONFRONTO DATI DI UMIDITÀ DEL SUOLO PER SATRIANO

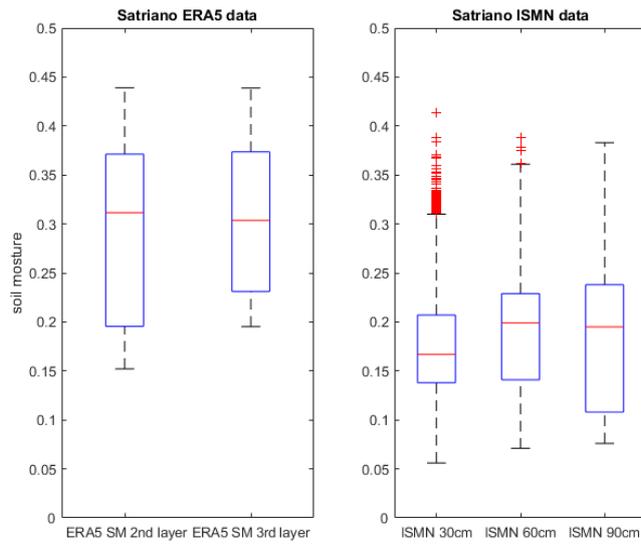


Figura A.16: Boxplot dati di umidità del suolo di Satriano

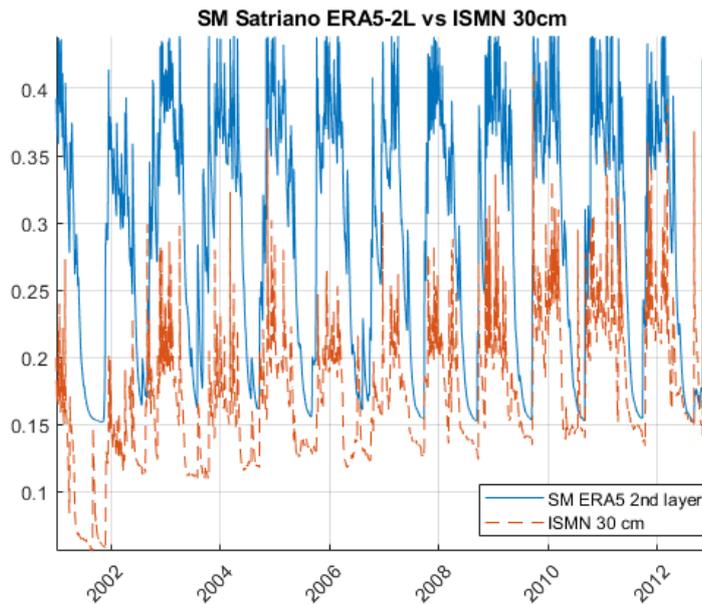


Figura A. 17: Confronto umidità del suolo tra secondo strato di ERA5-Land e i dati della ISMN alla profondità di 30cm per la stazione di Satriano

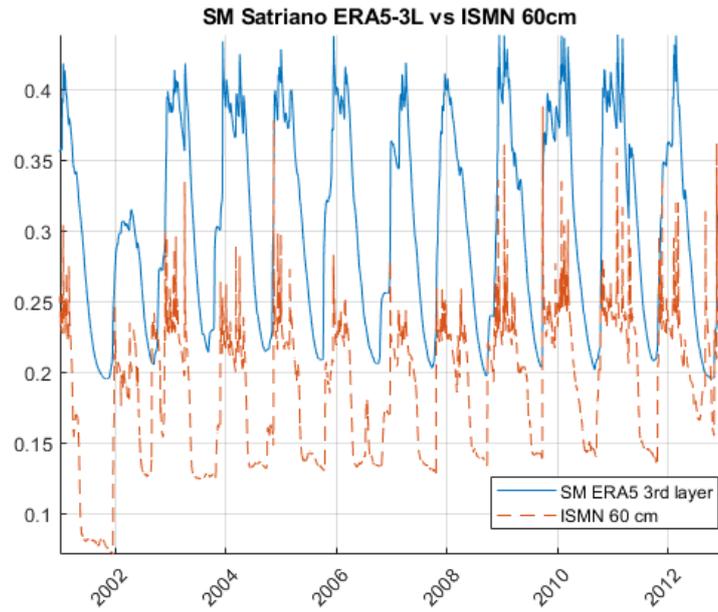


Figura A.18: Confronto umidità del suolo tra il terzo strato di ERA5-Land e i dati della ISMN alla profondità di 60cm per la stazione di Satriano

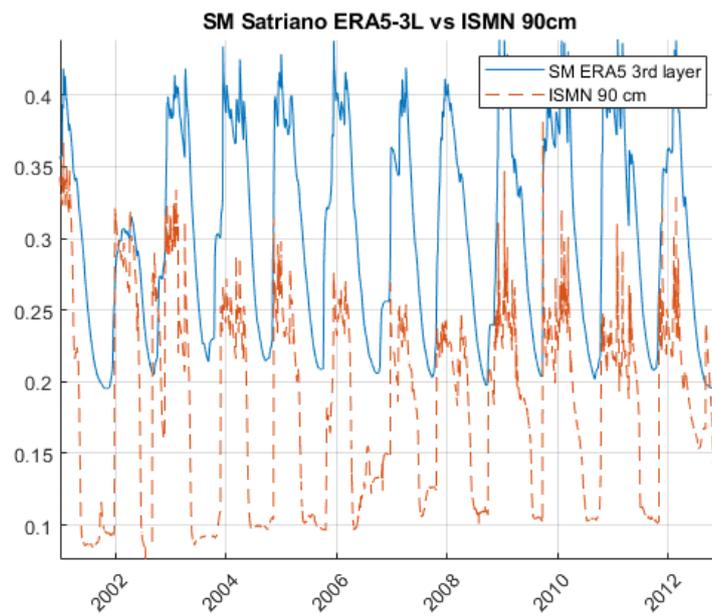


Figura A.19: Confronto umidità del suolo tra il terzo strato di ERA5-Land e i dati della ISMN profondità di 90cm per la stazione di Satriano

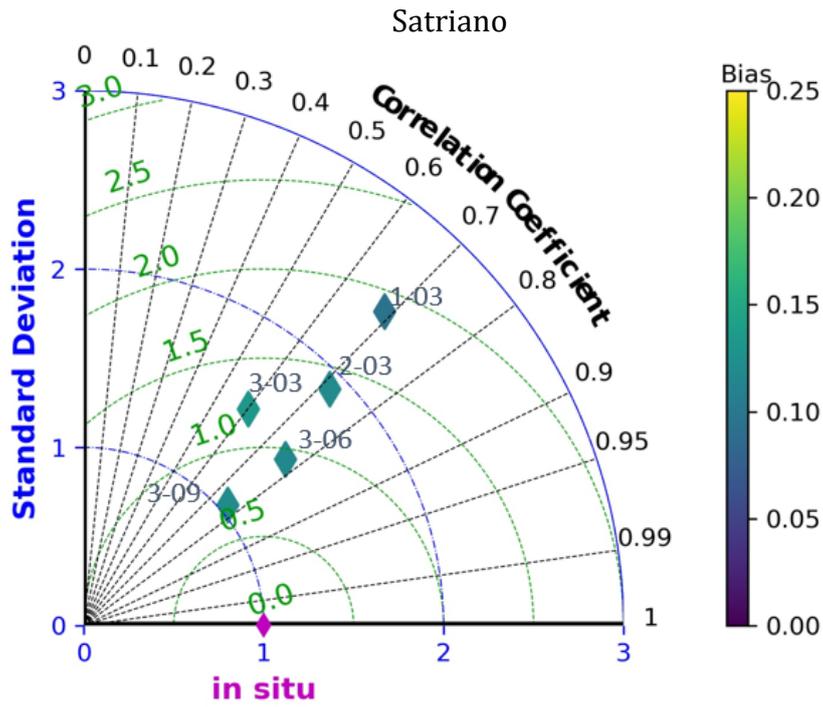


Figura A.20: Diagramma di Taylor per la stazione di Satriano

A.5 CONFRONTO DATI DI UMIDITÀ DEL SUOLO PER TORANO

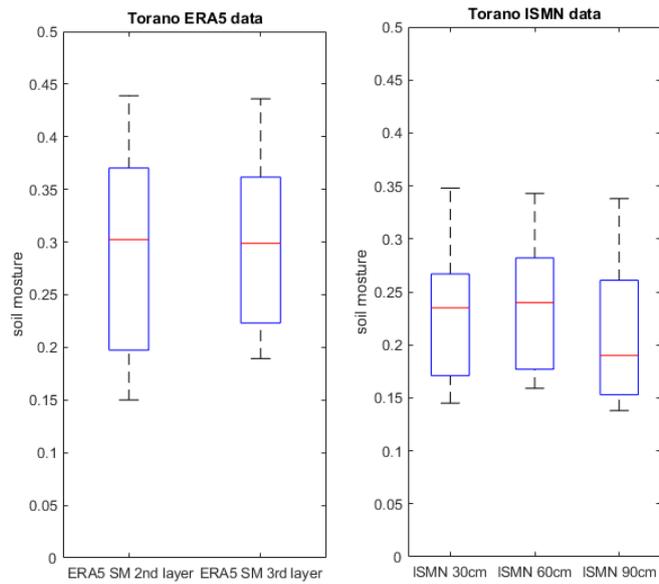


Figura A.21: Boxplot dati di umidità del suolo di Torano

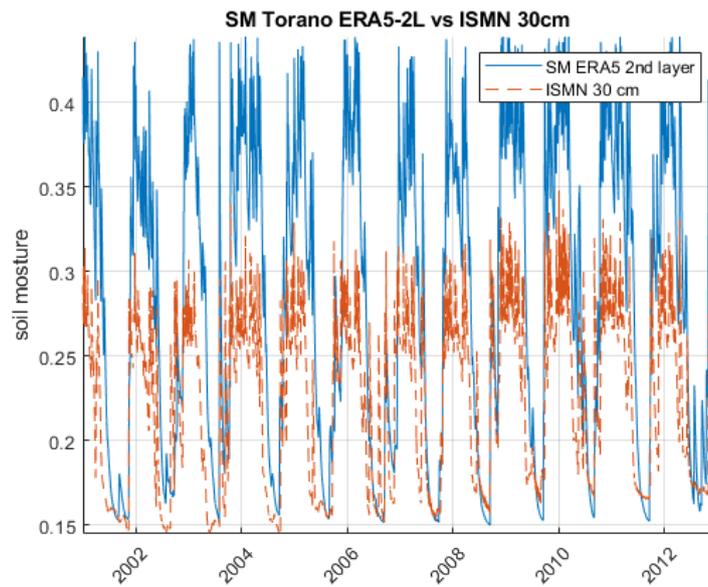


Figura A.22: Confronto umidità del suolo tra secondo strato di ERA5-Land e i dati della ISMN alla profondità di 30cm per la stazione di Torano

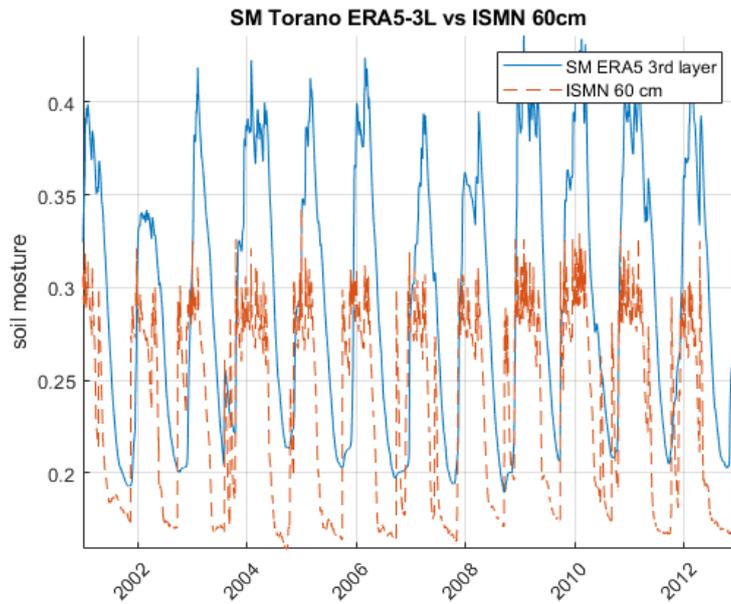


Figura A. 23: Confronto umidità del suolo tra il terzo strato di ERA5-Land e i dati della ISMN alla profondità di 60cm per la stazione di Torano

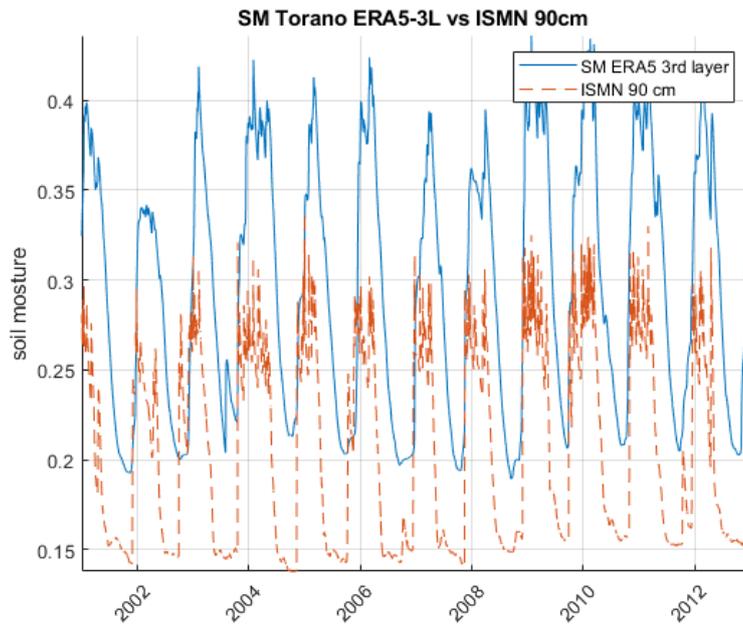


Figura A. 24: Confronto umidità del suolo tra il terzo strato di ERA5-Land e i dati della ISMN profondità di 90cm per la stazione di Torano

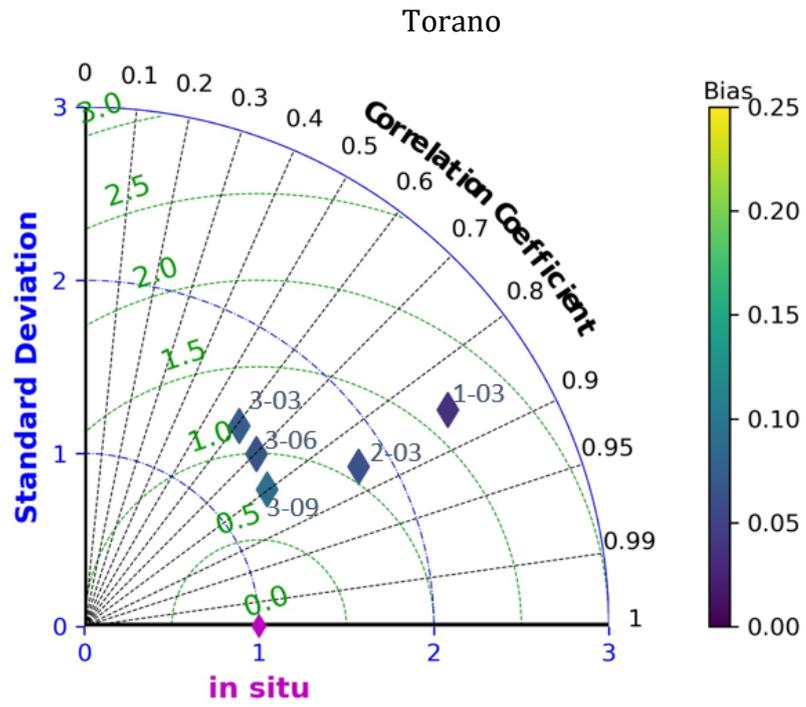


Figura A.25: Diagramma di Taylor per la stazione di Torano

A.6 CONFRONTO DATI DI UMIDITÀ DEL SUOLO PER EIDSVOLL

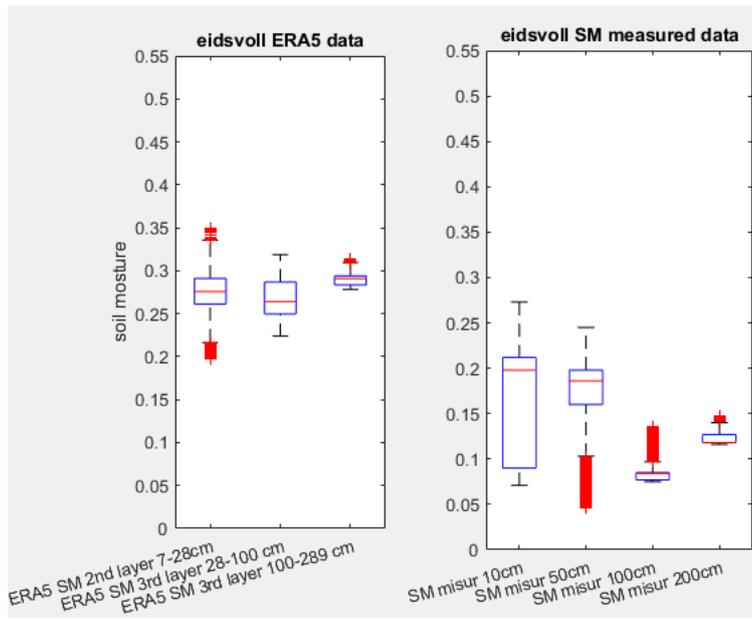


Figura A.26: Boxplot dati di umidità del suolo di Eidsvoll

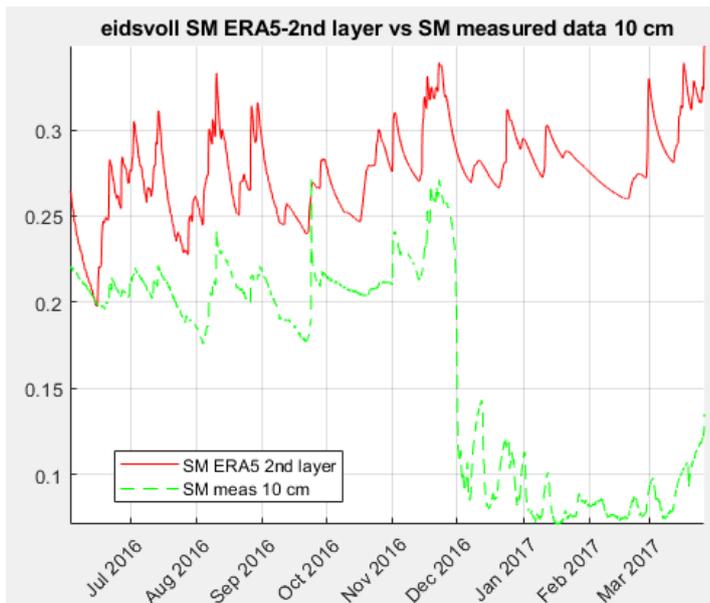


Figura A. 27: Confronto umidità del suolo tra il secondo strato di ERA5-Land e i dati della stazione di Eidsvoll alla profondità di 10cm

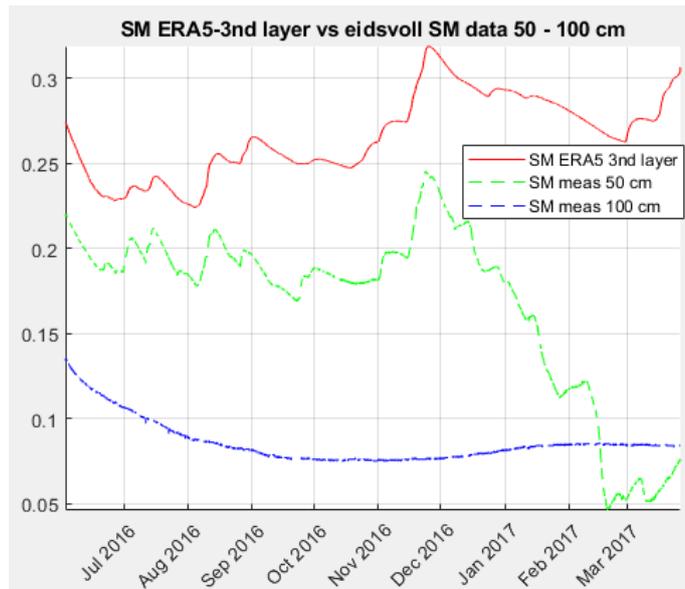


Figura A.28: Confronto umidità del suolo tra il terzo strato di ERA5-Land e i dati della stazione di Eidsvoll alla profondità di 50 e 100cm

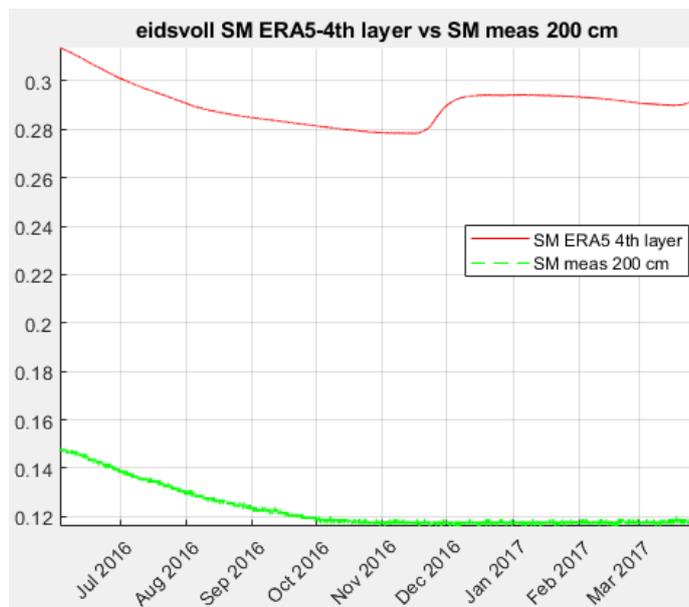
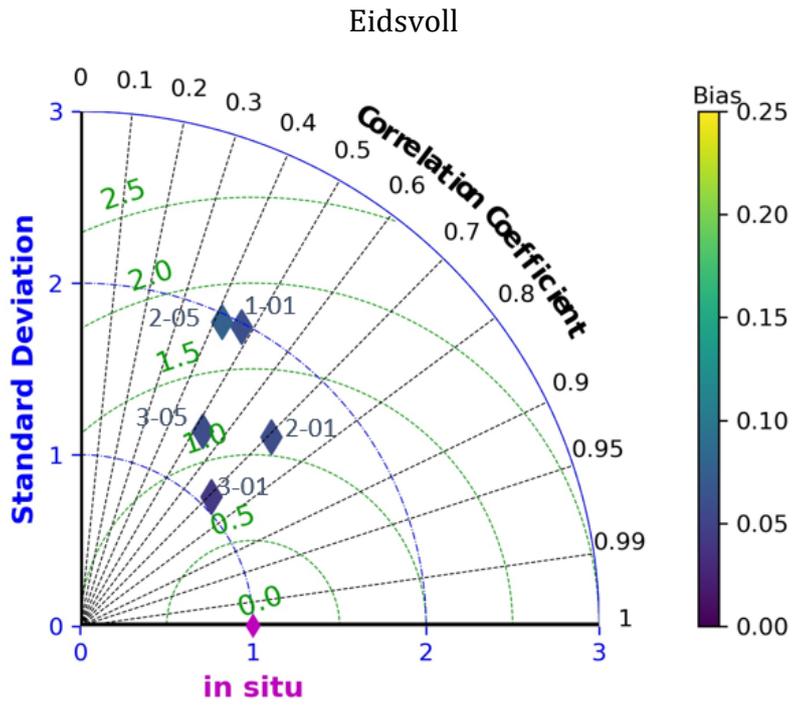


Figura A. 29: Confronto umidità del suolo tra il quarto strato di ERA5-Land e i dati della stazione di Eidsvoll alla profondità di 200cm



A.7 CONFRONTO DATI DI UMIDITÀ DEL SUOLO HOBØL

Nella stazione di monitoraggio nei pressi del fiume Hobøl si dispone di 3 diversi misuratori di umidità del suolo installati in corrispondenza di differenti tipologie di vegetazione. I grafici presentati tengono conto dei valori di umidità del suolo rispettivamente del punto di monitoraggio in corrispondenza della zona erbosa (dalla Figura A.30 alla Figura A.32), in presenza di arbusti (dalla Figura A.33 alla Figura A.35), e in presenza di alberi (dalla Figura A.36 alla Figura A.38).

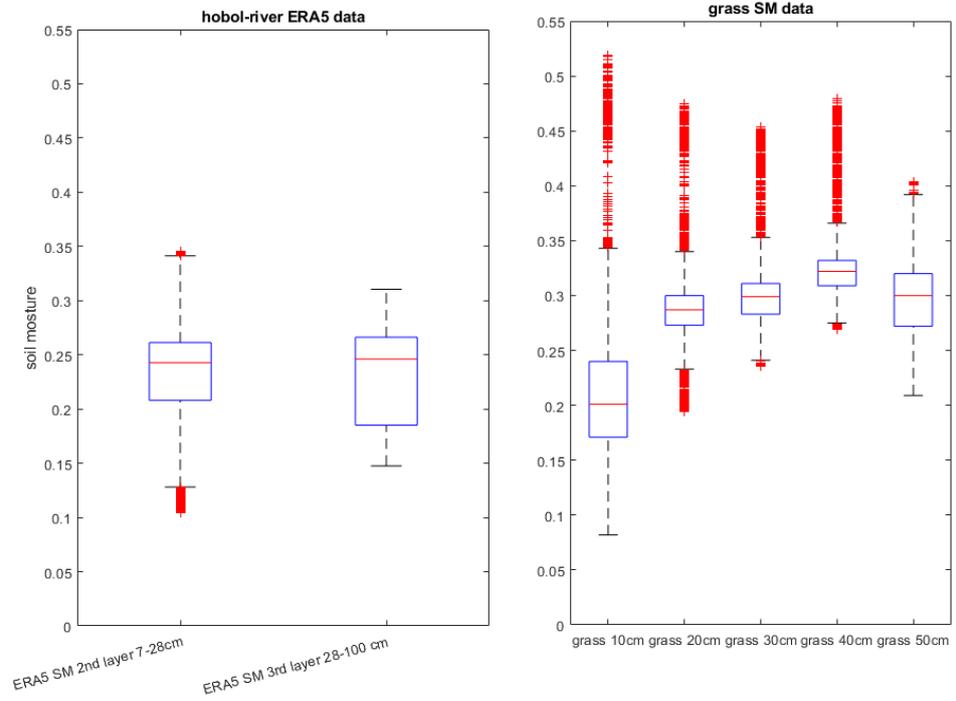


Figura A.30: Boxplot dati di umidità del suolo di Hobøl (zona erbosa)

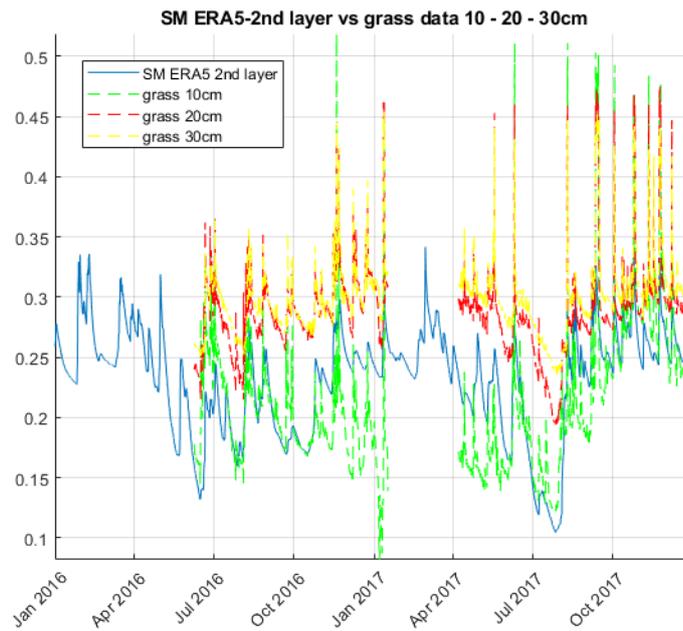


Figura A.31: Confronto umidità del suolo tra il secondo strato di ERA5e i dati della stazione di Hobøl (zona erbosa) alla profondità di 10,20 e 30 cm

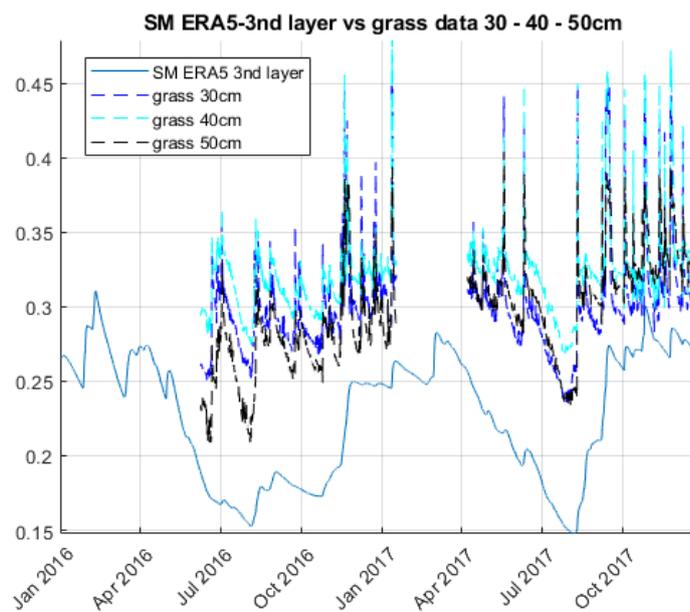


Figura A.32: Confronto umidità del suolo tra il terzo strato di ERA5 e i dati della stazione di Hobøl (zona erbosa) alla profondità di 30,40 e 50 cm

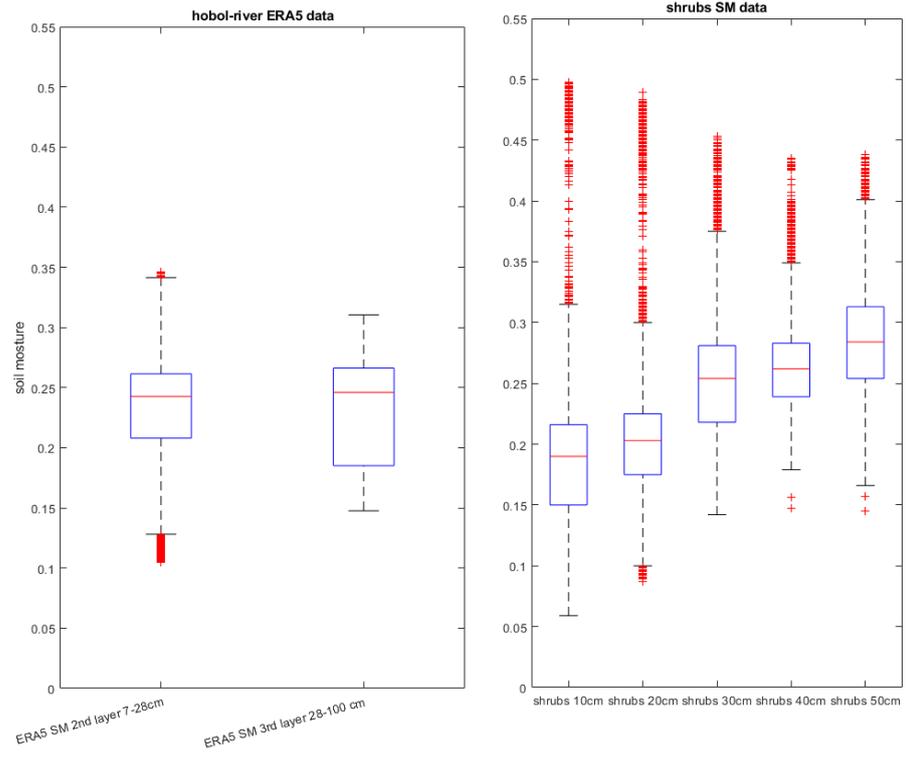


Figura A.33: Boxplot dati di umidità del suolo di Hobøl (arbusti)

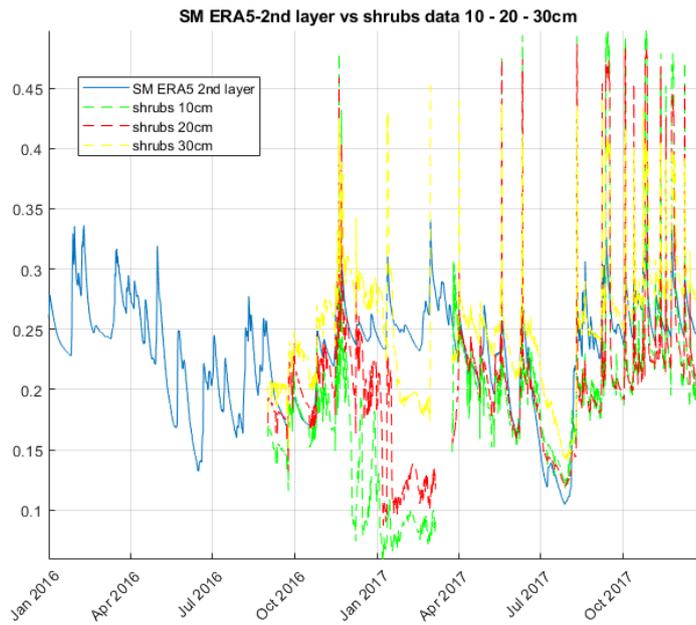


Figura A.34: Confronto umidità del suolo tra il secondo strato di ERA5e i dati della stazione di Hobøl (arbusti) alla profondità di 10,20 e 30 cm

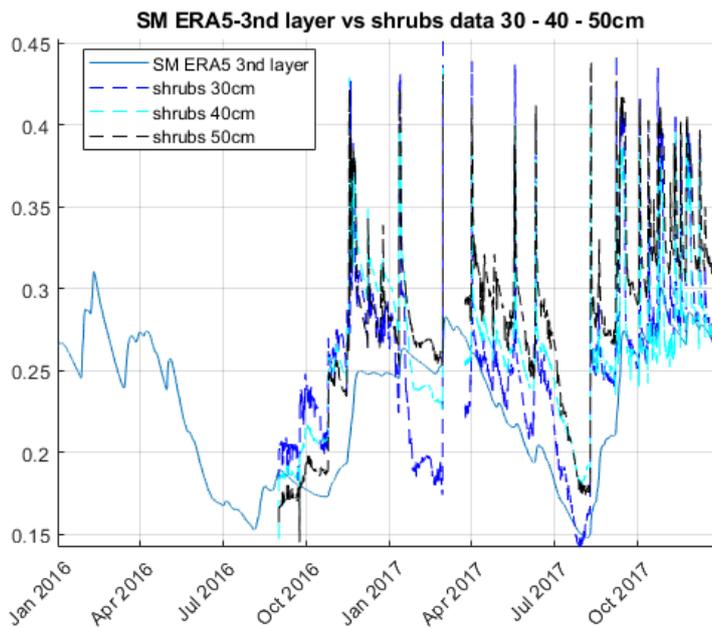


Figura A.35: Confronto umidità del suolo tra il terzo strato di ERA5 e i dati della stazione di Hobøl (arbusti) alla profondità di 30,40 e 50 cm

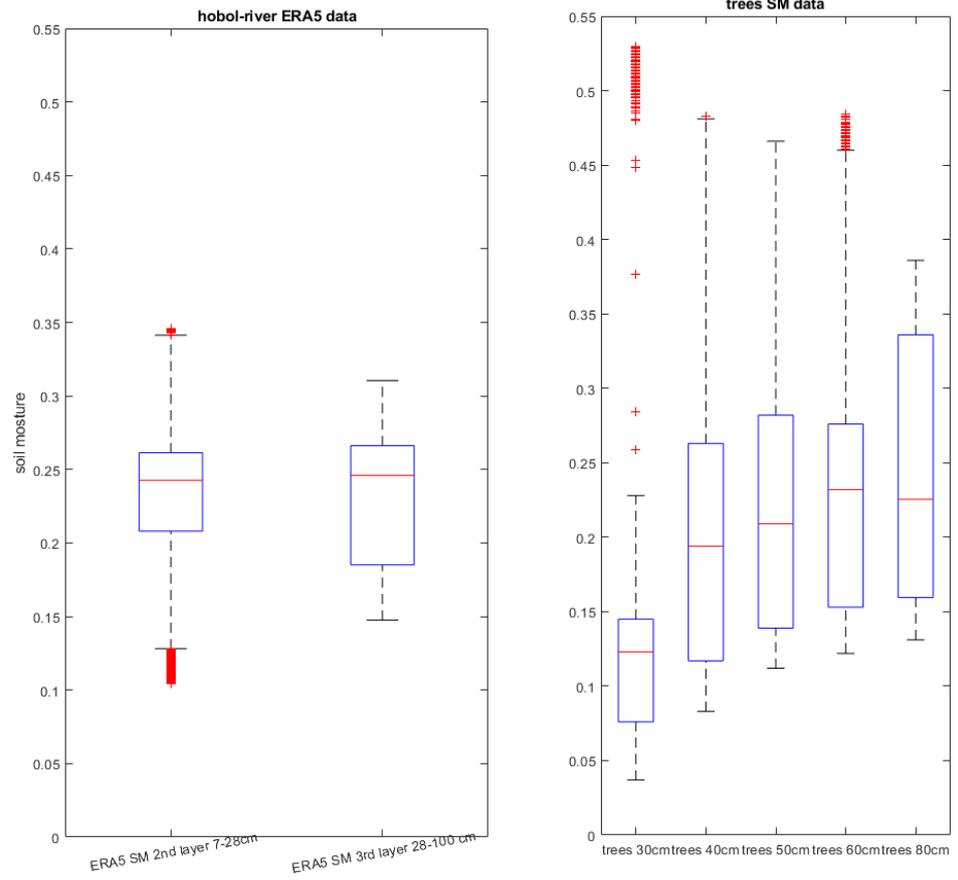


Figura A.36: Boxplot dati di umidità del suolo di Hobøl (alberi)

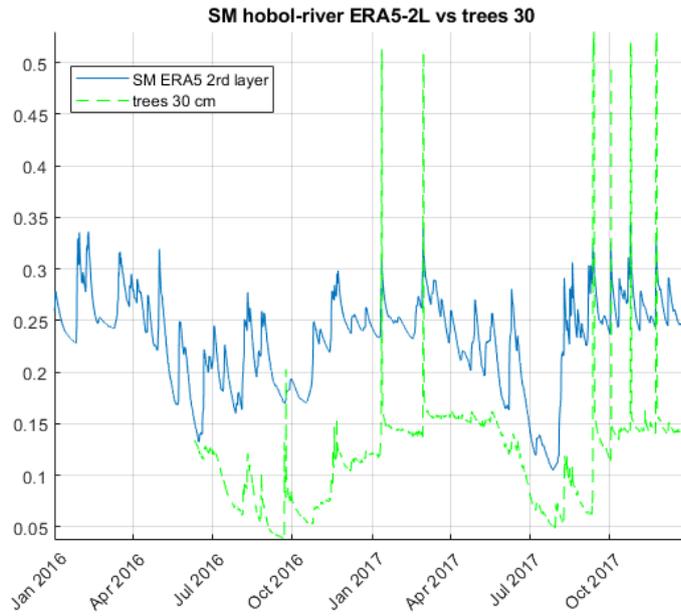


Figura A.37: Confronto umidità del suolo tra il secondo strato di ERA5e i dati della stazione di Hobøl (alberi) alla profondità di 30 cm

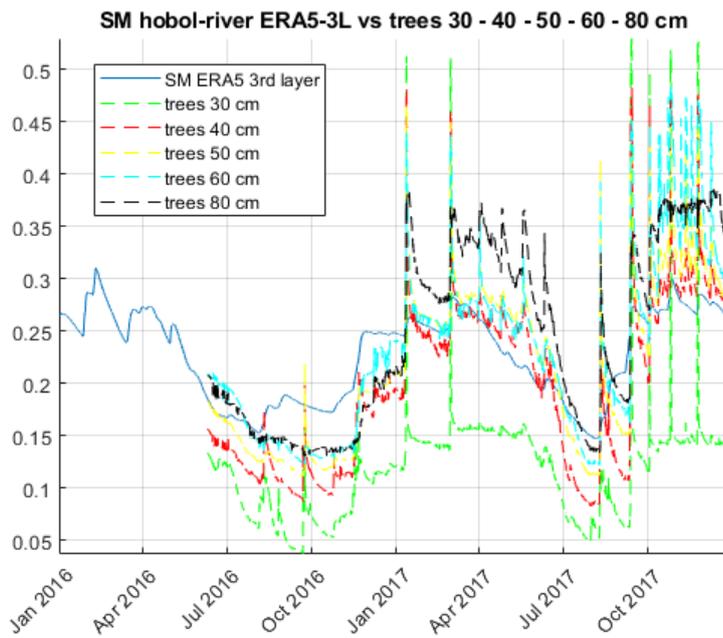


Figura A.38: Confronto umidità del suolo tra il terzo strato di ERA5 e i dati della stazione di Hobøl (alberi) alla profondità di 30,40, 50 e 80cm

APPENDICE B CODICI MATLAB

Preparazione dei dati di precipitazioni per CRTL-T

```

clc
clear all
AAA = importdata('CASTEL_DI_LUCIO_UNIONE.csv');
A = struct2cell(AAA);
x = A(1,1);
x = x{:,:};
x = x(length(x):-1:1,1);
y = A{2,1};
y = (y(2:1:end,1:1:3));
ID = str2num(char(y(:,1)));
location = char(y(1,2));
date = char(y(:,3));

% ----- ragionare per posizioni fissate
% a1 = (date(:,1)); % a2 = (date(:,2)); % a3 = (date(:,3));
% a4 = (date(:,4));
% m1 = (date(:,6)); % m2 = (date(:,7)); la serie inizia dal
31 Dicembre
% mese = str2num(strcat((date(length(date):-
1:1,6)), (date(length(date):-1:1,7))));
% in questo modo inizio dal primo Gennaio
year = str2num(strcat((date(length(date):-
1:1,1)), (date(length(date):-1:1,2)), (date(length(date):-
1:1,3)), (date(length(date):-1:1,4))));
mese = str2num(strcat((date(length(date):-
1:1,6)), (date(length(date):-1:1,7))));
gg = str2num(strcat((date(length(date):-
1:1,9)), (date(length(date):-1:1,10))));
ora = str2num(strcat((date(length(date):-
1:1,12)), (date(length(date):-1:1,13))));
min = str2num(strcat((date(length(date):-
1:1,15)), (date(length(date):-1:1,16))));
sec = str2num(strcat((date(length(date):-
1:1,18)), (date(length(date):-1:1,19))));
temp_matrix = [ gg, mese, year, ora, min,];

% ----- creare la matrice in cui allocare questi dati
% size la grandezza della serie riferit  all'anno di
partenza
% 5 slot per la data e uno per la precipitazione

DATE = nan(length(x),6);

```

```
DATE(1:1:end,6) = x;
DATE(1:1:end,1:1:5) = temp_matrix;

%--- codice di David

anno = 2009
praw=DATE;
% praw=load('prova.txt');

% praw(find(isnan(praw(:,6))==1),:)=[];

t=24*60*(datenum(praw(:,3), praw(:,2), praw(:,1),
praw(:,4), praw(:,5), zeros(size(praw(:,1))))); % da giorni
in minuti
dt=t(2:end)-t(1:end-1); % intervalli
%pause
prate=praw(1:end-1,6)./dt; %

n=length(dt);
k=1;
for i=1:n
    % k=k+1;

    vv=k:1:k-1+round(dt(i));
    pm(vv)=prate(i);
    k=k+round(dt(i));

end

% trova massimi annuali per fissata durata

twin=[10:10:60 180 360 720 1440]
for i=1:length(twin)
    [pmaxtwin(i) imax(i)]=maxanntlm(twin(i),pm,anno);

end
inmax=datestr(imax/60/24+datenum(anno, 1, 1, 00, 01, 00));

N=length(pm);

twin= 60 % aggregazione in minuti
ptwin=nan(N-twin-1,1);
for i = 1:N-twin-1
    ptwin(i)=sum(pm(i:1:i+twin-1));
end
[pmaxtwin imax]=nanmax(ptwin);
```

```
datestr(imax/60/24+datenum(anno, 1, 1, 00, 01, 00));

if sum(praw(:,6))~=sum(pm)
    disp('Error: Cumulative rainfall of praw different from
pm')
    sum(praw(:,6))
    sum(pm)
    sum(praw(:,6))-    sum(pm)
end

%% This one builds a series at aggregation agg [min]
agg= 60 % aggregation in minutes
m= floor(N/agg);
pagg= nan(m,1);
for i = 1:m
    pagg(i)=sum( pm(agg.*(i-1)+1: 1 : agg.*i));
end

num_in = datenum('2009-01-01 00:00:00');
num_fin = datenum('2014-12-31 23:00:00');
vettore_date = [num_in : 1/24 : num_fin]';
pierpaolo = datevec(datestr(vettore_date));
nuovoID = ones(length(pierpaolo),1).*ID(1,1);
final =[nuovoID, pierpaolo, pagg];

DD=strcat(datestr( final(:,2:7), 'yy-mm-dd HH:MM:SS'),
'+0000|', num2str(final(:,8)));
writematrix(DD, strcat('TimeSeries_Sensor_',
num2str(ID(1,1)), '.csv'));
```

Traccia soglia empirica eventi innescanti, non innescanti e calcolo del TSS.

```

function f = fpre(x, RI0, RI1, RD0, RD1)
% function that gives tss as a function of the parameters
% f = -pre is to be minimized (the optimization algorithms
of Matlab
% do only minimization
% RI0 = intensity of non triggering events
% RI1 = i of trig ev
% RD0, RD1 duration of non trig and trig
% ===
tp = sum(RI1 >= x(1).*RD1.^x(2));
tn = sum(RI0 <= x(1).*RD0.^x(2));
fp = sum(RI0 >= x(1).*RD0.^x(2));
fn = sum(RI1 <= x(1).*RD1.^x(2));
f = - tp./(fp+tp); %
function [bpartss, btss, bparpre, bpre, ...
        bguz50, bguz5, bgla50, bgla95, roc_tss, ...
        roc_pre, roc_guz5, roc_gla95, roc_guz50, roc_gla50,
optunc] = fthd_opt(intensity0, intensity1, time0, time1)
% thd_opt becomes a FUNCTION!!!
% intensity0= intensità piogge non innescanti (vettore
colonna)
%
%% ottimizza ROCTSS
[bpartss, btss, bparpre, bpre, optunc] =
opttsspre(intensity0, intensity1, time0, time1);
% threshold defined on triggering events
% ala "Guzzetti" guz vedi paper di marra HESS 2018
logint1 = log(intensity1');
logt1 = log(time1');
[bguz50, ~, rguz] = regress(logint1, [logt1
ones(size(logt1))]) ;
[murguz, sigrguz] = normfit(rguz) ;
bguz5=nan(1,2) ;
bguz5(2) = bguz50(2) + norminv(0.05, murguz, sigrguz) ;
bguz5(1) = bguz50(1) ;
bguz50(2)= exp(bguz50(2)) ;
bguz5(2) = exp(bguz5(2)) ;
bguz5 = [bguz5(2) bguz5(1)] ;
bguz50 = [bguz50(2) bguz50(1)] ;

% threshold defined on Non-triggering events
% ala "Glade" gla
http://www.cswcs.org.tw/AllDataPos/DownLoadPos/Root3/1.3-
Glade\_121106-print.pdf
logint0 = log(intensity0');

```

```

logt0 = log(time0');
[bgla50, ~, rgla] = regress(logint0, [logt0
ones(size(logt0))]) ;
[murgla, sigrgla] = normfit(rgla) ;
bgla95=nan(1,2) ;
bgla95(2) = bgla50(2) + norminv(0.95, murgla, sigrgla) ;
bgla95(1) = bgla50(1);
bgla50(2)= exp(bgla50(2)) ;
bgla95(2) = exp(bgla95(2)) ;
bgla95 = [bgla95(2) bgla95(1)] ;
bgla50 = [bgla50(2) bgla50(1)] ;

%% mi interessano le statistiche ROC per le soglie migliori

[tp_tss, tn_tss, fp_tss, fn_tss] = rocthr(bpartss(1),
bpartss(2), intensity0, intensity1, time0, time1);
roc_tss = [tp_tss, tn_tss, fp_tss, fn_tss];
[tp_pre, tn_pre, fp_pre, fn_pre] = rocthr(bparpre(1),
bparpre(2), intensity0, intensity1, time0, time1);
roc_pre = [tp_pre, tn_pre, fp_pre, fn_pre];
[tp_guz5, tn_guz5, fp_guz5, fn_guz5] = rocthr(bguz5(1),
bguz5(2), intensity0, intensity1, time0, time1);
roc_guz5 = [tp_guz5, tn_guz5, fp_guz5, fn_guz5];
[tp_gla95, tn_gla95, fp_gla95, fn_gla95] =
rocthr(bgla95(1), bgla95(2), intensity0, intensity1, time0,
time1);
roc_gla95 = [tp_gla95, tn_gla95, fp_gla95, fn_gla95];
[tp_guz50, tn_guz50, fp_guz50, fn_guz50] =
rocthr(bguz50(1), bguz50(2), intensity0, intensity1, time0,
time1);
roc_guz50 = [tp_guz50, tn_guz50, fp_guz50, fn_guz50] ;
[tp_gla50, tn_gla50, fp_gla50, fn_gla50] =
rocthr(bgla50(1), bgla50(2), intensity0, intensity1, time0,
time1);
roc_gla50 = [tp_gla50, tn_gla50, fp_gla50, fn_gla50];

function f = ftss(x, RI0, RI1, RD0, RD1)
% function that gives tss as a function of the parameters
% f = -tss is to be minimized (the optimization algorithms
of Matlab
% do only minimization
% RI0 = intensity of non triggering events
% RI1 = i of trig ev
% RD0, RD1 duration of non trig and trig
% ===

tp = sum(RI1 >= x(1).*RD1.^x(2));

```

```

tn = sum(RI0 <= x(1).*RD0.^x(2));
fp = sum(RI0 >= x(1).*RD0.^x(2));
fn = sum(RI1 <= x(1).*RD1.^x(2));
f = tp./(tp+fn) - fp./(fp+tn);
f = - 100.*f; % because Matlab minimizes

function [bpartssPS2, tssdef, bparprePS2, predef, optunc]
= opttsspre(intensity0, intensity1, time0, time1, nopts)
% [bpartssPS2, tssdef, bparprePS2, predef] =
opttsspre(intensity0, intensity1, time0, time1, nopts)%
if nargin == 4
    nopts = 1;
end
partssPS=zeros(nopts,2);
tssPS=zeros(nopts, 1);
options = optimoptions(@particleswarm, 'SwarmSize', 500,
'StallIterLimit', 20, 'Vectorized', 'off', 'UseParallel',
false);
for r = 1:nopts;
    [partssPS(r,:), tssPS(r)] = particleswarm(@(x) ftss(x,
intensity0, intensity1, time0, time1), 2, [0 -1.5], [200 -
0.2], options);
end
[btssPS ]= min(tssPS);
optunc = [std(partssPS) max(tssPS) min(tssPS)]; %
optimization uncertainty (standard deviation) added 18-04-
2020 covid19

% adesso trovami tutte le ottimizzazioni a cui corrisponde
il minimo -tss

iiPS = find(tssPS == btssPS);
bpartssPS2 = mean(partssPS(iiPS,:), 1)

tssdef = -ftss(bpartssPS2, intensity0, intensity1, time0,
time1);
tssdef=tssdef/100;
% fammi precision, ma solo una volta...
[bparprePS2, predef] = particleswarm(@(x) fpre(x,
intensity0, intensity1, time0, time1), 2);

function [tp, tn, fp, fn] = rocthr(a, b, RI0, RI1, RD0,
RD1)
tp = sum(RI1 >= a.*RD1.^b);
tn = sum(RI0 <= a.*RD0.^b);
fp = sum(RI0 >= a.*RD0.^b);
fn = sum(RI1 <= a.*RD1.^b);

```

```

load intensity1.txt
load intensity0.txt
load time1.txt
load time0.txt
%%
%intensity1=intensity1(and(time1>3, time1<=7*24))
%intensity0=intensity0(and(time0>3, time0<=7*24))
%time1=time1(and(time1>3, time1<=7*24))
%time0=time0(and(time0>3, time0<=7*24))
%%
figure
plot(time0,intensity0, '*g');
hold on
plot(time1,intensity1, '*r');
set(gca, 'xscale', 'log', 'yscale', 'log')
[bpartss, btss, bparpre, bpre, bguz50, bguz5, bgla50,
bgla95, roc_tss, roc_pre, roc_guz5, roc_gla95, roc_guz50,
roc_gla50, optunc] = fthd_opt(intensity0', intensity1',
time0', time1');
hold on
d=[1:1:1000]
plot(d, bpartss(1).*d.^bpartss(2), 'k', 'linewidth', 2)
plot(d, bguz5(1).*d.^bguz5(2), 'r', 'linewidth', 2)
plot(d, bgla95(1).*d.^bgla95(2), 'm', 'linewidth', 2)
legend('non trig', 'trig', 'max TSS', 'IRPI', 'Reverse
IRPI')
tss_tss=roc_tss(1)/(roc_tss(1)+roc_tss(4)) -
roc_tss(3)/(roc_tss(2)+roc_tss(3))
tss_guz5=roc_guz5(1)/(roc_guz5(1)+roc_guz5(4)) -
roc_guz5(3)/(roc_guz5(2)+roc_guz5(3))
tss_gla95=roc_gla95(1)/(roc_gla95(1)+roc_gla95(4)) -
roc_gla95(3)/(roc_gla95(2)+roc_gla95(3))

```

Rete neurale artificiale solo dati di precipitazione versione 1.

```

% Solve a Pattern Recognition Problem with a Neural Network
% Script generated by Neural Pattern Recognition app
% Created 26-Luglio-2021 12:02:10
%
% This script assumes these variables are defined:
%
%   XLog - input data.
%   T - target data.

opts = spreadsheetImportOptions("NumVariables", 27);
opts.Sheet = "Sheet1";
opts.DataRange = "A2:AA47543";
opts.VariableNames = ["PKEY", "ID_rain_gauge",
"index_pos1", "index_pos2", "RE_start_date_MAT",
"RE_end_date_MAT", "D_E", "E_E", "I_E", "IP_E", "Emax24_E",
"A_class", "Bin_landslide", "lon_rain", "lat_rain",
"lon_lsl", "lat_lsl", "lsl_date_MAT", "Numerofrane",
"SMLEV1tin", "SMLEV1tFin", "SMLEV2tin", "SMLEV2tFin",
"SMLEV3tin", "SMLEV3tFin", "SMLEV4tin", "SMLEV4tFin"];
opts.VariableTypes = ["double", "double", "double",
"double", "double", "double", "double", "double", "double",
"double", "double", "double", "double", "double", "double",
"double", "double", "double", "double", "double", "double",
"double"];
sm = readtable("C:\Users\Pierpaolo
Distefano\Documents\ANN_umidita_suolo\Dtab-
4lev_tin_tfin.xlsx", opts, "UseExcel", false);
clear opts

idx_in = find(sm.Bin_landslide == 1);
no_inne = sm(1:1:(idx_in-1),1:1:end);
innesco = sm(min(idx_in):1:max(idx_in),1:1:end);

%%
% innescante 1
% non innescante 0
int_peak_1 = innesco.IP_E;
int_peak_0 = no_inne.IP_E;
time1 = innesco.D_E;
time0 = no_inne.D_E;
cumulata1 = innesco.E_E;
cumulata0 = no_inne.E_E;
int1 = innesco.I_E;
int0 = no_inne.I_E;

```

```

%da inserire nel secondo tentativo nella
% matrice XLog
% XLog = [log(time0) log(cumulata0) log(int_peak_0);
log(time1) log(cumulata1) log(int_peak_1)];
% T_prova = [zeros(size(int_peak_0)) ;
ones(size(int_peak_1))]; % Target (Frana= 1, Non frana = 0]
XLog = [log(time0) log(cumulata0) log(int_peak_0) ;
log(time1) log(cumulata1) log(int_peak_1)];
T_prova = [zeros(size(time0)) ; ones(size(time1))]; %
Target (Frana= 1, Non frana = 0]

% T = [T_prova 1-T_prova]; %binario (non frana (0,1) ;
frana (1;0)
%%

x = XLog';
t = T_prova';

% Choose a Training Function
% For a list of all training functions type: help nntrain
% 'trainlm' is usually fastest.
% 'trainbr' takes longer but may be better for challenging
problems.
% 'trainscg' uses less memory. Suitable in low memory
situations.
trainFcn = 'trainscg'; % Scaled conjugate gradient
backpropagation.

% Create a Pattern Recognition Network ---- mettere un
ciclo for per far
% variare il numero di neuroni
% conta =[1:1:2]
% for ii =1:1:20

ii = [5:1:20]; % numero di neuroni nello strato nascosto

maxTSSall = nan(size(ii));
maxTSStrain = nan(size(ii));
maxTSSval = nan(size(ii));
maxTSSstest = nan(size(ii));
imaxTSSall = nan(size(ii));
imaxTSStrain = nan(size(ii));
imaxTSSval = nan(size(ii));
imaxTSSstest = nan(size(ii));
maxTPRtrain = nan(size(ii));
maxFPRtrain = nan(size(ii));

```

```
maxTPRval = nan(size(ii));
maxFPRval = nan(size(ii));
maxTPRtest = nan(size(ii));
maxFPRtest = nan(size(ii));
ContaElem = nan(numel(ii),11);

ntentmax = 40
for i=1:numel(ii)
    hiddenLayerSize = ii(i);
    net = patternnet(hiddenLayerSize, trainFcn);

    % Choose Input and Output Pre/Post-Processing Functions
    % For a list of all processing functions type: help
nnprocess
    net.input.processFcns =
{'removeconstantrows','mapminmax'};

    maxTSSstrain(i) = 0.5; % valore fittizio per far
partire il while
    maxTSSstest(i) = 1;
    maxTSSstest(i) = 1; % valore fittizio per far partire il
while

    k=0;
    while and(or(maxTSSstrain(i)<=maxTSSstest(i),
maxTSSstrain(i)<=maxTSSval(i)), k < ntentmax)

        % Setup Division of Data for Training, Validation,
Testing
        % For a list of all data division functions type:
help nndivision
        net.divideFcn = 'dividerand'; % Divide data
randomly
        net.divideMode = 'sample'; % Divide up every
sample
        net.divideParam.trainRatio = 70/100;
        net.divideParam.valRatio = 15/100;
        net.divideParam.testRatio = 15/100;

        % Choose a Performance Function
        % For a list of all performance functions type:
help nnperformance
        net.performFcn = 'crossentropy'; % Cross-Entropy

        % Choose Plot Functions
```

```

        % For a list of all plot functions type: help
nnplot
        net.plotFcns =
{'plotperform','plottrainstate','ploterrhist', ...
    'plotconfusion', 'plotroc'};

    % Train the Network
    [net,tr] = train(net,x,t);

    % Test the Network
    y = net(x);
    e = gsubtract(t,y);
    performance = perform(net,t,y);
    tind = vec2ind(t);
    yind = vec2ind(y);
    percentErrors = sum(tind ~= yind)/numel(tind);

    % Recalculate Training, Validation and Test
    Performance
    trainTargets = t .* tr.trainMask{1};
    valTargets = t .* tr.valMask{1};
    testTargets = t .* tr.testMask{1};
    trainPerformance = perform(net,trainTargets,y);
    valPerformance = perform(net,valTargets,y);
    testPerformance = perform(net,testTargets,y);

    train_elem_nonan = length(trainTargets) -
sum(isnan(trainTargets));
    elemtrain0 = numel(find(trainTargets == 0));
    elemtrain1 = numel(find(trainTargets == 1));

    val_elem_nonan = length(valTargets) -
sum(isnan(valTargets));
    elemval0 = numel(find(valTargets == 0));
    elemval1 = numel(find(valTargets == 1));

    test_elem_nonan = length(testTargets) -
sum(isnan(testTargets));
    elemtest0 = numel(find(testTargets == 0));
    elemtest1 = numel(find(testTargets == 1));

    ContaElem(i,1) = ii(i);
    ContaElem(i,2) = train_elem_nonan;
    ContaElem(i,3) = elemtrain0;
    ContaElem(i,4) = elemtrain1;
    ContaElem(i,5) = val_elem_nonan;

```

```

        ContaElem(i,6) = elemval0;
        ContaElem(i,7) = elemval1;
        ContaElem(i,8) = test_elem_nonan;
        ContaElem(i,9) = elemtest0;
        ContaElem(i,10) = elemtest1;
        ContaElem(i,11) = train_elem_nonan + val_elem_nonan
+ test_elem_nonan;

        % View the Network
        % view(net)
        % Plots
        % Uncomment these lines to enable various plots.
        %figure, plotperform(tr)
        %figure, plottrainstate(tr)
        %figure, ploterrhist(e)
        %figure, plotconfusion(t,y)
        %title(strcat('{Confusion matrix 40000 ANN num
neuron hidden layer}',' ', {num2str(ii)}))
        %figure, plotroc(t,y)
        [tprall,fprall,~] = roc(t,y);
        [maxTSSall(i),imaxTSSall(i)] = max(tprall-fprall);
        maxTPRall(i)= tprall(imaxTSSall(i));
        maxFPRall(i)= fprall(imaxTSSall(i));
        [tprtrain,fprtrain,thresholdstrain] = roc(t.*
tr.trainMask{1},y.* tr.trainMask{1});
        [tprval,fprval,~] = roc(t.* tr.valMask{1},y.*
tr.valMask{1});
        [tprtest,fprtest,~] = roc(t.* tr.testMask{1},y.*
tr.testMask{1});
        [maxTSSstrain(i),imaxTSSstrain(i)] = max(tprtrain-
fprtrain);
        [maxTSSval(i),imaxTSSval(i)] = max(tprval-fprval);
        [maxTSSstest(i),imaxTSSstest(i)] = max(tprtest-
fprtest);

        maxTPRtrain(i)= tprtrain(imaxTSSstrain(i));
        maxFPRtrain(i)= fprtrain(imaxTSSstrain(i));
        maxTPRval(i)= tprval(imaxTSSval(i));
        maxFPRval(i)= fprval(imaxTSSval(i));
        maxTPRtest(i)= tprtest(imaxTSSstest(i));
        maxFPRtest(i)= fprtest(imaxTSSstest(i));

        % Deployment
        % Change the (false) values to (true) to enable the
following code blocks.
        % See the help for each generation function for
more information.

```

```

        if (false)
            % Generate MATLAB function for neural network
            for application
                % deployment in MATLAB scripts or with MATLAB
                Compiler and Builder
                % tools, or simply to examine the calculations
                your trained neural
                % network performs.
                genFunction(net, 'myNeuralNetworkFunction');
                y = myNeuralNetworkFunction(x);
            end
            if (false)
                % Generate a matrix-only MATLAB function for
                neural network code
                % generation with MATLAB Coder tools.

                genFunction(net, 'myNeuralNetworkFunction', 'MatrixOnly', 'yes
                ');
                y = myNeuralNetworkFunction(x);
            end
            if (false)
                % Generate a Simulink diagram for simulation or
                deployment with.
                % Simulink Coder tools.
                gensim(net);

            end
            %

            k = k + 1;
            if k >= ntentmax
                disp(strcat('RAGGIUNTO LIMITE DI TENTATIVI PER
                UN NUMERO DI NEURONI PARI A', num2str(ii(i))))
            end
        end

end

% coerenza = and((maxTSSstrain>maxTSSval),
(maxTSSval>maxTSStest));

[bestTSSall, ibestall] = max(maxTSSall);
% [bestTSSall, ibestall] = max(maxTSSall.*coerenza); % se
metti 2 ti devi
% creare TAB_TSS_coerenza ...

```

```
TAB_TSS = [ii(ibestall) bestTSSall maxTSStrain(ibestall)
maxTSSval(ibestall) maxTSSstest(ibestall)];
TAB_TPR = [ii(ibestall) maxTPRall(ibestall)
maxTPRtrain(ibestall) maxTPRval(ibestall)
maxTPRtest(ibestall)];
TAB_FPR = [ii(ibestall) maxFPRall(ibestall)
maxFPRtrain(ibestall) maxFPRval(ibestall)
maxFPRtest(ibestall)];
```

Accostamento dati di umidità del suolo ai DB di eventi innescanti e non innescanti.

```

clear all
years = [2000:2021];
level = 4;
sm = [];
rawt = [];
lat = [];
lon = [];
for i =1:numel(years)
    sm = cat(3, sm, ncread(strcat('sv', num2str(level),
    '_ ', num2str(years(i)), '.nc'), strcat('swv1',
    num2str(level))));
    rawt = double(cat(1, rawt, ncread(strcat('sv',
    num2str(level), '_ ', num2str(years(i)), '.nc'), 'time')));
end
lat = double(ncread(strcat('sv', num2str(level), '_ ',
    num2str(years(i)), '.nc'), 'latitude'));
lon = double(ncread(strcat('sv', num2str(level), '_ ',
    num2str(years(i)), '.nc'), 'longitude'));
tsm_utc =
    (datenum(1900,01,01,00,00,00)*24+double(rawt))./24;
clear i
save sm4.mat

function [tabsmcomp] = smatt_f(D, smfile)
%%
% timesv1 = ncread('sv1_2012.nc', 'time');
%
newtime=(datenum(1900,01,01,00,00,00)*24+double(timesv1))./
24;
% newtimedt= datetime(datestr(newtime));
% dstflag = isdst(newtimedt, 'TimeZone', 'Europe/Rome');
% i dati di pioggia sono UTC come i dati di umidità del
suolo

% carica date di inizio della pioggia in UTC
% revin : date di inizio (formato matlab)
% revfin: date di fine evento (formato matlab)
% lsldate : date di innesco (formato matlab)
% revlatlon: coordinate lat, lon pluviometro
% latsm, lonsm: lat e lon dei dati di soil moisture
% tsm_utc: time in giorni UTC dei dati di soil moisture
% voglio prendere i valore di umidità del suolo a revin,
revfin, lsldate
% e anche a diversi lag prima di revin

```

```
% Algoritmo
% mi basta cercare il punto griglia più vicino e poi
prendere il
% valore con la data più prossima a quella cercata.

%%
% TEST-INPUT-FILES
%clear all
%clc
%
%load rain.txt
%revin = rain(:,3);
%revfin = datenum(rain(:,6), rain(:,5), rain(:,4),
rain(:,7), rain(:,8), rain(:,9));
%revlatlon = rain(:,1:2);
%
%load frane.txt
%lsldate = datenum(frane(:,5), frane(:,4), frane(:,3),
frane(:,6), frane(:,7), frane(:,8));
%lsllatlon = frane(:,1:2);

% addpath F:\NGI_UNICT\Analysis\smdata
addpath C:\Users\Nunziarita\Desktop\SICILIA\smdata\
revin = D(:,5);
revfin = D(:,6);
revlatlon = [D(:,15) D(:,14)];

lsldate = D(:,18);
lsllatlon = [D(:,17) D(:,16)];
lslbin = D(:,13);

load(strcat(smfile, '.mat'))
% sm<lev> files contain the following variables:
% lat, level, lon, rawt, sm, tsm_utc, years
latsm = lat;
lonsm = lon;
[LATSM, LONSM] = meshgrid(latsm, lonsm);
latsmvec = reshape(LATSM, numel(LATSM), 1);
lonsmvec = reshape(LONSM, numel(LONSM), 1);

inans = find(isnan(sm(:, :, 1)) == 1);
latsmvec(inans) = NaN;
lonsmvec(inans) = NaN;

nrev=numel(revfin);
```

```

distmin = nan(nrev,1);
idistmin = nan(nrev,1);
ilat = nan(nrev,1);
ilon = nan(nrev,1);
itime_in = nan(nrev,1);
itime_fin = nan(nrev,1);
itime_lsl = nan(nrev,1);
tabsm = nan(nrev,5);
presm1 = nan(nrev,1);
presm2 = nan(nrev,1);
presm3 = nan(nrev,1);

for i = 1: nrev

    %distance([latsmvec lonsmvec], revlatlon(i,:),[6378.137
0.0818191908426215])
    disp(strcat('Progress ', smfile , ': ',
num2str(100*i./nrev),'%.'))

    % pause

    [distmin(i), idistmin(i)] = nanmin(distance([latsmvec
lonsmvec], revlatlon(i,:),[6378.137 0.0818191908426215]));
% ellissoide WGS84
    ilat(i) = find(latsm == LATSM(idistmin(i)));
    ilon(i) = find(lonsm == LONSM(idistmin(i)));
    [~, itime_in(i)] = nanmin(abs(tsm_utc - revin(i)));
    [~, itime_fin(i)] = nanmin(abs(tsm_utc - revfin(i)));
    [~, itime_lsl(i)] = nanmin(abs(tsm_utc - lsldate(i)));

    % in questa parte del codice si vanno a
    % pescare i dati di umidità del suolo con un certo lag
prima
    % dell'inizio dell'evento di pioggia

    if itime_in(i)>1
        presm1(i) = sm(ilon(i), ilat(i), itime_in(i)-1);
    else
        presm1(i) = NaN;
    end
    if itime_in(i)>2
        presm2(i) = sm(ilon(i), ilat(i), itime_in(i)-2);
    else
        presm2(i) = NaN;
    end
end

```

```

    if itime_in(i)>3
        presm3(i) = sm(ilon(i), ilat(i), itime_in(i)-3);
    else
        presm3(i) = NaN;
    end

    % in questa parte del codice si costruisce una tabella
    che mette
    % insieme i dati pescati
    tabsm(i,:) = [sm(ilon(i), ilat(i), itime_in(i)) ...
        sm(ilon(i), ilat(i), itime_in(i)+1) ...
        sm(ilon(i), ilat(i), itime_in(i)+2) ...
        sm(ilon(i), ilat(i), itime_fin(i)) ...
        sm(ilon(i), ilat(i), itime_lsl(i))] ;

end

tabsm(lslbin==0,5)= NaN; % if no landslide occur t1sl makes
no sense
tabsmcomp = [presm3, presm2, presm1, tabsm];

dbfile = 'db_NOR_rain_sm_lsl_simplified_matdates'
D = load('db_NOR_rain_sm_lsl_simplified_matdates.txt');
[tabsmcomp1] = smatt_f(D, 'sm1')
[tabsmcomp2] = smatt_f(D, 'sm2')
[tabsmcomp3] = smatt_f(D, 'sm3')
[tabsmcomp4] = smatt_f(D, 'sm4')

% Dtab = array2table([D, tabsmcomp1, tabsmcomp2,
tabsmcomp3], 'VariableNames', ...
%     {'PKEY', 'ID_rain_gauge','index_pos1', 'index_pos2',
'RE_start_date_MAT', ...
%     'RE_end_date_MAT', 'D_E', 'E_E', 'I_E', 'IP_E',
'Emax24_E', 'A_class', ...
%     'Bin_landslide', 'lon_rain', 'lat_rain', 'lon_lsl',
'lat_lsl', ...
%     'lsl_date_MAT', 'Numero frane', ...
%     'SMLEV1(tin-3)', 'SMLEV1(tin-2)', 'SMLEV1(tin-1)',
'SMLEV1(tin)', ...
%     'SMLEV1(tin+1)', 'SMLEV1(tin+2)', 'SMLEV1(tFin)',
'SMLEV1(tlsl)', ...
%     'SMLEV2(tin-3)', 'SMLEV2(tin-2)', 'SMLEV2(tin-1)',
'SMLEV2(tin)', ...
%     'SMLEV2(tin+1)', 'SMLEV2(tin+2)', 'SMLEV2(tFin)',
'SMLEV2(tlsl)', ...
%     'SMLEV3(tin-3)', 'SMLEV3(tin-2)', 'SMLEV3(tin-1)',
'SMLEV3(tin)', ...

```

```

%      'SMLEV3(tin+1)', 'SMLEV3(tin+2)', 'SMLEV3(tFin)',
'SMLEV3(tlsl)'});

Dtab = array2table([D, tabsmcomp1, tabsmcomp2, tabsmcomp3,
tabsmcomp4], 'VariableNames', ...
    {'PKEY', 'ID_rain_gauge', 'index_pos1', 'index_pos2',
'RE_start_date_MAT', ...
'RE_end_date_MAT', 'D_E', 'E_E', 'I_E', 'IP_E',
'Emax24_E', 'A_class', ...
'Bin_landslide', 'lon_rain', 'lat_rain', 'lon_lsl',
'lat_lsl', ...
'ls1_date_MAT', 'Numero_frane', ...
'SMLEV1tin_meno3', 'SMLEV1tin_meno2', 'SMLEV1tin_meno1',
'SMLEV1_tin', ...
'SMLEV1tin_piu1', 'SMLEV1tin_piu2', 'SMLEV1_tFin',
'SMLEV1_tsl', ...
'SMLEV2tin_meno3',
'SMLEV2tin_meno2', 'SMLEV2tin_meno1', 'SMLEV2_tin', ...
'SMLEV2tin_piu1', 'SMLEV2tin_piu2', 'SMLEV2_tFin',
'SMLEV2_tsl', ...
'SMLEV3tin_meno3',
'SMLEV3tin_meno2', 'SMLEV3tin_meno1', 'SMLEV3_tin', ...
'SMLEV3tin_piu1', 'SMLEV3tin_piu2', 'SMLEV3_tFin',
'SMLEV3_tsl', ...
'SMLEV4tin_meno3',
'SMLEV4tin_meno2', 'SMLEV4tin_meno1', 'SMLEV4_tin', ...
'SMLEV4tin_piu1', 'SMLEV4tin_piu2', 'SMLEV4_tFin',
'SMLEV4_tsl'});

writetable(Dtab, 'Dtab-Nor_SM_2000_2021_data17_05_22.xlsx' )

```

Rete neurale artificiale con dati di precipitazione e umidità del suolo versione 2.

```

% Solve a Pattern Recognition Problem with a Neural Network
% This script assumes these variables are defined:
%
%   XLog - input data.
%   T - target data.

sm = readtable("C:\Users\Pierpaolo
Distefano\Desktop\dati_10_11\download_SM_norway\implemento_
SM\Dtab-Nor_SM_2000_2021_data17_05_22.xlsx", opts,
"UseExcel", false);
clear opts

idx_in = find(sm.Bin_landslide == 1);
no_inne = sm(1:1:(idx_in-1),1:1:end);
innesco = sm(min(idx_in):1:max(idx_in),1:1:end);

%%
% innescante 1
% non innescante 0
int_peak_1 = innesco.IP_E;
int_peak_0 = no_inne.IP_E;
time1 = innesco.D_E;
time0 = no_inne.D_E;
cumulata1 = innesco.E_E;
cumulata0 = no_inne.E_E;
int1 = innesco.I_E;
int0 = no_inne.I_E;

%---- dati umidità iniziale
sm1ltin_1 = innesco.SMLEV1_tin;
sm1ltin_0 = no_inne.SMLEV1_tin;
sm2ltin_1 = innesco.SMLEV2_tin;
sm2ltin_0 = no_inne.SMLEV2_tin;
sm3ltin_1 = innesco.SMLEV3_tin;
sm3ltin_0 = no_inne.SMLEV3_tin;
sm4ltin_1 = innesco.SMLEV4_tin;
sm4ltin_0 = no_inne.SMLEV4_tin;

%---- dati umidità finale
sm1ltfin_1 = double(innesco.SMLEV1_tsl);
sm1ltfin_0 = no_inne.SMLEV1_tFin;
sm2ltfin_1 = double(innesco.SMLEV2_tsl);
sm2ltfin_0 = no_inne.SMLEV2_tFin;
sm3ltfin_1 = double(innesco.SMLEV3_tsl);

```

```

sm3ltfin_0 = no_inne.SMLEV3_tFin;
sm4ltfin_1 = double(innesco.SMLEV4_tls1);
sm4ltfin_0 = no_inne.SMLEV4_tFin;

XLog = [log(time0) log(cumulata0) log(smlltin_0) ;
log(time1) log(cumulata1) log(smlltin_1)];
XsenzaLog = [(time0) (cumulata0) (smlltin_0) ; (time1)
(cumulata1) (smlltin_1)];
T_prova = [zeros(size(time0)) ; ones(size(time1))]; %
Target (Frana= 1, Non frana = 0]

% T = [T_prova 1-T_prova]; %binario (non frana (0,1) ;
frana (1;0)
%%

x = XLog';
t = T_prova';

% Choose a Training Function
trainFcn = 'trainscg'; % Scaled conjugate gradient
backpropagation.

% Create a Pattern Recognition Network ---- mettere un
ciclo for per far
% variare il numero di neuroni
% conta =[1:1:2]
% for ii =1:1:20

ii = [5:1:20]; % numero di neuroni nello strato nascosto

maxTSSall = nan(size(ii));
maxTSStrain = nan(size(ii));
maxTSSval = nan(size(ii));
maxTSSstest = nan(size(ii));
imaxTSSall = nan(size(ii));
imaxTSStrain = nan(size(ii));
imaxTSSval = nan(size(ii));
imaxTSSstest = nan(size(ii));
maxTPRtrain = nan(size(ii));
maxFPRtrain = nan(size(ii));
maxTPRval = nan(size(ii));
maxFPRval = nan(size(ii));
maxTPRtest = nan(size(ii));
maxFPRtest = nan(size(ii));
ContaElem = nan(numel(ii),11);

```

```

ntentmax = 40
for i=1:numel(ii)
    hiddenLayerSize = ii(i);
    net = patternnet(hiddenLayerSize, trainFcn);

    % Choose Input and Output Pre/Post-Processing Functions
    % For a list of all processing functions type: help
nnprocess
    net.input.processFcns =
{'removeconstantrows','mapminmax'};

    maxTSSstrain(i) = 0.5; % valore fittizio per far
partire il while
    maxTSSstest(i) = 1;
    maxTSSstest(i) = 1; % valore fittizio per far partire il
while

    k=0;
    while and(or(maxTSSstrain(i)<=maxTSSstest(i),
maxTSSstrain(i)<=maxTSSval(i)), k < ntentmax)

        % Setup Division of Data for Training, Validation,
Testing
        % For a list of all data division functions type:
help nndivision
        net.divideFcn = 'dividerand'; % Divide data
randomly
        net.divideMode = 'sample'; % Divide up every
sample
        net.divideParam.trainRatio = 70/100;
        net.divideParam.valRatio = 15/100;
        net.divideParam.testRatio = 15/100;

        % Choose a Performance Function
        % For a list of all performance functions type:
help nnperformance
        net.performFcn = 'crossentropy'; % Cross-Entropy

        % Choose Plot Functions
        % For a list of all plot functions type: help
nnplot
        net.plotFcns =
{'plotperform','plottrainstate','ploterrhist', ...
'plotconfusion', 'plotroc'};

        % Train the Network

```

```

[net,tr] = train(net,x,t);

% Test the Network
y = net(x);
e = gsubtract(t,y);
performance = perform(net,t,y);
tind = vec2ind(t);
yind = vec2ind(y);
percentErrors = sum(tind ~= yind)/numel(tind);

% Recalculate Training, Validation and Test
Performance
trainTargets = t .* tr.trainMask{1};
valTargets = t .* tr.valMask{1};
testTargets = t .* tr.testMask{1};
trainPerformance = perform(net,trainTargets,y);
valPerformance = perform(net,valTargets,y);
testPerformance = perform(net,testTargets,y);

train_elem_nonan = length(trainTargets) -
sum(isnan(trainTargets));
elemtrain0 = numel(find(trainTargets == 0));
elemtrain1 = numel(find(trainTargets == 1));

val_elem_nonan = length(valTargets) -
sum(isnan(valTargets));
elemval0 = numel(find(valTargets == 0));
elemvall1 = numel(find(valTargets == 1));

test_elem_nonan = length(testTargets) -
sum(isnan(testTargets));
elemtest0 = numel(find(testTargets == 0));
elemtest1 = numel(find(testTargets == 1));

ContaElem(i,1) = ii(i);
ContaElem(i,2) = train_elem_nonan;
ContaElem(i,3) = elemtrain0;
ContaElem(i,4) = elemtrain1;
ContaElem(i,5) = val_elem_nonan;
ContaElem(i,6) = elemval0;
ContaElem(i,7) = elemvall1;
ContaElem(i,8) = test_elem_nonan;
ContaElem(i,9) = elemtest0;
ContaElem(i,10) = elemtest1;
ContaElem(i,11) = train_elem_nonan + val_elem_nonan
+ test_elem_nonan;

```

```

% View the Network
% view(net)

% Plots
% Uncomment these lines to enable various plots.
%figure, plotperform(tr)
%figure, plottrainstate(tr)
%figure, ploterrhist(e)
%figure, plotconfusion(t,y)
%title(strcat('{Confusion matrix 40000 ANN num
neuroni hidden layer}',' '), {num2str(ii)}))
%figure, plotroc(t,y)
[tprall,fprall,~] = roc(t,y);
[maxTSSall(i),imaxTSSall(i)] = max(tprall-fprall);
maxTPRall(i)= tprall(imaxTSSall(i));
maxFPRall(i)= fprall(imaxTSSall(i));
[tprtrain,fprtrain,thresholdstrain] = roc(t.*
tr.trainMask{1},y.* tr.trainMask{1});
[tprval,fprval,~] = roc(t.* tr.valMask{1},y.*
tr.valMask{1});
[tprtest,fprtest,~] = roc(t.* tr.testMask{1},y.*
tr.testMask{1});
[maxTSSstrain(i),imaxTSSstrain(i)] = max(tprtrain-
fprtrain);
[maxTSSval(i),imaxTSSval(i)] = max(tprval-fprval);
[maxTSSstest(i),imaxTSSstest(i)] = max(tprtest-
fprtest);

maxTPRtrain(i)= tprtrain(imaxTSSstrain(i));
maxFPRtrain(i)= fprtrain(imaxTSSstrain(i));
maxTPRval(i)= tprval(imaxTSSval(i));
maxFPRval(i)= fprval(imaxTSSval(i));
maxTPRtest(i)= tprtest(imaxTSSstest(i));
maxFPRtest(i)= fprtest(imaxTSSstest(i));

% Deployment
% Change the (false) values to (true) to enable the
following code blocks.
% See the help for each generation function for
more information.
if (false)
    % Generate MATLAB function for neural network
for application
    % deployment in MATLAB scripts or with MATLAB
Compiler and Builder

```

```

        % tools, or simply to examine the calculations
your trained neural
        % network performs.
        genFunction(net, 'myNeuralNetworkFunction');
        y = myNeuralNetworkFunction(x);
    end
    if (false)
        % Generate a matrix-only MATLAB function for
neural network code
        % generation with MATLAB Coder tools.

genFunction(net, 'myNeuralNetworkFunction', 'MatrixOnly', 'yes
');
        y = myNeuralNetworkFunction(x);
    end
    if (false)
        % Generate a Simulink diagram for simulation or
deployment with.
        % Simulink Coder tools.
        gensim(net);

    end
    k = k + 1;
    if k >= ntentmax
        disp(strcat('RAGGIUNTO LIMITE DI TENTATIVI PER
UN NUMERO DI NEURONI PARI A', num2str(ii(i))))
    end
end

end

% coerenza = and((maxTSSStrain>maxTSSval),
(maxTSSval>maxTSSstest));

[bestTSSall, ibestall] = max(maxTSSall);
% [bestTSSall, ibestall] = max(maxTSSall.*coerenza); % se
metti 2 ti devi
% creare TAB_TSS_coerenza ...

TAB_TSS = [ii(ibestall) bestTSSall maxTSSStrain(ibestall)
maxTSSval(ibestall) maxTSSstest(ibestall)];
TAB_TPR = [ii(ibestall) maxTPRall(ibestall)
maxTPRtrain(ibestall) maxTPRval(ibestall)
maxTPRtest(ibestall)];

```

```
TAB_FPR = [ii(ibestall) maxFPRall(ibestall)
maxFPRtrain(ibestall) maxFPRval(ibestall)
maxFPRtest(ibestall)];
```

BIBLIOGRAFIA

- Aha, D.W., Kibler, D., Albert, M., 1991. Instance-Based Learning Algorithms. *Mach. Learn.* 6, 37--66.
- Alecci, S., Rossi, G., 2007. Controllo di qualità dei dati pluviometrici ed analisi delle serie temporali". In: Siccità, Rossi G. (Ed.), *Analisi, monitoraggio e mitigazione. Applicazioni in Sicilia*. Nuova Editoriale Bios, Cosenza (in Italian).
- Aleotti, P., 2004. A warning system for rainfall-induced shallow failures. *Eng. Geol.* 73, 247–265. <https://doi.org/10.1016/J.ENGGE0.2004.01.007>
- Alvioli, M., Baum, R.L., 2016. Parallelization of the TRIGRS model for rainfall-induced landslides using the message passing interface. *Environ. Model. Softw.* 81, 122–135. <https://doi.org/10.1016/j.envsoft.2016.04.002>
- Antinoro, C., Arnone, E., Noto, L. V., 2017. The use of soil water retention curve models in analyzing slope stability in differently structured soils. *Catena* 150, 133–145. <https://doi.org/10.1016/j.catena.2016.11.019>
- Ardizzone, F., Basile, G., Cardinali, M., Casagli, N., Del Conte, S., Del Ventisette, C., Fiorucci, F., Garfagnoli, F., Gigli, G., Guzzetti, F., Iovine, G., Mondini, A.C., Moretti, S., Panebianco, M., Raspini, F., Reichenbach, P., Rossi, M., Tanteri, L., Terranova, O., 2012. Landslide inventory map for the Briga and the Giampileri catchments, NE Sicily, Italy. *J. Maps* 8, 176–180. <https://doi.org/10.1080/17445647.2012.694271>
- Batta, M., 2020. Machine Learning Algorithms - A Review. *Int. J. Sci. Res. (IJ)* 9, 381–386. <https://doi.org/10.21275/ART20203995>
- Baum, R.L., Godt, J.W., Savage, W.Z., 2010. Estimating the timing and location of shallow rainfall-induced landslides using a model for transient, unsaturated infiltration. *J. Geophys. Res. Earth Surf.* 115. <https://doi.org/10.1029/2009JF001321>
- Bogaard, T., Greco, R., 2018. Invited perspectives: Hydrological perspectives on precipitation intensity-duration thresholds for landslide initiation: Proposing hydro-meteorological thresholds. *Nat. Hazards Earth Syst. Sci.* 18, 31–39. <https://doi.org/10.5194/nhess-18-31-2018>
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32. <https://doi.org/10.1201/9780367816377-11>
- Breznitz S (1984) The false alarm effect. *Cry wolf: The psychology of false alarms*. Lawrence Erlbaum, New Jersey, pp 9–16
- Brocca, L., Ciabatta, L., Moramarco, T., Ponziani, F., Berni, N., Wagner, W., 2016. Use of Satellite Soil Moisture Products for the Operational Mitigation of Landslides Risk in Central Italy, *Satellite Soil Moisture Retrieval: Techniques and Applications*. Elsevier Inc. <https://doi.org/10.1016/B978-0-12-803388-3.00012-7>
- Brunetti M. T., S. Peruccacci, M. Rossi, S. Luciani, D. Valigi, and F.G., 2010. Rainfall thresholds for the possible occurrence of landslides in Italy. *Nat. Hazards Earth Syst. Sci.* 47, 633–635. [https://doi.org/10.1016/S1387-6473\(03\)00110-6](https://doi.org/10.1016/S1387-6473(03)00110-6)
- Caine, N., 1980. The rainfall intensity-duration control of shallow landslides and debris flows. *Geogr. Ann. Ser. A, Phys. Geogr.* 62.1-2 23-27. 62, 23–27.

- Calvello, M., d'Orsi, R.N., Piciullo, L., Paes, N., Magalhaes, M., Lacerda, W.A., 2014. The Rio de Janeiro early warning system for rainfall-induced landslides: Analysis of performance for the years 2010-2013. *Int. J. Disaster Risk Reduct.* 12, 3–15. <https://doi.org/10.1016/j.ijdrr.2014.10.005>
- Calvello, M., Pecoraro, G., 2018. *FranItalia: a catalog of recent Italian landslides. Geoenvironmental Disasters* 5. <https://doi.org/10.1186/s40677-018-0105-5>
- Campbell, R.H., 1975. Debris flows originating from soil slips during rainstorms in Southern California. *Q. J. Eng. Geol.* 7, 339–349. <https://doi.org/10.1144/GSL.QJEG.1974.007.04.04>
- Cao, Y., Yin, K., Alexander, D.E., Zhou, C., 2016. Using an extreme learning machine to predict the displacement of step-like landslides in relation to controlling factors. *Landslides* 13, 725–736. <https://doi.org/10.1007/s10346-015-0596-z>
- Carrara, A., 1983. Multivariate models for landslide hazard evaluation. *J. Int. Assoc. Math. Geol.* 15, 403–426. <https://doi.org/10.1007/BF01031290>
- Chacon, J., Irigaray, C., Fernandez, T., El Hamdouni, R., 2006. Engineering geology maps: landslides and geographical information systems. *Bull. Eng. Geol. Environ.* 65, 341–411. <https://doi.org/10.1007/s10064-006-0064-z>
- Charalambous, C., 1992. Conjugate gradient algorithm for efficient training of artificial neural networks. *IEE Proceedings, Part G Circuits, Devices Syst.* 139, 301–310. <https://doi.org/10.1049/ip-g-2.1992.0050>
- Chien-Yuan, C., Tien-Chien, C., Fan-Chieh, Y., Wen-Hui, Y., Chun-Chieh, T., 2005. Rainfall duration and debris-flow initiated studies for real-time monitoring. *Environmental Geology* 47, 715–724.
- Conforti, M., Pascale, S., Robustelli, G., Sdao, F., 2014. Evaluation of prediction capability of the artificial neural networks for mapping landslide susceptibility in the Turbolo River catchment (northern Calabria, Italy). *CATENA* 113, 236–250. <https://doi.org/10.1016/J.CATENA.2013.08.006>
- Conrad, J.L., Morphew, M.D., Baum, R.L., Mirus, B.B., 2021. Hydromet: A new code for automated objective optimization of hydrometeorological thresholds for landslide initiation. *Water (Switzerland)* 13. <https://doi.org/10.3390/w13131752>
- Crow, W.T., Berg, A.A., Cosh, M.H., Loew, A., Mohanty, B.P., Panciera, R., De Rosnay, P., Ryu, D., Walker, J.P., 2012. Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products. *Rev. Geophys.* 50, 1–20. <https://doi.org/10.1029/2011RG000372>
- Crozier, M.J., Eyles, R.J., 1980. Assessing the probability of rapid mass movement, in: *Third Australia-New Zealand Conference on Geomechanics: Wellington*. p. 2.
- Cruden, D. M. and Varnes, D. J.: *Landslide types and processes, Spec. Rep. - Natl. Res. Council. Transp. Res. Board*, 247, 36–75, 1996.
- Dahal, R., Hasegawa, S., 2008. Representative rainfall thresholds for landslides in the Nepal Himalaya. *Geomorphology* 100, 429–443.
- Distefano, P., Peres, D.J., Scandura, P., Cancelliere, A., 2022. Brief communication: Introducing rainfall thresholds for landslide triggering based on artificial neural networks. *Nat. Hazards Earth Syst. Sci.* 1151–1157.

- Dorigo, W., Himmelbauer, I., Aberer, D., Schremmer, L., Petrakovic, I., Zappa, L., Preimesberger, W., Xaver, A., Annor, F., Ardö, J., Baldocchi, D., Bitelli, M., Blöschl, G., Bogena, H., Brocca, L., Calvet, J.C., Camarero, J.J., Capello, G., Choi, M., Cosh, M.C., Van De Giesen, N., Hajdu, I., Ikonen, J., Jensen, K.H., Kanniah, K.D., De Kat, I., Kirchengast, G., Kumar Rai, P., Kyrouac, J., Larson, K., Liu, S., Loew, A., Moghaddam, M., Martínez Fernández, J., Mattar Bader, C., Morbidelli, R., Musial, J.P., Osenga, E., Palecki, M.A., Pellarin, T., Petropoulos, G.P., Pfeil, I., Powers, J., Robock, A., Rüdiger, C., Rummel, U., Strobel, M., Su, Z., Sullivan, R., Tagesson, T., Varlagin, A., Vreugdenhil, M., Walker, J., Wen, J., Wenger, F., Wigneron, J.P., Woods, M., Yang, K., Zeng, Y., Zhang, X., Zreda, M., Dietrich, S., Gruber, A., Van Oevelen, P., Wagner, W., Scipal, K., Drusch, M., Sabia, R., 2021. The International Soil Moisture Network: Serving Earth system science for over a decade. *Hydrol. Earth Syst. Sci.* 25, 5749–5804. <https://doi.org/10.5194/hess-25-5749-2021>
- Dorigo, W.A., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., Gruber, A., Drusch, M., Mecklenburg, S., Van Oevelen, P., Robock, A., Jackson, T., 2011. The International Soil Moisture Network: A data hosting facility for global in situ soil moisture measurements. *Hydrol. Earth Syst. Sci.* 15, 1675–1698. <https://doi.org/10.5194/hess-15-1675-2011>
- Ermini, L., Catani, F., Casagli, N., 2005. Artificial Neural Networks applied to landslide susceptibility assessment. *Geomorphology* 66, 327–343. <https://doi.org/10.1016/j.geomorph.2004.09.025>
- Foti, E., Faraci, C., Scandura, P., Cancelliere, A., La Rocca, C., Musumeci, R. E., Nicolosi, V. M., Peres, D., Stancanelli, L. M., 2013. Da giampileri a saponara: analisi delle cause scatenanti e delle cause predisponenti. *ATTI DEI CONVEGNI LINCEI-ACCADEMIA NAZIONALE DEI LINCEI* 270, 45–64.
- Froude, M.J., Petley, D.N., 2018. Global fatal landslide occurrence from 2004 to 2016. *Nat. Hazards Earth Syst. Sci.* 18, 2161–2181. <https://doi.org/10.5194/nhess-18-2161-2018>
- Gariano, S.L., Brunetti, M.T., Iovine, G., Melillo, M., Peruccacci, S., Terranova, O., Vennari, C., Guzzetti, F., 2015. Calibration and validation of rainfall thresholds for shallow landslide forecasting in Sicily, southern Italy. *Geomorphology* 228, 653–665. <https://doi.org/10.1016/j.geomorph.2014.10.019>
- Gariano, S.L., Melillo, M., Peruccacci, S., Brunetti, M.T., 2020. How much does the rainfall temporal resolution affect rainfall thresholds for landslide triggering? *Nat. Hazards* 100, 655–670. <https://doi.org/10.1007/s11069-019-03830-x>
- Glade, T., Crozier, M., Smith, P., 2000a. Applying Probability Determination to Refine Landslide-triggering Rainfall Thresholds Using an Empirical “Antecedent Daily Rainfall Model.” *Pure Appl. Geophys.* 157, 1059–1079. <https://doi.org/10.1007/s000240050017>
- Glade, T., Crozier, M., Smith, P., 2000b. Applying probability determination to refine landslide-triggering rainfall thresholds using an empirical “Antecedent Daily Rainfall Model.” *Pure Appl. Geophys.* 157, 1059–1079. <https://doi.org/10.1007/s000240050017>

- Godt, J.W., Baum, R.L., Chleborad, A.F., 2006. Rainfall characteristics for shallow landsliding in Seattle, Washington, USA. *J. Br. Geomorphol. Res. Gr.* 110, 97–110. <https://doi.org/10.1002/esp.1237>
- Granger, C.W.J., Terasvirta, T., 1993. *Modelling Nonlinear Economic Relationships*. Oxford University Press, Oxford.
- Guzzetti, F., Peruccacci, S., Rossi, M., Stark, C.P., 2008. The rainfall intensity-duration control of shallow landslides and debris flows: An update. *Landslides* 5, 3–17. <https://doi.org/10.1007/s10346-007-0112-1>
- Guzzetti, F., Peruccacci, S., Rossi, M., Stark, C.P., 2007. Rainfall thresholds for the initiation of landslides in central and southern Europe. *Meteorol. Atmos. Phys.* 98, 239–267. <https://doi.org/10.1007/s00703-007-0262-7>
- Haque, U., da Silva, P.F., Devoli, G., Pilz, J., Zhao, B., Khaloua, A., Wilopo, W., Andersen, P., Lu, P., Lee, J., Yamamoto, T., Keellings, D., Jian-Hong, W., Glass, G.E., 2019. The human cost of global warming: Deadly landslides and their triggers (1995–2014). *Sci. Total Environ.* 682, 673–684. <https://doi.org/10.1016/j.scitotenv.2019.03.415>
- Haykin, S., 1999. *Neural Networks- A Comprehensive Foundation* 2nd Ed. Prentice Hall.
- Highland, L. M. and Bobrowsky, P.: The landslide Handbook - A guide to understanding landslides, *US Geol. Surv. Circ.*, 1–147, <https://doi.org/10.3133/cir1325>, 2008.
- Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. *Neural Networks* 2, 359–366
- Intrieri, E., Gigli, G., Mugnai, F., Fanti, R., Casagli, N., 2012. Design and implementation of a landslide early warning system. *Eng. Geol.* 147–148, 124–136. <https://doi.org/10.1016/j.enggeo.2012.07.017>
- John, G.H., Langley, P., 2013. Estimating Continuous Distributions in Bayesian Classifiers 338–345.
- Jibson, R. W. (1989). Debris flows in southern Puerto Rico. *Landslide processes of the eastern United States and Puerto Rico*, Geological Society of America special paper, 236, 29–55.
- Jung, M., Reichstein, M., Ciais, P., Seneviratne, S.I., Sheffield, J., Goulden, M.L., Bonan, G., Cescatti, A., Chen, J., De Jeu, R., Dolman, A.J., Eugster, W., Gerten, D., Gianelle, D., Gobron, N., Heinke, J., Kimball, J., Law, B.E., Montagnani, L., Mu, Q., Mueller, B., Oleson, K., Papale, D., Richardson, A.D., Roupsard, O., Running, S., Tomelleri, E., Viovy, N., Weber, U., Williams, C., Wood, E., Zaehle, S., Zhang, K., 2010. Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature* 467, 951–954. <https://doi.org/10.1038/nature09396>
- Keefer, D.K., Wilson, R.C., Mark, R.K., Brabb, E.E., Brown, W.M., Ellen, S.D., Harp, E.L., Wieczorek, G.F., Alger, C.S., Zatzkin, R.S., 1987. Real-time landslide warning during heavy rainfall. *Science* (80-). 238, 921–925. <https://doi.org/10.1126/science.238.4829.921>
- Klimasauskas, C. C. (1991). Applying neural networks. Part 3: Training a neural network. *Proceedings in Artificial Intelligence*, 20–24

- Kohonen, T., Oja, E., Simula, O., Visa, A., & Kangas, J., 2005. Engineering applications of the self-organizing map. *Proc. IEEE*, 84(10), 1358-1384. 1358-1384.
- Krkač, M., Špoljarić, D., Bernat, S., Arbanas, S.M., 2017. Method for prediction of landslide movements based on random forests. *Mach Learn* 14, 947-960. <https://doi.org/10.1007/s10346-016-0761-z>
- Krzeminska, D., Kerkhof, T., Skaalsveen, K., Stolte, J., 2019. Effect of riparian vegetation on stream bank stability in small agricultural catchments. *Catena* 172, 87-96. <https://doi.org/10.1016/j.catena.2018.08.014>
- Lagomarsino, D., Segoni, S., Fanti, R., Catani, F., 2013. Updating and tuning a regional-scale landslide early warning system. *Landslides* 10, 91-97. <https://doi.org/10.1007/s10346-012-0376-y>
- Lepore, C., Arnone, E., Noto, L. V., Sivandran, G., Bras, R.L., 2013. Physically based modeling of rainfall-triggered landslides: A case study in the Luquillo forest, Puerto Rico. *Hydrol. Earth Syst. Sci.* 17, 3371-3387. <https://doi.org/10.5194/hess-17-3371-2013>
- Li, C., Ma, T., Zhu, X., Li, W., 2011. The power-law relationship between landslide occurrence and rainfall level. *Geomorphology* 130, 221-229. <https://doi.org/10.1016/j.geomorph.2011.03.018>
- Liang, M., Wang, S.X., Luo, Y.-H., 1994. Fast learning algorithms for multi-layered feedforward neural network. *IEEE 1994 Natl. Aerosp. Electron. Conf. NAECON 2 (1994)*, 787-790. <https://doi.org/10.1109/naecon.1994.332959>
- Lin, G.W., Kuo, H.L., Chen, C.W., Wei, L.W., Zhang, J.M., 2020. Using a tank model to determine hydro-meteorological thresholds for large-scale landslides in Taiwan. *Water (Switzerland)* 12, 1-17. <https://doi.org/10.3390/w12010253>
- Lussana, C., Einar Tveito, O., Dobler, A., Tunheim, K., 2019. SeNorge_2018, daily precipitation, and temperature datasets over Norway. *Earth Syst. Sci. Data* 11, 1531-1551. <https://doi.org/10.5194/essd-11-1531-2019>
- Makridakis, S., Wheelwright, S.C., McGee, V.E., 1983. *Forecasting: Methods and Applications*, 2nd ed. John Wiley, New York.
- Marino, P., Peres, D.J., Cancelliere, A., Greco, R., Bogaard, T.A., 2020. Soil moisture information can improve shallow landslide forecasting using the hydrometeorological threshold approach. <https://doi.org/10.1007/s10346-020-01420-8>
- Marra, F., 2019. Rainfall thresholds for landslide occurrence: systematic underestimation using coarse temporal resolution data. *Nat. Hazards* 95, 883-890. <https://doi.org/10.1007/s11069-018-3508-4>
- Martelloni, G., Segoni, S., Fanti, R., Catani, F., 2012. Rainfall thresholds for the forecasting of landslide occurrence at regional scale. *Landslides* 9, 485-495. <https://doi.org/10.1007/s10346-011-0308-2>
- McCulloch, W. S., & Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* 5,4, 115-133.
- Melchiorre, C., Matteucci, M., Azzoni, A., Zanchi, A., 2008. Artificial neural networks and cluster analysis in landslide susceptibility zonation. *Geomorphology* 94, 379-400. <https://doi.org/10.1016/J.GEOMORPH.2006.10.035>

- Melillo, M., Brunetti, M.T., Peruccacci, S., Gariano, S.L., Guzzetti, F., 2016. Rainfall thresholds for the possible landslide occurrence in Sicily (Southern Italy) based on the automatic reconstruction of rainfall events. *Landslides* 13, 165–172. <https://doi.org/10.1007/s10346-015-0630-1>
- Melillo, M., Brunetti, M.T., Peruccacci, S., Gariano, S.L., Guzzetti, F., 2015. An algorithm for the objective reconstruction of rainfall events responsible for landslides. *Landslides* 12, 311–320. <https://doi.org/10.1007/s10346-014-0471-3>
- Melillo, M., Brunetti, M.T., Peruccacci, S., Gariano, S.L., Roccati, A., Guzzetti, F., 2018. A tool for the automatic calculation of rainfall thresholds for landslide occurrence. *Environ. Model. Softw.* 105, 230–243. <https://doi.org/10.1016/j.envsoft.2018.03.024>
- Mercogliano, P., Segoni, S., Rossi, G., Sikorsky, B., Tofani, V., Schiano, P., Catani, F., Casagli, N., 2013. Brief communication A prototype forecasting chain for rainfall induced shallow landslides. *Nat. Hazards Earth Syst. Sci.* 13, 771–777. <https://doi.org/10.5194/nhess-13-771-2013>
- Meyer, N.K., Dyrddal, A. V., Frauenfelder, R., Eitzelmann, B., Nadim, F., 2012. Hydrometeorological threshold conditions for debris flow initiation in Norway. *Nat. Hazards Earth Syst. Sci.* 12, 3059–3073. <https://doi.org/10.5194/nhess-12-3059-2012>
- Miao, F., Wu, Y., Xie, Y., Li, Y., 2018. Prediction of landslide displacement with step-like behavior based on multialgorithm optimization and a support vector regression model. *Landslides* 15, 475–488. <https://doi.org/10.1007/s10346-017-0883-y>
- Mirus, B.B., Becker, R.E., Baum, R.L., Smith, J.B., 2018a. Integrating real-time subsurface hydrologic monitoring with empirical rainfall thresholds to improve landslide early warning. *Landslides* 15, 1909–1919. <https://doi.org/10.1007/s10346-018-0995-z>
- Mirus, B.B., Morphew, M.D., Smith, J.B., 2018b. Developing hydro-meteorological thresholds for shallow landslide initiation and early warning. *Water (Switzerland)* 10, 1–19. <https://doi.org/10.3390/W10091274>
- Møller, M.F., 1993. A scaled conjugate gradient algorithm for fast supervised learning, *Neural Networks*. [https://doi.org/10.1016/S0893-6080\(05\)80056-5](https://doi.org/10.1016/S0893-6080(05)80056-5)
- Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D.G., Piles, M., Rodríguez-Fernández, N.J., Zsoter, E., Buontempo, C., Thépaut, J.N., 2021. ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data* 13, 4349–4383. <https://doi.org/10.5194/essd-13-4349-2021>
- Nikolopoulos, E.I., Borga, M., Creutin, J.D., Marra, F., 2015. Estimation of debris flow triggering rainfall: Influence of rain gauge density and interpolation methods. *Geomorphology* 243, 40–50. <https://doi.org/10.1016/j.geomorph.2015.04.028>

- Palau, R.M., Hürlimann, M., Berenguer, M., Sempere-Torres, D., 2021. Towards the use of hydrometeorological thresholds for the regional-scale LEWS of Catalonia (NE Spain). *Egu21* 8221.
- Palazzolo, N., Peres, D.J., Creaco, E., Cancelliere, A., 2022. Potential improvements of landslide prediction by hydro- meteorological thresholds : an investigation based on reanalysis soil moisture data and principal component analysis 1–22.
- Palazzolo, N., Peres, D.J., Creaco, E., Cancelliere, A., 2021. Exploring the potential of soil moisture reanalysis data for improving the identification of regional landslide triggering thresholds. *Egu2021* 5194.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., D. V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* 2825–2830.
- Pelosi, A., Terribile, F., D’Urso, G., Chirico, G.B., 2020. Comparison of ERA5-Land and UERRA MESCAN-SURFEX reanalysis data with spatially interpolated weather observations for the regional assessment of reference evapotranspiration. *Water (Switzerland)* 12. <https://doi.org/10.3390/W12061669>
- Pereira, S., Zêzere, J.L., Bateira, C., 2012. Technical Note: Assessing predictive capacity and conditional independence of landslide predisposing factors for shallow landslide susceptibility models. *Nat. Hazards Earth Syst. Sci.* 12, 979–988. <https://doi.org/10.5194/nhess-12-979-2012>
- Peres, D.J., Cancelliere, A., 2021. Comparing methods for determining landslide early warning thresholds: potential use of non-triggering rainfall for locations with scarce landslide data availability. *Landslides.* <https://doi.org/10.1007/s10346-021-01704-7>
- Peres, D.J., Cancelliere, A., 2016. Estimating return period of landslide triggering by Monte Carlo simulation. *J. Hydrol.* 541, 256–271. <https://doi.org/10.1016/J.JHYDROL.2016.03.036>
- Peres, D.J., Cancelliere, A., 2014. Derivation and evaluation of landslide-triggering thresholds by a Monte Carlo approach. *Hydrol. Earth Syst. Sci.* 18, 4913–4931. <https://doi.org/10.5194/hess-18-4913-2014>
- Peres, D.J., Cancelliere, A., Greco, R., Bogaard, T.A., 2018. Influence of uncertain identification of triggering rainfall on the assessment of landslide early warning thresholds. *Nat. Hazards Earth Syst. Sci.* 18, 633–646. <https://doi.org/10.5194/nhess-18-633-2018>
- Peruccacci, S., Brunetti, M.T., Gariano, S.L., Melillo, M., Rossi, M., Guzzetti, F., 2017. Rainfall thresholds for possible landslide occurrence in Italy. *Geomorphology* 290, 39–57. <https://doi.org/10.1016/J.GEOMORPH.2017.03.031>
- Peruccacci, S., Brunetti, M.T., Luciani, S., Vennari, C., Guzzetti, F., 2012. Lithological and seasonal control on rainfall thresholds for the possible initiation of landslides in central Italy. *Geomorphology* 139–140, 79–90. <https://doi.org/10.1016/J.GEOMORPH.2011.10.005>

- Piciullo, L., Calvello, M., Cepeda, J.M., 2018. Territorial early warning systems for rainfall-induced landslides. *Earth-Science Rev.* 179, 228–247. <https://doi.org/10.1016/j.earscirev.2018.02.013>
- Piciullo, L., Gilbert, G., 2022. Definition of soil water content and rainfall thresholds for landslide occurrence 16688.
- Ponziani, F., Pandolfo, C., Stelluti, M., Berni, N., Brocca, L., Moramarco, T., 2012. Assessment of rainfall thresholds and soil moisture modeling for operational hydrogeological risk prevention in the Umbria region (central Italy). *Landslides* 9(2), 229–237. <https://doi.org/10.1007/s10346-011-0287-3>
- Pumo, D., Francipane, A., Lo Conti, F., Arnone, E., Bitonto, P., Viola, F., La Loggia, G., Noto, L. V., 2016. The SESAMO early warning system for rainfall-triggered landslides. *J. Hydroinformatics* 18, 256–276. <https://doi.org/10.2166/hydro.2015.060>
- Ray, S., 2019. A Quick Review of Machine Learning Algorithms. *Proc. Int. Conf. Mach. Learn. Big Data, Cloud Parallel Comput. Trends, Perspectives Prospect. Com.* 2019 35–39. <https://doi.org/10.1109/COMITCon.2019.8862451>
- Reder, A., Rianna, G., 2021. Exploring ERA5 reanalysis potentialities for supporting landslide investigations: a test case from Campania Region (Southern Italy). *Landslides* 18, 1909–1924. <https://doi.org/10.1007/s10346-020-01610-4>
- Reichenbach, P., Rossi, M., Malamud, B.D., Mihir, M., Guzzetti, F., 2018. A review of statistically-based landslide susceptibility models. *Earth-Science Rev.* 180, 60–91. <https://doi.org/10.1016/j.earscirev.2018.03.001>
- Richards, L.A., 1931. Capillary conduction of liquids through porous mediums. *J. Appl. Phys.* 1, 318–333. <https://doi.org/10.1063/1.1745010>
- Rigon, R., Bertoldi, G., Over, T.M., 2006. GEOTop: A distributed hydrological model with coupled water and energy budgets. *J. Hydrometeorol.* 7, 371–388. <https://doi.org/10.1175/JHM497.1>
- Rosi, A., Segoni, S., Canavesi, V., Monni, A., Gallucci, A., Casagli, N., 2021. Definition of 3D rainfall thresholds to increase operative landslide early warning system performances. *Landslides* 18, 1045–1057. <https://doi.org/10.1007/s10346-020-01523-2>
- Rossi, G., Catani, F., Leoni, L., Segoni, S., Tofani, V., 2013. HIRESSS: A physically based slope stability simulator for HPC applications. *Nat. Hazards Earth Syst. Sci.* 13, 151–166. <https://doi.org/10.5194/nhess-13-151-2013>
- Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. *Nature* 323, 533–536.
- Saito, H., Nakayama, D., Matsuyama, H., 2010. Relationship between the initiation of a shallow landslide and rainfall intensity—duration thresholds in Japan. *Geomorphology* 118, 167–175. <https://doi.org/10.1016/j.geomorph.2009.12.016>
- Sarker, I.H., 2019. A machine learning based robust prediction model for real-life mobile phone data. *Internet of Things (Netherlands)* 5, 180–193. <https://doi.org/10.1016/j.iot.2019.01.007>

- Sarker, I.H., Kayes, A.S.M., Watters, P., 2019. Effectiveness analysis of machine learning classification models for predicting personalized context-aware smartphone usage. *J. Big Data* 6. <https://doi.org/10.1186/s40537-019-0219-y>
- Sathya, R., Abraham, A., 2013. Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification. *Int. J. Adv. Res. Artif. Intell.* 2, 34–38. <https://doi.org/10.14569/ijarai.2013.020206>
- Segoni, S., Leoni, L., Benedetti, A.I., Catani, F., Righini, G., Falorni, G., Gabellani, S., Rudari, R., Silvestro, F., Rebori, N., 2009. Towards a definition of a real-time forecasting network for rainfall induced shallow landslides. *Nat. Hazards Earth Syst. Sci.* 9, 2119–2133. <https://doi.org/10.5194/nhess-9-2119-2009>
- Segoni, S., Piciullo, L., Gariano, S.L., 2018a. A review of the recent literature on rainfall thresholds for landslide occurrence. *Landslides* 22, :478–494. <https://doi.org/10.1007/s10346-018-0966-4>
- Segoni, S., Rosi, A., Lagomarsino, D., Fanti, R., Casagli, N., 2018b. Brief communication: Using averaged soil moisture estimates to improve the performances of a regional-scale landslide early warning system. *Nat. Hazards Earth Syst. Sci.* 18, 807–812. <https://doi.org/10.5194/nhess-18-807-2018>
- Soci, C., Bazile, E., Besson, F.O., Landelius, T., 2016. High-resolution precipitation re-analysis system for climatological purposes. *Tellus, Ser. A Dyn. Meteorol. Oceanogr.* 68. <https://doi.org/10.3402/tellusa.v68.29879>
- Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res. Atmos.* 106.D7 106, 7183–7192.
- Thornthwaite, C.W., 1948. An approach toward a rational classification of climate. *Geogr. Rev.* 38 (1), 55e94. <https://doi.org/10.2307/210739>.
- Thornthwaite, C.W., Mather, J.R., 1957. Instructions and tables for computing potential evapotranspiration and the water balance. *Publ. Climatol. Lab. Climatol. Drexel Inst. Technol.* 10 (3), 185e311.
- Trigila A., Iadanza C., Lastoria B., Bussetini M., B.A., 2021. Dissesto idrogeologico in Italia: pericolosità e indicatori di rischio. - Ed. 2021. ISPRA, Rapp. 356/2021.
- Uwihirwe, J., Riveros, A., Wanjala, H., Schellekens, J., Weiland, F.S., Hrachowitz, M., Bogaard, T.A., 2022. Potential of satellite-derived hydro - meteorological information for landslide hazard assessment thresholds in Rwanda 1–32.
- van Natijne, A.L., Lindenbergh, R.C., Bogaard, T.A., 2020. Machine learning: New potential for local and regional deep-seated landslide nowcasting. *Sensors (Switzerland)* 20, 1–18. <https://doi.org/10.3390/s20051425>
- Varnes, D., 1978. Slope Movement Types and Processes. *Spec. Rep.* 176, 11–33.
- Wang, Q., Wang, D., Huang, Y., Wang, Z., Zhang, L., Guo, Q., Chen, Wei, Chen, Wengang, Sang, M., 2015. Landslide susceptibility mapping based on selected optimal combination of landslide predisposing factors in a large catchment. *Sustain.* 7, 16653–16669. <https://doi.org/10.3390/su71215839>
- Wicki, A., Jansson, P.E., Lehmann, P., Hauck, C., Stähli, M., 2021. Simulated or measured soil moisture: Which one is adding more value to regional landslide early warning? *Hydrol. Earth Syst. Sci.* 25, 4585–4610. <https://doi.org/10.5194/hess-25-4585-2021>

- Wicki, A., Lehmann, P., Hauck, C., Seneviratne, S.I., Waldner, P., Stähli, M., 2020. Assessing the potential of soil moisture measurements for regional landslide early warning. *Landslides* 17, 1881–1896. <https://doi.org/10.1007/s10346-020-01400-y>
- Widrow, B., Rumelhart, D. E., & Lehr, M. A. (1994). Neural networks: applications in industry, business and science. *Communications of the ACM*, 37(3), 93-106.
- Yin, Y., Wang, H., Gao, Y., Li, X., 2010. Real-time monitoring and early warning of landslides at relocated Wushan Town, the Three Gorges Reservoir, China. *Landslides* 7, 339–349. <https://doi.org/10.1007/s10346-010-0220-1>
- Yu, L., Zhou, C., Wang, Y., Cao, Y., Peres, D.J., 2022. Coupling Data-and Knowledge-Driven Methods for Landslide Susceptibility Mapping in Human-Modified Environments: A Case Study from Wanzhou County, Three Gorges Reservoir Area, China. *Remote Sens.* 14. <https://doi.org/10.3390/rs14030774>

RINGRAZIAMENTI

Il dottorato di ricerca è stato un periodo di crescita personale e professionale. Giunti alla conclusione di questo percorso è necessario e doveroso ringraziare chi ha messo la sua competenza e la sua umanità aiutandomi a realizzare questo importante obiettivo.

Un sentito ringraziamento va al Prof. Pietro Scandura che in qualità di Tutor mi ha permesso di lavorare, senza alcuna limitazione, sulle attività di mio interesse. I suoi spunti di riflessione sono sempre stati illuminanti.

Necessario ringraziare il Prof. Antonino Cancelliere, che già dai tempi della laurea magistrale mi permise di spendermi nel campo del rischio idrogeologico, e in qualità di Co-Tutor, e successivamente di coordinatore del dottorato mi ha coinvolto in numerose attività che hanno favorito il mio percorso di crescita.

Un doveroso ringraziamento è rivolto al Prof. David Peres, che è stato di stimolo e supporto durante tutto il dottorato, che sin da subito è riuscito a trasmettermi la passione nel campo della ricerca.

Un affettuoso ringraziamento va a tutto lo staff del Norwegian Geotechnical Institute che mi ha da subito fatto sentire un membro della loro grande famiglia, durante il periodo all'estero. Un sentito ringraziamento va, nello specifico al Dr. Luca Piciullo tutor durante il periodo all'estero, al direttore della sezione Natural Hazards Dominik Lang e ai due capo sezione Håkon Heyerdahl e Carl Bonnevie Harbitz.

È necessario, infine, ringraziare tutti i colleghi del dottorato con il quale è sempre esistito un meraviglioso dibattito e un continuo confronto costruttivo.

Il dottorato è finanziato dal "Bando 2/2019 per il finanziamento della borsa di dottorato regionale in Sicilia" nell'ambito del Programma Operativo dei Fondi Sociali Europei 2014–2020 (PO FSE 2014–2020, CUP E65E19000830002).