POLITECNICO DI TORINO

Facoltà di Ingegneria

Corso di Laurea Magistrale in Ingegneria per l'Ambiente e il Territorio

Tesi di Laurea Magistrale

Confronto su base non parametrica di curve regionalizzate di piogge estreme.

Relatori:

Prof. Pierluigi Claps

Ing. Andrea Libertino

Candidata:

Vittoria Bodi

Marzo 2019

Vorrei dedicare questo traguardo

alla persona che mi ha reso la donna forte che sono oggi.

Ti voglio bene Papà

Sommario

Indice delle Figure 7

Indice delle Tabelle 10

Introduzione 11

1 Metodi per la stima delle piogge estreme su area vasta 13

1.1 Classificazione dei metodi di analisi regionale 14

1.1.1 Procedura di Valutazione delle Piene 14

1.1.2 Il Kriging 15

1.1.3 Metodo della grandezza indice 16

1.2 Metodi di delimitazione delle regioni omogenee 18

1.2.1 La regionalizzazione per omogeneità geografica 19

1.2.2 La Cluster Analysis 20

1.2.3 La Region Of Influence 22

1.3 Verifica dell'omogeneità di una regione 24

1.4 La curva di crescita 26

1.5 Le curve di possibilità pluviometrica 27

2 Analisi regionale in Inghilterra 30

2.1 Selezione del sito di interesse 32

2.2 La variabile indice 33

2.3 La costruzione della curva di crescita 34

2.4 Stima del valore di progetto 35

3 Caratterizzazione geo-spaziale per la procedura FORGEX 37

3.1 Dati preliminari 38

3.2 Sistemi di rappresentazione territoriali 41

3.2.1 Coordinate Geografiche 41

3.2.2 Coordinate Cartografiche 42
3.3 Identificazione del sito oggetto dell'analisi regionale 45
4 Stima della curva di crescita con il metodo FORGEX 47
4.1 Identificazione delle serie di dati 49
4.1.1 Selezione e normalizzazione delle serie di dati adeguate 49
4.1.2 Identificazione e selezione del set di dati nell'intorno del sito 50
4.1.3 Metodo delle y-slices 53
4.1.4 Analisi dei valori massimi del network 56
4.2 Costruzione della curva di crescita 58
4.2.1 Costruzione della spezzata 58
4.2.2 Discretizzazione dell'intervallo 62
4.2.3 Selezione della spezzata ottimale 63
4.2.4 Fitting della spezzata con la migliore distribuzione 67
4.2.5 Il metodo FORGEX corretto 69
4.2.6 Validazione della distribuzione scelta 70
4.3 Confronto con altri metodi 72
4.3.1 Il metodo VAPI 72
4.3.2 Il metodo Kriging 73
4.3.3 Confronto tra le differenti curve di crescita 75
4.3.4 Confronto tra le differenti curve di possibilità pluviometrica 77
4.4 Particolarità del metodo FORGEX 79
5 Costruzione di un sistema informatizzato per la stima della curva di crescita con il metodo FORGEX 83
5.1 Identificazione dei dati 84
5.1.1 Analisi dei dati 84
5.1.2 Normalizzazione dei dati di precipitazione 89
5.1.3 Relazione tra i dati e il periodo di ritorno 91

5.2 Classificazione del set di dati 92
5.2.1 Metodo delle y-slices 92
5.2.2 Costruzione dei differenti network 97
5.3 Costruzione della curva di crescita 103
5.3.1 Costruzione del grafico y-slices 103
5.3.2 Calcolo dei gradienti 105
5.3.3 Costruzione della retta spezzata 107
5.3.4 Selezione della spezzata con il migliore adattamento ai punti 108
5.3.5 Calcolo dei parametri delle distribuzioni di probabilità 109
5.3.6 Costruzione della curva di crescita del metodo FORGEX 110
5.3.7 Costruzione delle curve di frequenza empirica 110
5.4 Costruzione delle curve di crescita di confronto 112
5.5 Costruzione delle curve di possibilità pluviometrica 113
6 Conclusioni 114
7 Appendice 117
7.1 Appendice A: Definizioni 117
7.1.1 Altezza di pioggia h_d 117
7.1.2 Durata delle precipitazioni 117
7.1.3 Tempo di ritorno 118
7.1.4 Curva di crescita 118
7.1.5 Regione omogenea. 119
7.1.6 Metodo Station Year 119
7.2 Appendice B: Distribuzioni di probabilità 120
7.2.1 Distribuzioni di probabilità di Gumbel 120
7.2.2 Distribuzione di probabilità GEV 122
7.3 Appendice C: Descrizione della piattaforma di sviluppo del software 125

7.3.1 Editor per la programmazione 125

7.3.2 Le principali librerie di Python 3.7 128

7.4 Appendice D: Codice dell'applicativo creato 134

7.4.1 Parte A 134

7.4.2 Parte B 142

7.5 Appendice E: Codice delle funzioni costruite per la realizzazione dell'applicativo 153

7.5.1 Parte A 153

7.5.2 Parte B 157

7.6 Appendice F: Schede delle località analizzate 160

7.6.1 Torino Ufficio Idrografico 160

7.6.2 Spigno Monferrato 166

7.6.3 Oropa Cervo 172

7.6.4 Varallo Sesia 178

7.6.5 Bardonecchia 184

7.7 Appendice G: mappe regionalizzate per il Kriging 190

8 Bibliografia 192

9 Sitografia 195

10 Ringraziamenti 196

Indice delle Figure

Figura 1. 1: metodi di delineazione delle regioni omogenee (Ouarda, 2001) 19
Figura 1. 2: Rappresentazione del risultato di una cluster analysis in un sito di interesse (Viglione, 2007) 22
Figura 1. 3: Rappresentazione dell'applicazione della region of influence in un sito di interesse (Viglione, 2007) 23
Figura 1. 4: rappresentazione di una generica distribuzione normale simmetrica rispetto al valore medio, (http://www.biostatistica.unich.it/mat_didattica/MENNA/A_Menna_UniCH_Lez2_2008_DistribuzioneNormale.pdf) 25
Figura 1. 5: rappresentazione delle differenti aree dell'Italia Occidentale, suddivise in regioni omogenee in funzione della tipologia di dati osservati. A destra si osservano le curve di crescita per ogni regione omogenea 27
Figura 1. 6: curva di possibilità pluviometrica in funzione delle differenti durate. 28
Figura 1. 7: curva di possibilità pluviometrica, una per ogni tempo di ritorno 29

Figura 2. 1: rappresentazione delle stazioni nei dintorni di un sito scelto per l'analisi. 32
Figura 2. 2: mappa dell'Inghilterra, rappresentativa del parametro indice (RMED), per la durata di 1 giorno, (Faulkner & Prudhomme, 1998) 33
Figura 2. 3: illustrazione della procedura di selezione del metodo FORGEX, anche utilizzata nel metodo 'FORGE', per la costruzione della curva di crescita, (Svensson & Jones , 2010) 34
Figura 2. 4: rappresentazione del fattore di crescita per T=100 anni, su tutto il territorio inglese (Faulker, 1999) 35
Figura 2. 5: rappresentazione della pioggia di progetto per durata 1 ora e T=100 anni, (Faulker, 1999) 36

Figura 3. 1: rappresentazione delle coordinate geografiche (tratto da http://profpozzi-classe004.blogspot.com/2017/11/le-coordinate-geografiche.html) 41
Figura 3. 2: rappresentazione del reticolo geografico, costruito dall'intersezione dei diversi meridiani e paralleli (tratto da http://profpozzi-classe004.blogspot.com/2017/11/le-coordinate-geografiche.html) 42
Figura 3. 3:rappresentazione della cartografia italiana, secondo la proiezione di Gauss-Boaga (tratto da https://geomappando.com/2016/07/07/sistemi-riferimento-coordinate-2-parte/) 43
Figura 3. 4: rappresentazione in coordinate UTM, (tratto da http://geomatica.como.polimi.it/corsi/cartografialaurea/cartografiaufficialeitaliana/cartografiaufficialeitalianaigmi.htm) 44
Figura 3. 5: rappresentazione delle stazioni presenti sul territorio piemontese, organizzate per quota. 45

Figura 4. 1: albero delle azioni del metodo FORGEX 48
Figura 4. 2: confronto del metodo di selezione dei network sul territorio inglese e piemontese, con differenti scale 51
Figura 4. 3: rappresentazione del grafico y-slices, con indicazione del valore mediano con un punto blu. 54
Figura 4. 4: Rappresentazione dei punti raggruppati per ogni network, nell'intorno del sito di Torino. 55

FIGURA 4. 5: RAPPRESENTAZIONE DELLE SERIE NETMAX SUL GRAFICO Y-SLICES, NELL'INTORNO DEL SITO DI TORINO. 58
FIGURA 4. 6: RAPPRESENTAZIONE DEL SET DI DATI DA UTILIZZARE NELLA COSTRUZIONE DELLA SPEZZATA, NELL'INTORNO DEL SITO DI TORINO. 59
FIGURA 4. 7: RAPPRESENTAZIONE DELLA CURVA DI CRESCITA COSTRUITA A TRATTI PER IL SITO DI TORINO PER LA DURATA DI 3H. 61
FIGURA 4. 8: RAPPRESENTAZIONE DELLA FUNZIONE A TRATTI SCELTA PER LA LOCALITÀ TORINO. 66
FIGURA 4. 9: RAPPRESENTAZIONE DELLE DISCRETIZZAZIONI RISPETTO AL SET DI DATI, IN FUNZIONE DEI TRE PARAMETRI. 66
FIGURA 4. 10: RAPPRESENTAZIONE DELLE DIFFERENTI DISTRIBUZIONI PER LA SCELTA DELLA PIÙ ADATTA. 68
FIGURA 4. 11: CONFRONTO TRA LA DISTRIBUZIONE GEV COSTRUITA CON I VALORI NORMALIZZATI RISPETTO ALLA MEDIANA, E ALLA MEDIA. 70
FIGURA 4. 12: CONFRONTO TRA LA DISTRIBUZIONE GEV E LE CURVE DI FREQUENZA EMPIRICA NEL SITO DI TORINO, PER VERIFICARE LA VALIDITÀ DELLA DISTRIBUZIONE SELEZIONATA, IN FUNZIONE DELLA VARIABILE RIDOTTA E DEL TEMPO DI RITORNO. 71
FIGURA 4. 13: RAPPRESENTAZIONE DELLA DISTRIBUZIONE GEV, A CONFRONTO CON LE CURVE DI FREQUENZA EMPIRICA, SIA IN FUNZIONE DI Y CHE DI T, NEL SITO DI TORINO. 73
FIGURA 4. 14: RAPPRESENTAZIONE DELLE DISTRIBUZIONI GEV OTTENUTE CON IL METODO KRIGING, A CONFRONTO CON LE CURVE DI FREQUENZA EMPIRICA, SIA IN FUNZIONE DELLA Y CHE DEL T, NEL SITO DI TORINO. 74
FIGURA 4. 15: RAPPRESENTAZIONE DELLE DISTRIBUZIONI GEV A CONFRONTO, IN BASE AI VALORI DI Y E DEI TEMPI DI RITORNO, NEL SITO DI TORINO. 76
FIGURA 4. 16: RAPPRESENTAZIONE DELLE CPP PER I TRE METODI DIFFERENTI, IN FUNZIONE DELLA Y E DEL TEMPO DI RITORNO, NEL SITO DI TORINO. 78
FIGURA 4. 17: RAPPRESENTAZIONE DELLA REGIONE PIEMONTE, CON IL NETWORK PIÙ AMPIO NELL'INTORNO DI BARDONECCHIA 79
FIGURA 4. 18: RAPPRESENTAZIONE DELLE TRE CURVE DI CRESCITA OTTENUTE CON LE DIFFERENTI METODOLOGIE PER IL SITO DI BARDONECCHIA. 80
FIGURA 4. 19: LOCALIZZAZIONE DEL SITO DI SPIGNO MONFERRATO E RAPPRESENTAZIONE DEL NETWORK PIÙ AMPIO NELL'INTORNO DEL SITO STESSO. 81
FIGURA 4. 20:RAPPRESENTAZIONE DELLE TRE CURVE DI CRESCITA OTTENUTE CON LE DIFFERENTI METODOLOGIE PER IL SITO DI SPIGNO MONFERRATO. 82

FIGURA 5. 1:DIAGRAMMA DI FLUSSO DEL BLOCCO DI CODICE INIZIALE 84
FIGURA 5. 2: SCHEMA GERARCHICO PER LA FUNZIONE 'NUMEROSITÀ' 86
FIGURA 5. 3: DIAGRAMMA DI FLUSSO DEL CODICE RELATIVO ALLA NORMALIZZAZIONE DELLE SERIE DI DATI. 89
FIGURA 5. 4: SCHEMA GERARCHICO DELLA FUNZIONE 'TEST' ESEGUITO SULLA DURATA. 90
FIGURA 5. 5: DIAGRAMMA DI FLUSSO DELLA TERZA ROUTINE PER IL CALCOLO DELLA FREQUENZA. 91
FIGURA 5. 6: DIAGRAMMA DI FLUSSO, RAPPRESENTATIVO DELLA ROUTINE DI CLASSIFICAZIONE DEL DATO. 92
FIGURA 5. 7: DIAGRAMMA DI FLUSSO DELLA FUNZIONE 'CLASSIFY' PER LA CLASSIFICAZIONE DELLE VARIABILI RIDOTTE Y. 94
FIGURA 5. 8: DIAGRAMMA DI FLUSSO DELLA FUNZIONE 'NUM_PER_SLICE'. 95
FIGURA 5. 9: DIAGRAMMA DI FLUSSO, DELLA FUNZIONE 'DIVERSI', PER UNIFORMARE LE MATRICI DELLE NUMEROSITÀ. 96
FIGURA 5. 10: DIAGRAMMA DI FLUSSO DELLA ROUTINE 4 97
FIGURA 5. 11: DIAGRAMMA DI FLUSSO, DELLA FUNZIONE 'NETWORK', UTILIZZATA PER LA RICERCA DEI DATI. 98
FIGURA 5. 12: DIAGRAMMA DI FLUSSO DELLA FUNZIONE CHE PERMETTE DI RICAVARE LA SERIE DI NETMAX. 99
FIGURA 5. 13: SET DI DIAGRAMMI DI FLUSSO DELLA FUNZIONE 'OFFSET'. 100

FIGURA 5. 14: RAPPRESENTAZIONE DEL DIAGRAMMA DI FLUSSO DELLA FUNZIONE 'OFFSET' PER IL CALCOLO DELLA DISTANZA DI OFFSET EFFETTIVA. 101
FIGURA 5. 15: DIAGRAMMA DI FLUSSO DELLA FUNZIONE 'FREQUENZA'. 102
FIGURA 5. 16: DIAGRAMMA DI FLUSSO PER LA COSTRUZIONE DELLA CURVA DI CRESCITA. 103
FIGURA 5. 17: DIAGRAMMA DI FLUSSO RAPPRESENTATIVO DELLA SUDDIVISIONE NEI DIVERSI INTERVALLI DI DATI. 104
FIGURA 5. 18: RAPPRESENTAZIONE DELLE OPERAZIONI ESEGUITE PER LA SUDDIVISIONE IN DIFFERENTI SLICES. 105
FIGURA 5. 19: DIAGRAMMA DI FLUSSO DELLA FUNZIONE 'GRADIENTE'. 106
FIGURA 5. 20: RAPPRESENTAZIONE DEL DIAGRAMMA DI FLUSSO DELLA FUNZIONE 'MODEL', PER LA RICERCA DEI VALORI DELLE ORDINATE DEGLI ESTREMI DELLE SLICES. 107
FIGURA 5. 21: RAPPRESENTAZIONE DELLA PROCEDURA DI CALCOLO DEI VALORI DELLE ORDINATE DI MODELLO 108
FIGURA 5. 22: RAPPRESENTAZIONE DELLA FUNZIONE 'EQ1', PER LA SELEZIONE DELLA MIGLIORE SPEZZATA. 109
FIGURA 5. 23: DIAGRAMMA DI FLUSSO DELLA FUNZIONE 'CURVE_FIT'. 109
FIGURA 5. 24: DIAGRAMMA DI FLUSSO PER L'APPLICAZIONE DELLA CORREZIONE AL SET DI DATI OTTENUTO CON IL FORGEX. 110
FIGURA 5. 25: PROCEDURA PER LA COSTRUZIONE DELLA CURVA DI FREQUENZA EMPIRICA DI STAZIONE 111
FIGURA 5. 26: DIAGRAMMA DI FLUSSO PER LA COSTRUZIONE DELLA CURVA 'STATION YEAR'. 111
FIGURA 5. 27:DIAGRAMMA DI FLUSSO PER LA STIMA DELLA CURVA DI POSSIBILITÀ PLUVIOMETRICA. 113

FIGURA 7. 1:SCHERMATA DELL'INTERFACCIA GRAFICA DEL SOFTWARE SPYDER UTILIZZATO PER PROGETTARE IL PROGRAMMA, DETTAGLIO DELL'EDITOR 126
FIGURA 7. 2: SCHERMATA DELL'INTERFACCIA GRAFICA DEL SOFTWARE SPYDER UTILIZZATO PER PROGETTARE IL PROGRAMMA, DETTAGLIO DELLA CONSOLE. 126
FIGURA 7. 3: SCHERMATA DELL'INTERFACCIA GRAFICA DEL SOFTWARE SPYDER UTILIZZATO PER PROGETTARE IL PROGRAMMA, DETTAGLIO DELLA FINESTRA 'FILE EXPLORER'. 127
FIGURA 7. 4: SCHERMATA DELL'INTERFACCIA GRAFICA DEL SOFTWARE SPYDER UTILIZZATO PER PROGETTARE IL PROGRAMMA, DETTAGLIO DELLA FINESTRA 'VARIABLE EXPLORER'. 127
FIGURA 7. 5: SCHERMATA DELL'INTERFACCIA GRAFICA DEL SOFTWARE SPYDER UTILIZZATO PER PROGETTARE IL PROGRAMMA. 128
FIGURA 7. 6: RAPPRESENTAZIONE DELLE MAPPE REGIONALIZZATE DEI PARAMETRI θ_1 θ_2 θ_3 UTILIZZATI PER IL CALCOLO DELLA DISTRIBUZIONE CON IL METODO KRIGING 190
FIGURA 7. 7: RAPPRESENTAZIONE DELLE MAPPE CONTENENTI LA REGIONALIZZAZIONE DEL VALORE MEDIO, UTILIZZATO PER LA COSTRUZIONE DELLA CURVA CPP 191

Indice delle Tabelle

TABELLA 1. 1: TEMPI DI RITORNO UTILIZZATI PER LA STIMA DELLE GRANDEZZE DI PROGETTO. 27

TABELLA 3. 1: ESTRATTO DELLA TABELLA RIASSUNTIVA DEI MASSIMI ANNUI DI PRECIPITAZIONE PER LE DIVERSE STAZIONI. 38
TABELLA 3. 2: ESTRATTO DELLA TABELLA RIASSUNTIVA DELLE INFORMAZIONI RELATIVE ALLE STAZIONI PRESENTI SUL TERRITORIO PIEMONTESE. 40
TABELLA 3. 3: LOCALITÀ NELLE QUALI È POSSIBILE TESTARE L'APPLICATIVO CREATO. 46

TABELLA 4. 1:VALORI DELLE VARIABILI RIDOTTE DI GUMBEL, UTILIZZATE NEL METODO Y-SLICES. 53
TABELLA 4. 2: VALORI DELL'EQUAZIONE (12) CALCOLATI PER IL SITO DI TORINO. 65
TABELLA 4. 3: TABELLA RIASSUNTIVA DEI PARAMETRI OTTENUTI PER LE DISTRIBUZIONI GEV E GUMBEL. 69
TABELLA 4. 4: PARAMETRI CARATTERISTICI PER LA COSTRUZIONE DELLA CURVA DI CRESCITA CON IL METODO VAPI. 72
TABELLA 4. 5: PARAMETRI CARATTERISTICI DEL METODO KRIGING ESTRATTI DALLE MAPPE REGIONALIZZATE 74
TABELLA 4. 6: VALORI DEI QUANTILI OTTENUTI DALLE TRE DIFFERENTI CURVE DI CRESCITA. 76
TABELLA 4. 7: VALORI DELLE ALTEZZE DI PRECIPITAZIONE OTTENUTE PER LE DIFFERENTI CURVE DI POSSIBILITÀ PLUVIOMETRICA 78

TABELLA 5. 1: INTESTAZIONE DEL FILE INIZIALE CON LE INFORMAZIONI DELLE STAZIONI 85
TABELLA 5. 2: INTESTAZIONE DEL FILE INIZIALE DEI DATI 85
TABELLA 5. 3: INTESTAZIONE DEL FILE 'ANAGRAFICA'. 88
TABELLA 5. 4: INTESTAZIONE DEL FILE RAPPRESENTATIVO DELLE PRECIPITAZIONI. 88
TABELLA 5. 5: INTESTAZIONE DEL FILE OUTPUT DELLA SECONDA ROUTINE DI CODICE. 91
TABELLA 5. 6: INTESTAZIONE DEL FILE CHE CONTIENE I VALORI DELLA VARIABILE RIDOTTA Y, DENOMINATO 'VALORI_Y'. 92
TABELLA 5. 7: INTESTAZIONE DEL FILE OTTENUTO DALL'APPLICAZIONE DELLA FUNZIONE 'CLASSIFY' 93
TABELLA 5. 8: INTESTAZIONE DELLA TABELLA OTTENUTA DALLA FUNZIONE 'NUM_PER_SLICE'. 95
TABELLA 5. 9: INTESTAZIONE DELLA MATRICE GENERALE 'ANAGRAFICA'. 97
TABELLA 5. 10: INTESTAZIONE DELLA TABELLA OTTENUTA A SEGUITO DELLE OPERAZIONI INTERSEZIONE TRA FILE. 99
TABELLA 5. 11: INTESTAZIONE DELLE DIFFERENTI MATRICI DENOMINATE 'PLOT_NUM_SLICES' 102
TABELLA 5. 12: INTESTAZIONE DELLE TABELLE 'PSX' E 'PSY' UTILIZZATE PER LA COSTRUZIONE DELLA SPEZZATA. 105
TABELLA 5. 13: INTESTAZIONE DELLA TABELLA RIELABORATA DELLA RETE DI STAZIONI PIÙ AMPIA. 111

Introduzione

Le precipitazioni estreme sono fenomeni molto interessanti dal punto di vista ambientale. L'abbattimento al suolo di grandi quantità di acqua, può provocare al territorio alcuni problemi, come frane o alluvioni, che nel peggiore dei casi, comportano la perdita di vite umane. Nel corso degli anni sono stati sviluppati studi differenti, per stimare la frequenza di questi fenomeni, cercare di limitare e, ove possibile, prevenire i danni ad essi associati.

In Italia, la stima della pioggia di progetto viene effettuata in funzione della disponibilità dei dati presenti sul territorio in esame. In presenza di stazioni strumentate, è prevista l'applicazione di una stima locale: si tratta di un'analisi basata sull'inferenza statistica delle serie storiche di massimi annuali di precipitazione. Il problema sorge in presenza di aree prive di dati sufficienti, oppure di stazioni installate di recente; quest'ultime infatti, non possiedono ancora serie storiche efficacemente robuste, da garantire un'adeguata precisione dell'analisi.

La stima della precipitazione di progetto in aree non strumentate, viene eseguita attraverso i metodi di analisi regionale, che si sviluppano sostituendo il valore temporale con un valore spaziale. Questo significa, che l'analisi si concentrerà sulla definizione di un'area, all'interno della quale, i dati raccolti dovranno garantire similarità tra loro, in modo da fornire una stima adeguata del valore richiesto, per il sito non strumentato.

Lo scopo di questa tesi è costruire un sistema informatizzato, fondato sulla metodologia applicata in Inghilterra, proposta dal 'Flood Estimation Handbook'. Il 'FEH' suggerisce un metodo di analisi regionale basato sull'uso di un particolare software, denominato 'FOcused Rainfall Growth curves EXtension'. Il 'FORGEX' permette la costruzione di curve di crescita per un sito privo di dati, in funzione esclusivamente della numerosità delle stazioni limitrofe.

Il metodo applicato in Inghilterra prevede una serie di accorgimenti, per garantire una corretta applicazione ai dati italiani. La prima differenza importante su cui porre l'accento, è la posizione geografica delle due nazioni: il clima differente a cui sono sottoposte, sicuramente, influenzerà la quantità delle informazioni relative alle precipitazioni osservate nei due paesi. In Inghilterra, si osserva una concentrazione di stazioni più elevata di quella presente sul territorio piemontese, come a dimostrazione della maggiore frequenza di eventi di precipitazione, che si riversano sul territorio inglese.

In secondo luogo, è necessario evidenziare il valore indice differente utilizzato nelle due aree. Il metodo inglese, si basa completamente sulla grandezza di riferimento, identificata con il valore mediano: in Inghilterra, infatti, sono state costruite una serie di mappe, rappresentanti questo elemento regionalizzato sull'intero territorio. In Italia, la grandezza di riferimento usata, è il valore medio: per tale motivo, sarà necessario imporre al sistema inglese una correzione, così da poter garantire il confronto con i risultati di altre analisi regionali, per verificare l'attendibilità del metodo 'FORGEX'.

Lo scopo di un'analisi regionale è valutare l'entità dell'evento più raro in grado di riversarsi in un determinato luogo. Per raggiungere questo risultato, si ricorre a metodi parametrici: grazie alla stima di particolari parametri statistici, il professionista è in grado di generare dei modelli, da cui ricavare il valore di progetto cercato. La tecnica descritta nel corso dell'elaborato, consente di ricavare lo stesso risultato senza l'ausilio di alcun tipo di parametro, ma basandosi esclusivamente sulle informazioni dellle stazioni adiacenti al sito in esame.

Capitolo 1

Metodi per la stima delle piogge estreme su area vasta

La stima delle precipitazioni estreme in aree molto ampie è uno degli obiettivi principali di un'analisi idrologica. La caratteristica sostanziale su cui è importante porre l'accento è la presenza o meno di stazioni strumentate. La stima locale, non essendo sempre applicabile, risulta di più difficile realizzazione, specialmente in casi in cui i dati non sono facilmente reperibili o non sono utili per una stima adeguata.

Nel corso degli anni, sono state sviluppate numerose tecniche di stima regionale. Un'analisi regionale, infatti, basa la sua applicazione sulla stima della grandezza richiesta nel sito scelto, in base alle osservazioni nelle aree limitrofe ad essa. Questa procedura risulta piuttosto complessa, sia per la stima di un evento di precipitazione, che per quella del picco di piena.

1.1 *Classificazione dei metodi di analisi regionale*

Un'analisi regionale viene applicata in determinate zone in cui non si hanno sufficienti osservazioni. Questa tipologia di analisi prevede di ottenere una stima della frequenza di accadimento dell'evento, esclusivamente sulla base dei dati provenienti dalle stazioni adiacenti al sito scelto. La chiave di un metodo regionale, appare proprio la sostituzione del tempo con lo spazio: questo vuol dire che invece di considerare le serie di eventi in funzione dell'anno in cui sono avvenuti, verranno creati degli insiemi di dati, raggruppati per caratteri simili, come la posizione sul territorio, o alcuni parametri statistici comuni.

L'analisi regionale in Italia si basa su alcuni dei seguenti metodi fondamentali:

1) La procedura di Valutazione delle Piene in Italia, espressa nel rapporto di De Michele e Rosso, del 2001;
2) Il metodo diretto di interpolazione geo-statistica denominato Kriging;
3) Il metodo della grandezza indice, più comunemente utilizzato per la valutazione delle piene.

Queste metodologie prevedono la scelta di una distribuzione di probabilità, la cui stima dei parametri, attraverso un approccio statistico, garantisce il valore della grandezza di progetto, nel sito oggetto d'esame.

1.1.1 *Procedura di Valutazione delle Piene*

Il metodo VAPI, così denominato nel corso degli anni, consente una stima della grandezza di progetto in differenti aree del territorio, basandosi sull'applicazione di una distribuzione a quattro parametri, la TCEV[1]. Tale distribuzione, è utilizzata poiché suddivide la serie di massimi annui in esame, in due popolazioni di eventi, rispettivamente ordinari e straordinari. In tal modo i due campioni tendono ad essere confrontati e analizzati in base all'intensità dei loro eventi.

[1] La distribuzione Two Components Extreme Values, è una distribuzione di probabilità a 4 parametri utilizzata in campo idrologico, definita anche come *distribuzione asintotica del massimo valore a due componenti.*

La procedura VAPI, si basa su diversi livelli di regionalizzazione, che garantiscono una stima dei parametri caratteristici sulla base di set di dati, la cui numerosità è molto elevata. Ad ogni livello di regionalizzazione, si valuta quanto il parametro considerato sia omogeneo nell'area in esame: ad ogni livello successivo, si cerca di incrementare le zone omogenee garantendo una stima più precisa dei parametri in oggetto.

I differenti livelli sono classificati come segue:

1) Nel primo livello di regionalizzazione si valuta il coefficiente di asimmetria, solitamente sull'intera area oggetto di studio;
2) Nel secondo livello di regionalizzazione si valuta il coefficiente di variazione, cercando di suddividere l'area in almeno due sottozone omogenee;
3) Nel terzo livello di regionalizzazione si valuta il parametro direttamente proporzionale alla media del set di dati.

I primi due livelli di regionalizzazione forniscono una stima della curva di crescita oggetto dell'analisi. Il terzo livello di regionalizzazione, invece, produce il valore della grandezza indice per il sito desiderato. Il risultato di un'analisi regionale eseguita attraverso il metodo VAPI, garantisce una stima della grandezza di progetto nell'area desiderata.

1.1.2 *Il Kriging*

Il metodo Kriging si basa sull'interpolazione geo-statistica di determinate grandezze note sul territorio, in modo da garantire la stima di quest'ultime in aree prive di informazioni. Il concetto di base di questa metodologia prevede la minimizzazione degli scarti quadratici medi tra le varie grandezze appartenenti a siti differenti. La stima di tali valori, in località non strumentate, avviene attraverso la valutazione dell'autocorrelazione tra le differenti variabili: a seconda della posizione spaziale del sito in esame sul territorio, si andrà a stimare la grandezza in oggetto in funzione di ciò che è presente nei dintorni.

Il grado di correlazione tra le differenti variabili deve essere misurato in modo da garantire una precisione adeguata degli elementi da stimare: questi valori vengono determinati in base ad una serie di variogrammi, che permetteranno di assegnare ad ogni elemento il suo giusto peso. Questa metodologia, non solo garantisce una stima delle grandezze in aree prive di osservazioni, ma permette all'utente di visualizzare

determinati errori, associati ad ogni operazione svolta. Il risultato di queste procedure, prevede una serie di mappe rappresentative dei caratteri principali dell'area in esame, in funzione della topografia di quest'ultima.

Le tecniche principalmente utilizzate per la stima delle piene di progetto sono le seguenti:

- Ordinary kriging che stima i pesi da attribuire ad ogni sito in funzione dei valori osservati nei differenti variogrammi;
- Topologikal Kriging che invece stima i parametri per ogni sito in modo arbitrario ed in funzione della topografia del sito stesso.

Le mappe in [Appendice G: mappe regionalizzate per il Kriging], mostrano i risultati che si ottengono per la regionalizzazione dei parametri tipici delle distribuzioni e dei valori medi ricavati nelle differenti località. Il criterio principale su cui si basano queste interpolazioni prevede la costruzione di un set di dati virtuali: nel caso delle precipitazioni, viene realizzata una rete di stazioni molto più fitta di quella esistente, con una maglia di circa 250 m tra le differenti stazioni. Lo scopo principale è aumentare il numero delle stazioni presenti sul territorio e allo stesso tempo incrementare la numerosità delle osservazioni per ogni sito.

In questo modo sarà possibile conoscere i valori medi, o i parametri, per esempio di una GEV, in qualunque punto del territorio, anche se in quell'area non sono presenti stazioni strumentate.

1.1.3 *Metodo della grandezza indice*

Il metodo indice è quello più diffuso nel campo dell'analisi regionale. L'ipotesi di base di tale metodologia è la stima della grandezza indice, che può essere identificata come il valore medio o mediano della serie di dati in esame. Tale elemento è utilizzato per la stima del valore di progetto dall'analisi.

L'elemento fondamentale, identificato con questa analisi, è il fattore di crescita, per il quale la grandezza indice andrà moltiplicata. Il fattore di crescita esprime un legame tra ogni elemento della serie di dati e il valore medio della stessa. La stima del fattore di

crescita produce una curva denominata curva di crescita, il cui obiettivo è la misura dell'incremento degli eventi estremi per un dato periodo di ritorno.

La caratteristica fondamentale di tale metodologia è l'utilizzo di una distribuzione a tre parametri, denominata GEV[2], che garantisce la costruzione in modo lineare della curva di crescita ricercata.

La curva di crescita, quindi, viene espressa in funzione di tre grandezze fondamentali:

- La media, o la mediana, che varia in funzione della numerosità della serie di dati;
- Il parametro di scala rappresentato da un momento del secondo ordine;
- Il parametro di forma, indicato dal momento del terzo ordine.

Il metodo indice prevede una serie di passaggi fondamentali per il corretto svolgimento di un'analisi regionale:

1) Ipotesi di delimitazione della regione omogenea sulla quale eseguire l'analisi;
2) Calcolo della grandezza indice;
3) Normalizzazione del campione in ogni sito, secondo la grandezza indice;
4) Stima dei parametri per la costruzione della curva di crescita dell'area;
5) Verifica dell'omogeneità dell'area omogenea.

L'obiettivo principale di ogni analisi regionale, eseguita con il metodo indice, è la stima della curva di crescita: a seconda della zona di appartenenza del sito si potranno osservare andamenti differenti.

Il metodo indice risulta di più semplice applicazione, anche grazie all'utilizzo di una funzione a tre parametri, spesso riconducibile ad una funzione a due parametri, come la Gumbel. Il punto cruciale di questo approccio è la ricerca di zone omogenee, tali da garantire la stima dei parametri idonea al sito in esame.

[2] La distribuzione 'Generalised Extreme Values', definita GEV, è una distribuzione a tre parametri, approfondita in [Appendice B: Distribuzioni di probabilità].

1.2 *Metodi di delimitazione delle regioni omogenee*

Nel corso degli anni sono state definite una serie di metodologie di delimitazione delle regioni, fondate sul concetto della continuità geografica: i metodi applicati consideravano l'omogeneità delle aree in base ai limiti amministrativi. In tal modo si ottenevano regioni climaticamente omogenee, ma con lacune legate all'ipotesi di similarità statistica. I primi problemi legati a questo approccio, emersero in presenza di stazioni poste ai confini della regione considerata, e di conseguenza, non in grado di dimostrare, in alcun modo, la loro appartenenza ad una regione piuttosto che ad un'altra. Vennero studiati ulteriori sistemi di delimitazione, che non consideravano le regioni in base alla continuità geografica, ma ricercando i caratteri statistici simili tra le stazioni differenti.

Il metodo utilizzato per uniformare i dati provenienti da zone non adiacenti prende il nome di 'pooling group'. Questa tecnica è molto utilizzata in ambito idrologico, in quanto permette di costruire gruppi di elementi differenti, sulla base di caratteri della stazione di appartenenza e della sua serie di dati.

I sistemi basati sul principio del 'pooling group' sono fondamentalmente di due tipi, la 'cluster analysis' e la 'region of influence'. Tali metodologie permettono di analizzare i dati raccolti da stazioni diverse, e raggrupparli in modo da considerare un unico macro campione di dati osservati. I metodi sopra citati, risultano molto differenti tra loro, ma non per questo inaffidabili.

La [Figura 1. 1] illustra i metodi accennati in precedenza, ponendo l'accento sui differenti risultati ottenibili. La mappa a), presenta una delimitazione per confini geografici, la mappa b) mostra il metodo della 'cluster analysis', mentre la mappa c) evidenzia la procedura delle 'region of influence'. Non si osservano da questa immagine i problemi associati all'applicazione di un metodo piuttosto che di un altro, ma è ben evidente come ciò che si ottiene risulti notevolmente differente.

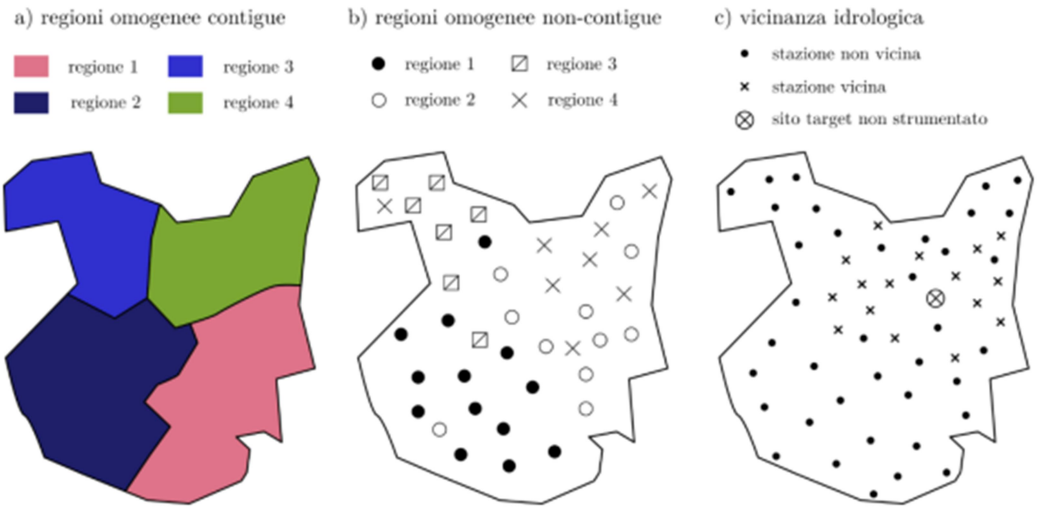


Figura 1. 1: metodi di delineazione delle regioni omogenee (Ouarda, 2001)

1.2.1 *La regionalizzazione per omogeneità geografica*

Questa particolare tecnica, consiste nel raggruppare i dati provenienti da stazioni limitrofe, considerando eventuali limiti amministrativi già noti. Non vengono considerati in alcun modo, gli eventuali caratteri statistici simili, o le caratteristiche climatiche differenti, che possono presentarsi tra le diverse serie di dati. L'unico criterio di selezione delle stazioni sarà la localizzazione in termini di confini geografici.

L'utilizzo di tale procedura potrebbe causare numerosi errori, dovuti alle differenze tra le serie di dati raggruppate in un'unica area: si possono osservare delle situazioni nelle quali, stazioni presenti in una data area, non abbiano alcuna caratteristica comune tra loro, a meno della loro posizione sul territorio. Per tale motivo, spesso selezionare le stazioni in base alla loro localizzazione, potrebbe non essere un vantaggio a discapito dei risultati dell'analisi regionale.

Questo approccio potrebbe fornire determinati errori di stima dei parametri associati alla distribuzione scelta a rappresentare la regione omogenea. Non sempre infatti la posizione simile sul territorio, può considerarsi un criterio sufficiente a garantire l'omogeneità dei dati in un sito privo di informazioni.

1.2.2 *La Cluster Analysis*

Tale procedura si basa sul principio di delimitazione delle regioni omogenee su base non geografica: il risultato di un'analisi di questo tipo produce, delle regioni fisse, valutando determinati caratteri tra differenti stazioni strumentate.

Il metodo prevede la definizione di un criterio di classificazione delle serie storiche appartenenti alle stazioni. Deve essere identificativo e pertinente ai valori che si andranno ad analizzare, garantendo inoltre, una classificazione efficace. La scelta dei parametri di classificazione è soggettiva, poiché ogni serie potrebbe presentare caratteri unici, non riscontrabili in altre sequenze di valori. In seguito all'identificazione del criterio di classificazione, sarà necessario eseguire una normalizzazione delle serie, in modo da rendere i dati, tra loro uniformi e analizzarli sulla base di una scala adeguata.

Il criterio di classificazione maggiormente utilizzato si basa sul concetto della distanza Euclidea, espressa dalla formula (1).

$$dist_{ij} = \sqrt{\sum_{k=1}^{n} (V_{k,i} - V_{k,j})^2} \qquad 1$$

Dove

- $V_{k,i}$ indica il valore della variabile di classificazione k per la i-esima stazione
- $V_{k,j}$ indica il valore della variabile di classificazione k per la j-esima stazione.

In base ai risultati ottenuti, verrà costruita una matrice delle distanze: in questo modo sarà possibile effettuare i differenti gruppi in modo da definire le diverse regioni.

La tecnica alla base della cluster analysis prevede due grandi categorie di selezione delle stazioni:

1) le tecniche gerarchiche prevedono un'andamento decrescente dell'analisi. Analizzano prima un cluster, dal criterio di classificazione più elevato, fino ad arrivare ad uno più ristretto e garantire risultati più precisi.

2) le tecniche non gerarchiche prevedono un'ipotesi iniziale, come il numero di sottogruppi per la classificazione. In base a questa informazione, l'analisi procede in maniera coerente.

Solitamente, nelle analisi regionali, si predilige l'utilizzo di un algoritmo agglomerativo, appartenente alla classe delle analisi gerarchiche: tale sistema prevede l'associazione delle stazioni, in funzione della loro distanza da un punto centrale, identificato come il centro del cluster. Minore sarà la distanza dal centro, maggiore sarà la probabilità che quella determinata stazione verrà associata al cluster in esame.

Ogni qualvolta l'algoritmo unisce stazioni differenti, sarà necessario ricalcolare la stazione centrale del cluster e il suo baricentro: dal momento che non è sempre possibile rivalutare per ogni stazione aggiuntiva, l'intero set di caratteri del cluster, si procederà semplicemente verificando l'omogeneità dell'intera regione. Non essendo noto a priori il numero esatto di regioni in cui si vuole suddividere l'intera area, è necessario studiare i valori delle serie storiche e osservare in quanti macro gruppi è possibile suddividere il set di dati.

Questa tecnica produrrà come risultato la rappresentazione di una serie di punti rappresentativi delle differenti stazioni in un piano cartesiano X,Y, nel quale i due assi identificheranno, rispettivamente, le coordinate geografiche di ogni stazione. Ogni gruppo di stazioni identificato, sugli assi cartesiani, presenta una curva di crescita e un'uguale distribuzione di frequenza, per ogni punto dell'area. In tal modo scegliere una distribuzione di probabilità, per un set di dati numeroso, risulterà una pratica molto più semplice e veloce, rispetto all'identificazione di una distribuzione di probabilità differente per ogni stazione.

La [Figura 1. 2], rappresenta un'applicazione della 'cluster analysis' in un sito di interesse: si osserva come il punto nero identifichi il sito di interesse nel quale si desidera eseguire l'analisi. La presenza di stazioni, poste ai confini di una regione, implica, a livello dell'analisi, determinate incongruenze di stima legate alla possibile variazione del criterio di scelta. Infatti, al variare del criterio di classificazione, il sito selezionato, potrebbe essere associato ad un'altra regione, con una differente curva di crescita. Al variare della regione di appartenenza, si osserverà una differente curva di crescita e di conseguenza una diversa distribuzione di probabilità associata.

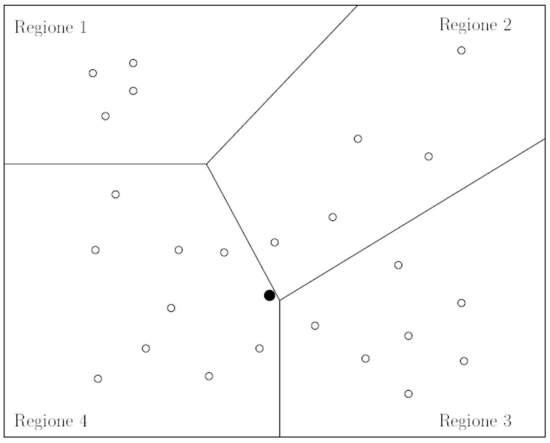


Figura 1. 2: Rappresentazione del risultato di una cluster analysis in un sito di interesse (Viglione, 2007)

1.2.3 *La Region Of Influence*

Un'ulteriore tecnica di delimitazione delle regioni omogenee è la 'Region Of Influence', per abbreviazione definita ROI. È una procedura in grado di fornire delle regioni, i cui confini non sono fissi a priori: i gruppi di stazioni realizzati con questo metodo, si trovano nell'intorno del sito di interesse dell'analisi, rigorosamente privo di dati. Il metodo ROI permette di eliminare le incongruenze che si sviluppano durante la tecnica della cluster analysis.

La tecnica di realizzazione di una ROI, si basa sul criterio di similarità, che vede come chiave la distanza Euclidea, descritta dalla formula (1). Le stazioni appartenenti ad una data 'region of influence', verranno classificate in base alla loro vicinanza con il punto di interesse. Per valutare l'ampiezza di una ROI, è necessario considerare eventuali parametri, come, ad esempio, i cambiamenti climatici o di quota, in modo da rendere la regione il più omogenea possibile.

Successivamente sarà necessario verificare che non siano presenti, all'interno delle serie storiche, dei valori 'outlines', ovvero non conformi con l'andamento della serie. Per ogni ROI, l'utente dovrà costruire, in funzione delle serie di dati, una curva di crescita differente, alla quale verrà associata una determinata distribuzione di probabilità.

Il principio potrebbe sembrare molto simile alla tecnica della cluster analysis. La caratteristica è l'assenza di due ROI uguali: al variare del sito di interesse, si modificheranno i set di dati da analizzare, e di conseguenza varieranno le curve di

crescita. Lo scopo della curva di crescita, è la valutazione dell'uniformità dei dati raggruppati all'interno della regione.

La rappresentazione di una regione omogenea, mediante il metodo 'Region Of Influence', permette di avere per ogni sito di interesse non strumentato, un numero di osservazioni maggiore e conforme alle stazioni presenti nei dintorni. In tal modo si costruirà un insieme di dati osservati, grazie al quale, sarà possibile effettuare la stima della grandezza di progetto in quel luogo. Questa tecnica permette di costruire delle regioni, in cui il sito d'interesse, risulterà sempre posizionato al centro. Per ogni sito non strumentato, si otterranno curve di crescita diverse l'una dall'altra.

Dal punto di vista pratico, risulta di difficile applicazione, poiché per ogni sito non strumentato, si dovrà ricostruire l'intorno desiderato e analizzare correttamente la distribuzione di frequenza associata. In termini di precisione e stabilità del risultato, permette di ottenere stime di valori di progetto in località prive di osservazioni, in modo coerente con le stazioni adiacenti.

La [Figura 1. 3] presenta un esempio di analisi con il metodo delle ROI, in cui il raggio intorno al sito di interesse, è fissato a priori. Si osserva come il punto centrale da cui far partire l'analisi sia esattamente il sito in esame: la scelta di un altro luogo, varierà l'intera distribuzione e di conseguenza anche la disposizione della ROI nello spazio in esame. Il risultato di un'analisi regionale con il metodo ROI, fornisce un'unica regione identificativa del sito di interesse.

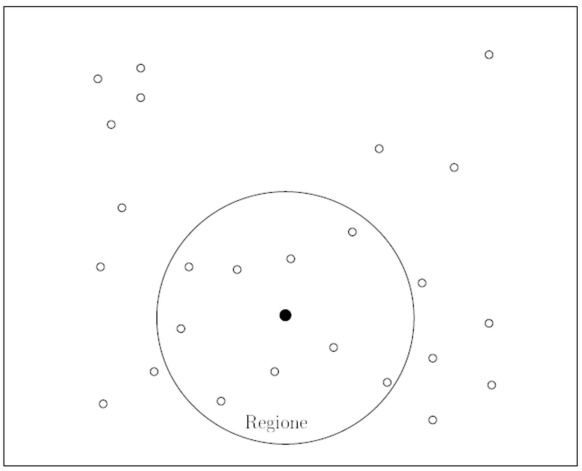


Figura 1. 3: Rappresentazione dell'applicazione della region of influence in un sito di interesse (Viglione, 2007)

1.3 *Verifica dell'omogeneità di una regione*

In un'analisi regionale, è necessario verificare la validità e l'omogeneità dei dati selezionati all'interno di una data regione. Nel corso degli anni sono stati sviluppati numerosi test di verifica dell'omogeneità, con risultati piuttosto soddisfacenti per le analisi regionali. Il principio di verifica dell'omogeneità di una regione vede le basi nel metodo di Hosking e Wallis, sviluppato nel 1997.

Tale sistema consente di stimare l'omogeneità della regione in esame, in base al parametro dell'eterogeneità della stessa, misurando, dunque, quanto i parametri in essa contenuti siano differenti. È una procedura che si basa sul confronto tra un campione del gruppo in esame e un campione di riferimento appartenente ad una regione sicuramente omogenea. Il parametro di riferimento utilizzato è la variabilità campionaria calcolata con il metodo degli L-momenti. Tale valore verrà confrontato con la variabilità della serie in esame.

Lo scopo principale del test è valutare un coefficiente identificativo della misura dell'eterogeneità della serie, come indicato nella formula (2):

$$\theta_{H\&W} = \frac{V - \mu_V}{\sigma_V} \qquad 2$$

I parametri presenti nell'equazione (2) sono i seguenti:

1) V rappresenta la varianza campionaria, calcolata con il metodo degli L-momenti;
2) μ_V indica la media delle serie in esame;
3) σ_V è la varianza delle serie in esame.

Il modo migliore per analizzare il parametro $\theta_{H\&W}$ è eseguire la rappresentazione grafica, come di una distribuzione normale. La verifica consiste nell'osservare la posizione del valore della varianza calcolata: nel caso quest'ultima ricada vicino alla media della distribuzione, la regione risulterà omogenea; nel caso contrario invece la regione verrà assunta come eterogenea, e sarà necessario ridefinire nuovamente la procedura di scelta delle stazioni.

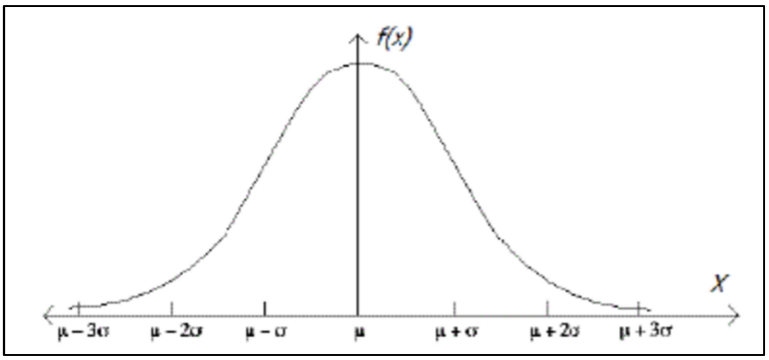


Figura 1. 4: rappresentazione di una generica distribuzione normale simmetrica rispetto al valore medio, (http://www.biostatistica.unich.it/mat_didattica/MENNA/A_Menna_UniCH_Lez2_2008_DistribuzioneNormale.pdf)

La [Figura 1. 4] evidenzia il concetto dell'omogeneità della regione: i tre casi riportati sotto, possono essere osservati nel grafico di una distribuzione normale, considerando gli intervalli simmetrici, rispetto al valore medio, posto nel centro. Il coefficiente $\theta_{H\&W}$, utilizzato per eseguire il test di Hosking e Wallis, potrà assumere quindi i seguenti valori:

- Nel caso in cui $\theta_{H\&W} < \mu + \sigma$ allora la regione sarà omogenea;
- Nel caso in cui $\mu + \sigma < \theta_{H\&W} < \mu + 2\sigma$ la regione potrebbe essere eterogenea, in quanto la variabilità osservata è inferiore a quella attesa;
- Nel caso in cui, invece, $\theta_{H\&W} > \mu + 2\sigma$ la regione sarà sicuramente eterogenea.

Il test di Hosking e Wallis, non è un test infallibile, anzi spesso per determinate ipotesi potrebbe non essere completamente attendibile. Sono riassunti alcuni casi per cui, il metodo non può essere utilizzato per la verifica:

1) Non è stato raggiunto un adeguato valore di $\theta_{H\&W}$: quindi la regione viene considerata eterogenea;
2) La distribuzione di probabilità scelta per eseguire l'analisi, non è adeguata.

In certi casi, prima di escludere l'ipotesi di verifica con il metodo di Hosking e Wallis, si può procedere modificando il parametro di riferimento, utilizzando il coefficiente di Skewness. Così facendo è possibile che il test assuma un risultato più preciso e utilizzabile per proseguire l'analisi.

1.4 *La curva di crescita*

L'analisi regionale svolta con il metodo della grandezza indice produrrà una serie di curve di crescita, una per ogni durata, corrispondenti al comportamento delle precipitazioni nel sito scelto. Tale curva, infatti, mostra all'utente l'andamento dei valori normalizzati nel corso del tempo, in funzione della variabile ridotta. La relazione che fornisce questo tipo di curva lega la probabilità di verificarsi dell'evento estremo, con il fattore di crescita individuato dalla precipitazione adimensionalizzata.

In ogni area omogenea, è importante che la curva di crescita sia uguale in ogni punto della regione e sia idonea all'insieme di valori analizzato al suo interno. In tal modo sarà possibile attribuire ad un unico set di dati, la distribuzione di probabilità più adeguata.

Come si osserva nel rapporto di De Michele e Rosso, del 2001, ogni area può corrispondere ad una differente curva di crescita: le differenze presentate nella [Figura 1. 5], sono legate alla posizione sul territorio. Per tale motivo è importante definire se il sito in esame si trova in un ambiente montano, quindi soggetto a maggiori precipitazioni nevose e venti forti, piuttosto che in un ambiente dove le precipitazioni sono di diversa entità, quindi consentono analisi più approfondite.
La [Figura 1. 5] illustra le curve di crescita delle diverse zone dell'Italia Occidentale, realizzate con la procedura VAPI, illustrata nel De Michele e Rosso, del 2001. Le differenti zone sono state analizzate e realizzate in corrispondenza della tipologia di dati osservati, per realizzare delle regioni omogenee idonee.

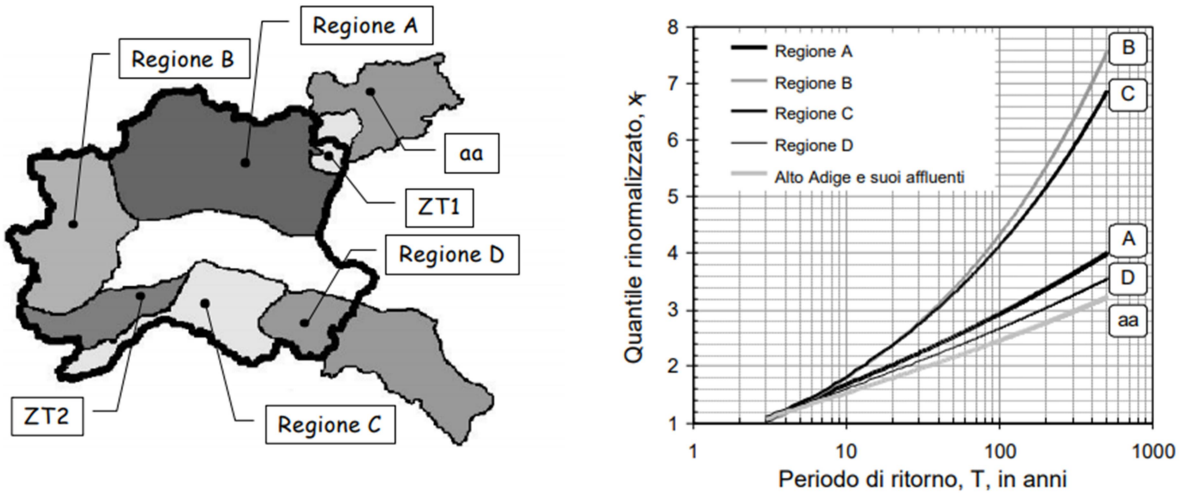


Figura 1. 5: rappresentazione delle differenti aree dell'Italia Occidentale, suddivise in regioni omogenee in funzione della tipologia di dati osservati. A destra si osservano le curve di crescita per ogni regione omogenea.

1.5 *Le curve di possibilità pluviometrica*

Le curve di possibilità pluviometrica sono il risultato finale di un'analisi regionale, e permettono di ottenere il valore della grandezza di progetto del sito scelto. Possono essere rappresentate in funzione del tempo di ritorno, per garantire all'utente una più veloce stima del valore da ottenere. I tempi di ritorno principali a cui è possibile voler stimare il valore della precipitazione di progetto, sono riassunti nella [Tabella 1. 1].

Tabella 1. 1: tempi di ritorno utilizzati per la stima delle grandezze di progetto.

Tempo di ritorno (anni)	Opera da realizzare
50	Ambiente urbano
200	Opere di protezione
1000	Dighe

A seconda della tipologia di opera per la quale è necessario conoscere la grandezza di progetto, si osservano differenti periodi di ritorno:

1) Per le dighe, è necessario stimare dei valori di progetto con tempi di ritorno molto elevati, che abbiano dunque frequenze di accadimento molto basse, in modo da garantirgli sicurezza per una vita utile dell'opera pari a 1000 anni.

2) Per quanto riguarda le opere di protezione, come le casse di laminazione, è necessario garantire una vita utile pari ad almeno 200 anni, che comporta delle probabilità di accadimento modeste;
3) In ambiente urbano, è possibile realizzare degli argini o delle rettifiche fluviali, che proteggano l'ambiente dagli eventi più pericolosi, ma la frequenza di accadimento in questo caso sarà più elevata.

Le curve di possibilità pluviometrica, definite anche come 'CPP', sono caratterizzate da due possibili tipologie di rappresentazione differente:

1) Nel caso in cui, si raffigurasse l'altezza di pioggia in funzione del tempo di ritorno, allora si osserveranno differenti curve, ognuna in funzione di una durata differente, come in [Figura 1. 6];
2) Nel caso in cui si costruisca un grafico in funzione della durata di un evento, si osserveranno differenti curve, una per ogni tempo di ritorno considerato come caratteristico, come in [Figura 1. 7].

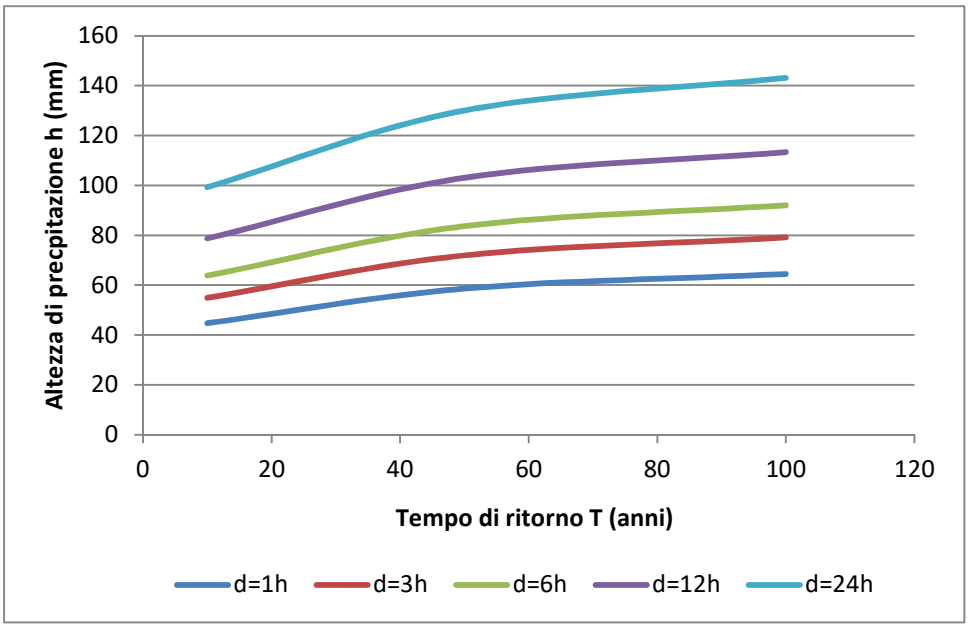


Figura 1. 6: curva di possibilità pluviometrica in funzione delle differenti durate.

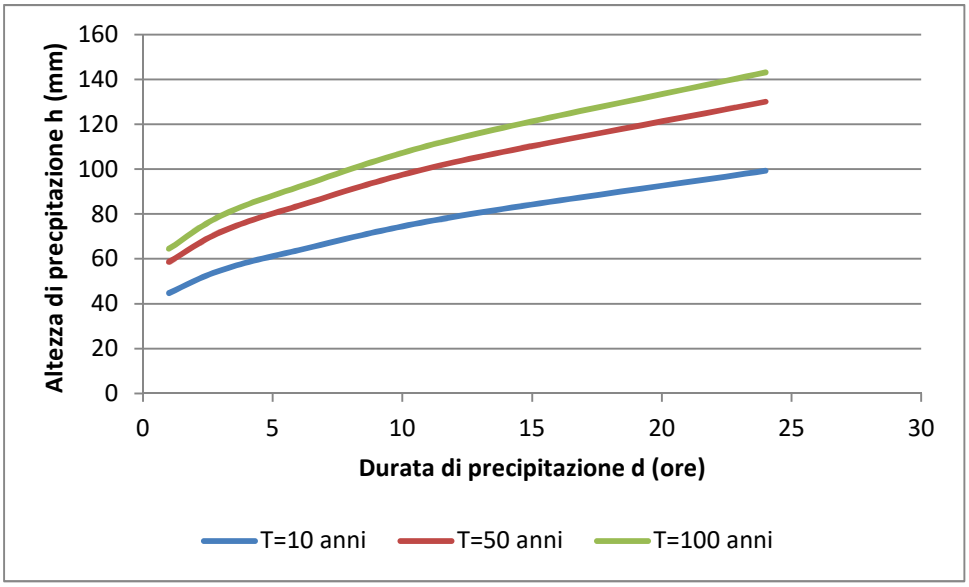


Figura 1. 7: curva di possibilità pluviometrica, una per ogni tempo di ritorno

Capitolo 2

Analisi regionale in Inghilterra

La tecnica di analisi regionale delle precipitazioni applicata in Inghilterra alla fine degli anni ’90, ha seguito un percorso nettamente diverso rispetto a quelli utilizzati in Italia. Si osservano, infatti, numerose differenze, prima fra tutte la quantità delle informazioni presenti. La nazione inglese appare colma di stazioni pluviometriche, in grado di fornire osservazioni in qualunque parte del territorio: a livello climatico, l’Inghilterra, poiché circondata dall’Oceano Atlantico, risulta più esposta ad eventi atmosferici frequenti e di entità differente da quelle osservate in Italia.

Per tale motivo, gli inglesi hanno definito una procedura automatizzata, di analisi regionale, basata sul metodo indice, che garantisce la stima della pioggia di progetto in ogni punto del loro territorio. Il ‘Flood Estimation Handbook’ di Faulker, (1999), presenta diverse procedure di stima della frequenza di un evento, sia in caso di piena che, nel volume due, nel caso di eventi piovosi.

Nel corso degli anni, si sono susseguite differenti tecniche di analisi, che hanno permesso di generare la procedura attuale, in grado di stimare quantili di progetto per tempi di ritorno molto elevati.

Il primo metodo che è stato sviluppato in Inghilterra per l'analisi regionale è il 'FSR', il cui acronimo indica 'Flood Studies Report', sviluppato dal NERC nel 1975. Questo permetteva una stima della precipitazione di progetto per tempi di ritorno superiori a 1000 anni. La differenza sostanziale si osserva nella variabile indice: tale valore era identificato come la precipitazione associata ad un tempo di ritorno pari a 5 anni. La suddivisione del territorio in due grandi regioni fisse, garantiva la costruzione di una curva di crescita, dipendente dall'intensità dell'evento, scelto come grandezza indice.

Successivamente, per aumentare la precisione di stima del modello e allungare l'analisi a tempi di ritorno più elevati, nel 1989, Reed e Stewart svilupparono una nuova tecnica, definita 'FORGE' ('FOcused Rainfall Growth Estimation'). Tale procedura venne applicata successivamente anche in altre parti del mondo, come in Australia, producendo curve di crescita con tempi di ritorno fino a 2000 anni.

Il metodo 'FORGE', molto simile al metodo attuale applicato in Inghilterra, prevedeva, come variabile indice, una precipitazione con tempo di ritorno pari a 50 anni. La scelta di aumentare il valore di tale grandezza è legata alla volontà di ricavare i quantili per periodi di ritorno più elevati. La modalità di costruzione della curva di crescita venne migliorata, garantendo una costruzione progressiva del set di dati da analizzare.

In Australia, la procedura di costruzione della curva di crescita, prevedeva un'espansione graduale del cerchio, concentrico al sito desiderato: la dimensione dell'area era influenzata dal numero delle stazioni da considerare, che tendeva a raddoppiare per ogni cerchio consecutivo. La serie di dati così ottenuta veniva rappresentata con un approccio 'station year', in grado di evidenziare un decrescere della dipendenza spaziale tra le stazioni, all'aumentare del periodo di ritorno.

L'attuale procedura di analisi regionale permette di ricavare valori di progetto con tempi di ritorno compresi tra 150 anni e 10000 anni: il software FORGEX, 'FOcused Rainfall Growth curves EXtension', è in grado di estendere la costruzione della curva di crescita oltre i 1000 anni. Questa procedura è stata utilizzata anche in altri paesi del mondo, come la Corea, fornendo risultati molto soddisfacenti e coerenti con altre metodologie di analisi di frequenza. In Italia, i metodi regionali prevedono una verifica dell'omogeneità in modo da garantire uniformità alla serie analizzata: tale verifica viene eseguita a seguito della selezione dei dati appartenenti ad ogni regione costruita.

In Inghilterra, la verifica dell'omogeneità di una regione è affidata alla distanza di ogni stazione dal sito in esame.

L'analisi regionale applicata in Inghilterra attualmente prevede la seguente serie di passaggi, per la stima della precipitazione di progetto:

1) Selezione del sito di interesse;
2) Scelta della grandezza indice;
3) Costruzione della curva di crescita;
4) Stima del quantile di progetto.

2.1 *Selezione del sito di interesse*

L'Inghilterra dispone di un territorio molto vasto, sul quale sono disposte in modo molto meticoloso, una serie di stazioni pluviometriche, in grado di fornire le osservazioni necessarie a garantire un'analisi regionale molto precisa. Il numero delle stazioni strumentate, osservate sul territorio inglese, risulta molto elevato e la loro concentrazione appare molto fitta, come si osserva in [Figura 2. 1]. La peculiarità del metodo inglese, fonda le basi sulla numerosità delle stazioni strumentate, ma anche sulla lunghezza delle serie associate ad ogni stazione. Infatti, basandosi completamente sulla numerosità, è necessario possedere per ogni stazione, set di dati di almeno 10 anni di osservazioni: in tal modo, si garantirà continuità e affidabilità all'analisi.

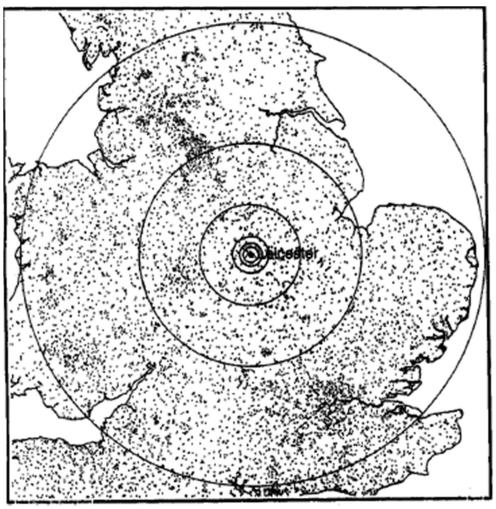


Figura 2. 1: rappresentazione delle stazioni nei dintorni di un sito scelto per l'analisi.

2.2 *La variabile indice*

Il valore indice dell'analisi regionale inglese, è il valore mediano. La scelta di tale elemento è legata alle caratteristiche statistiche del parametro: infatti, questo valore è meno influenzato dagli estremi della serie di dati rispetto al valore medio. Inoltre, sulla base dei metodi precedenti, l'evento con tempo di ritorno pari a due anni, garantisce all'analisi e alla serie di dati normalizzata, maggiore stabilità.

La numerosità degli eventi piovosi non è l'unica prerogativa del metodo: si osserva sul territorio inglese una fitta rete di stazioni, che offre la possibilità di costruire una serie di mappe identificative del valore mediano in ogni sito, per ogni durata caratteristica, come in [Figura 2. 2].

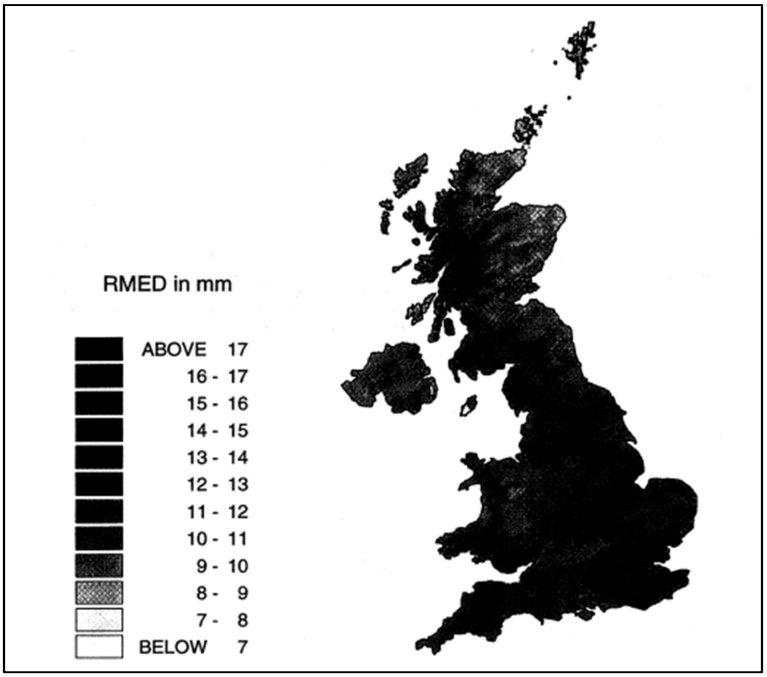


Figura 2. 2: mappa dell'Inghilterra, rappresentativa del parametro indice (RMED), per la durata di 1 giorno, (Faulkner & Prudhomme, 1998)

Il valore indice di un sito, è calcolato con una regressione statistica su base geografica: i valori mediani saranno stimati in funzione di alcuni parametri morfologici, come la quota e la distanza dal livello del mare. La mappa in [Figura 2. 2], presenta una regionalizzazione del valore mediano sul territorio inglese: in funzione dei parametri di sito, sarà possibile ricavare il valore della mediana nella località scelta per l'analisi.

I modelli di regressione applicati in Inghilterra, per la stima del valore indice su scala nazionale, prevedono due formule differenti a seconda della durata scelta. Si può calcolare il valore mediano, per durate comprese tra 1 e 8 giorni, oppure per durate inferiori, comprese tra 1 ora e 12 ore. Le variabili topografiche, utilizzate per ricavare i valori mediani, sono ottenute attraverso delle regressioni lineari su base statistica.

2.3 *La costruzione della curva di crescita*

Il metodo 'FORGEX', è il fulcro dell'analisi regionale per le piogge estreme in Inghilterra. La curva di crescita ottenuta con questo metodo, tende a fornire la stima dei valori fino a 10000 anni: non solo estende il limite superiore per quel che riguarda i valori del tempo di ritorno, ma permette di eseguire una stima degli eventi con periodi di ritorno a partire da 150 anni. Il software, in grado di fornire curve di crescita per qualunque set di dati fornitogli, è un applicativo a pagamento, che viene utilizzato dai principali enti inglesi.

Il metodo costruisce le curve di crescita in base al principio del 'pooling group' ovvero raggruppando i dati, in funzione di un carattere similare. Nel caso del FORGEX, il criterio di similarità è la distanza dal sito scelto. Come viene citato nell'articolo di Svensson e Jones, (2010), il FORGEX '*produce una curva di crescita su base empirica, combinando l'approccio di analisi di frequenza regionale, con l'analisi dei valori massimi del network'*. Questo approccio, farà corrispondere ad ogni tratto della curva di crescita, un dato network, cioè un insieme di stazioni nell'intorno del punto considerato, come è mostrato nella [Figura 2. 3].

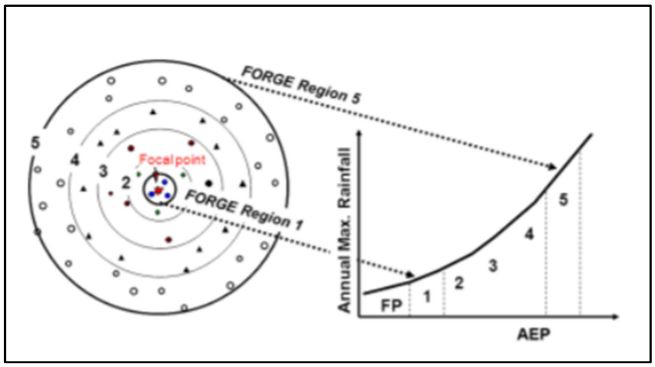


Figura 2. 3: illustrazione della procedura di selezione del metodo FORGEX, anche utilizzata nel metodo 'FORGE', per la costruzione della curva di crescita, (Svensson & Jones , 2010)

In seguito alla stima della curva di crescita, in Inghilterra è possibile costruire delle ulteriori mappe del territorio rappresentanti i risultati ottenuti in ogni luogo, come illustra la [Figura 2. 4]. In tal modo la caratterizzazione dei differenti fattori di crescita sarà più semplice da reperire.

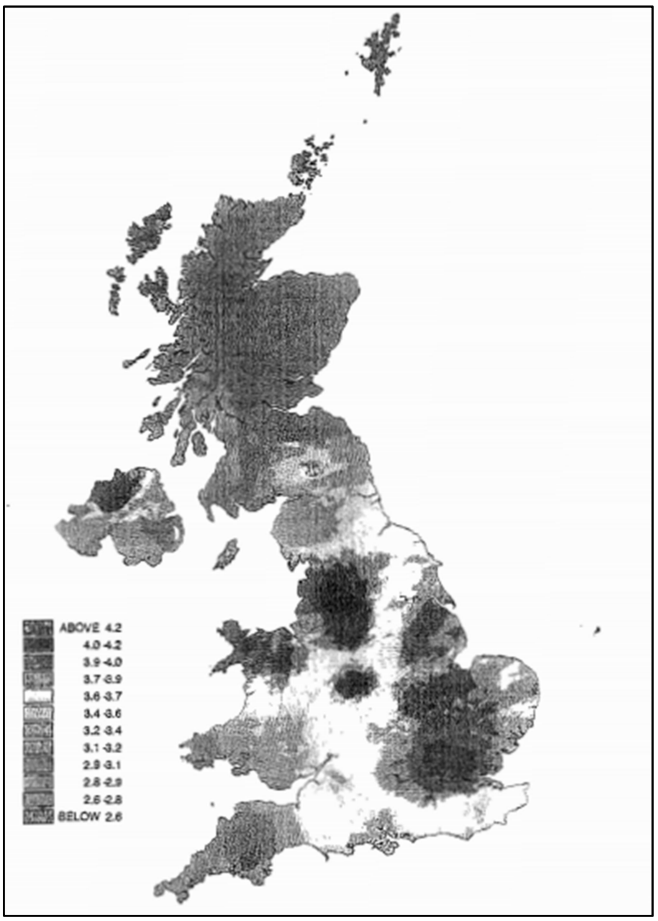


Figura 2. 4: rappresentazione del fattore di crescita per T=100 anni, su tutto il territorio inglese (Faulker, 1999)

2.4 *Stima del valore di progetto*

L'obiettivo di un'analisi regionale è la stima della pioggia di progetto in una località, in funzione delle osservazioni nell'intorno del sito stesso. A tal punto, osservando i risultati ottenuti, un'analisi regionale di precipitazione in Inghilterra appare molto più precisa e meticolosa, rispetto ad una applicata in Italia. Tale aspetto è legato alla realizzazione di diverse mappe territoriali, in grado di evidenziare con esattezza i valori indice e le curve di crescita luogo per luogo. Il prodotto tra queste due mappe,

consente di ottenere la stima del valore di progetto in ogni sito analizzato, in funzione di una determinata durata.

Ogni durata di precipitazione analizzata verrà trattata in maniera differente in modo da poter garantire un'analisi precisa. L'obiettivo finale dell'analisi regionale inglese è la costruzione di mappe in grado di rappresentare, per determinate durate, gli eventi di precipitazione e la loro frequenza. Ognuna di queste mappe racchiude le informazioni derivanti dal sito strumentato e, grazie alla tecnica implementata con il 'FORGEX', anche quelle provenienti da zone prive di dati. Un esempio delle mappe che si ottengono come risultato di un'analisi regionale delle precipitazioni in Inghilterra, è rappresentato dalla [Figura 2. 5].

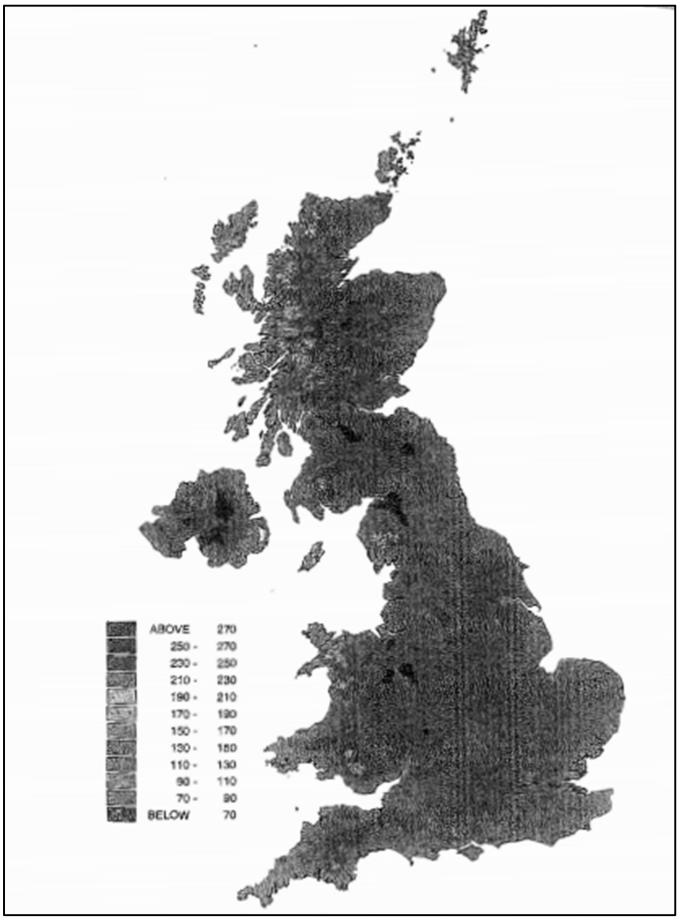


Figura 2. 5: rappresentazione della pioggia di progetto per durata 1 ora e T=100 anni, (Faulker, 1999)

A differenza della procedura di stima delle precipitazioni, in Inghilterra la stima delle portate di piena risulta molto più complicata e rudimentale rispetto a quella applicata in Italia, dove il valore della portata di piena di progetto viene stimato accuratamente.

Capitolo 3

Caratterizzazione geo-spaziale per la procedura FORGEX

Lo scopo pratico di questa tesi è cercare di riprodurre il metodo FORGEX per l'analisi regionale, applicandolo sul territorio italiano, prima fra tutte nella regione Piemonte. Per garantire un risultato ottimale e applicabile in altre regioni e/o paesi europei, si è deciso di utilizzare il software open source Python, come supporto per lo sviluppo dell'applicativo.

La scelta di utilizzare questo programma è legata all'utenza a cui si pensa di poter affiancare questo dispositivo. Nel corso degli ultimi anni, infatti, si è investito molto in programmi come Python, in quanto permettono di costruire applicativi molto complessi, sulla base di alcune semplici nozioni di informatica. Inoltre, nel mondo dell'idrologia, si è sviluppato molto come software di programmazione, a sostituzione di programmi a pagamento e più legati al calcolo matriciale, come per esempio Matlab.

3.1 *Dati preliminari*

Per il corretto svolgimento di un'analisi regionale, è necessario essere a conoscenza dei massimi annui di precipitazione e degli identificativi di tutte le stazioni adiacenti al sito scelto per lo studio. I dati pluviometrici sono organizzati negli annali, in base alla stazione di appartenenza, e in funzione delle durate caratteristiche di osservazione di un dato evento (1,3,6,12,24 ore), come si osserva nell'estratto in [Tabella 3. 1]. Le osservazioni sono ordinate in maniera crescente, in funzione dell'anno di accadimento dell'evento.

Tabella 3. 1: estratto della tabella riassuntiva dei massimi annui di precipitazione per le diverse stazioni.

Numero	Anno	Max 1h	Max3h	Max6h	Max12h	Max24h
12001	1994	18.7	31.3	35.5	41.2	43.7
12001	1995	19.2	21.9	24.8	35.4	36.8
12001	1996	27.8	30.4	30.4	48.9	81.2
12001	1997	27.5	38.6	45.4	46.4	46.4
12001	1998	33.5	64.6	65	65	74.2
12001	1999	18.4	21.5	26.5	30.7	35.4
12001	2000	23.2	31.4	33.2	35.4	45.7
12001	2001	14.5	21.3	35.7	43.3	46.3
12001	2002	41.3	41.5	42.9	49.1	82.5
12001	2003	18.4	25.8	43.3	47.1	50.5
12001	2004	18.57	22.87	27.71	33.47	48.11
12001	2005	21.71	35.96	45.38	56.21	65.75
12001	2006	25.59	28.8	42.71	63.8	67.93
12001	2007	22.93	36.18	50.89	52.29	52.29
12001	2008	13.51	13.51	19.84	32.27	34.88
12001	2009	13.45	23.06	32.08	37.11	53.75
12001	2010	19.9	26.02	33.59	41.41	58.57
12002	1987	32.2	66.2	99.2	111.8	135.2
12002	1989	9.4	19	22.8	35	40
12002	1992	22	34.4	40.8	40.8	41.6
12002	1994	9.2	14	23	41	44.8
12003	1989	30.53	30.53	30.53	33.5	34.48
12003	1990	58.4	63.2	63.2	63.2	64.2
12003	1991	15.4	27.3	35.9	40.5	53.9
12003	1992	50.1	55.5	56.5	56.5	58.3
12003	1993	24.9	29.1	45.1	63.4	99.1
12003	1994	35.1	35.4	40.2	51.1	55.3
12003	1995	39.8	50.7	50.7	52.8	53.8

Ogni pluviometro è contenuto all'interno di una stazione meteorologica, che può essere collocata in una qualsiasi posizione sul territorio, in modo da essere esposta ai fenomeni atmosferici. Una stazione meteorologica può trasmettere i dati raccolti in quel sito, in due modi diversi:

1) Manualmente: i dati sono spediti all'ente di riferimento attraverso un responsabile della stazione, con una frequenza ben determinata;
2) Automaticamente: i dati sono spediti all'ente di riferimento in modo automatizzato, senza l'ausilio di personale autorizzato.

I massimi annui di precipitazione osservati nel corso degli anni, vengono raggruppati in base alla durata dell'evento e in base alla stazione di raccolta: possono presentarsi, per ovvi motivi, delle serie di dati non sufficientemente lunghe da poter essere utilizzate per l'analisi. Questo potrebbe essere correlato con la presenza di stazioni recentemente installate, non in grado di fornire serie di dati robuste.

Ogni stazione ha una serie di caratteristiche identificative, come le coordinate o la quota. È molto importante definire esattamente la posizione della stazione sul territorio per garantire una corretta analisi dei dati: infatti le precipitazioni sono molto influenzate dalla quota alla quale si trova la stazione.

Questo è legato, in parte, alla topografia del sito in esame e, in parte, alla presenza ad alte quote di precipitazioni solide e venti che possono influenzare l'evento o danneggiare la stazione, rendendo la lettura poco attendibile. Questo fenomeno è maggiormente osservabile in casi di precipitazioni di lunga durata: per durate brevi, solitamente questi effetti risultano meno apparenti.

La [Tabella 3. 2] mostra un estratto dell'insieme di stazioni presenti sul territorio piemontese. Le stazioni sono ordinate in maniera crescente, in funzione del numero della stazione.

Tabella 3. 2: estratto della tabella riassuntiva delle informazioni relative alle stazioni presenti sul territorio piemontese.

Numero Stazione	Nome Stazione	UTM x	UTM y	quota
12001	Isola_S_Antonio	488564	4985460	77
12002	Sale	484376	4981210	83
12003	Alessandria_Lobbi	476727	4976201	90
12004	Alessandria_Aeroporto	467261	4973870	95
12005	Masio_Tanaro	453525	4969048	105
12006	Sezzadio	465851	4959096	110
12007	Casale_Monferrato_(Is_Piopp_)	455589	4998010	113
12008	Asti_Tanaro	437873	4970662	117
12009	Casale_Monferrato	461092	4997958	118
12010	Tortona	488787	4971060	120
12011	Cerano	484546	5029197	125
12012	Sezzadio_Bormida	465865	4959060	127
12013	Basaluzzo	474833	4956759	128
12014	Vercelli	454936	5019310	130
12015	Vercelli	452237	5019386	132
12016	Vercelli_Ist_Ris_	450518	5020260	135
12017	Nizza_Monferrato	446070	4957350	137
12018	Nizza_Monferrato	448319	4957203	138
12019	Tricerro_(2)	447480	5009461	139
12020	Castellalfero	438822	4982122	140
12021	Tricerro	447817	5009170	140
12022	Govone	429922	4959859	145

Per ogni stazione, è importante conoscere le seguenti caratteristiche:

1) Codice di identificazione della stazione, che permette di associarla in maniera rapida alle osservazioni di precipitazione;
2) Nome della stazione, che richiama il nome della località in cui la stazione è collocata;
3) Coordinata UTM x, che ne identifica la distanza dal meridiano di riferimento;
4) Coordinata UTM y, che ne identifica la distanza dall'equatore;
5) Quota sul livello del mare, che rappresenta l'altitudine della stazione.

3.2 *Sistemi di rappresentazione territoriali*

3.2.1 *Coordinate Geografiche*

Per identificare una data posizione nello spazio, esistono due tipologie di coordinate diverse: cartografiche e geografiche. Notevolmente più semplici da ricavare, le coordinate geografiche sono espresse in gradi, rispetto al centro di massa, e si dividono in latitudine, longitudine e altitudine. Per rappresentare un punto sull'ellissoide, come mostra la [Figura 3. 1], è necessario conoscere la definizione per ognuna delle componenti:

1) La latitudine corrisponde alla distanza angolare di un punto dall'equatore;
2) La longitudine indica, invece, la distanza di un punto dal meridiano di Greenwich, identificato come meridiano zero;
3) L'altitudine, invece, rappresenta la distanza verticale di un oggetto sul livello del mare.

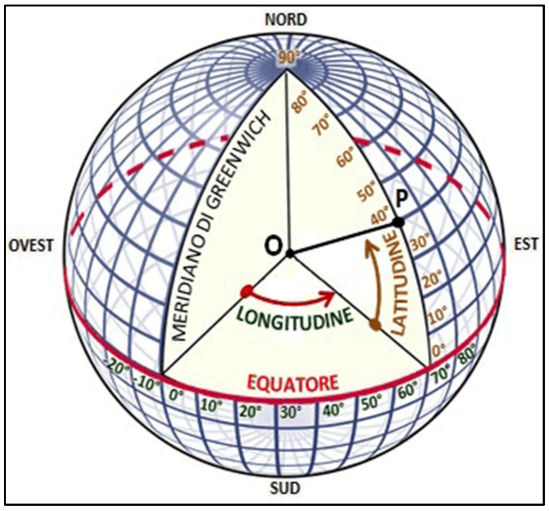


Figura 3. 1: rappresentazione delle coordinate geografiche (tratto da http://profpozzi-classe004.blogspot.com/2017/11/le-coordinate-geografiche.html)

Questo genere di coordinate non è utilizzabile per la rappresentazione di un insieme di punti in un piano. A tal proposito, si ricorre ad una serie di proiezioni cartografiche, che permettono di rappresentare la superficie terrestre su un piano, e di identificarne i punti richiesti. Trasformare le coordinate geografiche in un insieme di coordinate cartografiche, crea alcune distorsioni a livello della rappresentazione, che possono

essere risolte suddividendo la superficie terrestre in aree più piccole e di più semplice rappresentazione.

Ogni zona della superficie terrestre viene costruita grazie all'intersezione tra le linee congiungenti i due poli e le linee parallele all'equatore, che formano un reticolo, come si osserva in [Figura 3. 2]. Le linee verticali prendono il nome di 'meridiani', e quelle orizzontali sono definite 'paralleli'.

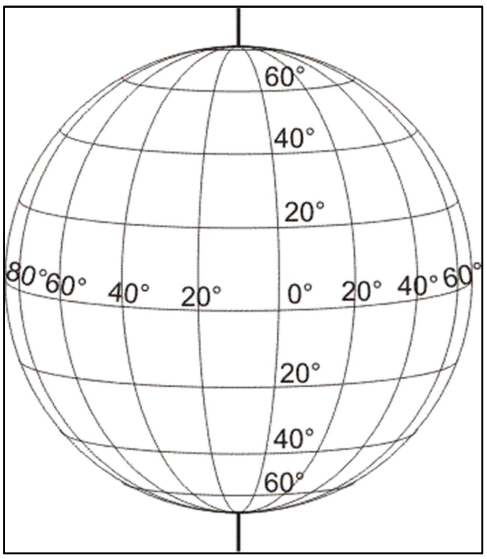


Figura 3. 2: rappresentazione del reticolo geografico, costruito dall'intersezione dei diversi meridiani e paralleli (tratto da http://profpozzi-classe004.blogspot.com/2017/11/le-coordinate-geografiche.html)

3.2.2 *Coordinate Cartografiche*

Le principali proiezioni cartografiche, utilizzate nell'ambito dei software territoriali, sono le proiezioni cilindriche. Si suddividono in due diverse tipologie, a seconda della posizione del cilindro rispetto alla superficie terrestre:

1) La proiezione di Gauss-Boaga, che pone il cilindro tangente all'ellissoide stesso;
2) La proiezione di Mercatore, che posiziona il cilindro tangente ad un meridiano.

3.2.2.1 *Proiezione di Gauss-Boaga*

La proiezione di Gauss-Boaga è utilizzata per la rappresentazione della cartografia ufficiale italiana. Questo tipo di cartografia, si basa sulla proiezione eseguita su due fusi particolari, il 32 e il 33, con un meridiano di riferimento modificato da un

fattore di contrazione. Questo tipo di rappresentazione presenta una falsa origine, per evitare che possano presentarsi alcuni punti con coordinate negative.

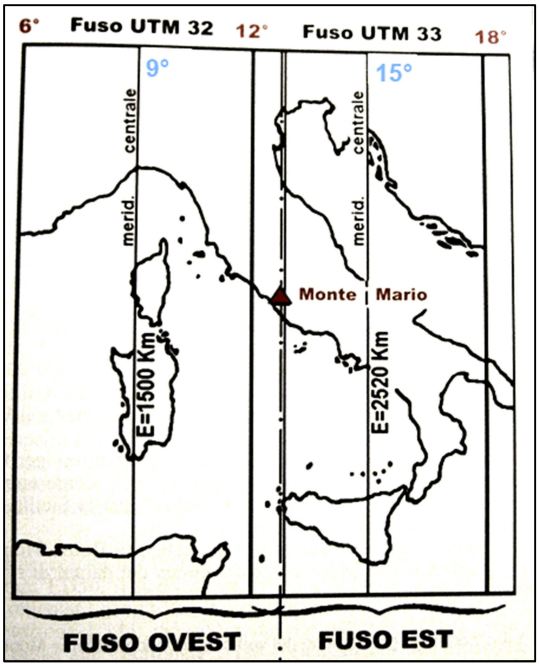


Figura 3. 3:rappresentazione della cartografia italiana, secondo la proiezione di Gauss-Boaga (tratto da https://geomappando.com/2016/07/07/sistemi-riferimento-coordinate-2-parte/)

3.2.2.2 *Proiezione di Mercatore*

La proiezione di Mercatore, definita anche come la proiezione Universale Trasversa di Mercatore, da cui U.T.M., esprime le coordinate di un punto in metri. Questo metodo di proiezione suddivide la superficie terrestre in 60 zone, ognuna delle quali sarà data dall'intersezione tra i 'fusi', le linee verticali, e le 'fasce', le linee orizzontali. L'Italia vede la sua collocazione nella zona 32esima e 33esima.

Il modo di rappresentazione di queste coordinate avviene attraverso una coppia di valori, associabili alle coordinate X e Y di un piano cartesiano:

1) Il primo elemento, rappresenta la distanza dal meridiano di riferimento. Si imposta un' origine fittizia, pari a 500000, per evitare di ottenere coordinate minori di zero, per i punti a sinistra dell'origine. Il valore delle coordinate sarà contato a partire dall'origine verso destra: i valori inferiori al meridiano di riferimento, corrisponderanno ai punti a sinistra di quest'ultimo.

2) Il secondo elemento, indica, il valore della distanza dall'equatore, identificato come il punto di origine. Le coordinate si inizieranno a contare dall'equatore verso il polo Nord.

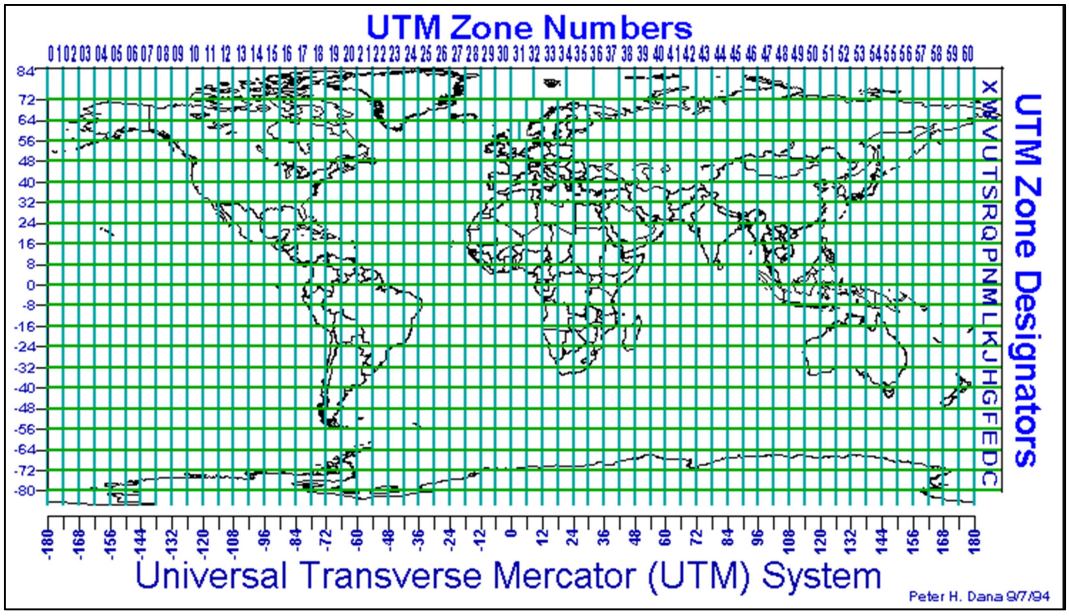


Figura 3. 4: rappresentazione in coordinate UTM, (tratto da http://geomatica.como.polimi.it/corsi/cartografialaurea/cartografiaufficialeitaliana/cartografiaufficialeitalianaigmi.htm)

Questo tipo di coordinate viene usato come sistema di riferimento per i software territoriali. Grazie alla facilità di rappresentazione delle coordinate, appare di semplice utilizzo, anche per calcolare le distanze tra due punti differenti.

3.3 *Identificazione del sito oggetto dell'analisi regionale*

In Italia i dati raccolti dalle stazioni pluviometriche, sono espressi in coordinate UTM, per uniformità di rappresentazione. La stima della curva di crescita, attraverso la metodologia FORGEX, prevede l'identificazione di un sito selezionato dall'utente, in cui si desidera eseguire le analisi. Il database utilizzato per verificare il funzionamento dell'applicativo, si limita alla regione Piemonte, in modo da permettere un confronto con alcuni risultati già ottenuti.

Le stazioni presenti sul territorio vengono rappresentate nella [Figura 3. 5]. In questo modo è possibile osservare, come la concentrazione dei siti strumentati sull'intero territorio piemontese non sia molto elevata.

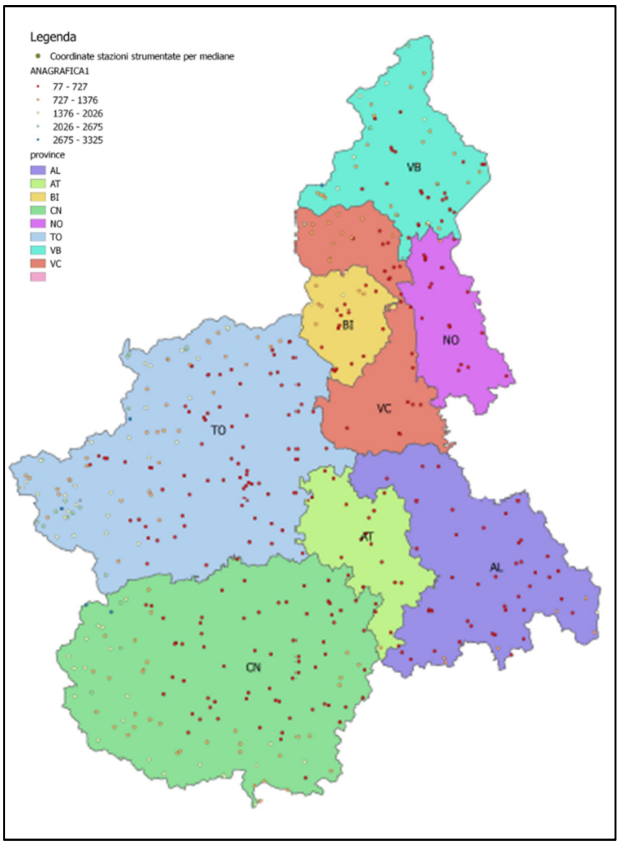


Figura 3. 5: rappresentazione delle stazioni presenti sul territorio piemontese, organizzate per quota.

Il Piemonte si trova tra il 32esimo e il 33esimo fuso: quindi le coordinate UTM dovranno essere limitate entro un dato intervallo di punti. È importante definire due limitazioni alle coordinate: in primo luogo sarà necessario garantire che appartengano al Piemonte e non ricadano in altre regioni. In secondo luogo, bisognerà evitare che le coordinate inserite si trovino nei pressi del bordo della regione.

Nel caso in cui il 'focal point' si trovasse nei limiti della regione Piemonte, l'analisi potrebbe comprendere troppe stazioni, e perdere la sua affidabilità. Infatti, nell'ipotesi di un'analisi ai valori estremi, si osserverebbe come dati posti a distanze molto elevate dal sito scelto, possano influenzare negativamente lo studio, contribuendo a diminuirne la sua attendibilità. È importante che il sito scelto per l'osservazione, garantisca una densità elevata di stazioni strumentate nei dintorni per evitare incongruenze di stima.

Per verificare la correttezza della procedura, ed eseguire un confronto immediato con altre tecniche, l'applicativo viene testato nelle stazioni strumentate la cui numerosità è molto elevata. Nel database della regione Piemonte, si è osservato tra le stazioni più numerose il sito di Oropa, con 63 dati. Il problema legato a tale località risulta la sua stretta vicinanza con la Valle d'Aosta, di cui non si conoscono dati per implementare l'analisi.

Le altre stazioni, che presentano numerosità elevate, in cui è stato possibile applicare la procedura FORGEX, sono riassunte nella [Tabella 3. 3].

Tabella 3. 3: località nelle quali è possibile testare l'applicativo creato.

Nome stazione	Codice	Numerosità
Oropa	12367	63
Spigno Monferrato	12203	61
Torino Ufficio Idrografico	12082	60
Varallo Sesia	12196	59
Bardonecchia	12387	50

Capitolo 4

Stima della curva di crescita con il metodo FORGEX

Lo scopo principale di tale elaborato è la ricerca di una procedura basata sulla falsa riga della metodologia inglese, in grado di identificare la curva di crescita in zone prive di informazioni. La metodologia FORGEX, che sorge in ausilio dell'analisi regionale in Inghilterra, è un approccio empirico che si basa esclusivamente sui dati osservati.

Questa tecnica è stata applicata ai dati italiani, presentando alcuni adattamenti rispetto alla procedura originale applicata in Inghilterra:

1) I dati italiani essendo in quantità inferiori rispetto ai dati inglesi, consentono un minor numero di 'network' nell'intorno del sito scelto;
2) La scelta della migliore spezzata che si adatta alla serie di dati, viene eseguita dall'utente, e come tale appare molto soggettiva, rispetto al criterio di selezione utilizzato in Inghilterra;
3) Il fitting della migliore distribuzione avviene sulla base di un processo iterativo di stima dei parametri caratteristici delle distribuzioni di probabilità più comuni.

La [Figura 4. 1] illustra sinteticamente i passaggi fondamentali che vengono affrontati nel corso di tale procedura: in special modo si evidenzia come le caratteristiche tipiche del metodo FORGEX vengano mantenute uguali e permettano una stima empirica della curva di crescita.

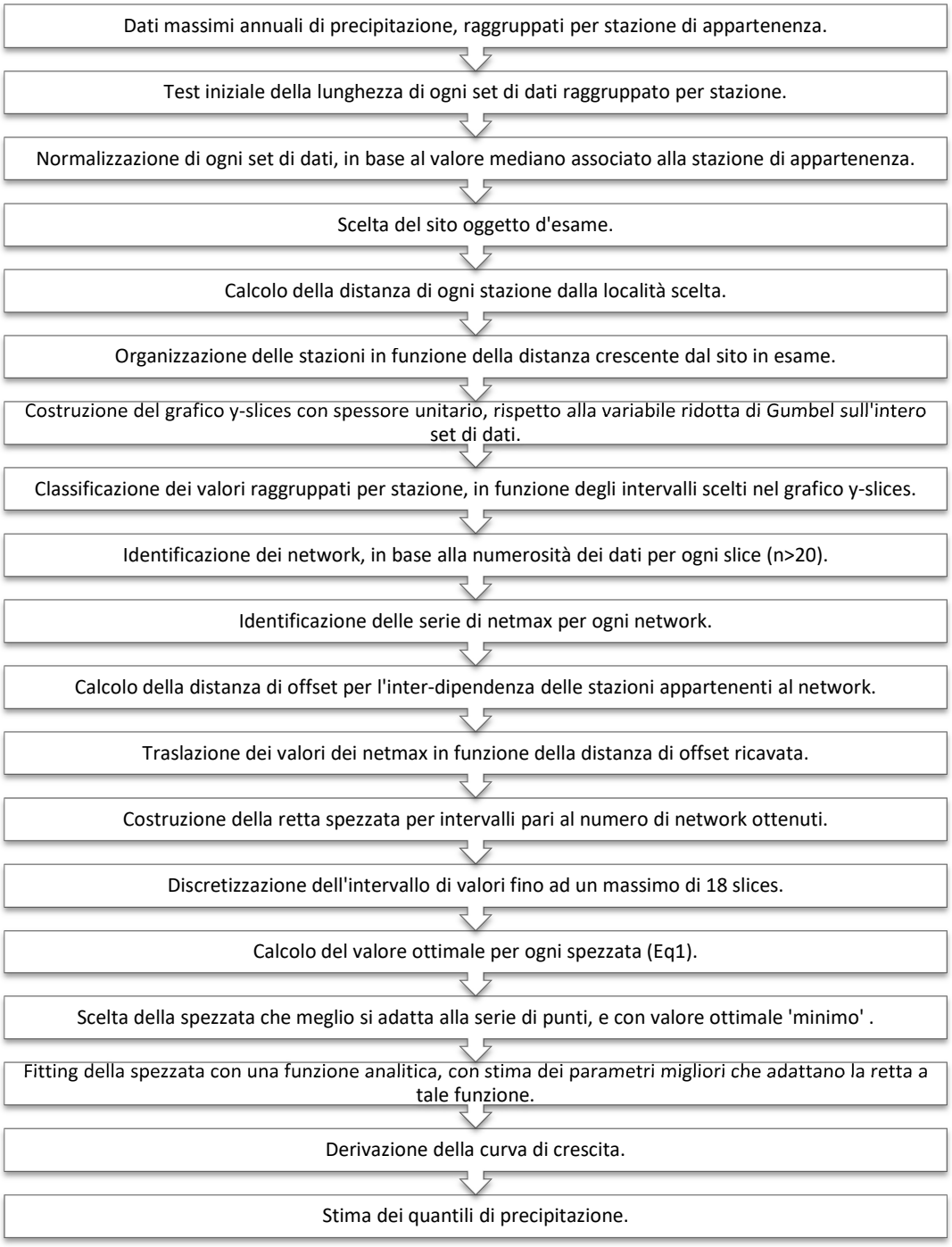


Figura 4. 1: albero delle azioni del metodo FORGEX

4.1 *Identificazione delle serie di dati*

4.1.1 *Selezione e normalizzazione delle serie di dati adeguate*

Il metodo FORGEX basa la sua analisi su una serie di massimi annuali di precipitazione raggruppati per stazione. Non tutti i set di dati possono essere utilizzati per la costruzione della curva di crescita: infatti, la prima valutazione, che viene eseguita su tali dati, riguarda la lunghezza della serie.

Per garantire un'analisi adeguata è necessario che ogni stazione appartenente al set di dati abbia una numerosità pari o superiore a 10. Questo parametro è legato al valore indice del metodo FORGEX utilizzato per la normalizzazione dell'insieme di dati. La mediana, infatti, assume un comportamento stabile su serie robuste e numerose. Per tale motivo è necessario eliminare quelle stazioni installate recentemente o che non forniscono serie sufficientemente lunghe per il proseguimento dell'analisi.

La scelta di utilizzare il valore mediano come parametro per garantire un'esatta normalizzazione degli elementi delle stazioni, fornisce all'analisi maggiore stabilità rispetto al valore medio. Reed , Faulkner e Stewart (1999) notarono come '*una serie di massimi annui di precipitazione, risulta già tipicamente distorta, dunque un valore estremo in questo insieme, potrebbe provocare un grosso cambiamento nel valore medio'*. A tal proposito, quindi, la normalizzazione del set di dati deve avvenire secondo la formula (3):

$$X_{i,j} = \frac{R_{i,j}}{RMED_j} \qquad 3$$

In cui si osservano i seguenti valori:

- $X_{i,j}$ indica il valore normalizzato del dato nella stazione *j*, alla posizione *i*;
- $R_{i,j}$ rappresenta il valore da normalizzare della serie nella stazione *j*, alla posizione *i*;
- $RMED_j$, che secondo la nomenclatura utilizzata in Inghilterra, indica il valore mediano di riferimento nella stazione *j*.

È importante evidenziare come per ogni stazione, verrà calcolato un differente valore mediano. In tal modo, non sarà costruita una serie normalizzata crescente, tipica dell'approccio 'station-year'[3]. Ogni set di dati, appartenente ad una diversa stazione, sarà indipendente l'uno dall'altro.

Il valore mediano non solo rende l'analisi più robusta, ma garantisce un legame con il periodo di ritorno: per definizione, il valore della mediana di qualunque serie di dati, corrisponde ad un evento con periodo di ritorno pari a 2 anni. Tale caratteristica permette di correlare la serie di dati normalizzata con la variabile ridotta, utilizzata nel grafico oggetto del metodo. La particolarità del metodo FORGEX, rispetto ai metodi regionali italiani, è la stima del valore normalizzato a partire da un range di valori di periodi di ritorno, noto a priori.

4.1.2 *Identificazione e selezione del set di dati nell'intorno del sito*

Successivamente, è necessario scegliere la località in cui si desidera ottenere una stima della curva di crescita. Il sito oggetto d'esame prenderà il nome di 'focal point', e garantirà continuità all'analisi: a partire da questo punto verranno tracciati diversi cerchi concentrici, denominati 'network', che conterranno tutte le stazioni entro i 200 chilometri dal sito scelto, valore imposto come distanza massima.

Il numero di network identificati dipenderà dalla numerosità dell'intero campione di dati. La procedura inglese stima un numero di network pari a 8: nel caso dei dati italiani, si ottiene un valore pari a 6 network, in quanto si possiede una numerosità inferiore di osservazioni per stazione, con una frequenza elevata degli eventi. I network identificati consentiranno la costruzione di un grafico, con pari numero di intervalli di classificazione sull'asse delle ascisse.

Il criterio di selezione delle stazioni, all'interno di ogni network, dipende esclusivamente dalla distanza calcolata rispetto al 'focal point'. È importante calcolare, per ogni stazione, la distanza dal sito scelto, eliminando a priori le località poste oltre la distanza massima. Questo limite viene imposto dalla metodologia inglese, per garantire all'interno dei 'network', l'uniformità dei dati ed evitare di incorrere in cambiamenti

[3] Il metodo 'station-year' considera le serie di dati indipendentemente dalla stazione a cui appartengono, fornendo set di dati ordinati, in modo crescente in funzione dell'anno di accadimento.

climatici elevati. Nel caso italiano, tale distanza non viene mai superata, poiché i dati relativi al territorio in esame, sono limitati alla regione Piemonte, come si nota in [Figura 4. 2].

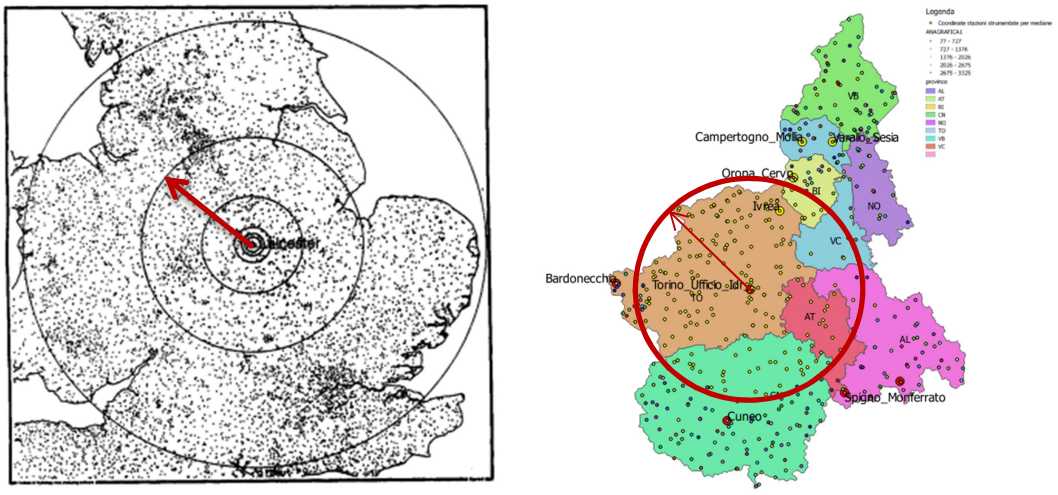


Figura 4. 2: confronto del metodo di selezione dei network sul territorio inglese e piemontese, con differenti scale

Nella [Figura 4. 2], si osserva la tecnica di selezione delle reti di stazioni nell'intorno del sito scelto, sia in Inghilterra che in Piemonte. Si osservano immediatamente le seguenti differenze:

- La densità complessiva delle stazioni italiane è inferiore a quella osservata in Inghilterra;
- Le stazioni strumentate nell'intorno del sito italiano, Torino, hanno una concentrazione inferiore rispetto a quelle presenti nell'intorno di Leicester in Inghilterra;
- Il cerchio massimo ottenuto nell'intorno di Torino è pari a 80 km, corrispondente, nella mappa di sinistra, al cerchio indicato dalla freccia in rosso, identificativo del penultimo network esaminato. Le due mappe non sono rappresentate in scala, ma noto il valore del cerchio alla massima distanza nell'intorno di Leicester, è stato possibile creare il legame tra le due differenti aree e stimare l'ampiezza dei cerchi concentrici.

Ogni network, selezionato durante questo processo, appare omogeneo grazie al criterio base dell'analisi, che prevede di valutare le stazioni in funzione della loro distanza dal 'focal point'. In tal modo, non sarà necessario eseguire alcun test di omogeneità sulla

regione stessa, in quanto si osserveranno eventi molto simili all'interno del medesimo network. L'omogeneità della regione è garantita, poichè risulterebbe strana la presenza di eventi di precipitazione molto diversi tra loro, localizzati nella stessa area.

Nell'intorno del punto centrale, in base al principio della distanza euclidea, saranno costruiti diversi network: ognuno di essi racchiuderà un dato numero di stazioni. Il criterio inglese per selezionare le varie stazioni, dipende dal numero di dati presente in ogni slices corrispondente ad un determinato network. L'ipotesi iniziale prevede che il network sia completo al raggiungimento di almeno 20 osservazioni per ogni slices. I network più vicini al 'focal point' saranno rappresentati da una serie di dati più fitti, a giustificare la presenza di eventi con periodi di ritorno più basso. Lo scopo di tale rappresentazione è cercare di uniformare l'intero set di dati, in modo da garantire per ogni slices una numerosità simile.

Il metodo FORGEX non fornisce un valore di raggio minimo da utilizzarsi, in quanto desidera una costruzione progressiva del set di dati, variabile in funzione delle stazioni limitrofe. Finché il valore minimo di dati per slice non verrà raggiunto, il network dovrà ampliarsi, comprendendo ulteriori stazioni.

Svensson & Jones (2010), hanno dichiarato, in termini generali che *'la dimensione di una regione è definita come il compromesso tra mantenere una regione piccola e dunque pertinente, ed evitare di costruire regioni comprendenti stazioni molto distanti, per ottenere delle stime di precipitazioni ad alto rendimento'.* In tal modo si osserveranno network più piccoli, ma maggiormente adeguati rispetto al sito scelto come oggetto di studio.

Questo tipo di classificazione appare molto simile alla metodologia 'region of influence', se non che, in questo caso, si realizzano le regioni 'omogenee', una di seguito all'altra. Ovviamente l'ultimo network comprenderà tutti i precedenti, con il problema di non raggiungere mai la numerosità minima imposta per ogni slices. Tale risultato è legato alla presenza di valori in quantità ridotta, associati a periodi di ritorno più elevati.

4.1.3 *Metodo delle y-slices*

Il metodo utilizzato per la rappresentazione dei dati raggruppati nell'intorno del sito in esame prende il nome di 'y-slices' e fa riferimento alla presenza di una serie di strisce verticali, in corrispondenza di determinati valori sull'asse delle ascisse. L'insieme di stazioni compreso in ogni network garantisce una serie di dati in funzione della durata caratteristica dell'evento. Come tale, è importante associare ad ogni network, in ordine di definizione, la relativa slice sul piano cartesiano.

Il legame tra il valore mediano e il periodo di ritorno viene espresso nella rappresentazione grafica del metodo 'y-slices'. Infatti, la suddivisione in network avviene in base a determinati intervalli definiti sull'asse delle ascisse: si ottengono *n* intervalli, quanti network si costruiscono. È importante definire gli estremi delle slices, a partire dal punto identificativo del valore mediano, le cui coordinate corrisponderanno a (0.3665, 1). Non è un requisito del metodo inglese garantire lo stesso numero di slices pari al numero di network esaminati, ma questo legame, semplifica notevolmente le operazioni successive.

I dati raggruppati in ogni network saranno suddivisi nelle differenti slices, ma non verranno rappresentati tutti sul piano cartesiano in esame: infatti, i soli valori inseriti nel grafico, saranno quelli compresi tra gli estremi della slice ad esso associata. Nel caso in cui si analizzasse il primo network, gli unici valori raffigurati, saranno quelli compresi tra il minimo valore dell'intervallo e il suo successivo (y_0 e y_1). La [Tabella 4. 1] riassume i valori delle variabili ridotte di Gumbel, fissate come gli estremi delle diverse slices nel grafico in [Figura 4. 3].

Tabella 4. 1:valori delle variabili ridotte di Gumbel, utilizzate nel metodo y-slices.

y_0	y_1	y_2	y_3	y_4	y_5	y_6
-0,6335	0,3665	1,3665	2,3665	3,3665	4,3665	5,3665

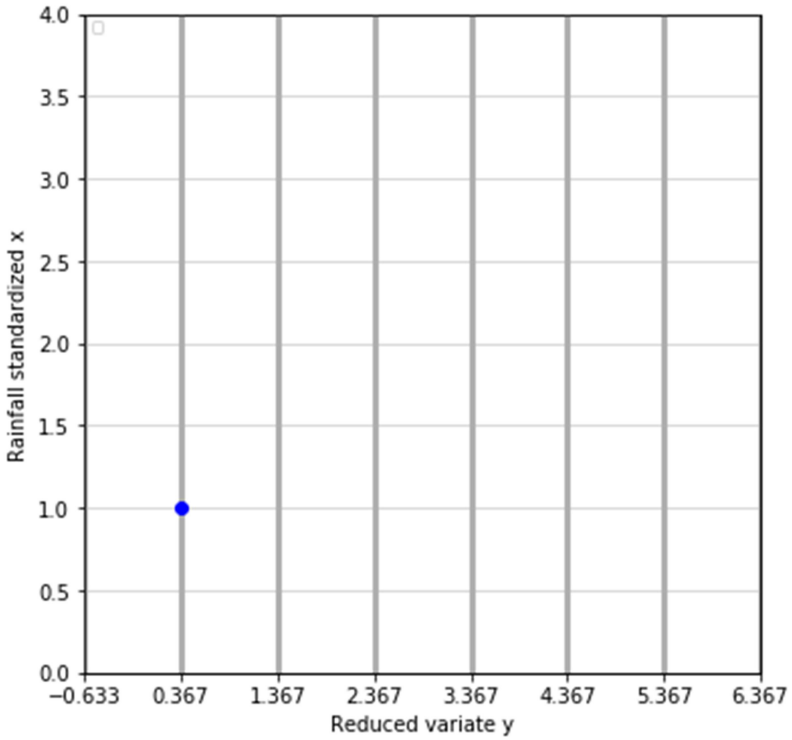


Figura 4. 3: rappresentazione del grafico y-slices, con indicazione del valore mediano con un punto blu.

Il grafico in [Figura 4. 3], mostra un esempio della suddivisione in differenti slices. Il diagramma prevede due assi, X e Y, che rappresentano rispettivamente la variabile ridotta di Gumbel, ricavata dalla formula (4), e la serie normalizzata rispetto al parametro di riferimento, ottenuta con la formula (3).

$$y = -\ln(-\ln F) \qquad 4$$

Per ottenere il valore esatto della probabilità di accadimento, il metodo utilizza la funzione di Gringorten, data dalla formula (5). In base alla posizione dell'i-esimo valore, è possibile ricavare la frequenza ad esso associata, dalla quale ottenere il valore della variabile ridotta y.

$$F = \frac{i - 0.44}{N + 0.12} \qquad 5$$

Dove:

- i rappresenta la posizione del valore nella serie di dati;
- N la numerosità totale dell'intera serie di dati.

È necessario porre l'accento sull'applicazione della formula (5): la frequenza F, verrà calcolata in funzione dei dati, raggruppati per stazione di appartenenza. Per tale motivo, la rappresentazione dei dati vedrà la presenza di differenti valori normalizzati, in corrispondenza della stessa variabile sulle ascisse, come si può osservare in [Figura 4. 4].

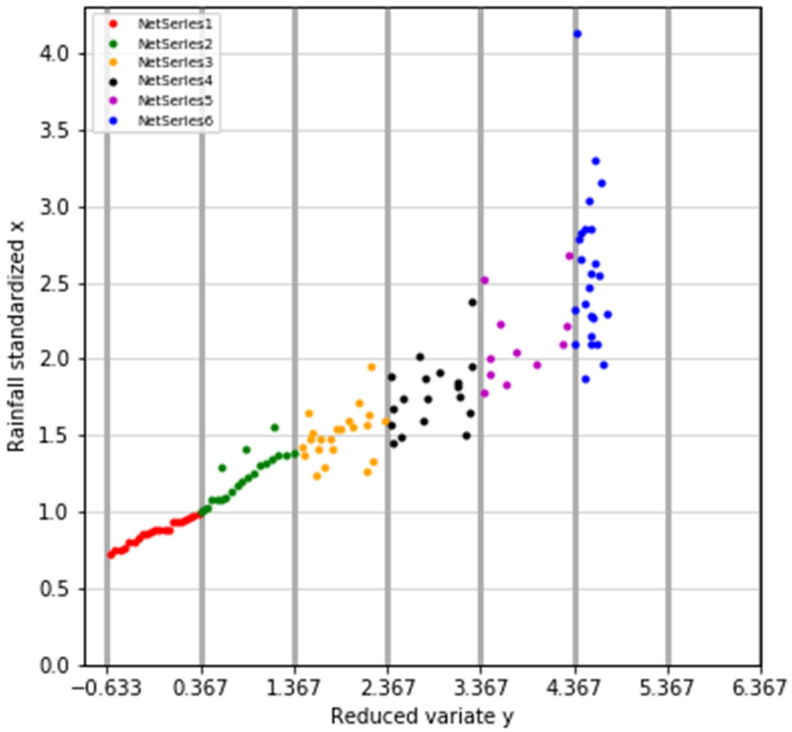


Figura 4. 4: Rappresentazione dei punti raggruppati per ogni network, nell'intorno del sito di Torino.

La definizione delle strisce verticali avverrà a partire dal valore mediano, che come già accennato, corrisponde ad un periodo di ritorno pari a 2 anni. Il metodo FORGEX prevede che si sviluppino, a destra del valore mediano, da 4 a 6 intervalli, a seconda della densità delle stazioni nell'intorno del sito di interesse. Ma non si osserveranno mai elementi inferiori al valore minimo del campione, posizionato intorno a -0.6335. In [Figura 4. 4], si osserva che, per i dati italiani, sono state ottenute un numero di slices pari a 6, corrispondenti a 6 network esaminati. Inoltre, si nota che, per l'intervallo superiore al valore di y_6 , non sono stati rilevati dati in grado di contribuire all'analisi. Questo è correlato con la presenza di serie di dati, la cui frequenza di accadimento è molto elevata.

4.1.4 *Analisi dei valori massimi del network*

I network più ampi non presenteranno dati sufficienti per soddisfare il requisito minimo di 20 osservazioni, all'interno della slice di riferimento. Per ovviare a tale problema, il metodo FORGEX suggerisce di considerare per ogni network, le serie di massimi annui di precipitazione, eliminando la dipendenza dalla stazione di appartenenza. I punti raccolti in questi insiemi verranno denominati 'netmax'.

La selezione di queste serie di dati, avviene considerando gli eventi esclusivamente in funzione dell'anno di accadimento: in tal modo, per ogni network si avranno delle serie ordinate in modo crescente, dal primo anno di osservazione all'ultimo. Ovviamente l'ultimo network, avrà una serie di netmax, che sarà comprensiva di tutte le precedenti.

Così facendo, ogni rete di stazioni avrà una serie aggiuntiva di valori, ordinati in modo crescente, utile a determinare la coda della curva di crescita. Studi precedenti, hanno dimostrato che queste serie, rappresentate insieme ai dati ottenuti dal 'pooling', erano leggermente spostate verso sinistra rispetto alle altre. Tale spostamento, secondo Dales e Reed, (1989), *'sembrava dovuto alla presenza della distribuzione GEV, utilizzata per rappresentare la serie stessa',* successivamente però Reed e Stewart, (1994), dichiararono *'come non fosse legato assolutamente alla distribuzione in sé, ma all'effetto dell'interdipendenza spaziale tra elementi appartenenti allo stesso network'.* La rappresentazione di tali serie di dati, dovrà effettivamente essere leggermente traslata verso destra, per garantire una visualizzazione uniforme.

L'articolo di Dales & Reed ,(1989), esplicita come la *"dipendenza spaziale, sia semplicemente associata all'analisi di precipitazione: infatti il metodo combinando serie di dati di lunghezza in termini di anni diversi, e da stazioni differenti, ricrea una serie unica di valori. Questa dipendenza, è evidenziata dal numero effettivo di stazioni indipendenti, che risulta essere inferiore al numero totale di stazioni presenti nel network".*

La dipendenza spaziale si verifica in modo più evidente per quei dati provenienti da stazioni molto vicine, come accade nei primi network. Infatti, maggiore sarà la distanza tra le differenti stazioni, minore sarà la dipendenza spaziale: questo significa che più le stazioni sono vicine una all'altra, minore sarà lo spostamento dei 'netmax value'.

Viceversa, maggiore sarà il valore della distanza tra le diverse stazioni di un network, quindi reti più ampie, maggiore sarà lo spostamento della serie di 'netmax'. Il modello della dipendenza spaziale varia al modificarsi di una serie di fattori che dipendono dal network selezionato.

Per garantire una corretta stima della curva di crescita sarà necessario calcolare, per ogni serie di 'netmax', un valore denominato distanza di offset, espresso dall'equazione (6):

$$\frac{lnN_e}{lnN} = 0.081 + 0.085 * lnAREA - 0.051lnN - 0.027lnD \quad 6$$

Dove:

- N_e rappresenta il numero esatto di stazioni indipendenti;
- lnN_e indica il valore della distanza di 'offset', ovvero quanto il 'netmax' deve essere spostato in funzione della dipendenza spaziale;
- D indica la durata della precipitazione, espressa in ore:
  - Nel caso in cui la durata sia superiore alla durata giornaliera, allora sarà espressa in giorni;
  - Nel caso in cui la durata sia inferiore a 15 ore, è necessario dividere il valore della durata per 18 ore, in modo da garantire l'uniformità della rappresentazione.
- *AREA* rappresenta la parte di spazio coperta dal network di riferimento, ed è calcolata con la formula (7):

$$AREA = 2.5 * \left(\bar{d}\right)^2 \quad 7$$

  Dove :
  - d indica la media della distanza tra le diverse coppie di stazioni interdipendenti.

La rappresentazione delle serie di 'netmax' garantisce un incremento della numerosità delle slices più a destra, corrispondenti a periodi di ritorno più elevati, come si vede nella [Figura 4. 5].

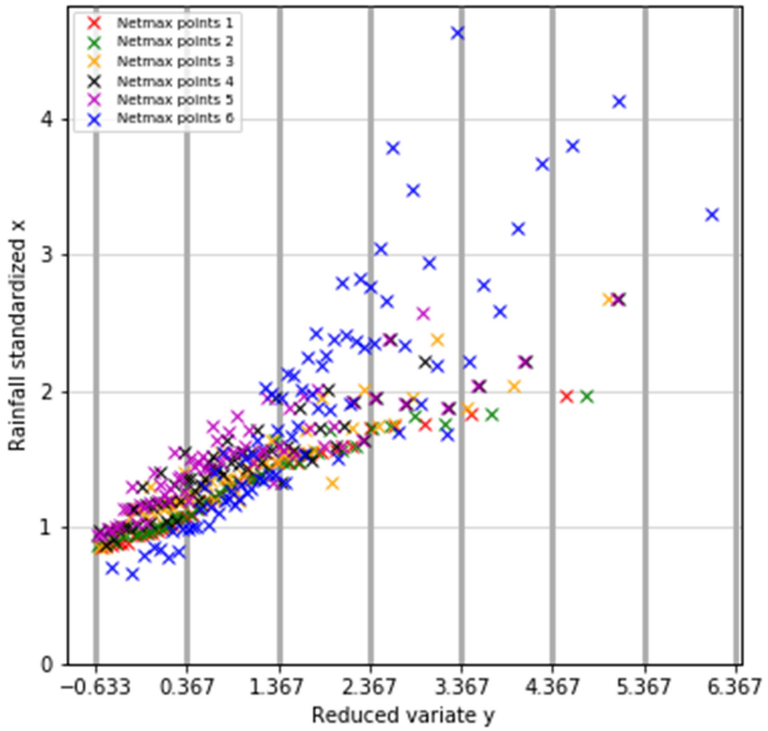


Figura 4. 5: rappresentazione delle serie netmax sul grafico y-slices, nell'intorno del sito di Torino.

4.2 *Costruzione della curva di crescita*

4.2.1 *Costruzione della spezzata*

La rappresentazione dell'insieme di punti per network differenti, garantisce la costruzione della curva di crescita come una funzione a tratti. Ogni segmento corrisponderà ad un dato network analizzato. All'interno di ciascun intervallo si osservano i valori ottenuti dal 'pooling', classificati per slices, e le serie di 'netmax', che dovranno essere suddivise in funzione del loro intervallo di appartenenza. Ogni slices possiederà dunque un set di dati, sul quale sarà possibile procedere con l'analisi.

La serie di passaggi che è necessario eseguire per costruire la spezzata da un insieme di punti sparsi, è la seguente:

1) Determinazione del gradiente del segmento associato all'intervallo di punti, classificato in base al valore delle ascisse;
2) Applicazione della funzione a tratti per la ricerca dei valori di modello associati ai valori delle ascisse.

La [Figura 4. 6] mostra per il sito di Torino, l'insieme di dati sul quale sarà necessario lavorare:

- I puntini rappresentano i valori individuali osservati, raggruppati per ogni stazione: sono raffigurati con differenti colori, ad identificare il network di appartenenza;
- Le crocette invece, rappresentano le serie di netmax, raggruppate per anno e raffigurate con i colori corrispondenti al network di appartenenza.

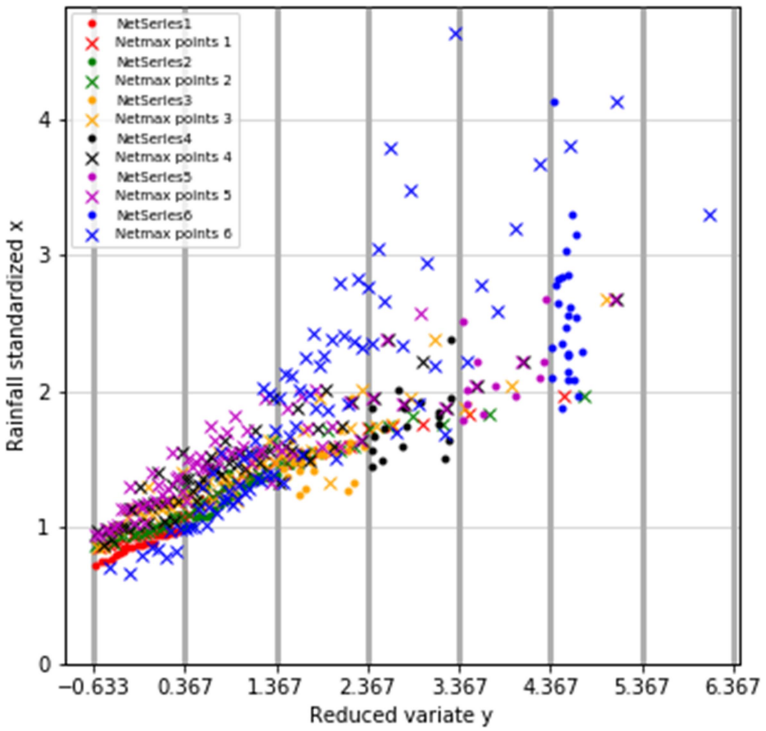


Figura 4. 6: rappresentazione del set di dati da utilizzare nella costruzione della spezzata, nell'intorno del sito di Torino.

La costruzione della prima funzione a tratti avverrà sulla base di un numero di intervalli pari al numero di network selezionati. Il modo più semplice per determinare il gradiente di una retta di un insieme sparso, è una regressione ai minimi quadrati. L'equazione (8) rappresenta la formula che si andrà ad applicare per ottenere il valore del gradiente richiesto:

$$m = \frac{\overline{yx} - \bar{y} * \bar{x}}{\overline{y^2} - \bar{y}^2} \quad (8)$$

Dove:

- m indica il coefficiente angolare del tratto in esame;
- $\bar{y}$ è il valore medio della variabile ridotta di Gumbel, per l'intervallo desiderato;
- $\bar{x}$ è il valore medio della serie di punti normalizzati, associati all'intervallo.

Noti i valori delle variabili ridotte per ogni intervallo, e noti i coefficienti angolari di ognuno di essi, è possibile stimare le ordinate, corrispondenti ai valori della variabile ridotta.

Per eseguire tale operazione, è necessario identificare la posizione del segmento in esame, rispetto al valore della mediana:

1) Noti:

   x_1=1 ,

   y_1=0.3665 ,

   y_0=-0.6335 ,

   Per slices con valori -0.6335<y<0.3665, si calcola x_0 come:

$$m = \frac{x_1 - x_0}{y_1 - y_0} \rightarrow x_0 = x_1 - m_{01}(y_1 - y_0) \quad (9)$$

2) Per slices con valori y>0.3665, la formula da applicare sarà, invece, la seguente:

$$m = \frac{x_2 - x_1}{y_2 - y_1} \rightarrow x_2 = x_1 + m_{12}(y_2 - y_1) \quad (10)$$

Le due equazioni (9) e (10), vengono utilizzate per garantire la continuità dei differenti segmenti.

L'equazione (11) fornisce invece lo schema della funzione a tratti, nell'ipotesi di considerare quattro segmenti, come è esemplificato nell'articolo di Reed , Faulkner, e Stewart, (1999). Il numero delle equazioni varierà al modificarsi del numero di slice e quindi in funzione dei network realizzati.

La funzione a tratti sarà definita nel seguente modo:

$$x_{model} = \begin{cases} 1 + m_1(y - y_1) & per\ y_0 < y_1 \\ 1 + m_2(y - y_2) & per\ y_1 < y < y_2 \\ 1 + m_2(y_2 - y_1) + m_3(y - y_3) & per\ y_2 < y < y_3 \\ 1 + m_2(y_2 - y_1) + m_3(y_3 - y_2) + m_4(y - y_4) & per\ y > y_3 \end{cases} \quad (11)$$

Dove:

- x_{model} rappresenta la serie di valori normalizzati da stimare;
- I coefficienti m, numerati da 1 a 4, indicano i gradienti dei segmenti, costruiti per ogni slices;
- I valori delle y numerati, indicano gli estremi di ogni slices, a partire da $y_1 = 0.3665$, corrispondente al valore mediano.
- I valori di y, privi di numero al pedice, rappresentano, le componenti sull'asse delle ascisse dei valori osservati.

L'equazione (11), produrrà una serie di segmenti continui, caratterizzati da diversi cambi di pendenza, come si evince dalla [Figura 4. 7].

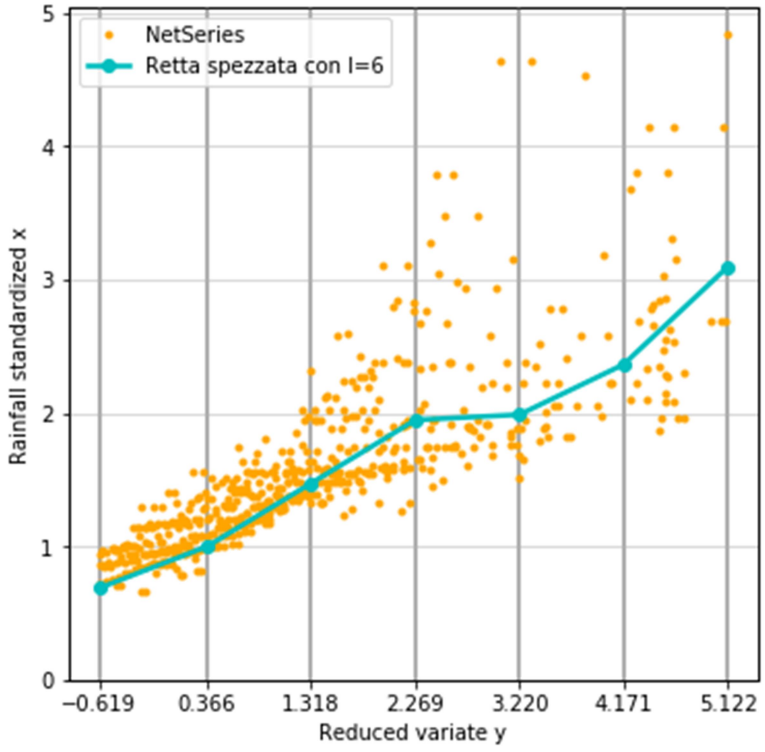


Figura 4. 7: rappresentazione della curva di crescita costruita a tratti per il sito di Torino per la durata di 3h.

Questa prima interpolazione viene costruita in modo da avere un'idea iniziale della possibile curva di crescita per l'insieme di dati selezionato. Successivamente sarà necessario incrementare il numero di intervalli, per aumentare la precisione della curva e renderla il più lineare possibile.

4.2.2 *Discretizzazione dell'intervallo*

Il metodo FORGEX prevede la costruzione di una curva di crescita in funzione di un modello, in grado di smorzare i cambi di pendenza, e rendere tale curva il più omogenea possibile. Il metodo inglese utilizzato segue l'idea della regressione lineare ai minimi quadrati, con l'aggiunta della funzione penalità. Tale elemento, permetterà di smussare le congiunzioni tra i differenti segmenti.

La formula (12) illustra la funzione utile per tracciare al meglio i diversi segmenti oggetto del metodo.

$$\sum_{i=1}^{M}(x_i - x_{model})^2 + 0.05\left(\frac{M}{l-1}\right)\sum_{j=2}^{l}\left(m_j - m_{j-1}\right)^2 \qquad 12$$

Dove:

- x_i indica il valore osservato della serie, normalizzato;
- x_{model} è il valore stimato della serie di dati secondo la formula (11);
- M è la numerosità totale dei dati osservati;
- l è il numero di segmenti utilizzati per costruire la funzione;
- m_j indica il gradiente del j-esimo segmento;
- $0{,}05$ è un parametro utilizzato in Inghilterra, per rendere migliore l'aspetto della funzione a tratti.

Questo passaggio è stato leggermente modificato per il caso italiano: vengono realizzate una serie di discretizzazioni del macro intervallo, in modo da aumentare la precisione della stima dei valori di modello e garantire una funzione a tratti che sia il più lineare possibile. Tale processo viene eseguito fino ad un numero di intervalli non superiore a 16: infatti si nota come, superando tale valore, la spezzata non segua più l'andamento

del set di punti in esame, ma tenderà a divergere verso l'alto, non garantendo credibilità al metodo.

Per ogni discretizzazione dell'intervallo, si ricalcolano i valori di modello associati alle variabili presenti sull'asse delle ascisse, relative ai dati osservati. In tal modo sarà possibile ricostruire, in modo iterativo, una spezzata che aumenti la sua omogeneità all'aumentare del numero di intervalli.

Nel corso dell'analisi, si possono osservare alcune località poste nei bordi della regione. In tali aree, l'adattamento della funzione a tratti diminuisce, poiché la numerosità complessiva del set di dati è ridotta, a causa della difficoltà di reperire informazioni utili nell'intorno desiderato.

4.2.3 *Selezione della spezzata ottimale*

La discretizzazione eseguita sull'intervallo di valori dovrebbe garantire un andamento più affine al set di dati in esame, fornendo una serie di risultati dell'equazione (12) decrescenti all'aumentare del numero di intervalli. Tale equazione mostra due componenti principali, che si comportano in un modo leggermente differente da quanto atteso.

Il primo addendo calcola gli scarti quadratici minimi tra i valori osservati e quelli stimati in base al modello. Si osserva, come per le slices limitrofe al valore mediano, gli scarti siano inferiori all'unità, a testimonianza di un migliore adattamento. Più le slices si allontanano da tale valore, diminuendo la concentrazione di dati al loro interno, maggiore sarà il risultato di questo primo addendo, a dimostrazione della divergenza che si viene a creare tra i valori stimati e quelli osservati.

Il secondo addendo, invece, rappresenta gli scarti tra i gradienti di ogni slices: come tale risulterà sempre di un ordine di grandezza inferiore rispetto al precedente, dunque non molto di aiuto a riequilibrare la situazione. Tale valore verrà moltiplicato per una quantità variabile, in funzione del numero di intervalli e di un parametro utilizzato per l'ottimizzazione della curva.

In Inghilterra, il migliore adattamento si ottiene ponendo tale parametro pari a 0,05. In Italia, modificando questo valore, si sono presentate due situazioni:

1) Nel caso in cui il parametro aumenti, i valori dell'equazione (12), non si modificano eccessivamente, mantenendo l'andamento crescente dei risultati;
2) Nel caso in cui il parametro diminuisca, la serie di risultati assumerebbe un comportamento crescente, ma in quantità molto ridotta rispetto a quanto accade con 0,05.

Per questo motivo, l'equazione (12), verrà utilizzata dall'utente, per selezionare la migliore discretizzazione da utilizzare per rappresentare la curva di crescita.

Osservando i risultati dell'equazione per ognuna delle discretizzazioni eseguite, si è notato come in certi casi, tale valore, risulti non coerente con gli altri, come ad evidenziare un migliore adattamento alla serie di dati. In questo modo è stato possibile ipotizzare un criterio di scelta della funzione a tratti. L'utente, dovrà analizzare i differenti risultati, e selezionare il numero di intervalli di discretizzazione, per il quale il risultato dell'equazione (12) potrà essere considerato come un 'minimo locale'[4]

La scelta della migliore discretizzazione non sempre appare come un procedimento semplice e rapido: più volte si osservano risultati, che potrebbero garantire un perfetto adattamento, ma che in realtà, rappresentati, mostrano, nella coda della funzione, un andamento divergente verso l'alto. Questo fenomeno è legato alla scarsa numerosità di eventi con periodi di ritorno molto alti, tali da poter essere inseriti nelle slices più destra.

La [Tabella 4. 2] rappresenta i risultati dell'equazione (12), ottenuti per l'analisi eseguita nel sito di Torino, per la durata di 3 ore, modificando il parametro caratteristico. Si osserva come l'andamento crescente dei risultati venga mantenuto, nonostante si modifichi il valore del parametro P.

Si nota un'improvvisa diminuzione del valore per un numero di slices pari a 11: in questo modo sarà possibile utilizzare la discretizzazione con 11 intervalli, come la più adatta.

[4] Un minimo locale è un valore di una serie crescente di dati, che si presenta come non uniforme con tale serie: solitamente risulta molto inferiore rispetto agli altri elementi.

Tabella 4. 2: valori dell'equazione (12) calcolati per il sito di Torino.

Numero di slices L	Risultato EQ1 (P=0.05)	Risultato EQ1 (P=0.08)	Risultato EQ1 (P=0.02)
6	101.69	103.25	100.12
8	117.46	120.54	114.38
9	123.82	131.84	115.80
10	267.38	292.73	242.02
11	**146.38**	**152.194**	**140.57**
12	217.60	256.05	179.14
13	217.01	257.43	180.59
14	240.00	304.26	175.74
15	852.68	1059.20	646.16
16	725.70	972.28	479.13

La [Figura 4. 8] rappresenta la funzione a tratti selezionata secondo il criterio del minimo locale, che si presenta come risultato dell'equazione (12). Si osserva come l'ultimo tratto non sia uniforme con il resto della funzione: questo problema è legato alla numerosità dei dati. Come si può notare, nella parte in alto a destra del grafico, la concentrazione dei punti è ridotta.

Per eseguire una verifica della correttezza della discretizzazione scelta, è possibile visualizzare le funzioni a tratti ottenute nei tre casi differenti, rispetto alla nuvola di punti, come mostra la [Figura 4. 9]. Per garantire uniformità con il metodo applicato in Inghilterra, si continua l'analisi con il valore del parametro P pari a 0,05.

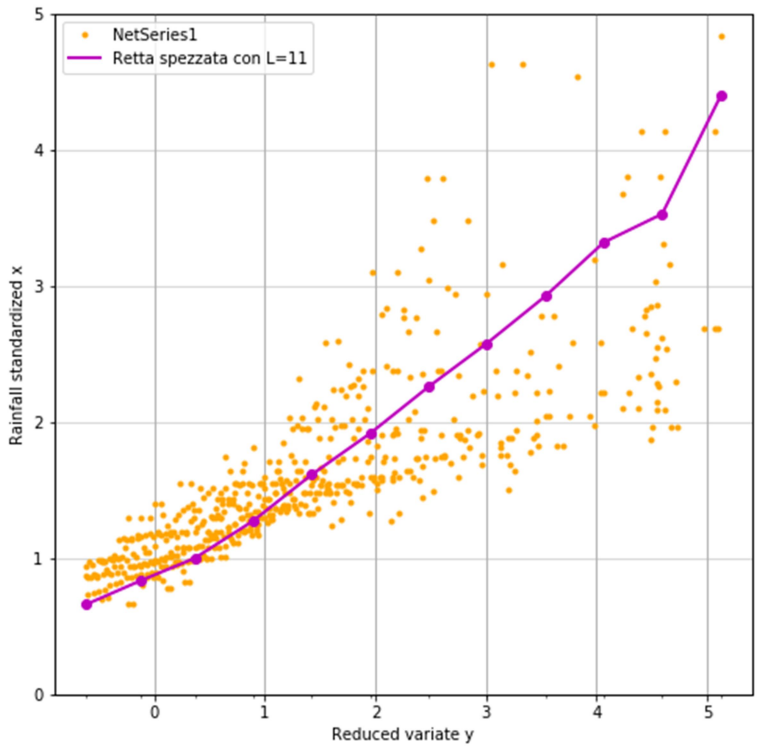


Figura 4. 8: rappresentazione della funzione a tratti scelta per la località Torino.

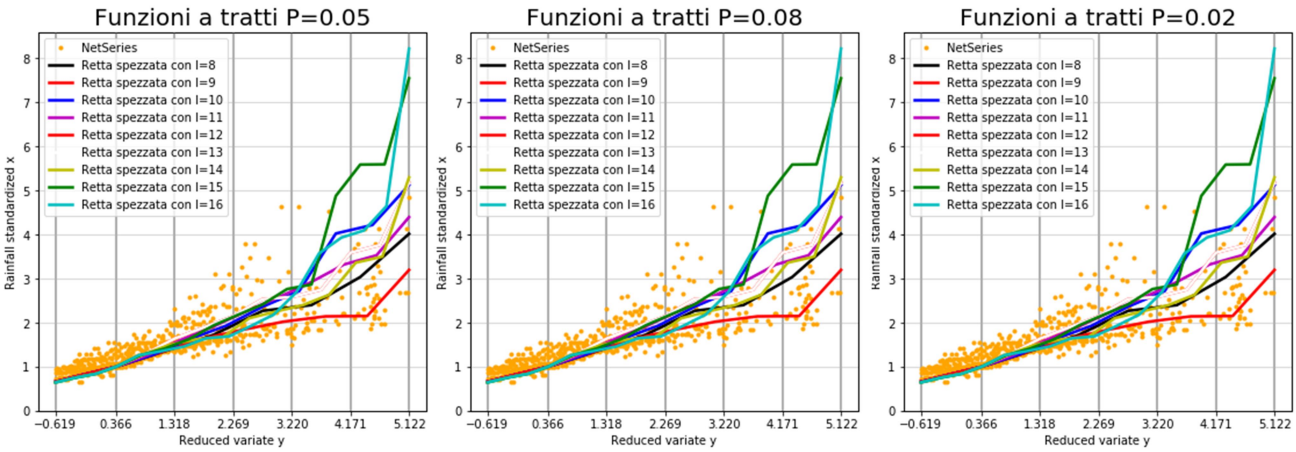


Figura 4. 9: rappresentazione delle discretizzazioni rispetto al set di dati, in funzione dei tre parametri.

4.2.4 *Fitting della spezzata con la migliore distribuzione*

La scelta della spezzata, che meglio si adatta alla serie di punti, risulta il passaggio più delicato dell'intera metodologia. Successivamente, è necessario ricercare una funzione analitica, in grado di adattarsi alla retta costruita per tratti, così da rendere la curva di crescita indipendente dal set di dati e confrontabile con altri metodi.

Nel campo dell'analisi regionale, solitamente le funzioni utilizzate maggiormente sono la distribuzione di Gumbel e la Gev, che vengono rappresentate dalle formule seguenti, in funzione dei parametri caratteristici:

$$x_{GEV} = \varepsilon + \frac{\alpha}{k}\left(1 - e^{-ky}\right) \quad (13)$$

$$x_{GUMBEL} = \varepsilon + \alpha y \quad (14)$$

Dove:

- ε è il parametro di posizione, che rappresenta lo spostamento della distribuzione in un piano: se tale parametro aumenta, la distribuzione si muoverà verso l'alto, altrimenti la distribuzione si sposterà verso il basso;
- α è il parametro di scala, che influenza la distribuzione allungandola verso l'alto, se aumenta, e comprimendola verso il basso se diminuisce;
- k è il parametro di forma, che assumendo esclusivamente valori negativi, controlla lo sviluppo verso l'alto della distribuzione;
- y identifica il valore della variabile ridotta, legata al periodo di ritorno.

I parametri incogniti delle equazioni (13) e (14), vengono stimati in modo da modellare la migliore distribuzione sulla spezzata ottenuta. Solitamente si osserva come la distribuzione GEV sia quella che garantisca un migliore adattamento alla funzione a tratti.

Il risultato ottenuto con la stima dei migliori parametri garantisce alla procedura una curva di crescita identificativa di quel sito, ottenuta con la metodologia FORGEX. Questa curva viene utilizzata per la stima del fattore di crescita in funzione di un dato periodo di ritorno. Il quantile normalizzato, associato al tempo di ritorno, moltiplicato per la grandezza indice, garantirà la stima della precipitazione di progetto.

Nella [Figura 4. 10], sono rappresentate le due distribuzioni generate in base ai valori della funzione a tratti. Si osserva come la distribuzione GEV, in verde, si adatti decisamente meglio alla spezzata, in viola, ad eccezione della coda della distribuzione che non segue correttamente la funzione a tratti. Questo problema, può essere legato all'ottimizzazione della curva sulla funzione a tratti, che nell'ultima slices possiede poche informazioni.

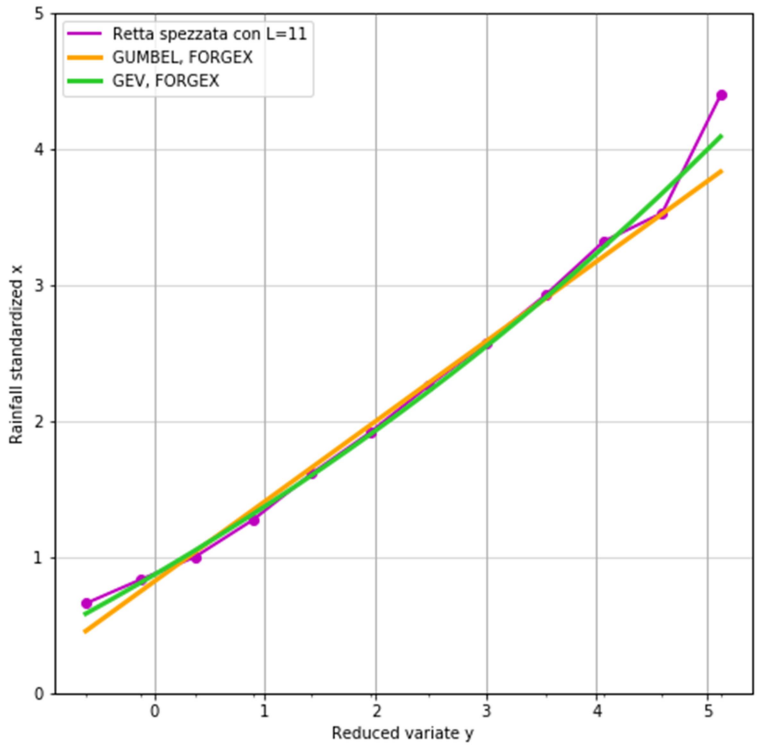


Figura 4. 10: Rappresentazione delle differenti distribuzioni per la scelta della più adatta.

La [Figura 4. 10] rappresenta il risultato della metodologia FORGEX, al momento della scelta della migliore funzione analitica. Si osservano tre elementi:

1) in viola, è rappresentata la spezzata, con i valori estremi degli intervalli, raffigurati dai punti, per evidenziare le differenti slices;
2) in arancione, è rappresentata la distribuzione di Gumbel, i cui parametri sono stati stimati per garantire il migliore adattamento alla spezzata;
3) in verde, è evidenziata la distribuzione Gev, che meglio si adatta alla spezzata oggetto dell'analisi.

La [Tabella 4. 3] mostra i parametri delle distribuzioni di Gumbel e GEV, stimati per il sito di Torino, in funzione del set di dati modello. I parametri ottenuti risultano per entrambe le distribuzioni simili tra loro.

Tabella 4. 3: tabella riassuntiva dei parametri ottenuti per le distribuzioni Gev e Gumbel.

Distribuzione	α	ε	k
Gumbel	0.588	0.820	-
GEV	0.476	0.869	-0.103

La distribuzione che maggiormente si associa alla spezzata, per la maggior parte delle località analizzate, è la GEV, che verrà utilizzata come curva di crescita. Le schede nella [Appendice F: Schede delle località analizzate] mostrano le applicazioni del metodo in località diverse.

4.2.5 *Il metodo FORGEX corretto*

Il metodo FORGEX, prevede di costruire la curva di crescita sulla base del valore indice, considerato come il valore mediano. Per uniformare questo risultato con le metodologie applicate in Italia, è necessario applicare ai valori dei quantili normalizzati, una determinata correzione, espressa nella formula seguente:

$$x_{norm} = \frac{x_{norm} * mediana}{media} \quad 15$$

In questo modo, il set di dati sarà coerente con la procedura inglese, e potrà essere confrontato con le curve di crescita ottenute dai metodi utilizzati in Italia.

La [Figura 4. 11] mostra il confronto tra le due curve utilizzando due differenti valori indice:

1) La curva in verde chiaro mostra il set di dati normalizzato rispetto al valore mediano. Si nota come tale normalizzazione produca dei quantili più elevati per variabili ridotte maggiori.
2) La curva in verde scuro illustra l'insieme di valori, in seguito alla correzione rispetto al valore medio. Come si nota, i valori dei quantili massimi sono inferiori rispetto a quelli ottenuti con la normalizzazione rispetto alla mediana. Per uniformità tra i metodi, l'analisi prosegue con i valori corretti.

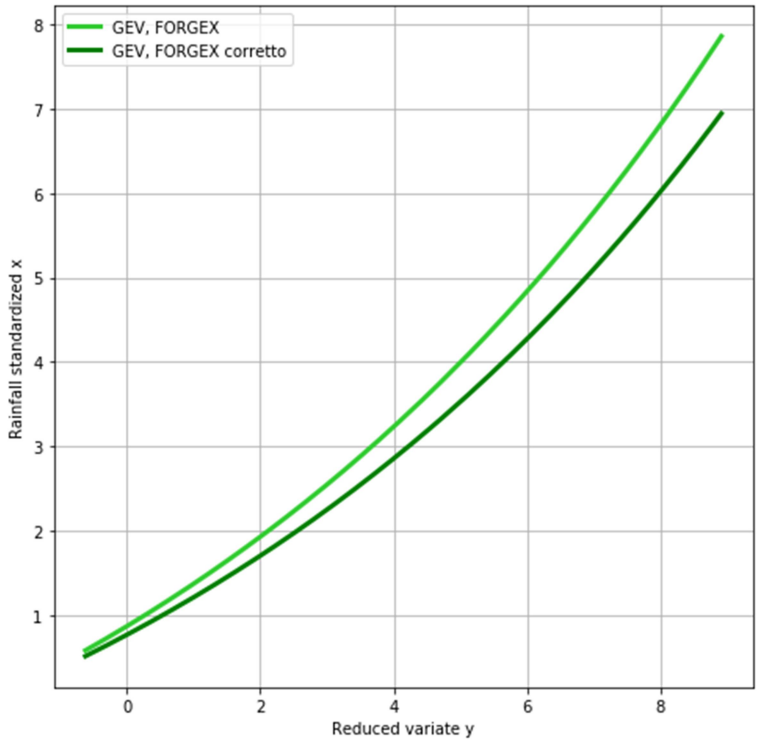


Figura 4. 11: confronto tra la distribuzione GEV costruita con i valori normalizzati rispetto alla mediana, e alla media.

4.2.6 *Validazione della distribuzione scelta*

I metodi di analisi regionale prevedono la verifica della validità della distribuzione scelta nei siti strumentati, per eseguire un controllo in modo diretto: il test eseguito assume maggiore affidabilità per le stazioni la cui numerosità è molto elevata. Il confronto viene effettuato costruendo una curva empirica di stazione e una curva empirica 'station year'.

La curva empirica di stazione viene creata in base ai dati appartenenti alla stazione strumentata, in modo da verificare immediatamente l'adattamento della distribuzione ai dati osservati. Tale curva è prodotta ricercando il sito desiderato nel database della regione Piemonte: in tal modo, la normalizzazione della serie avverrà in funzione del suo valore medio.

La curva empirica 'station year', invece, viene realizzata in funzione dell'intero set di dati nell'intorno del sito considerato. Quest'ultima è utilizzata per verificare l'adattamento della distribuzione scelta ai dati del 'pooling'. Questa curva, invece, prevede la normalizzazione dei dati, rispetto al valore medio di ogni stazione. In questo modo, sarà

possibile rappresentare l'andamento delle due curve, sia in funzione della variabile ridotta, che del tempo di ritorno T.

La [Figura 4. 12], mostra la distribuzione GEV selezionata per il miglior adattamento, a confronto con le curve empiriche di stazione e 'station year' del sito di Torino: si osserva come la distribuzione segua un andamento coerente con le altre curve, per valori del tempo di ritorno molto bassi. La coda della distribuzione, invece, dimostra una sovrastima del quantile normalizzato per periodi di ritorno molto elevati, rispetto a quanto atteso dai dati del pooling.

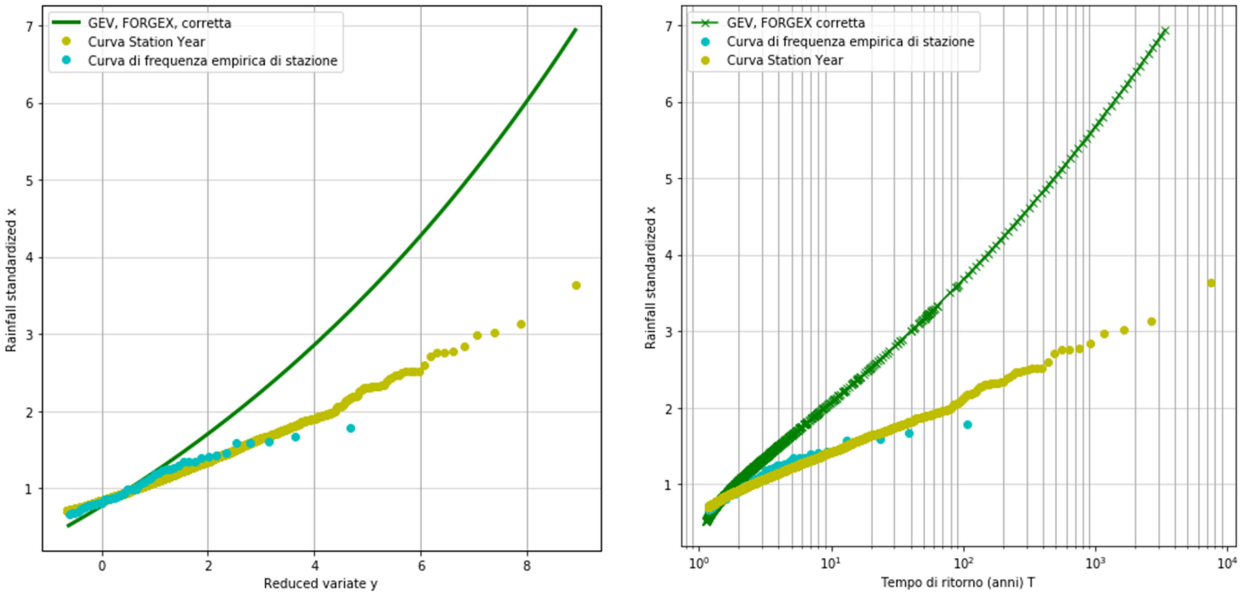


Figura 4. 12: confronto tra la distribuzione GEV e le curve di frequenza empirica nel sito di Torino, per verificare la validità della distribuzione selezionata, in funzione della variabile ridotta e del tempo di ritorno.

Le discrepanze piuttosto evidenti, saranno esaminate nel corso dei prossimi capitoli, in relazione anche dell'ampiezza dell'area considerata in funzione del massimo tempo di ritorno di interesse.

4.3 *Confronto con altri metodi*

La curva di crescita, costruita come una funzione a tratti, non garantisce la linearità richiesta nell'ambito dell'analisi regionale. Per tale motivo, si è eseguita la stima dei parametri di alcune distribuzioni, per cercare di rendere la spezzata una funzione analitica, e garantire un confronto con le altre metodologie applicate in Italia, come, per esempio, il metodo VAPI e la tecnica del Kriging.

4.3.1 *Il metodo VAPI*

Il progetto VAPI, descritto nel rapporto di De Michele e Rosso,(2001), analizza per ogni comune sul territorio piemontese, i parametri della distribuzione più adatta, rappresentativa dell'area in esame. In funzione di tali elementi è stato possibile costruire la curva di crescita, a partire dai valori delle variabili ridotte appartenenti alla spezzata in oggetto. La curva di crescita ottenuta consente una stima dei quantili normalizzati, in funzione della seguente equazione :

$$x_{GEV} = \varepsilon + \frac{\alpha}{k}\left(1 - e^{-ky}\right) \qquad 16$$

A seconda del sito in esame, i parametri di scala, di posizione e di forma assumeranno valori differenti. Per il sito di Torino, sono riassunti i parametri caratteristici nella [Tabella 4. 4], estratti dal manuale di De Michele e Rosso.

Tabella 4. 4: parametri caratteristici per la costruzione della curva di crescita con il metodo VAPI.

Località	α	ε	k
Torino	0.262	0.836	-0.048

La curva di crescita rappresentata in [Figura 4. 13] mostra l'andamento della distribuzione GEV costruita in funzione dei parametri riportati in [Tabella 4. 4]. Si osserva come le curve di frequenza empirica si adattino perfettamente alla curva di crescita realizzata con il metodo VAPI: nella coda della distribuzione si nota una lieve sovrastima dei quantili della curva ottenuta con il metodo VAPI, a confronto con la 'station year'.

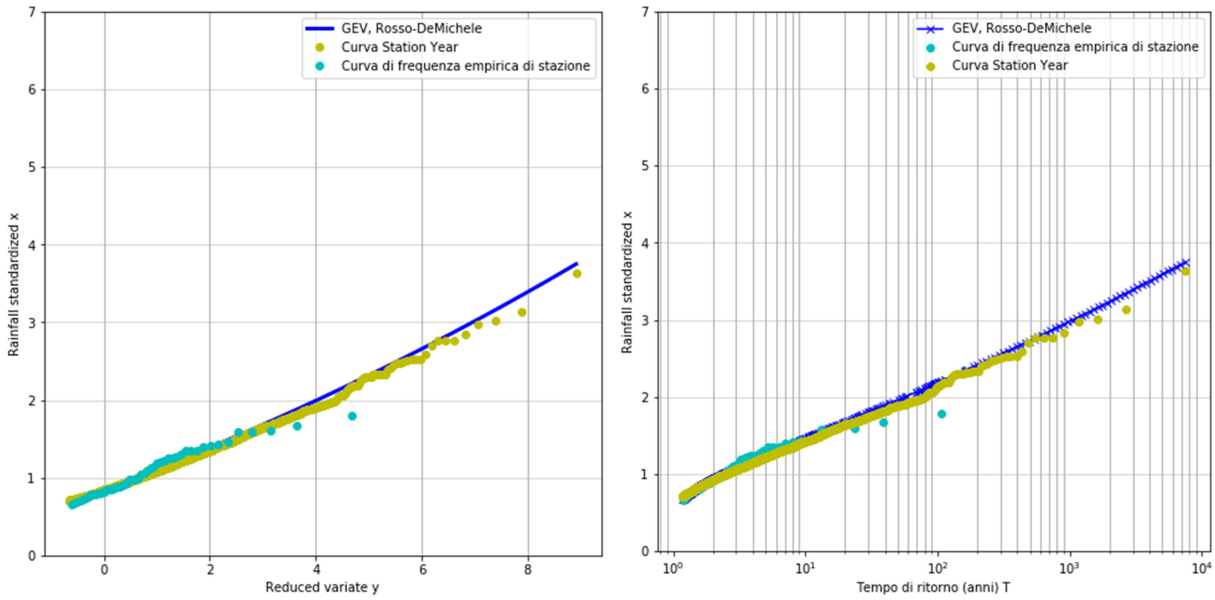


Figura 4. 13: rappresentazione della distribuzione GEV, a confronto con le curve di frequenza empirica, sia in funzione di y che di T, nel sito di Torino.

4.3.2 *Il metodo Kriging*

Un'ulteriore procedura utilizzata per costruire la curva di crescita, da confrontare con i risultati ottenuti con il metodo FORGEX, è il Kriging. Questa tecnica prevede di ricavare i parametri caratteristici per la costruzione della curva di crescita, da alcune mappe di regionalizzazione, riportate in [Appendice G: mappe regionalizzate per il Kriging].

Per ricavare il set di dati da utilizzare per la costruzione della curva di crescita più idonea del sito stesso, è necessario applicare la seguente formula (Jenkinson, 1955) :

$$x_{GEV} = 1 - \frac{\theta_2^*}{\theta_3}\left[\Gamma(1-\theta_3) - \left(-\ln\left(\frac{T-1}{T}\right)\right)^{-\theta_3}\right] \quad 17$$

Dove:

- θ_2^* indica il parametro di scala, normalizzato rispetto al valore medio;
- θ_3 rappresenta il parametro di forma, adimensionale, che deve necessariamente essere cambiato di segno, in seguito all'estrazione dalla mappa raster rappresentativa;
- $\Gamma(1-\theta_3)$ è la distribuzione di probabilità Gamma, applicata per la stima del valore del quantile;

- T è il periodo di ritorno.

I parametri caratteristici di ogni sito, saranno estratti dalle mappe raster, in funzione delle coordinate della località: per ognuna di esse si otterranno dei valori differenti dei parametri. I parametri estratti per il sito di Torino, sono riassunti in [Tabella 4. 5].

Tabella 4. 5: parametri caratteristici del metodo Kriging estratti dalle mappe regionalizzate

Località	$\boldsymbol{\theta_2^*}$	$\boldsymbol{\theta_3}$
Torino	0.27755	-0.04354

I grafici rappresentati in [Figura 4. 14], mostrano il confronto tra la distribuzione costruita in funzione dei parametri carattestici del sito scelto, e le curve di frequenza empirica. In entrambi i casi si osserva come l'andamento delle tre serie di valori risulti adeguato: per periodi di ritorno elevati, si osserva, nuovamente, una lieve sovrastima del quantile normalizzato rispetto alla curva 'station year'.

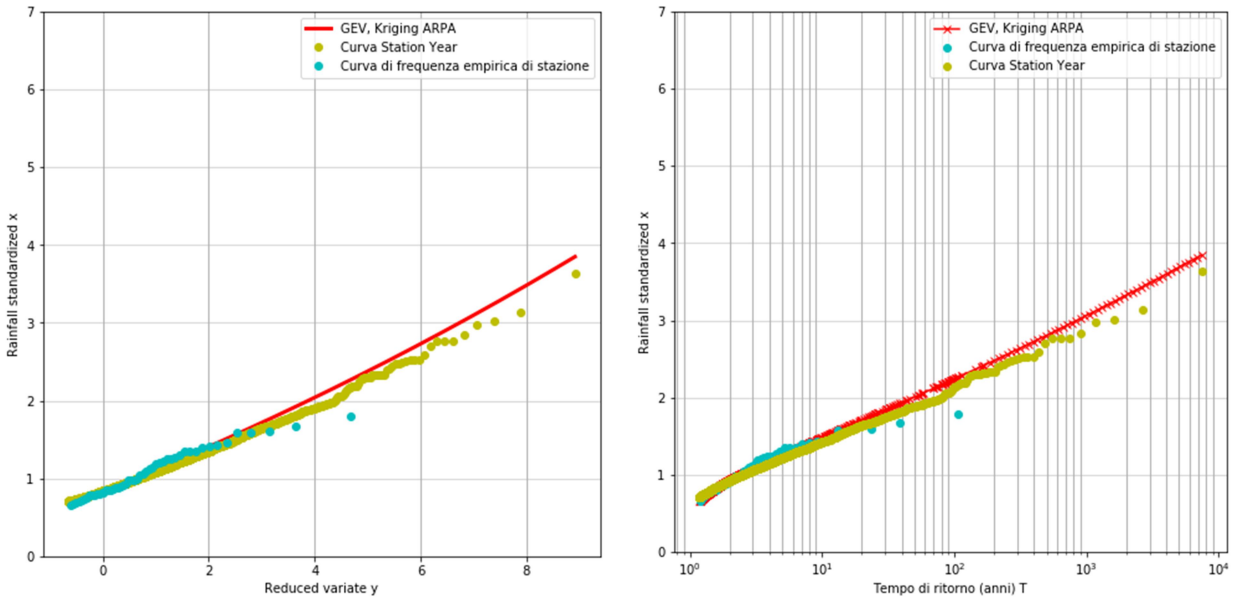


Figura 4. 14: rappresentazione delle distribuzioni GEV ottenute con il metodo Kriging, a confronto con le curve di frequenza empirica, sia in funzione della y che del T, nel sito di Torino.

4.3.3 *Confronto tra le differenti curve di crescita*

La funzione analitica, che si adatta alla spezzata ottenuta, rappresenta la curva di crescita per il sito oggetto dell'analisi. Da tale funzione, si potranno ricavare i quantili normalizzati dell'insieme di dati, così da stimare successivamente la precipitazione di progetto. La curva di crescita illustra, come in funzione del periodo di ritorno, l'andamento di una data quantità di precipitazione normalizzata, aumenti. Come tale, è utile costruire queste curve come il più possibile rappresentative del set di dati analizzato.

La modalità di utilizzo di tale curva avviene stimando il valore del periodo di ritorno al quale si desidera eseguire l'analisi. Infatti, in base a tale elemento, sarà possibile ricavare il valore della variabile ridotta associata e successivamente il fattore di crescita. La funzione analitica, con il migliore adattamento alla spezzata oggetto del FORGEX, consente all'utente di ricavare il valore del quantile ricercato, inserendo nella funzione, il periodo di ritorno o la variabile ridotta ad esso associata.

Il set di dati ottenuto con il metodo FORGEX è un gruppo di valori misto, ottenuto da diverse stazioni: risalire all'esatto quantile, sulla base della funzione a tratti, potrebbe essere piuttosto dispendioso e poco preciso. Per tale motivo, si è scelto di utilizzare una funzione analitica, in modo da rendere le operazioni di stima dei quantili, più rapide e affidabili possibile.

Nella [Figura 4. 12], si osserva la distribuzione GEV, ottenuta con il metodo FORGEX, rappresentata in funzione della variabile ridotta y e del tempo di ritorno T. Tale curva verrà confrontata con le curve ottenute come risultato dei due metodi analizzati: il metodo VAPI e il metodo Kriging.

La [Figura 4. 15] mostra la differenza tra queste tre curve principali: si osserva come la distribuzione GEV, ottenuta con i parametri del metodo FORGEX, tenda a divergere verso l'alto, separandosi, per tempi di ritorno molto elevati, dalle altre due distribuzioni. All'interno di tale grafico si osservano le curve di frequenza empirica, che vengono utilizzate per la verifica di attendibilità del metodo stesso. Come già accennato, la distribuzione, ottenuta con il metodo inglese, produce un risultato molto simile alle altre curve, per periodi di ritorno bassi. All'aumentare della variabile ridotta, dunque del

periodo di ritorno, si osserva come le tre curve tendano ad allargarsi e a divergere verso direzioni differenti.

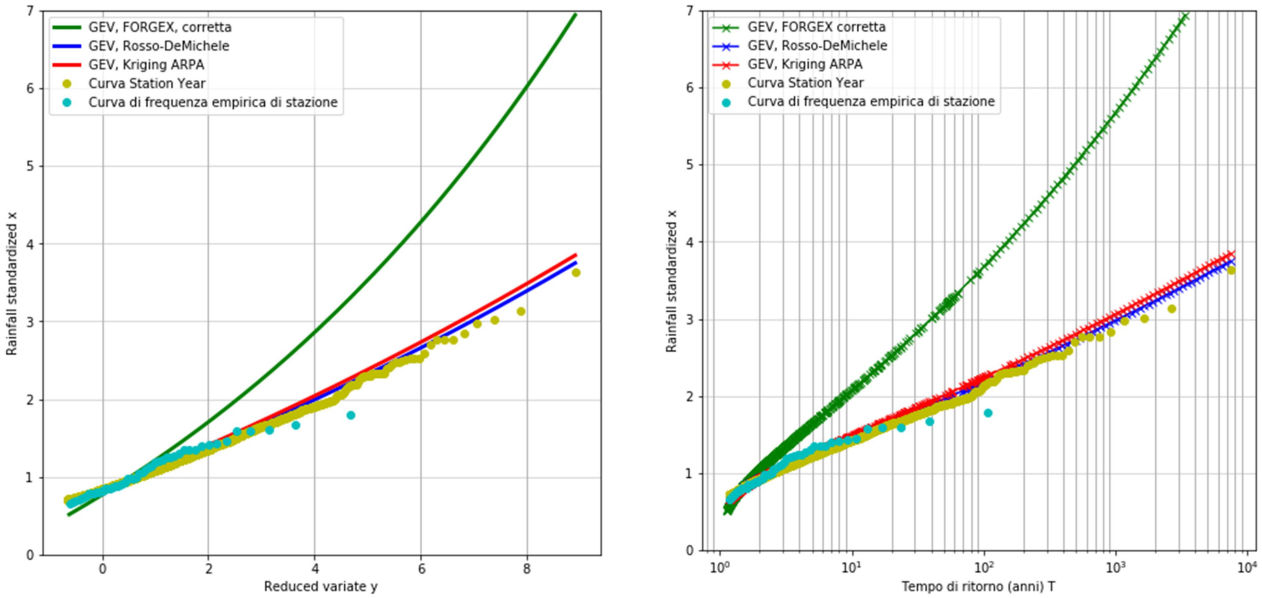


Figura 4. 15: rappresentazione delle distribuzioni GEV a confronto, in base ai valori di y e dei tempi di ritorno, nel sito di Torino.

È possibile, dalla [Figura 4. 15], ricavare il valore del quantile in funzione della variabile ridotta oppure del periodo di ritorno ad essa associato. Per evidenziare le differenze tra i tre metodi, sono stati ricavati i valori dei quantili per tre periodi di ritorno caratteristici, riassunti in [Tabella 4. 6].

Tabella 4. 6: valori dei quantili ottenuti dalle tre differenti curve di crescita.

Variabile ridotta y	Tempo di ritorno T (anni)	Valore quantile FORGEX	Valore quantile VAPI	Valore quantile KRIGING
3.9	50	3.13	1.95	2.03
5.3	200	4.23	2.41	2.48
7	1000	5.77	3.01	3.06

Come si osserva dalla [Tabella 4. 6], i valori dei quantili, ottenuti con il metodo FORGEX, appaiono molto più elevati rispetto alle altre due curve: per tempi di ritorno superiori a 100 anni, la curva tende a produrre risultati che sono il doppio dei quantili ottenuti con gli altri metodi. Questo fenomeno potrebbe essere legato alla modalità di ottimizzazione della distribuzione rispetto alla funzione a tratti: la sovrastima dei quantili per periodi di

ritorno molto elevati, potrebbe essere correlata alla scarsità di eventi rari, tali da poter essere rappresentati.

4.3.4 *Confronto tra le differenti curve di possibilità pluviometrica*

L'obiettivo di un'analisi regionale è la stima della precipitazione di progetto in aree prive di osservazioni. Per tale motivo, è necessario ottenere la relazione analitica che lega le variabili ridotte, associate ad un determinato tempo di ritorno, alle altezze di precipitazione, in modo da ricavare il valore di progetto desiderato. La curva che mette in relazione queste due grandezze prende il nome di curva di possibilità pluviometrica, e assume la seguente forma:

$$h_D = \bar{h} * K(T) \quad 18$$

Dove:

- h_D indica la quantità di progetto da stimare;
- $\bar{h}$ rappresenta la variabile indice, che nell'equazione viene identificata come il valore medio, ma potrebbe essere associata anche alla mediana del sito scelto, come avviene in Inghilterra;
- $K(T)$ rappresenta il fattore di crescita, ottenuto in funzione del tempo di ritorno.

Per quanto riguarda i metodi applicati in Italia, è noto il valore indice regionalizzato sul territorio piemontese, pari al valore medio. La procedura inglese, prevede di normalizzare la serie di dati in funzione del valore mediano, in quanto la regionalizzazione di tale parametro è stata eseguita sull'intero territorio.

Per poter garantire un confronto diretto tra le differenti curve, si è applicata la correzione ai dati italiani analizzati con il metodo inglese, in modo da poter utilizzare il valore medio della serie di dati, come valore indice.

In [Figura 4. 16], è rappresentata la curva di possibilità pluviometrica CPP, per il sito di Torino: i valori delle altezze di pioggia ottenuti per tempi di ritorno caratteristici sono riassunti in [Tabella 4. 7].

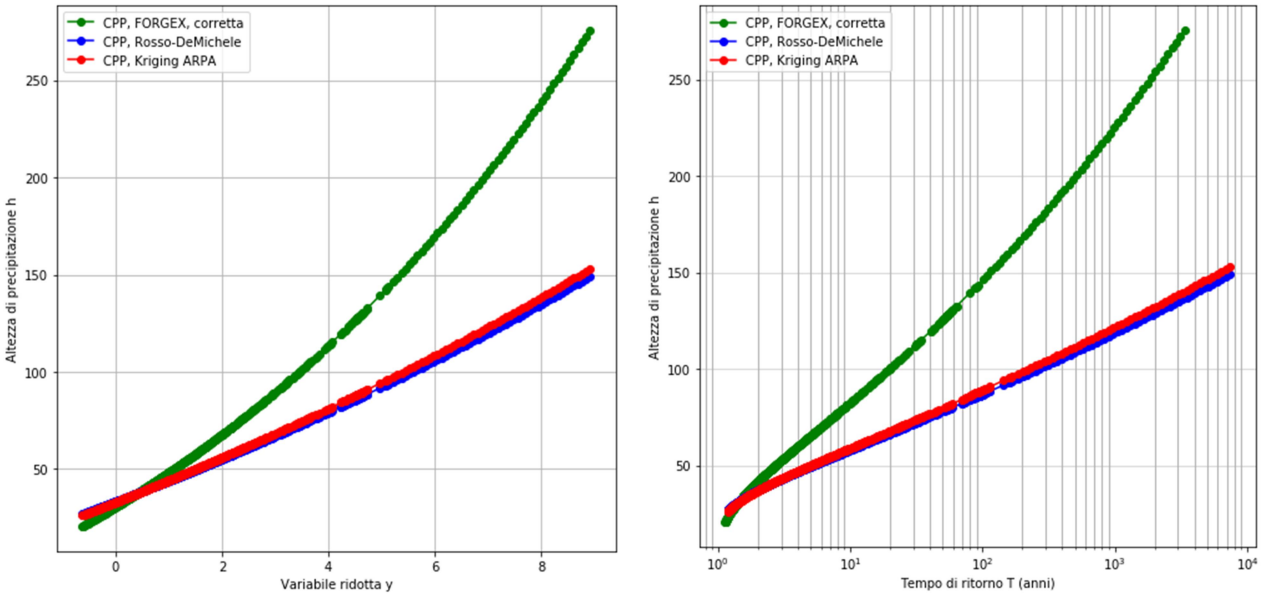


Figura 4. 16: rappresentazione delle CPP per i tre metodi differenti, in funzione della y e del tempo di ritorno, nel sito di Torino.

Come si osserva in [Tabella 4. 7], i valori delle altezze di pioggia, nel sito di Torino, per una durata di 3 ore, ottenute con il metodo FORGEX tendono ad essere molto più elevati rispetto a quelli ottenuti con gli altri metodi, specialmente per valori del tempo di ritorno molto alti.

Tabella 4. 7: valori delle altezze di precipitazione ottenute per le differenti curve di possibilità pluviometrica

Variabile ridotta y	Tempo di ritorno T (anni)	Altezza di pioggia FORGEX	Altezza di pioggia VAPI	Altezza di pioggia KRIGING
3.9	50	124.34	77.46	79.64
5.3	200	168.04	95.74	98.52
7	1000	229.22	119.57	121.56

4.4 *Particolarità del metodo FORGEX*

L'analisi regionale eseguita con il metodo FORGEX è stata applicata in differenti zone del Piemonte, per analizzare il comportamento del metodo. In [Appendice F: Schede delle località analizzate], sono presenti alcune schede illustrative dei risultati ottenuti nelle località principali, per le differenti durate. La caratteristica principale del metodo FORGEX è la selezione dei differenti network: in base all'ampiezza dell'ultimo cerchio costruito, si potranno osservare risultati differenti.

Per qualche località si sono presentati alcuni problemi, durante la selezione delle stazioni. Una tra queste è Bardonecchia, per la quale è stato necessario creare gruppi di stazioni più ampi, rispetto a quelli realizzati in altre zone. Come si osserva dalla [Figura 4. 17], il cerchio nell'intorno del sito scelto si presenta con un raggio di circa 145 km: questa situazione si verifica poiché l'area da analizzare si trova nei pressi dei bordi della regione. Nei dintorni del sito selezionato non sono presenti grandi quantità di stazioni, tali da garantire un'analisi adeguata. Il numero di stazioni considerate per questa analisi è pari a 240.

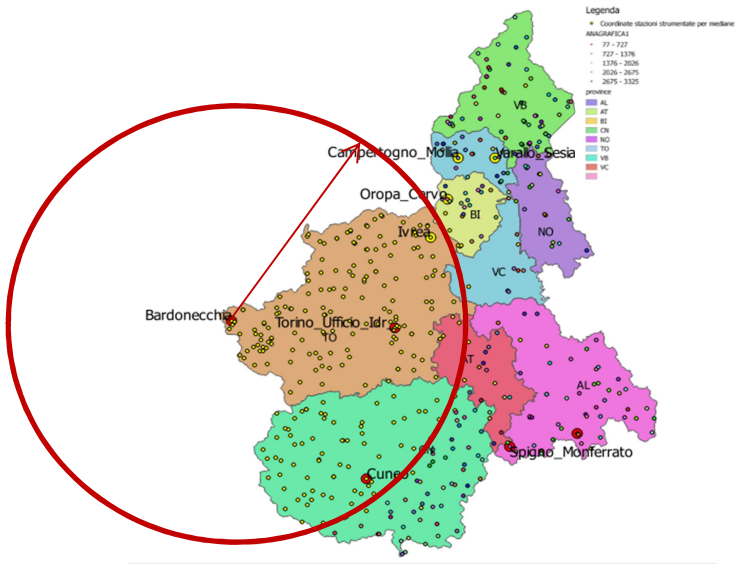


Figura 4. 17: rappresentazione della regione Piemonte, con il network più ampio nell'intorno di Bardonecchia

I risultati a cui si è giunti con questo approccio sono molto differenti rispetto a quanto atteso: si osserva dalla [Figura 4. 18], che la curva ottenuta con il metodo Kriging ha un andamento molto simile a quella ottenuta con il metodo FORGEX. Questo aspetto ha posto numerosi dubbi legati all'ampiezza del network: la curva ottenuta con il Kriging, rispetto a ciò che si presenta in altre località, appare divergente verso l'alto.

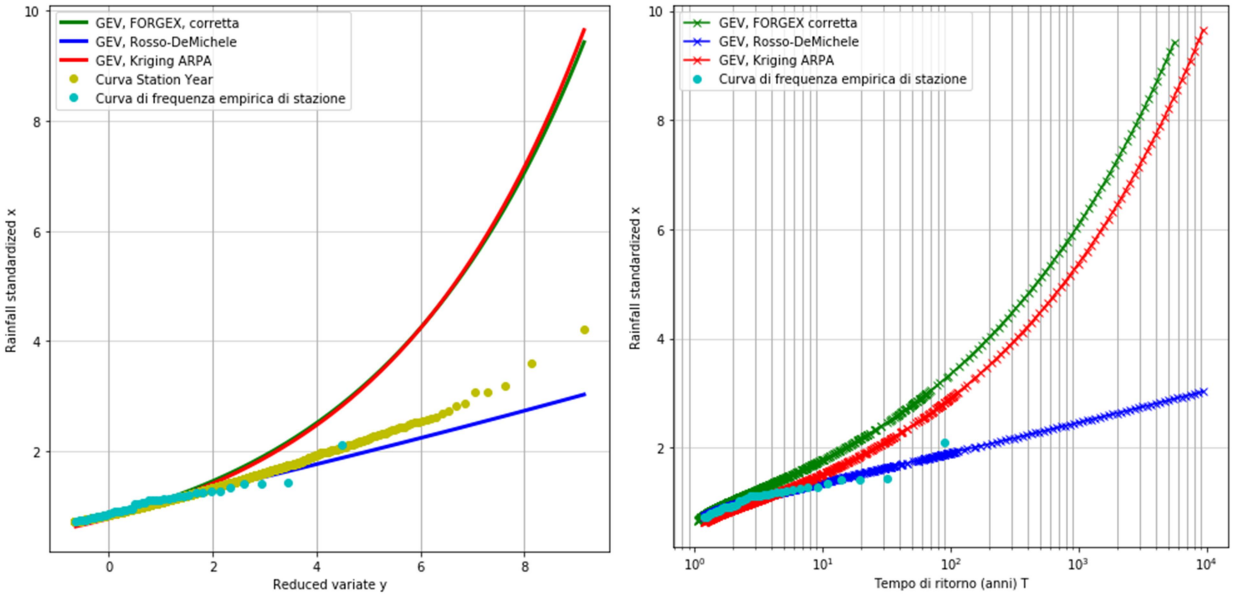


Figura 4. 18: rappresentazione delle tre curve di crescita ottenute con le differenti metodologie per il sito di Bardonecchia.

Questo comportamento nel sito di Bardonecchia per la durata di 6 ore, ha scaturito una certa curiosità, a tal punto da voler verificare se il problema alla base di tale divergenza, fosse legato alla numerosità della serie, o alla disposizione delle stazioni sul territorio. A tal proposito, sono state eseguite alcune analisi, utilizzando una banca dati virtuale: le stazioni, infatti, si trovano a 250 m di distanza una dall'altra, con una lunghezza delle serie più elevata.

I risultati raggiunti hanno corretto il comportamento della curva, ottenuta con il metodo FORGEX, ma non hanno risolto il problema legato alla curva costruita con i parametri del Kriging. Incrementando il numero di stazioni e di dati da analizzare, si osserva come l'andamento della curva FORGEX, sia conforme a ciò che si ottiene nelle altre località in cui si è applicata la procedura. Si nota comunque una sovrastima dei quantili nella coda della distribuzione, ma per valori di periodi di ritorno molto piccoli, il metodo fornisce risultati coerenti con la procedura VAPI.

Un'altra località nella quale sono stati ottenuti risultati non attesi, è il sito di Spigno Monferrato. Come si osserva dalla [Figura 4. 19], l'area in esame si trova nei bordi della regione Piemonte. Come tale, la determinazione dei differenti network poteva creare qualche scompenso nell'analisi: i risultati ottenuti si sono rivelati come intermedi tra ciò che succede ponendosi nel centro della regione Piemonte, e ciò che accade all'estremo occidentale della regione, Bardonecchia. L'ampiezza del network più ampio costruito nell'intorno di Spigno Monferrato, è pari a 56 km, con un numero di stazioni raccolte pari a 80. Rispetto a ciò che accade per il sito di Bardonecchia, l'ampiezza dell'ultimo network appare molto ridotta: questo aspetto potrebbe essere legato alla differente localizzazione delle stazioni nell'intorno desiderato, e alla lunghezza delle differenti serie di dati.

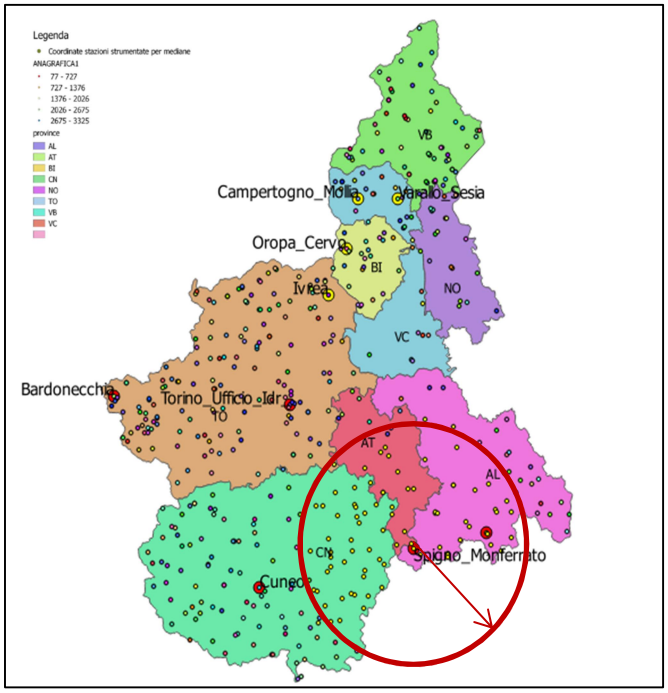


Figura 4. 19: localizzazione del sito di Spigno Monferrato e rappresentazione del network più ampio nell'intorno del sito stesso.

Come si osserva in [Figura 4. 20], la curva di crescita ottenuta con l'approccio empirico si presenta molto similare alle curve parametriche. La stima del quantile della curva FORGEX, per un tempo di ritorno pari a 200 anni, non appare eccessiva rispetto alla curva VAPI o alla curva Kriging. Questa lieve sovrastima si osserva esclusivamente per durate inferiori alle 6 ore: infatti per durate di analisi superiori a tale valore, il risultato ottenuto nell'intorno del sito di Spigno Monferrato, produce nella curva un andamento divergente verso l'alto, molto simile a ciò che si presenta in località come Torino. Inoltre, come si osserva in [Figura 4. 20], la curva di crescita ottenuta con il metodo FORGEX, tende a garantire una sovrastima del quantile rispetto alla curva 'station year', a dimostrazione della verifica dell'attendibilità del metodo nel sito analizzato.

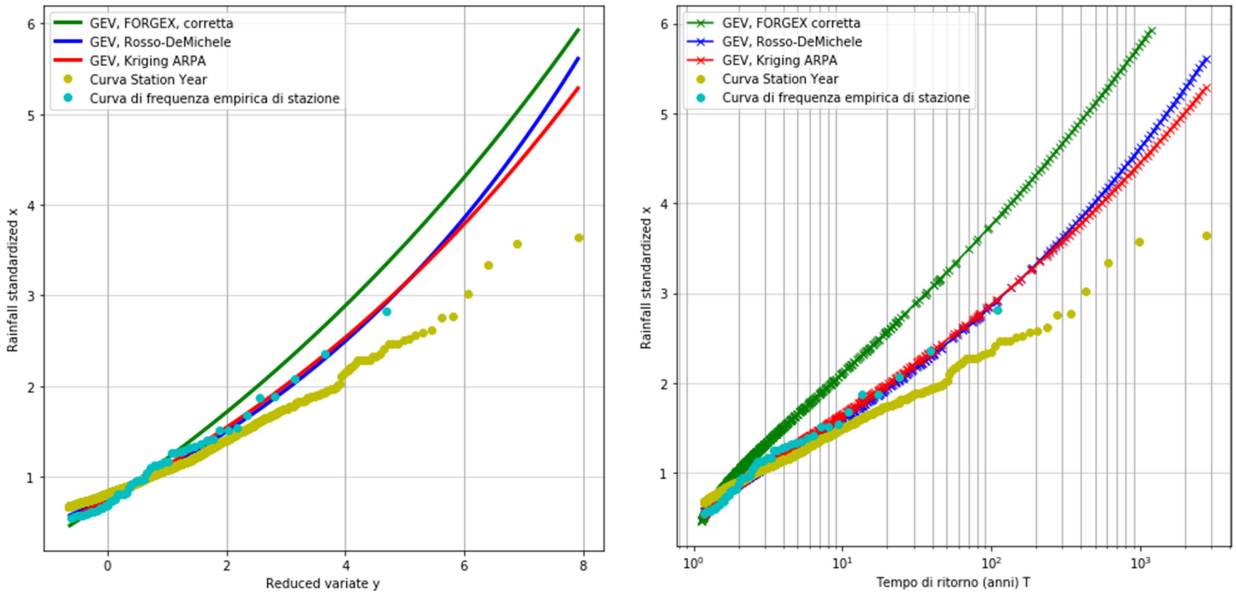


Figura 4. 20:rappresentazione delle tre curve di crescita ottenute con le differenti metodologie per il sito di Spigno Monferrato.

Capitolo 5

Costruzione di un sistema informatizzato per la stima della curva di crescita con il metodo FORGEX

L'analisi regionale in Inghilterra, per garantire stime per tempi di ritorno molto elevati, viene supportata dal software FORGEX. A causa della scarsa diffusione di tale metodologia, è stato costruito un applicativo che permettesse l'analisi dei dati attraverso questa tecnica. In tal modo, è stato possibile verificare il grado di affidabilità del metodo e la sua predisposizione all'utilizzo.

La procedura provvede ad una elaborazione dei dati, generando una curva di crescita. La caratteristica principale di tale curva è la costruzione a tratti, che non garantisce la linearità del set di dati. Questa metodologia analitica, non convenzionale, produce un output di facile successo, ma non confrontabile con le altre tecniche. Per tale motivo, a valle della costruzione della curva stimata come una funzione a tratti, sono stati eseguiti alcuni accorgimenti al metodo originale, per rendere i risultati confrontabili con le metodologie di analisi regionale più comuni.

L'applicativo eseguito può essere suddiviso in cinque macro-blocchi, in modo da rendere lo sviluppo e successivamente la lettura, divisa per operazioni. È indispensabile identificare, per ogni blocco di codice, i dati di input, l'output e il loro formato.

Nel corso della costruzione della procedura, sono state utilizzate alcune funzioni importate da librerie presenti nell'ambiente integrato di Python. Per alcune operazioni, più mirate, non sono state riscontrate funzioni esistenti, ma si è cercato di crearne alcune, per velocizzare le operazioni da svolgere. La definizione delle funzioni caratteristiche viene eseguita prima dell'applicativo, immediatamente di seguito all'importazione delle librerie principalmente utilizzate nel codice.

5.1 *Identificazione dei dati*

5.1.1 *Analisi dei dati*

Nella [Figura 5. 1] è rappresentato lo schema del primo blocco di codice costruito, per l'analisi regionale.

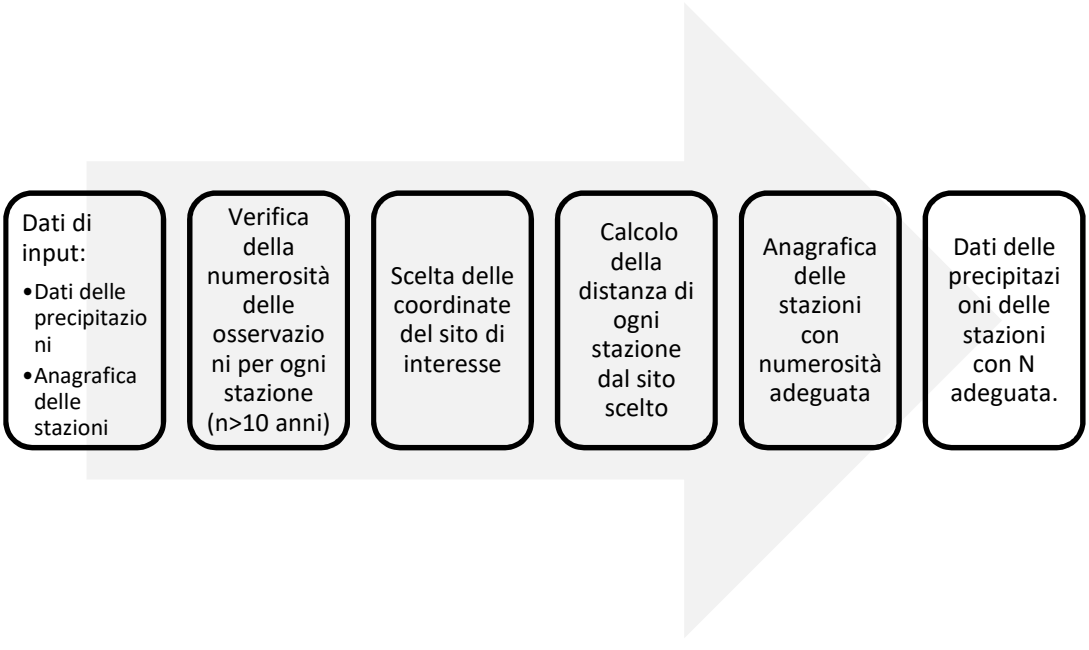


Figura 5. 1:Diagramma di flusso del blocco di codice iniziale

5.1.1.1 *Gestione dei dati*

La prima fase di un'analisi regionale è selezionare i dati necessari per il suo svolgimento. A tal proposito, è importante fornire all'applicativo due file di dati contenenti sia le informazioni relative alle stazioni da analizzare, sia i dati delle precipitazioni per le durate caratteristiche. In questo modo, i due file saranno correttamente inseriti e processati nel corso della procedura.

È molto importante che il formato dei file sia esattamente quello richiesto dall'applicazione: si utilizza un formato Excel, per la serie di informazioni delle diverse stazioni, mentre per i dati delle precipitazioni si predilige un file Excel con dati in formato virgola mobile. Questa scelta deve essere rispettata, in modo da evitare un messaggio di errore, per mancata correttezza nel formato dei file inseriti.

Il programma utilizza la libreria Pandas come supporto iniziale. Per inserire i dati all'interno di una matrice multidimensionale, è necessario, infatti, conoscere la numerosità delle osservazioni a disposizione. Per eseguire questa operazione, il file verrà inizialmente importato in un 'DataFrame', che permette di definire l'esatta dimensione da utilizzare per costruire la matrice.

Il punto cruciale, risulta la parametrizzazione di tale dimensione, in modo da slegare completamente l'utilizzo di un dato file dall'applicativo: in tal modo, modificando il file di input, il programma riuscirà ad adattare il file inserito in una matrice adeguata. I file importati per l'utilizzo del programma saranno organizzati nel modo seguente:

Tabella 5. 1: intestazione del file iniziale con le informazioni delle stazioni

Nome stazione	Numero stazione	UTM x	UTM y	Quota

Tabella 5. 2: intestazione del file iniziale dei dati

Numero stazione	Anno	Max1h	Max3h	Max6h	Max12h	Max24h

5.1.1.2 *Verifica della validità dei dati*

In ogni database, esiste una serie di dati che non può essere considerata per lo scopo dell'analisi: spesso questo è legato alla scarsa numerosità del set, oppure alla presenza di valori non utilizzabili. Il metodo FORGEX prevede un controllo a priori, in modo da evitare di procedere con l'analisi considerando stazioni, i cui dati sono

inutilizzabili. Seconda la procedura inglese, il numero di osservazioni per ogni stazione deve essere superiore a 10 anni, in modo da garantire stabilità alla serie stessa.

Per questo tipo di operazione non è presente una funzione in Python, in grado di svolgere esattamente la selezione. Per tale motivo, è stata costruita da zero una funzione utile alla verifica della numerosità dei dati. Tale elemento, è stato creato per controllare il numero di dati per durata, associati ad una stazione: il suo sviluppo avviene con il supporto della libreria Pandas, utilizzata per il conteggio dei valori. Questa scelta è motivata da una velocizzazione del processo, che garantisce un risultato ottimale, e con una semplice riga di codice, come si osserva in [Appendice E: Codice delle funzioni costruite per la realizzazione dell'applicativo].

Questa funzione, oltre a verificare la numerosità del set di dati in esame, controlla la presenza di valori inesistenti, o che non sono stati riportati: nel database questo genere di dati viene rappresentato con la sigla 'nan'. Effettuando un controllo generale sull'intero set di dati, il programma è in grado di definire quali elementi non possono proseguire l'analisi, associandogli un valore pari a '-1', in modo da identificarli velocemente ed eliminarli.

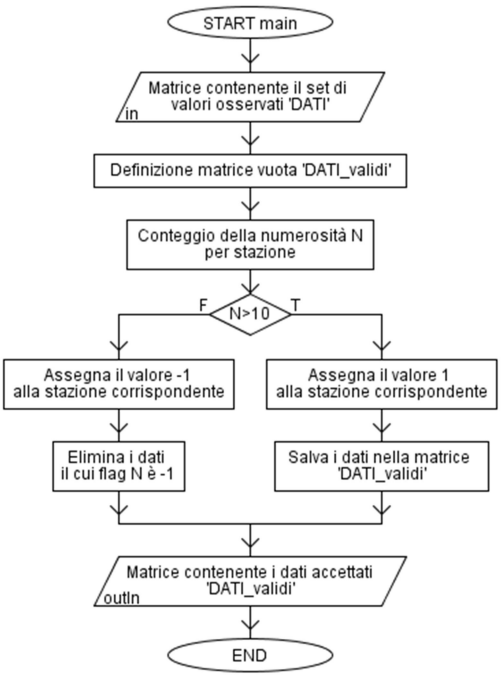


Figura 5. 2: Schema gerarchico per la funzione 'numerosità'

Come si osserva dalla [Figura 5. 2], la funzione potrà fornire due caratteri che permettono al programma di classificare quali siano le stazioni idonee all'analisi, eliminando le altre.

In tal modo sarà eseguita una prima classificazione per verificare quali dati possano o meno, essere utilizzati per l'analisi regionale: le stazioni che presenteranno valori di numerosità inferiore al limite consentito, non proseguiranno l'elaborazione, in quanto saranno escluse dal set di dati principali.

5.1.1.3 *Localizzazione del dato*

L'analisi regionale si svolge nei dintorni di un punto, denominato 'focal point', che viene individuato attraverso due coordinate cartografiche nel sistema Universale Trasverso di Mercatore, detto UTM. Queste due coordinate sono inserite dall'utente, e possono essere casuali, ovvero non corrispondenti ad una stazione precisa, in modo da garantire il corretto svolgimento dell'analisi regionale. Metodi come la regionalizzazione delle precipitazioni, prevedono che il sito scelto come 'focal point', sia non strumentato. Lo scopo è costruire una serie di dati, sulla base delle informazioni presenti nei dintorni della zona selezionata.

Nell'applicazione creata, si fa riferimento ad un insieme di stazioni, poste nella regione Piemonte. Si osserva che nelle zone nei pressi dei bordi della regione, dove non si hanno dati provenienti dalla zona francese, o dalla Lombardia, l'analisi potrà essere eseguita solo in parte. In corrispondenza dei limiti della regione, il cerchio intorno al 'focal point' dovrà svilupparsi esclusivamente da una parte, considerando stazioni poste a distanze molto elevate dal centro. In tali casi, il cerchio sarà molto ampio, e potrebbe non garantire credibilità ai risultati.

5.1.1.4 *Individuazione delle serie di dati*

Le stazioni che non hanno superato il limite di almeno 10 anni di osservazioni, verranno eliminate dal database generale e non saranno considerate per l'analisi regionale. Le rimanenti, invece, dovranno essere valutate e ordinate in base al criterio chiave dell'intera metodologia: per ogni stazione sarà necessario calcolare la distanza dal sito di interesse, identificato come origine di un ipotetico sistema di coordinate.

Il valore della distanza da ogni stazione è utilizzato come cartina di tornasole per la verifica dell'omogeneità del network in esame. Gli eventi, che si verificano a maggiore distanza dal sito di interesse, non influenzeranno l'analisi come se si trovassero nelle immediate vicinanze di quest'ultimo. Il calcolo della distanza è di notevole importanza e affidabilità per l'analisi con il metodo FORGEX: garantisce un grado di uniformità e coerenza tra le differenti stazioni selezionate.

In seguito alla valutazione della distanza, il primo blocco di codice produrrà in output due importanti risultati:

1) Il file 'anagrafica', contenente non solo più le coordinate e la quota delle singole stazioni, ma anche la loro distanza dal punto centrale dell'analisi. L'intestazione è mostrata nella [Tabella 5. 3]:

Tabella 5. 3: Intestazione del file 'anagrafica'.

Codice	Nome	UTM x	UTM y	Quota	Distanza

2) Il file 'precipitazioni', che in seguito ad un'intersezione con il file 'anagrafica', riporta i dati per le cinque durate caratteristiche, associati alle stazioni che hanno superato il requisito di validità. L'intestazione del file è raffigurata in [Tabella 5. 4]:

Tabella 5. 4: Intestazione del file rappresentativo delle precipitazioni.

Codice	Anno	Max1h	Max3h	Max6h	Max12h	Max24h

Qualunque tipologia di file può subire una prima analisi per garantire una corretta regionalizzazione secondo il metodo FORGEX: è assolutamente indispensabile, per questa prima procedura convertire i file nei formati corretti.

5.1.2 *Normalizzazione dei dati di precipitazione*

Lo schema in [Figura 5. 3], rappresenta le fasi delle operazioni che vengono eseguite in questo secondo blocco di codice.

Figura 5. 3: Diagramma di flusso del codice relativo alla normalizzazione delle serie di dati.

5.1.2.1 *Scelta della durata caratteristica di analisi*

Come si nota dalla [Figura 5. 3], il primo passaggio di questo insieme di operazioni è la scelta di una durata caratteristica. A tale proposito viene costruita una funzione, denominata 'test', il cui compito è verificare che l'utente inserisca o meno un valore corretto della durata di analisi. La [Figura 5. 4] racchiude il funzionamento della funzione 'test ', il cui codice è riportato in [Appendice E: Codice delle funzioni costruite per la realizzazione dell'applicativo].

Il processo di scelta della durata è affidato all'utente. La funzione 'test' ha il compito di confrontare il valore inserito con un vettore contenente le cinque durate possibili. Possono presentarsi due casi differenti:

1) L'utente inserisce il valore corretto di durata, quindi viene estratto il set di dati corrispondente a quel valore;
2) L'utente inserisce un valore assente nel set di dati di confronto, dunque si ottiene un messaggio di errore, che permetterà all'utente di capire quali sono le durate per le quali l'analisi può continuare. Successivamente verrà richiesto di inserire nuovamente il valore della durata.

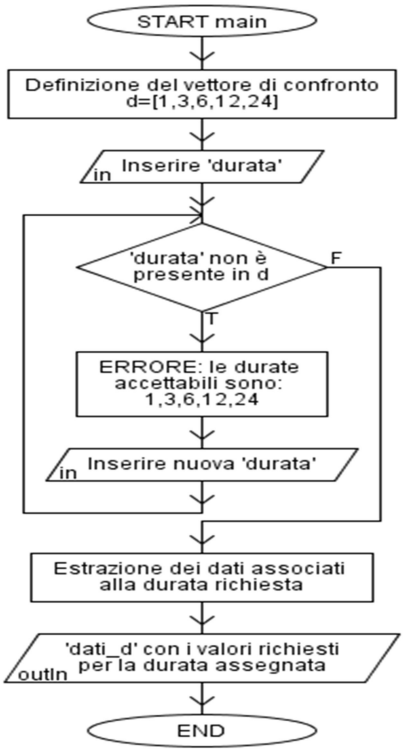


Figura 5. 4: Schema gerarchico della funzione 'test' eseguito sulla durata.

La scelta della durata caratteristica è un passaggio importante dell'analisi, in quanto estrae il set di dati adeguato, che si desidera analizzare. Attraverso la funzione di [Figura 5. 4], la matrice dei dati delle precipitazioni viene alleggerita, poiché riporterà esclusivamente la colonna di dati, per la durata richiesta dall'utente.

5.1.2.2 *Scelta del parametro di riferimento e normalizzazione*

In seguito all'identificazione dei dati su cui eseguire l'analisi, sarà necessario definire il parametro di riferimento, rispetto al quale verranno normalizzati. Il processo di normalizzazione, secondo quanto scritto nell'articolo di Reed , Faulkner, & Stewart, del 1999, è di fondamentale importanza, in quanto consente all'intero set di dati di poter essere confrontato in modo uniforme. Il parametro di riferimento utilizzato in Inghilterra è il valore mediano, poiché meno influenzabile dagli estremi della serie.

La normalizzazione non avviene rispetto ad un singolo valore mediano, ma si otterrà per ogni stazione un valore differente. Questa procedura è piuttosto insolita: solitamente, si predilige l'utilizzo di una serie unica crescente, come per l'approccio 'station-year'.

Il valore della mediana per ogni singola serie, non verrà riportato nella tabella riassuntiva dei dati normalizzati. Tale valore sarà utilizzato esclusivamente nei calcoli e si riporterà, nell'ultima colonna della matrice, il valore ottenuto dalla normalizzazione, come appare in [Tabella 5. 5].

Tabella 5. 5: Intestazione del file output della seconda routine di codice.

Codice	Anno	Max D ore	Val. Normalizzato

5.1.3 *Relazione tra i dati e il periodo di ritorno*

Questo terzo blocco di codice può essere eseguito contemporaneamente al secondo, poichè entrambi considerano i dati delle precipitazioni delle stazioni soddisfacenti i requisiti iniziali. In [Figura 5. 5] è rappresentato lo schema per questa parte del codice.

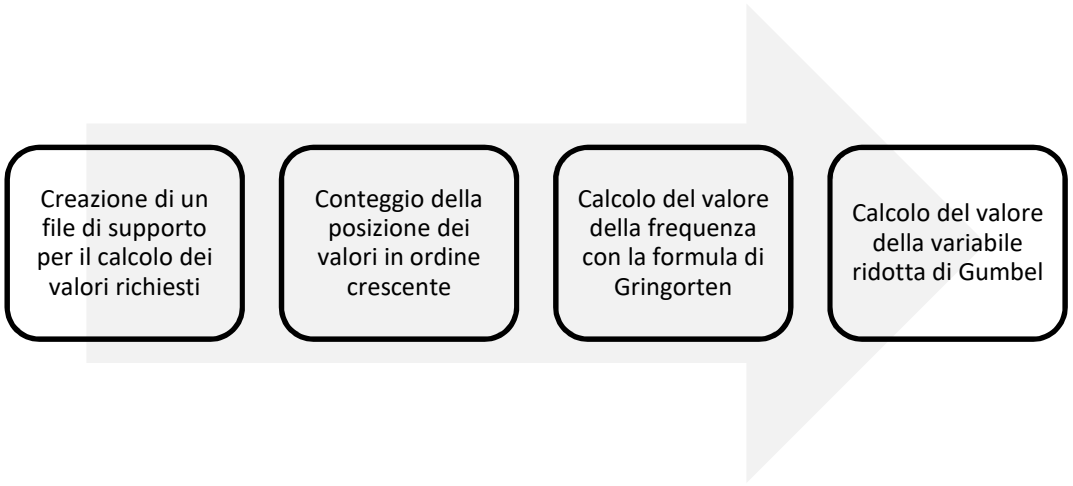


Figura 5. 5: Diagramma di flusso della terza routine per il calcolo della frequenza.

5.1.3.1 *Plotting position del set di dati*

L'organizzazione di una matrice, contenente i valori della frequenza associata ad ogni evento, si basa sulla lunghezza delle serie appartenenti ad ogni stazione. Il calcolo della frequenza avviene attraverso la funzione di Gringorten, espressa dalla formula (5), e viene eseguito sulle differenti serie di valori, una per ogni stazione. Associata ad ogni valore di frequenza, sarà necessario calcolare il valore della variabile ridotta corrispondente, con la formula (4).

Il file ottenuto, a seguito della procedura descritta dalla [Figura 5. 5], conterrà i valori delle variabili ridotte di Gumbel calcolati per ogni stazione. Si presenta una serie di dati con valore inferiore a -0.6335, corrispondente al valore mediano normalizzato, sottratto di un'unità. Al di sotto di tale valore si trovano alcuni dati, i cui tempi di ritorno non sono richiesti nell'analisi regionale.

Tali dati non vengono considerati e si possono eliminare dalla matrice generale, alleggerendone la dimensione per le operazioni successive. Il file ottenuto avrà la seguente organizzazione:

Tabella 5. 6: Intestazione del file che contiene i valori della variabile ridotta y, denominato 'valori_y'.

Codice stazione	*Anno*	*Posizione i*	*Frequenza*	*Var. ridotta*

A tal punto, l'applicativo può procedere secondo il metodo FORGEX, selezionando i differenti gruppi di stazioni nell'intorno del 'focal point'.

5.2 *Classificazione del set di dati*

5.2.1 *Metodo delle y-slices*

Alla base dell'analisi regionale con il metodo FORGEX, si osserva il metodo delle y-slices: in questo blocco di codice, si procederà con la classificazione, rispetto alla variabile y, del file ottenuto in output dalla routine precedente. Lo schema in [Figura 5. 6], riassume brevemente i passaggi che vengono eseguiti.

Figura 5. 6: Diagramma di flusso, rappresentativo della routine di classificazione del dato.

5.2.1.1 *Classificazione dei dati*

I valori delle variabili ridotte vengono classificati in modo da poter essere inseriti nella slice di riferimento corretta. Per questa operazione, è stata costruita la funzione 'classify', in [Figura 5. 7]. Tale applicativo, per ogni elemento corrispondente ad un dato intervallo, associa un numero identificativo della slice di riferimento: il criterio alla base di tale funzione è lo 'switch', consistente in una serie di casi differenti, a seconda del valore in esame.

Il numero delle slices è definito in funzione del numero di network costruiti. In tal modo, si esegue una prima classificazione, che permette di raggruppare i valori in base alla variabile ridotta, posta nell'ultima colonna della matrice in [Tabella 5. 6].

Come si osserva in [Figura 5. 7], l'insieme di valori inutilizzabili per proseguire l'analisi, viene associato ad una sigla 'NaN': in questo modo saranno più semplici da elaborare. La libreria Pandas consente di eliminare, con un'unica riga di istruzioni, i valori inaccettabili identificati con la sigla caratteristica.

La [Figura 5. 7] produce in output una matrice, contentente per ogni stazione, un numero identificativo della slices di riferimento. In [Tabella 5. 7] è riproposta l'intestazione del file in output:

Tabella 5. 7: intestazione del file ottenuto dall'applicazione della funzione 'classify'

Codice stazione	Distanza	Val_norm	Valori y	Slices

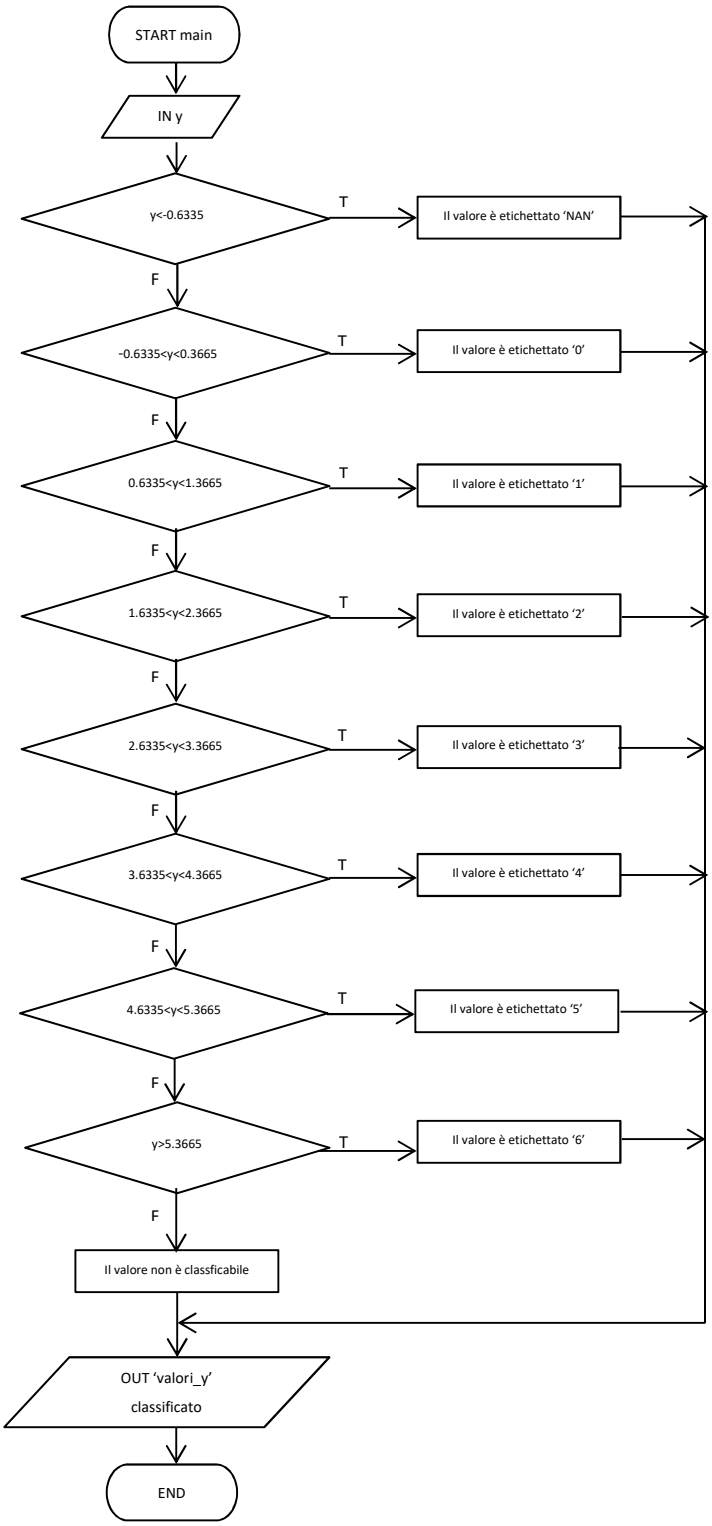


Figura 5. 7: Diagramma di flusso della funzione 'classify' per la classificazione delle variabili ridotte y.

Questa prima classificazione permette un conteggio più rapido dei valori per slices. La ricerca della numerosità per ogni slice, viene eseguita attraverso la funzione 'num_per_slice'. La [Figura 5. 8] mostra il diagramma di flusso della funzione il cui codice è riportato in [Appendice E: Codice delle funzioni costruite per la realizzazione dell'applicativo].

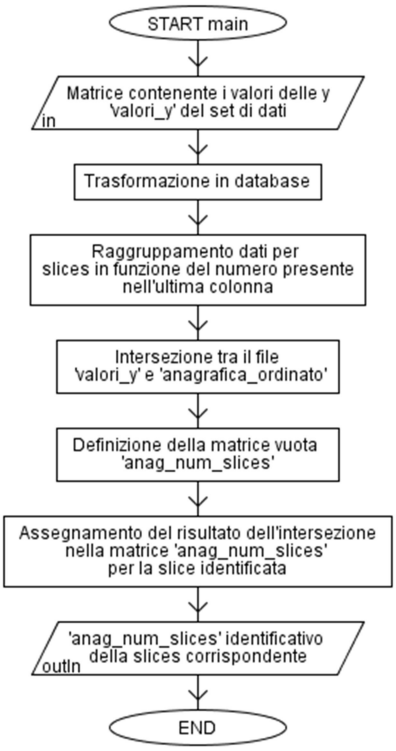


Figura 5. 8: Diagramma di flusso della funzione 'num_per_slice'.

La funzione in [Figura 5. 8] permette di ottenere per ogni slice, l'insieme di stazioni e il numero di dati corrispondente. Per ogni intervallo si otterrà una matrice, organizzata come in [Tabella 5. 8].

Tabella 5. 8: intestazione della tabella ottenuta dalla funzione 'num_per_slice'.

Codice	Distanza	Numerosità

L'obiettivo è ottenere una matrice unica, che metta in relazione il codice di ogni stazione, con la numerosità dei valori per ogni slices. In questo modo, sarà più rapido identificare le diverse stazioni appartenenti a ciascun network. La matrice in [Tabella 5. 8] sarà ordinata in funzione della distanza crescente.

Per ottenere una matrice di questo tipo, è importante che le differenti tabelle ottenute, una per ogni slices, abbiano tutte la stessa dimensione del file 'anagrafica'. Solo in questo modo sarà possibile ottenere un'unica tabella da cui estrarre le informazioni utili. Le matrici, che si ottengono con la funzione in [Figura 5. 8], avranno tutte lunghezze diverse: per tale motivo è stata costruita la funzione in grado di uniformare le dimensioni delle differenti matrici, il cui diagramma di flusso è rappresentato in [Figura 5. 9]. Tale funzione è in grado di determinare i valori di scarto tra una matrice e l'altra. In questo modo, ad ogni stazione presente nella matrice di dati, sarà associato un certo valore: nel caso in cui non siano presenti osservazioni per una determinata slices, tale valore sarà pari a zero.

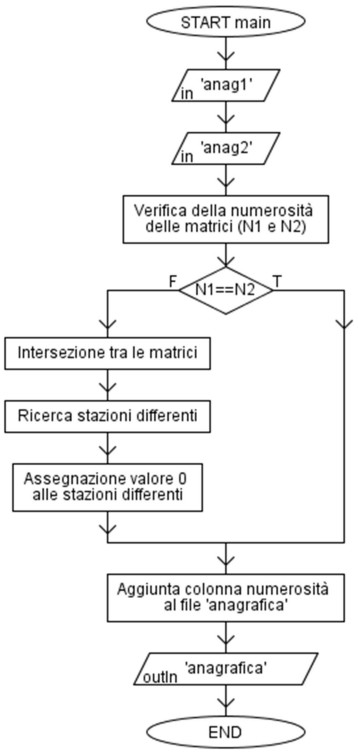


Figura 5. 9: diagramma di flusso, della funzione 'diversi', per uniformare le matrici delle numerosità.

A seguito dell'applicazione della funzione in [Figura 5. 9], si otterrà esattamente la matrice ricercata, che raggrupperà la numerosità di ogni slice per ciascuna stazione presente nei dintorni del punto scelto. Questa operazione viene eseguita per rendere la ricerca dei network il più semplice e iterativa possibile.

La matrice 'anagrafica' ottenuta in [Tabella 5. 9] sarà l'output di questo blocco di codice, che conclude la procedura di identificazione del dato, pronto per essere elaborato al fine di garantire la costruzione della curva di crescita.

Tabella 5. 9: intestazione della matrice generale 'anagrafica'.

Codice	*Distanza*	*Num1*	*Num2*	*Num3*	*Num4*	*Num5*	*Num6*

5.2.2 *Costruzione dei differenti network*

Lo schema di questo blocco di codice viene rappresentato in [Figura 5. 10].

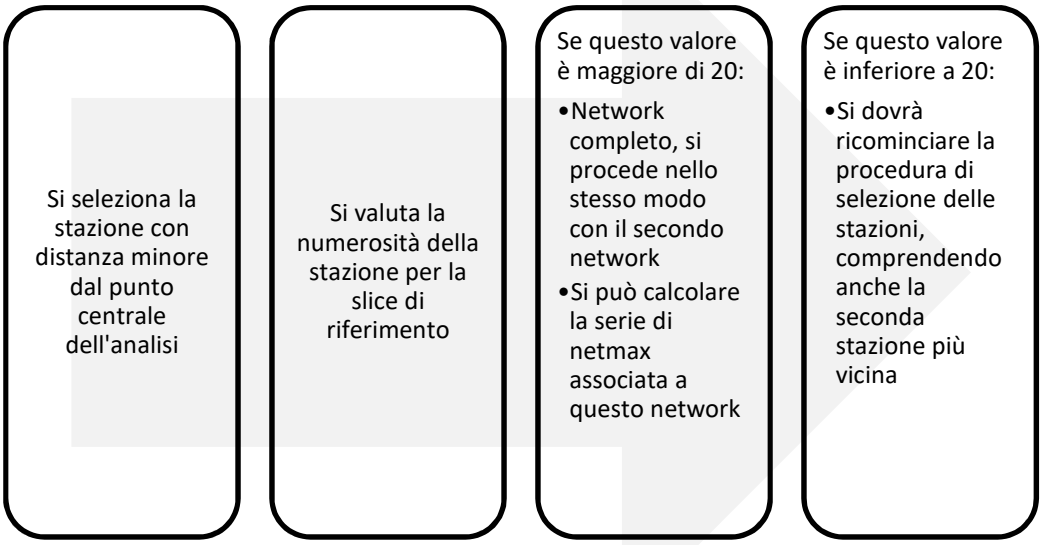


Figura 5. 10: diagramma di flusso della routine 4

Il metodo FORGEX è caratterizzato dalla creazione di network progressivi: ognuno di questi raggruppa un insieme di dati nell'intorno del 'focal point'. La caratteristica principale di ogni network è la sua definizione: ciascun insieme viene costruito in base alla numerosità delle singole stazioni, per ogni slices.

5.2.2.1 *Selezione iterativa delle stazioni*

La matrice in [Tabella 5. 9] è il punto di partenza per la definizione dei network. Infatti, al raggiungimento della numerosità minima per ogni slices, il network sarà completo e non si dovranno considerare ulteriori stazioni. Questa procedura è stata eseguita attraverso la funzione 'network', che procede in modo iterativo sul file 'anagrafica'.

Nella [Figura 5. 11] è rappresentato il diagramma di flusso della funzione 'network'. Per costruire questa funzione è stato utilizzato il ciclo 'while', che garantisce l'iterazione finché la condizione, posta tra parentesi, risulta vera. In tal modo, il ciclo procederà, fintanto che la somma delle numerosità della slices in esame, non supererà la numerosità minima, imposta pari a 20 elementi. Per un'ottimizzazione del ciclo, è stato utilizzato un flag[5], in modo da evitare la selezione delle stesse stazioni più volte.

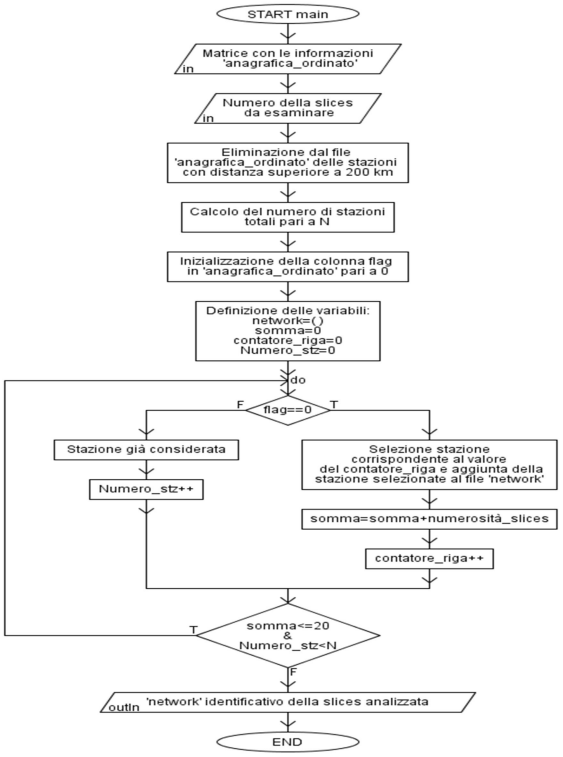


Figura 5. 11: diagramma di flusso, della funzione 'network', utilizzata per la ricerca dei dati.

[5] Un valore flag, viene identificato come un elemento che può assumere solo due valori a seconda della condizione, vera o falsa: si utilizza per separare una situazione verificata da quella ancora da controllare.

Al completamento della formazione di tutti i network, si otterranno le matrici riassuntive delle stazioni presenti in ognuno di essi. Le osservazioni relative alle stazioni selezionate, verranno ricavate intersecando ogni matrice ottenuta, con il file contenente tutti i dati delle precipitazioni. Non solo è importante definire un network per ogni slices, ma è necessario verificare che l'ultimo network costruito raggruppi tutti i precedenti. L'obiettivo finale è la costruzione di matrici con la seguente organizzazione:

Tabella 5. 10: intestazione della tabella ottenuta a seguito delle operazioni intersezione tra file.

Codice stazione	***Anno***	***MaxDh***	***Val_norm***	***y***

5.2.2.2 *Selezione delle serie di massimi annui per network*

Poiché per le stazioni a maggiore distanza non sarà possibile raggiungere la numerosità minima imposta per ogni slice, verranno costruite delle serie di dati, denominate 'netmax series'. Questi insiemi di elementi saranno ricavati grazie ad una funzione 'netmax', costruita per tale operazione. Ogni network deve fornire una serie di massimi annui, continua, utilizzata per aumentare la numerosità dei dati nelle slices. Maggiore sarà la numerosità dei dati all'interno di una slices, più precisa sarà la costruzione della curva di crescita. Il diagramma di flusso della funzione 'netmax' è rappresentato nella [Figura 5. 12].

Figura 5. 12: diagramma di flusso della funzione che permette di ricavare la serie di netmax

5.2.2.3 *Dipendenza spaziale della serie di netmax*

La serie di 'netmax' identificata sarà normalizzata rispetto al valore mediano di ogni stazione. Inoltre, per costruire tali serie, la funzione 'netmax' selezionerà per ogni anno il valore massimo, eliminando la correlazione con la stazione di appartenenza.

Quando si ricavano i valori dei massimi annui per ogni network, il codice della stazione non viene più preso in considerazione, ma si fa riferimento esclusivamente all'anno di accadimento dell'evento: per tale motivo, è possibile che per lo stesso anno si abbiano diversi dati associati a stazioni differenti, poste nelle immediate vicinanze.

Il problema della dipendenza spaziale è stato risolto traslando i dati appartenenti alla serie di netmax, di una quantità denominata distanza di offset. La formula (6) viene inserita all'interno di una funzione omonima, che consente per ogni network, di calcolare il valore esatto di tale distanza. La [Figura 5. 14] mostra il diagramma di flusso della funzione, riportata in [Appendice E: Codice delle funzioni costruite per la realizzazione dell'applicativo]. La [Figura 5. 13] mostra le diverse componenti che servono per il calcolo della distanza di offset: i numeri in alto a sinistra, in ogni riquadro, indicano l'ordine con il quale le azioni vengono svolte.

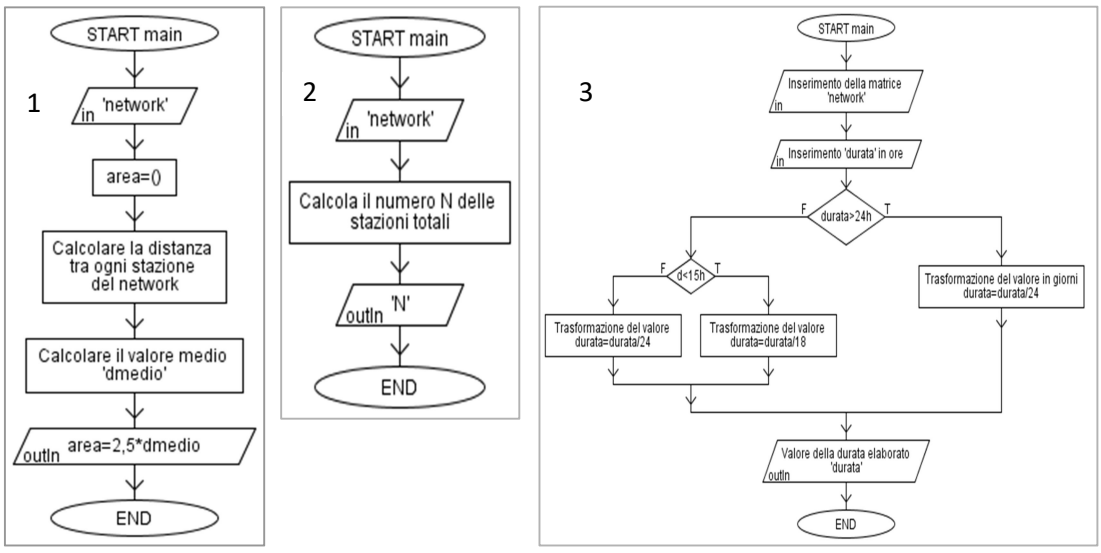


Figura 5. 13: set di diagrammi di flusso della funzione 'offset'.

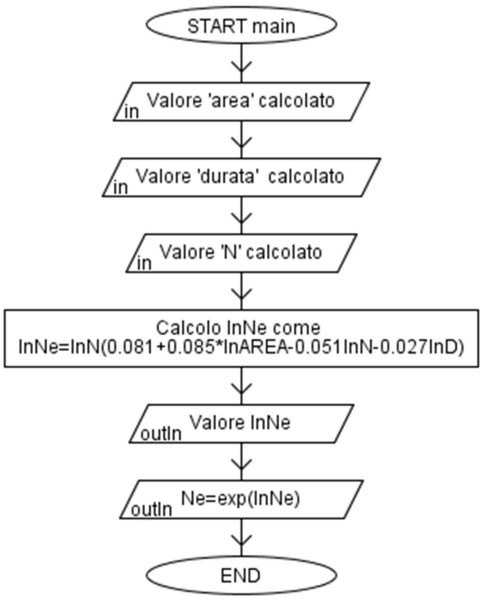


Figura 5. 14: rappresentazione del diagramma di flusso della funzione 'offset' per il calcolo della distanza di offset effettiva.

5.2.2.4 *Classificazione della serie di netmax*

In seguito, all'identificazione delle serie di netmax, è necessario porre l'accento sul metodo di rappresentazione utilizzato per questi dati. Non trattandosi di serie raggruppate per codice della stazione, questi valori saranno raffigurati secondo la tecnica 'station year'.

Poiché ogni serie di netmax non è suddivisa per stazioni differenti, il calcolo della frequenza subirà un cambiamento: ogni network possiede una serie di variabili ridotte, y, raggruppate in base alla stazione di riferimento. La serie di netmax, invece, conterrà un intervallo di valori y, dipendente dalla lunghezza totale della serie. Per costruire questo insieme, si è creata la funzione riportata in [Appendice E: Codice delle funzioni costruite per la realizzazione dell'applicativo], il cui diagramma di flusso è riportato in [Figura 5. 15].

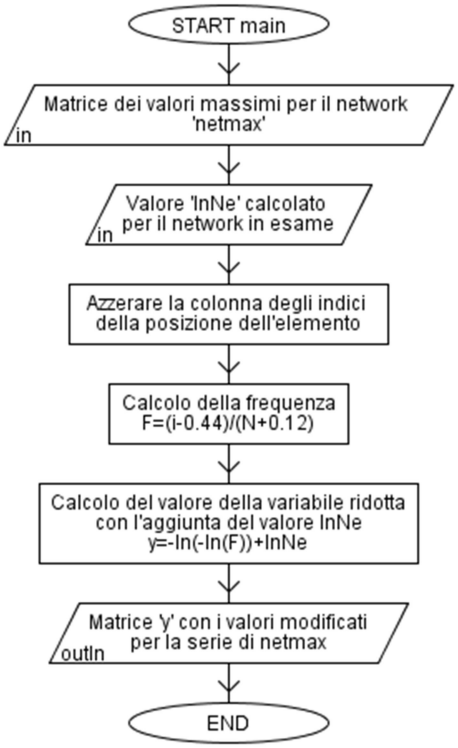


Figura 5. 15: Diagramma di flusso della funzione 'frequenza'.

Tabella 5. 11: intestazione delle differenti matrici denominate 'plot_num_slices'

Codice stazione	Y	X

In seguito a queste elaborazioni l'applicativo produrrà una serie di matrici, in [Tabella 5. 11], che conterranno sia i valori delle stazioni individuali, che i valori dei netmax. A tal punto le differenti matrici verranno unite, per rendere la costruzione della spezzata interpolante la nuvola di punti, più fluida.

5.3 *Costruzione della curva di crescita*

Il risultato finale della procedura è una curva di crescita stimata in base ai dati raccolti network per network. La procedura di costruzione di tale curva, vede il suo sviluppo in diversi passaggi rappresentati nello schema in [Figura 5. 16].

Figura 5. 16: diagramma di flusso per la costruzione della curva di crescita.

Nel corso dell'applicativo è stata creata la funzione 'model', che garantisce la costruzione della spezzata in funzione dei punti osservati. Poiché tale funzione risulta complessa e di difficile applicazione, per semplicità è stato suddiviso il suo diagramma di flusso a seconda dell'operazione che viene effettuata.

5.3.1 *Costruzione del grafico y-slices*

La costruzione del grafico rappresentativo dell'interpolazione avviene in modo da garantire, per ogni slices, l'esatta posizione degli estremi in funzione del numero di intervalli in cui viene suddiviso il set di dati. Per la prima interpolazione, si otterrà un numero di slices pari al numero di network.

La caratteristica principale è la suddivisione in slices differenti: infatti, poiché il punto corrispondente al valore mediano risulta fisso a priori, è bene costruire le slices a partire

da quest'ultimo. Per tale motivo, si è ipotizzato di avere un 20% del totale degli intervalli a sinistra del valore mediano, e i rimanenti a destra: in questo modo si ottengono dei grafici che garantiscono una corretta suddivisione in slices, in base alle quali avviene la stima dei valori di modello. Il diagramma di flusso in [Figura 5. 17], illustra la determinazione dei diversi valori delle y, in base al numero di intervalli selezionato.

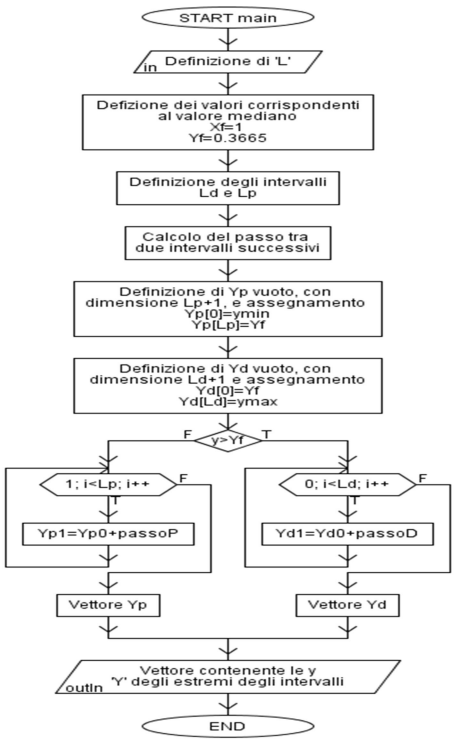


Figura 5. 17: diagramma di flusso rappresentativo della suddivisione nei diversi intervalli di dati.

Successivamente, è necessario che la matrice di dati completa venga classificata in funzione dei nuovi intervalli. Questa procedura è illustrata nel diagramma di flusso in [Figura 5. 18]. La classificazione a tal punto risulta automatizzata: al variare del numero di intervalli, si modificheranno i set di dati raggruppati per ogni slices.

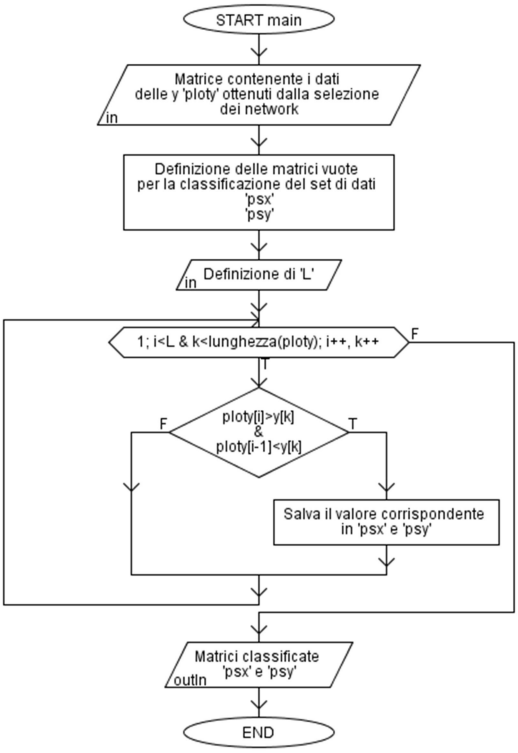


Figura 5. 18: rappresentazione delle operazioni eseguite per la suddivisione in differenti slices.

A questo punto, le matrici ottenute consentiranno il calcolo del gradiente per ogni slices. La composizione di queste matrici è illustrata nella [Tabella 5. 12]. In tal modo, è stato possibile unire in un'unica matrice, sia i dati rilevati dal 'pooling', che le serie di netmax associate ad ogni rete di stazioni.

Tabella 5. 12: intestazione delle tabelle 'psx' e 'psy' utilizzate per la costruzione della spezzata.

Slice1	Slice2	Slice3	Slice4	Slice5	Slices6

È importante classificare i valori delle x e delle y mantenendo le due matrici distinte, per evitare di incorrere in ulteriori errori.

5.3.2 *Calcolo dei gradienti*

Il primo elemento necessario per la costruzione di una retta è la stima del gradiente ad essa associato: questa procedura viene eseguita attraverso una regressione ai minimi quadrati. Le matrici in [Tabella 5. 12] rappresentano il set di dati classificato per le differenti slices: al variare del numero degli intervalli, il numero delle colonne della matrice tende ad aumentare.

Così facendo, l'applicazione della funzione 'gradiente' è semplificata. La funzione viene costruita in modo tale da permettere alle coordinate x e y di rimanere indipendenti per ogni stazione, come mostra la [Figura 5. 19]:

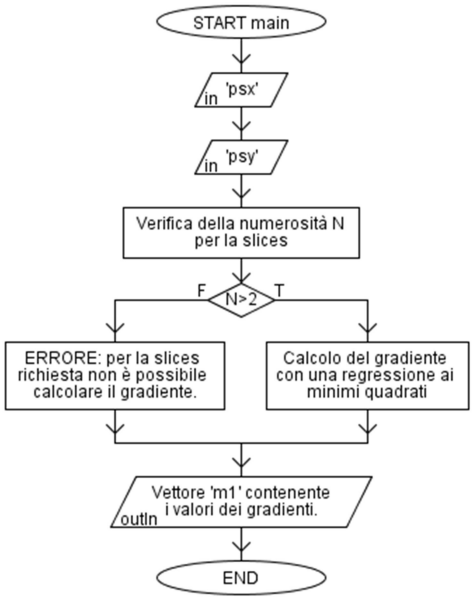


Figura 5. 19: diagramma di flusso della funzione 'gradiente'.

In questo modo, per ogni slices, si otterrà un gradiente diverso: per facilitarne l'utilizzo, i risultati saranno raccolti in un vettore unico, in modo da richiamare semplicemente il valore interessato. Non sempre è stato possibile ricavare il valore del gradiente: in tali casi la costruzione della curva non è riuscita a proseguire. Questo inconveniente si è presentato nelle slices più a destra, in corrispondenza dei network costruiti a distanza maggiore dal 'focal point'.

5.3.3 *Costruzione della retta spezzata*

Per costruire la retta interpolante, è necessario calcolare i valori delle ordinate, in funzione delle variabili ridotte note. Nella [Figura 5. 20] è rappresentato il diagramma di flusso della funzione 'model': questa prima parte, permette di stimare i valori delle ordinate degli estremi delle slices definite in precedenza.

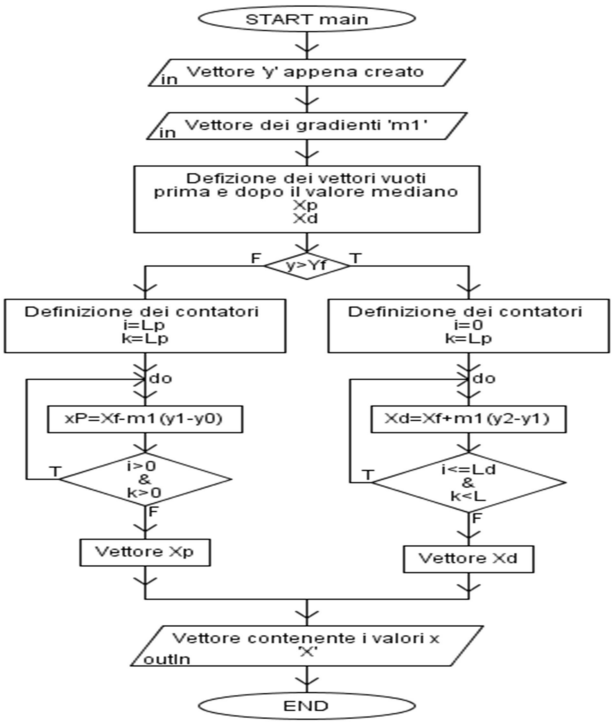


Figura 5. 20: rappresentazione del diagramma di flusso della funzione 'model', per la ricerca dei valori delle ordinate degli estremi delle slices.

Successivamente, note le coordinate degli estremi di ogni intervallo, sarà possibile calcolare l'intero set di dati interpolanti la nuvola di punti osservati. Per eseguire questa operazione, la funzione 'model' sviluppa la tecnica in [Figura 5. 21].

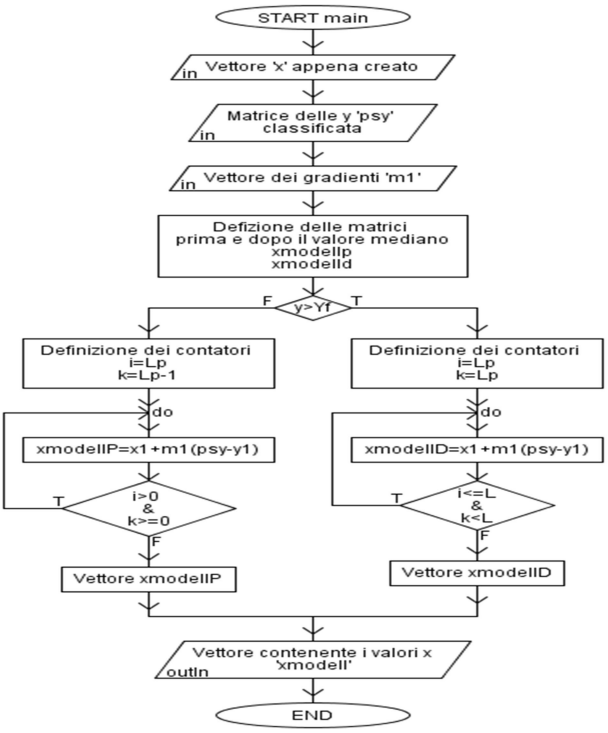


Figura 5. 21: rappresentazione della procedura di calcolo dei valori delle ordinate di modello

Grazie a questa funzione, è possibile stimare i punti di modello in funzione di una qualsiasi discretizzazione degli intervalli. La procedura permette di eseguire una prima interpolazione, considerando un numero di slices pari al numero di network ottenuti. Successivamente verrà incrementato il numero di intervalli, in modo da visualizzare un migliore adattamento della funzione a tratti sull'insieme di punti.

5.3.4 *Selezione della spezzata con il migliore adattamento ai punti*

La scelta della migliore discretizzazione viene effettuata in funzione di una differenza ai minimi quadrati. La funzione 'EQ1' è stata definita, per implementare la seguente formula:

$$\sum_{i=1}^{M}(x_i - x_{model})^2 + 0.05\left(\frac{M}{l-1}\right)\sum_{j=2}^{l}\left(m_j - m_{j-1}\right)^2 \qquad 19$$

L'applicativo fornirà in output una tabella riassuntiva dei risultati ottenuti per ciascuna discretizzazione effettuata. In questo modo, l'utente potrà selezionare in modo autonomo la suddivisione in intervalli, che minimizza tale risultato.

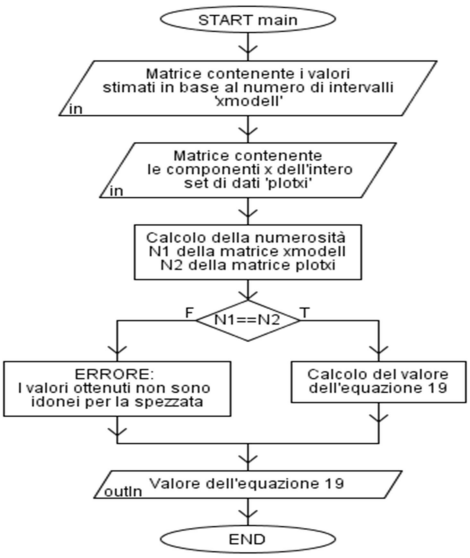


Figura 5. 22: rappresentazione della funzione 'EQ1', per la selezione della migliore spezzata.

5.3.5 *Calcolo dei parametri delle distribuzioni di probabilità*

Successivamente, è necessario identificare la funzione analitica che meglio si adatta alla spezzata scelta a rappresentare il set di dati. Per garantire il migliore adattamento a tale curva sono state costruite le due funzioni 'GEV' e 'GUMBEL', che consentono di ricavare i valori delle ordinate, associate alle variabili ridotte. In tal modo, la stima dei parametri α, ε e k avverrà attraverso l'ausilio di una funzione implementata in Python, denominata 'curve_fit', appartenente alla libreria 'scipy.optimize', descritta nell'[Appendice C: Descrizione della piattaforma di sviluppo del software]. Il diagramma di flusso in [Figura 5. 23] illustra come è possibile utilizzare questa funzione e quali sono i parametri di input della stessa.

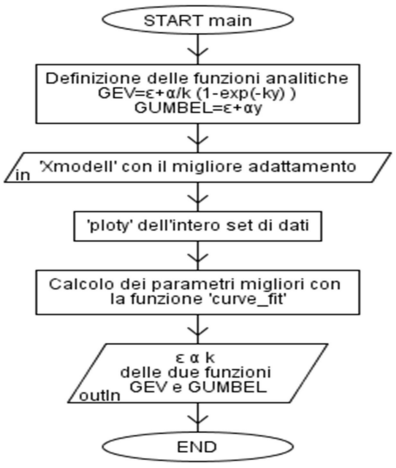


Figura 5. 23: diagramma di flusso della funzione 'curve_fit'.

5.3.6 *Costruzione della curva di crescita del metodo FORGEX*

I parametri che garantiscono un miglior adattamento di una distribuzione di probabilità nota alla funzione a tratti ottenuta con il metodo FORGEX, vengono costruiti sulla base di un set di dati normalizzato rispetto al valore mediano. Per garantire un confronto con le altre tecniche di analisi regionale, è necessario applicare un fattore di correzione a queste serie di valori, così da rendere la normalizzazione in funzione del valore medio. Tale correzione viene applicata esclusivamente sul set di dati relativi alla curva FORGEX.

Il diagramma di flusso in [Figura 5. 24] illustra la metodologia utilizzata per applicare questa correzione.

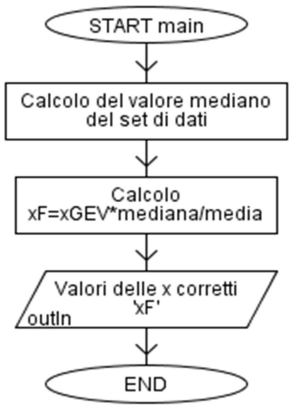


Figura 5. 24: diagramma di flusso per l'applicazione della correzione al set di dati ottenuto con il FORGEX.

5.3.7 *Costruzione delle curve di frequenza empirica*

Le curve di frequenza empirica vengono utilizzate per verificare la validità della distribuzione scelta, rispetto al set di dati in esame. Per la costruzione di queste due curve è necessario essere in possesso dei dati appartenenti alla stazione strumentata e dei dati raggruppati dal pooling nell'intorno della stessa.

La curva empirica di stazione viene costruita estraendo i dati necessari dal database generale: la normalizzazione della serie avverrà rispetto al valore medio del sito scelto, in modo da poter confrontare tale curva con i grafici principali. La [Figura 5. 25] illustra la tecnica per la costruzione di tale curva.

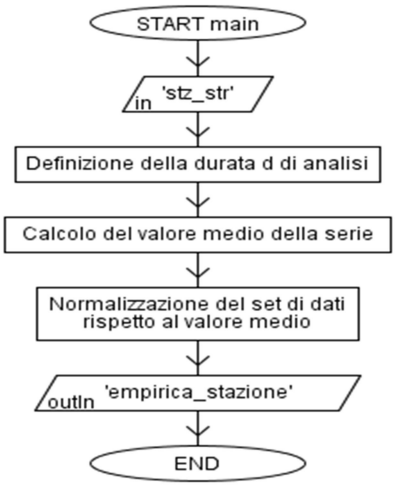


Figura 5. 25: procedura per la costruzione della curva di frequenza empirica di stazione

La curva 'station year' verrà costruita in base a tutti i dati raggruppati nell'intorno del sito in esame. Il network più ampio verrà rielaborato per verificare quanto la distribuzione selezionata si adatti all'insieme di dati. Questi verranno normalizzati rispetto al valore medio di ogni stazione. Il calcolo della frequenza e della variabile ridotta, invece, avverrà considerando un'unica serie di dati, eliminando le differenti stazioni. L'insieme di operazioni eseguite è rappresentato nel diagramma di flusso in [Figura 5. 26]. Il risultato ottenuto consentirà di rappresentare la curva 'station year', in funzione dei dati racchiusi nella seguente tabella:

Tabella 5. 13: intestazione della tabella rielaborata della rete di stazioni più ampia.

Codice stazione	***Anno***	***MaxDh***	***Val_norm_media***	***Y***	***F***

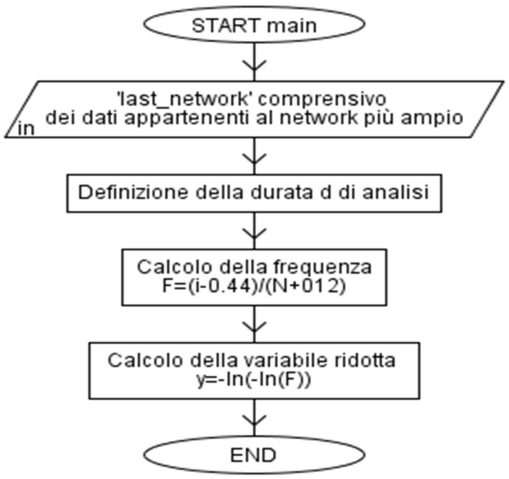


Figura 5. 26: diagramma di flusso per la costruzione della curva 'station year'.

5.4 *Costruzione delle curve di crescita di confronto*

In seguito alla costruzione della curva di crescita del metodo FORGEX, è possibile effettuare qualche confronto con le principali metodologie di regionalizzazione applicate in Italia. I metodi utilizzati per il confronto sono i seguenti:

1) Il metodo VAPI che prevede di calcolare i valori dei quantili con la formula seguente, in base ai parametri caratteristici del sito, estratti dal rapporto di De Michele e Rosso, (2001):

$$x_R = \varepsilon + \frac{\alpha}{k}\left(1 - e^{-ky}\right) \quad (20)$$

2) Il metodo del Kriging , che ricava i valori dei parametri utilizzati nella seguente formula, dalle mappe di regionalizzazione del Piemonte:

$$x_K = 1 - \frac{\theta_2^*}{\theta_3}\left[\Gamma(1-\theta_3) - \left(-\ln\left(\frac{T-1}{T}\right)\right)^{-\theta_3}\right] \quad (21)$$

I parametri presenti in entrambe le fomule, poiché variabili da un sito ad un altro, sono inseriti dall'utente. Quet'ultima parte dell'applicativo richiede all'utente di interagire con il programma, in modo da ottenere i risultati desiderati.

5.5 *Costruzione delle curve di possibilità pluviometrica*

La costruzione delle curve di possibilità pluviometrica è l'obiettivo di un'analisi regionale. Tale curva viene ricavata moltiplicando il valore indice per il quantile normalizzato della serie di dati associata al sito in esame.

Le differenti altezze di precipitazione, saranno calcolate in funzione del fattore di crescita ottenuto. Il diagramma di flusso in [Figura 5. 27] mostra la procedura che consente di costruire i grafici rappresentativi delle curve di possibilità pluviometrica, per ricavare i valori di progetto desiderati.

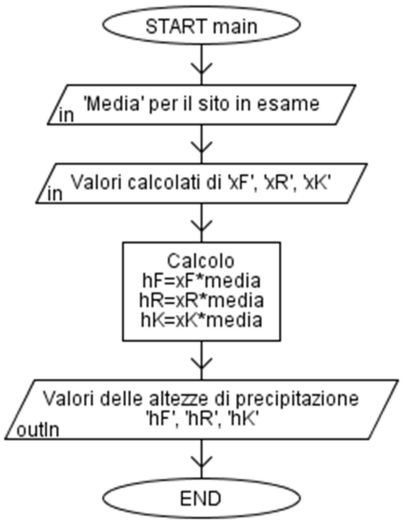


Figura 5. 27:diagramma di flusso per la stima della curva di possibilità pluviometrica.

Il risultato di un'analisi regionale eseguita con questo tipo di approccio empirico, fornisce i risultati riassunti in [Appendice F: Schede delle località analizzate] per differenti aree del Piemonte.

Capitolo 6

Conclusioni

L'analisi regionale di frequenza delle precipitazioni, viene eseguita in Italia in funzione di una serie di parametri: questi consentono di determinare le curve di crescita, utili alla stima della grandezza di progetto. L'utilizzo di un metodo, come quello inglese, slega completamente il legame tra i parametri di sito statistici e l'altezza di precipitazione. La determinazione dei valori di progetto, basata esclusivamente sulla numerosità delle serie storiche delle stazioni pluviometriche, ha prodotto risultati accettabili.

L'applicazione di questo metodo ai dati italiani ha fatto emergere qualche osservazione, in merito alle tecniche principali di analisi regionale, che si utilizzano tradizionalmente in Italia. Sono stati messi in evidenza, alcuni aspetti legati alla numerosità e alla concentrazione delle stazioni italiane, rispetto a quelle inglesi: le infomazioni, riguardanti le precipitazioni inglesi, appaiono molto più dettagliate e numerose, rispetto a quelle italiane. Questo aspetto ha influenzato negativamente i risultati ottenuti, obbligando l'applicativo ad alcuni adattamenti.

Un'ulteriore differenza riscontrata nel corso dell'elaborazione della tesi, è stata la grandezza indice alla base del metodo: in Inghilterra, infatti, si utilizza il valore mediano, che è stato regionalizzato sull'intero territorio. Questo aspetto, molto differente da ciò che si utilizza in Italia, ha prodotto a livello del metodo implementato, un punto di distacco dalla procedura inglese. Per riuscire ad ottenere risultati confrontabili con le curve di crescita dei metodi regionali italiani è stato necessario applicare una correzione ai dati ottenuti con l'applicazione del metodo FORGEX.

L'implementazione attraverso un sistema informatizzato ha riproposto un'applicativo, sulla base del FORGEX inglese, che procede in modo semi-automatico, garantendo non solo le curve di crescita tipiche del sito scelto, ma anche la curva di possibilità pluviometrica, in grado di determinare la stima della pioggia di progetto. L'utente, con il quale questo programma potrà affiancarsi, dovrà inserire nel corso dello svolgimento dell'analisi, alcune grandezze tipiche, variabili da un sito all'altro.

All'utente saranno infatti richieste alcune informazioni, come le coordinate del sito in cui si vuole eseguire l'analisi, la durata per la quale si desidera conoscere il valore di progetto, il valore indice della località in esame, ricavato dalle mappe regionalizzate, e alcuni parametri fissi del sito, in modo da costruire le curve di crescita con i comuni metodi di analisi regionale utilizzati in Italia. L'unico vincolo che si impone all'utente, riguarda il valore della durata inserita: al momento, l'applicativo garantisce una stima dei valori di progetto per le cinque durate caratteristiche, 1, 3, 6 , 12 e 24 ore.

L'applicativo prodotto garantisce, indipendentemente dalle statistiche e dagli aspetti del territorio, una stima dell'altezza di pioggia, proprio come ciò che si ottiene applicando i metodi VAPI e Kriging, con esso confrontati. I principali problemi riscontrati riguardano i dati presenti all'interno dei database italiani. Poiché uno tra i requisiti del metodo è l'elevata numerosità di dati per stazione, sono stati rilevati determinati inconvenienti di stima in corrispondenza di alcune località.

Dal punto di vista tecnico è possibile rendere il software più automatico, e sicuramente ridurre i tempi computazionali, specialmente in presenza di quantità di dati molto elevate. Per quel che riguarda, invece, il suo funzionamento teorico, i risultati ottenuti hanno prodotto l'andamento atteso. In certi casi, si sono ottenute stime dei quantili molto simili ai classici metodi comunemente usati. Tale metodo potrebbe fornire ai

professionisti uno strumento utile per ricavare una stima dei quantili in determinate località caratteristiche, senza l'ausilio di alcun parametro. Essendo un metodo piuttosto dipendente dal numero delle osservazioni, potrebbe essere affiancato ad un'analisi statistica, volta a verificare i risultati ottenuti.

Appendice

7.1 *Appendice A: Definizioni*

7.1.1 *Altezza di pioggia h_d*

È definita come la quantità di pioggia, espressa in [mm], che si è abbattuta su un determinato punto dello spazio, durante un dato intervallo di tempo *d*. Lo strumento che consente la misura di questo valore, è il pluviometro, che permette di misurare anche il valore di durata associato ad essa.

7.1.2 *Durata delle precipitazioni*

È definita come il periodo di tempo durante il quale la precipitazione ha luogo, e si misura in [ore]. L'altezza di pioggia è funzione della durata stessa. Le misure delle durate di precipitazioni, variano al variare dello strumento che recupera le informazioni pluviometriche. Le precipitazioni si suddividono in due grandi categorie:

- Le piogge giornaliere: sono caratterizzate da una misura in continuo, nell'arco di 24 ore, dalle 9.00 del mattino alle 9.00 del mattino seguente. Non considerano l'orario in cui è iniziata la precipitazione, ma solo quale è l'altezza di pioggia in quell'intervallo specifico.
- Le piogge brevi: sono misurate dall'inizio della precipitazione su un intervallo mobile. Possono essere classificate in funzione della durata come:
    - Piogge di 1 ora

- Piogge di 3 ore
- Piogge di 6 ore
- Piogge di 12 ore
- Piogge di 24 ore

La pioggia giornaliera e la pioggia breve di 24 ore, hanno la stessa durata, ma vengono rilevate con modalità diverse, assumendo caratteri molto differenti.

7.1.3 *Tempo di ritorno*

Il periodo di ritorno è definito come il tempo medio, in cui un dato valore assegnato alla grandezza di progetto, viene uguagliato o superato almeno una volta. È possibile definire il periodo di ritorno come il tempo medio che intercorre tra il verificarsi di due eventi di precipitazioni successive, che uguagliano o superano un valore soglia. In altre parole, il tempo di ritorno è definito anche come il reciproco della probabilità di superamento 'p' del valore assegnato.

$$T = \frac{1}{p} \qquad 22$$

Spesso, questa grandezza viene considerata in modo non adeguato. Infatti, può essere utilizzato questo valore, come la probabilità che un dato evento si verifichi ogni T anni: in realtà fissando, per la costruzione di opera il T a 100 anni, non è detto che per quell'intervallo di tempo, l'evento non si verifichi.

7.1.4 *Curva di crescita*

La curva di crescita indica la misura della variabilità delle precipitazioni estreme in una data zona, per diversi periodi di ritorno. Lo scopo della curva di crescita è mostrare l'andamento delle precipitazioni per una data durata. Il parametro identificato della curva di crescita è il fattore di crescita, solitamente espresso come K(T), in funzione del tempo di ritorno.

Tale elemento garantisce il legame tra il valore della precipitazione osservata, e il valore indice utilizzato per l'analisi, come mostra la formula seguente:

$$K(T) = \frac{h_{oss}}{\bar{h}} \quad (23)$$

Dove:

- K(T) è il fattore di crescita;
- h_{oss} è la precipitazione osservata;
- $\bar{h}$ indica il valore indice, che può essere espresso come il valore medio o mediano della serie.

7.1.5 *Regione omogenea.*

Una regione omogenea, composta da un numero N di stazioni pluviometriche, prevede che i dati associati ad ogni stazione, abbiano una stessa distribuzione di frequenza, e uguale curva di crescita, in modo da rendere l'analisi nel sito in esame uniforme. Questo tipo di regioni viene utilizzato per la stima del valore di precipitazione in aree non strumentate.

7.1.6 *Metodo Station Year*

Il metodo station year è un metodo di analisi delle serie di dati che raggruppa le serie storiche in funzione dell'anno di accadimento dell'evento. La peculiarità di tale metodologia è la rappresentazione di tale serie: la costruzione delle curve 'station year', prevede che i valori appartenenti a differenti stazioni vengano ordinati in modo crescente, così da avere un unico macro campione di frequenze osservate. I valori osservati non presenteranno valori che si ripetono, in modo da garantire continuità alla serie.

7.2 *Appendice B: Distribuzioni di probabilità*

7.2.1 *Distribuzioni di probabilità di Gumbel*

La distribuzione di Gumbel, viene impiegata per modellare la distribuzione del massimo numero di campioni, che possono essere estratti da distribuzioni diverse. Essa risulta molto utile nel predire la probabilità di accadimento di eventi estremi.

La distribuzione di Gumbel, o distribuzione del massimo valore del primo tipo (Extreme value type 1 - EV1) può essere espressa in queste due forme canoniche:

$$F(x) = e^{e^{-y}} \tag{24}$$

$$y = \frac{1}{\theta_2}(x - \theta_1) \tag{25}$$

7.2.1.1 *Forme della distribuzione di Gumbel*

$$F(x) = e^{-e^{\frac{1}{\theta_2}(x-\theta_1)}} \tag{26}$$

$$f(x) = e^{-\frac{1}{\theta_2}(x-\theta_1)} e^{-e^{\frac{1}{\theta_2}(x-\theta_1)}} \tag{27}$$

Da cui è possibile ricavare:

$$x(F) = \theta_1 - \theta_2 \ln\left(-ln\left(\frac{F-1}{F}\right)\right) \tag{28}$$

Successivamente alla stima dei quantili è possibile definire alcuni parametri caratteristici per semplificare le equazioni soprastanti:

α→ Fattore di scala: $\alpha = \frac{1}{\theta_2}$

ε→ Fattore di posizione: $\varepsilon = \theta_1$

Con i quali è possibile esprimere il valore del quantile come:

$$x(F) \equiv K(T) = \varepsilon + \alpha y \quad (29)$$

7.2.1.2 *Parametri delle distribuzione di Gumbel*

Ogni distribuzione di probabilità prevede la stima dei parametri in differenti modi, riassunti in seguito:

7.2.1.2.1 Il metodo dei momenti

$$\begin{cases} \theta_1 = \mu(x) - 0{,}5772 * \theta_2 \\ \theta_2 = \dfrac{\sqrt{6} * \sigma(x)}{\pi} \end{cases} \quad (30)$$

Dove:

$$\mu = \theta_1 + 0.5772\theta_2 \quad (31)$$

$$\sigma^2 = \pi^2 \left(\frac{\theta_2^2}{6}\right) \quad (32)$$

$$\gamma = 1.1396 \quad (33)$$

$$k = 5 + \frac{2}{5} \quad (34)$$

7.2.1.2.2 Il metodo degli L-momenti

$$\begin{cases} \theta_1 = \gamma_1 - 0{,}5772 * \theta_2 \\ \theta_2 = \dfrac{\gamma_2}{\ln(2)} \end{cases} \quad (35)$$

Dove:

$$\gamma_1 = \theta_1 + 0.5772\theta_2 \quad (36)$$

$$\gamma_2 = \theta_2 \ln(2) \quad (37)$$

$$\tau_3 = 0.1699 \quad (38)$$

$$\tau_4 = 0.1504 \quad (39)$$

7.2.2 *Distribuzione di probabilità GEV*

La distribuzione GEV, anche definita distribuzione del valore estremo di Gumbel) è una distribuzione a 3 parametri di tipo continuo. È importante evidenziare, come nel caso in cui il terzo parametro fosse pari a 0, allora questa distribuzione convergerà su di una Gumbel. Tale distribuzione può essere espressa come:

$$F(x) = e^{e^{-y}} \quad (40)$$

$$y = \begin{cases} \frac{1}{\theta_3}\ln\left(1 - \frac{\theta_3(x-\theta_1)}{\theta_2}\right) & Per\ \theta_3 \neq 0 \\ \frac{x-\theta_1}{\theta_2} & Per\ \theta_3 = 0 \end{cases} \quad (41)$$

7.2.2.1 *Forme della distribuzione GEV*

$$F(x) = e^{-\left(1-\frac{\theta_3(x-\theta_1)}{\theta_2}\right)^{\frac{1}{\theta_3}}} \quad (42)$$

$$f(x) = \frac{1}{\theta_2} e^{(-(1-\theta_3)\left(\frac{1}{\theta_3}\right)\ln\left(1-\frac{\theta_3(x-\theta_1)}{\theta_2}\right) - e^{\left(\frac{1}{\theta_3}\right)\ln\left(1-\frac{\theta_3(x-\theta_1)}{\theta_2}\right)}} \quad (43)$$

Da cui è possibile ricavare:

$$x(F) = \begin{cases} \theta_1 + (\frac{\theta_2}{\theta_3})(1 - (-\ln(F))^{\theta_3} & Per\ \theta_3 \neq 0 \\ \theta_1 - \theta_2 \ln\left(-ln\left(\frac{F-1}{F}\right)\right) & Per\ \theta_3 = 0 \end{cases} \quad (44)$$

Successivamente alla stima dei quantili è possibile definire alcuni parametri caratteristici per semplificare le equazioni soprastanti:

α→ Fattore di scala: $\alpha = \frac{1}{\theta_2}$

ε→ Fattore di posizione: $\varepsilon = \theta_1$

k→Fattore di forma: $k = \frac{1}{\theta_3}$

Con i quali è possibile esprimere il valore del quantile come:

$$x(F) \equiv K(T) = \varepsilon + \frac{k}{\alpha}\left(1 - \left(-\ln\left(1 - \frac{1}{T}\right)\right)^{\frac{1}{k}}\right) \quad 45$$

7.2.2.2 *Parametri delle distribuzione GEV*

Ogni distribuzione di probabilità prevede la stima dei parametri in differenti modi, riassunti in seguito:

7.2.2.2.1 Il metodo dei momenti

$$\begin{cases} \theta_1 = \mu - 0{,}5772 * \theta_2 \\ \theta_2 = \dfrac{\sqrt{6} * \sigma}{\pi} \end{cases} \quad 46$$

Dove:

$$\mu = \theta_1 + \frac{\theta_2\left(1 - \Gamma(1 + \theta_3)\right)}{\theta_3} \quad 47$$

$$\sigma^2 = \left(\frac{\theta_2}{\theta_3}\right)^2 \left(\Gamma(1 + 2\theta_3) - \gamma^2(1 + \theta_3)\right) \quad 48$$

7.2.2.2.2 Il metodo degli L-momenti

$$\begin{cases} \theta_1 = \gamma_1\left(\dfrac{\theta_2}{\theta_3}\right)\left(1 + \Gamma(1 + \theta_3)\right) \\ \theta_2 = \dfrac{\gamma_2\,\theta_3}{(1 - 2^{-\theta_3})\Gamma(1 + \theta_3)} \\ \theta_3 \cong 7.8590c + 2.9554c^2 \\ c = \dfrac{2}{3 + \tau} - \dfrac{ln2}{ln3} \end{cases} \quad 49$$

Dove:

$$\gamma_1 = \theta_1 + \theta_2 \frac{\left(1 - \Gamma(1+\theta_3)\right)}{\theta_3} \quad (50)$$

$$\gamma_2 = \theta_2 \left(1 - 2^{-\theta_3}\right) \left(\frac{\Gamma(1+\theta_3)}{\theta_3}\right) \quad (51)$$

$$\tau_3 = \frac{2\left(1 - 3^{-\theta_3}\right)}{1 - 2^{-\theta_3}} - 3 \quad (52)$$

$$\tau_4 = \frac{5\left(1 - 4^{-\theta_3}\right) - 10\left(1 - 3^{-\theta_3}\right) + 6\left(1 - 2^{-\theta_3}\right)}{1 - 2^{-\theta_3}} \quad (53)$$

7.3 *Appendice C: Descrizione della piattaforma di sviluppo del software*

Nell'ambiente di sviluppo di Python è possibile osservare tre tipi di errori fondamentali, a cui bisogna prestare notevole attenzione:

1) Syntax Error: sono degli errori di sintassi, legati ad un'errata costruzione nella forma del programma;
2) Value Error: sono degli errori di valore, ovvero si presentano quando il valore a cui si fa riferimento non è corretto per l'operazione che si desidera svolgere;
3) Type Error: sono degli errori di tipo, legati alla dichiarazione del valore in modo errato.

Al momento in cui si presenta l'errore, l'utente è in grado di capire quale riga andare a correggere: spesso il tipo di errore è affiancato dalla motivazione per cui si è presentato quel problema.

7.3.1 *Editor per la programmazione*

Python 3.7 è un software che può lavorare su qualunque tipo di editor di testo; per comodità si è scelto di utilizzare Spyder, che è un ambiente di sviluppo scientifico, in grado di consentire una programmazione guidata. La piattaforma di editor si presenta con un'interfaccia grafica suddivisa in tre parti, interscambiabili e modificabili a seconda della richiesta dell'utente.

La prima finestra è un editor di testo, in cui l'utente scrive e progetta il file di testo rappresentativo del programma richiesto, come si può notare nella [Figura 7. 1].

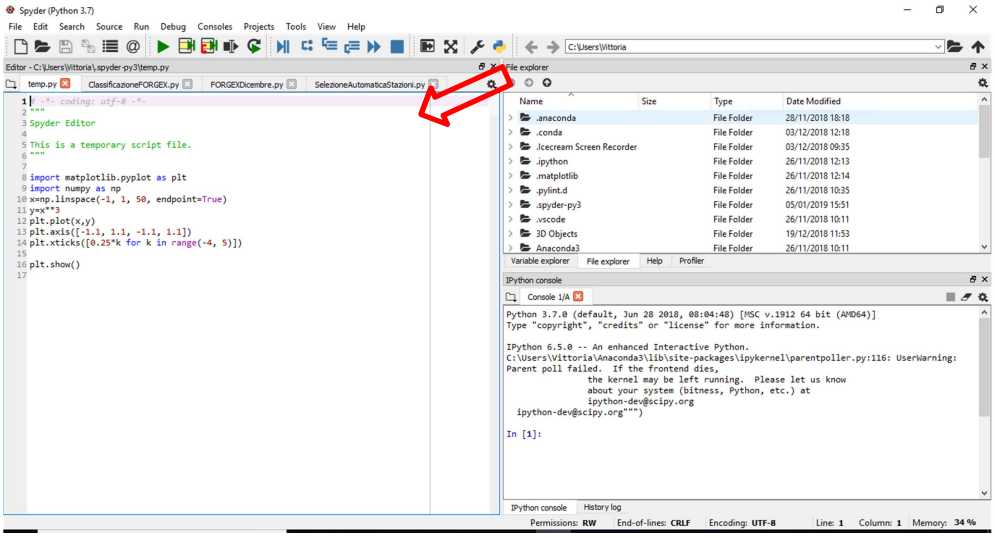


Figura 7. 1:schermata dell'interfaccia grafica del software Spyder utilizzato per progettare il programma, dettaglio dell'editor

La seconda finestra, in [Figura 7. 2], è la console di Python, in cui si osservano i risultati del programma nella finestra di testo, e si osservano eventuali messaggi di errori riguardanti il programma.

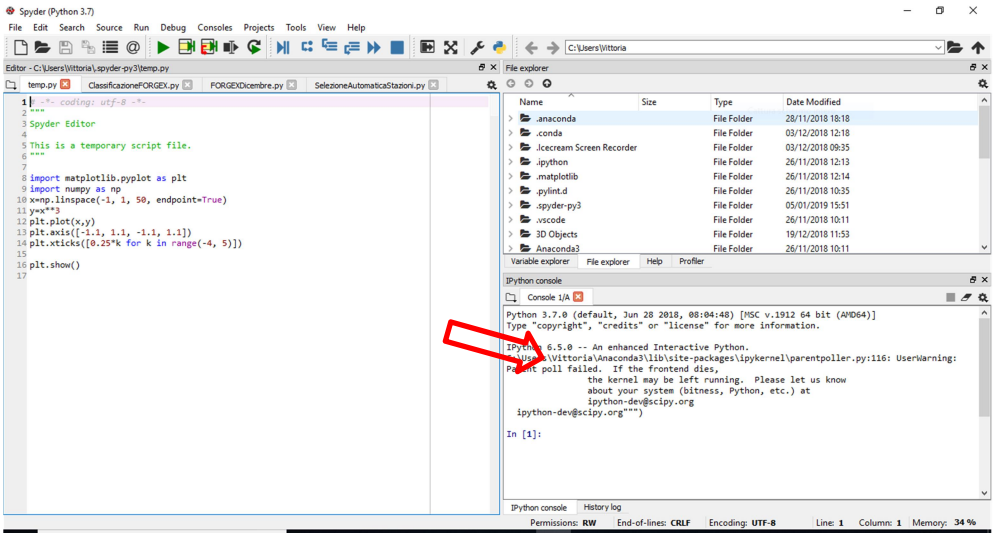


Figura 7. 2: schermata dell'interfaccia grafica del software Spyder utilizzato per progettare il programma, dettaglio della console.

La terza area è interscambiabile:

1) è possibile osservare la casella che racchiude tutte le variabili che sono state utilizzate nel corso del programma, come si vede dalla [Figura 7. 3];
2) potrebbe presentare un riassunto dei diversi percorsi per ogni file utilizzato o utilizzabile nel programma, come si osserva dalla [Figura 7. 4];

3) presenta un'interfaccia di aiuto, 'Help';
4) potrebbe essere presente una casella denominata 'Profiler', in cui sono riassunte le volte che una data funzione o metodo sono state utilizzate nel programma stesso.

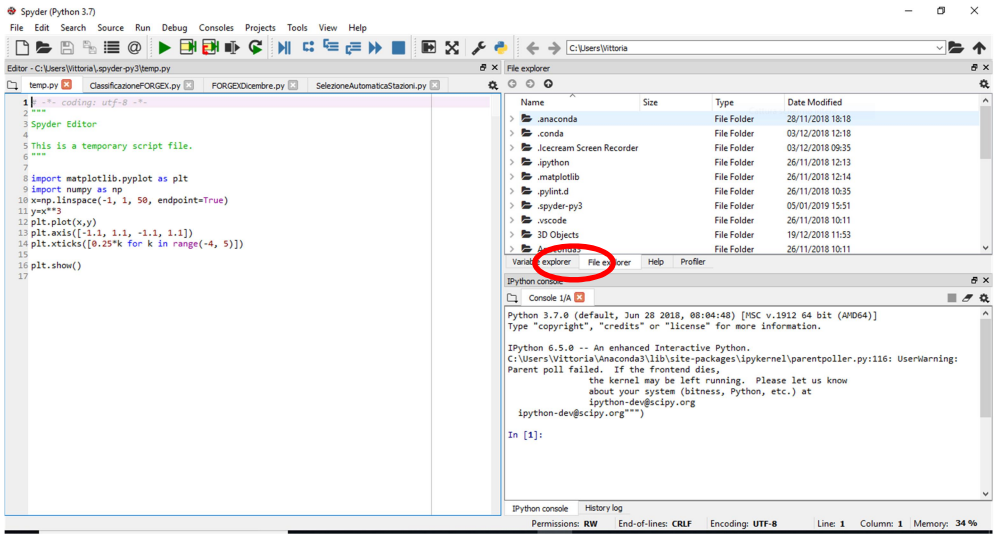


Figura 7. 3: schermata dell'interfaccia grafica del software Spyder utilizzato per progettare il programma, dettaglio della finestra 'File explorer'.

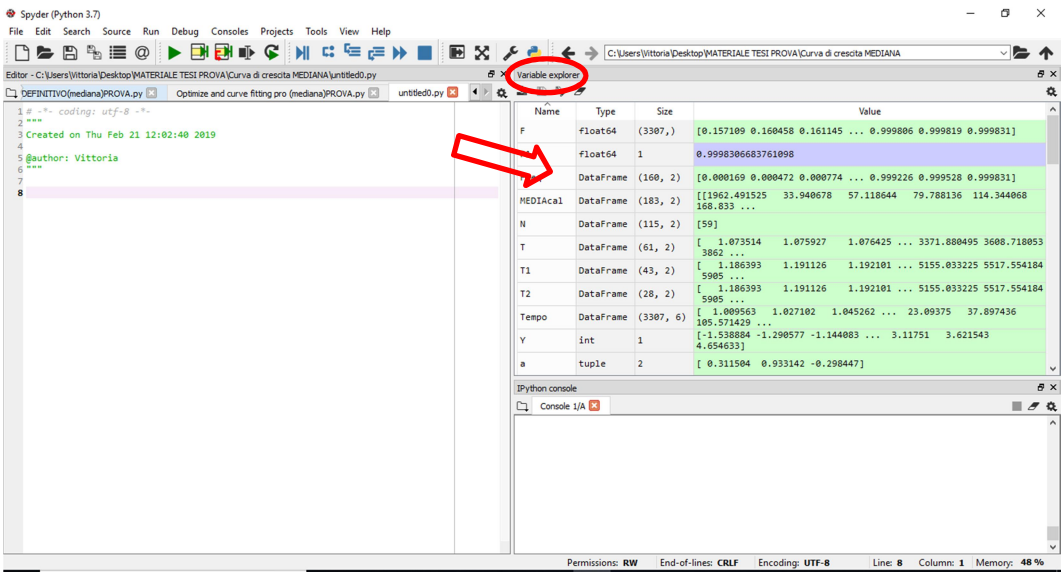


Figura 7. 4: schermata dell'interfaccia grafica del software Spyder utilizzato per progettare il programma, dettaglio della finestra 'Variable explorer'.

L'editor presenta, nella parte alta della sua interfaccia, una serie di tasti (nel riquadro in rosso nella [Figura 7. 5]), che richiamano funzioni particolari: ognuno di essi può essere richiamato anche dai menù estraibili (nel rettangolo in giallo della [Figura 7. 5]), che brevemente riassumono anche il compito della funzione.

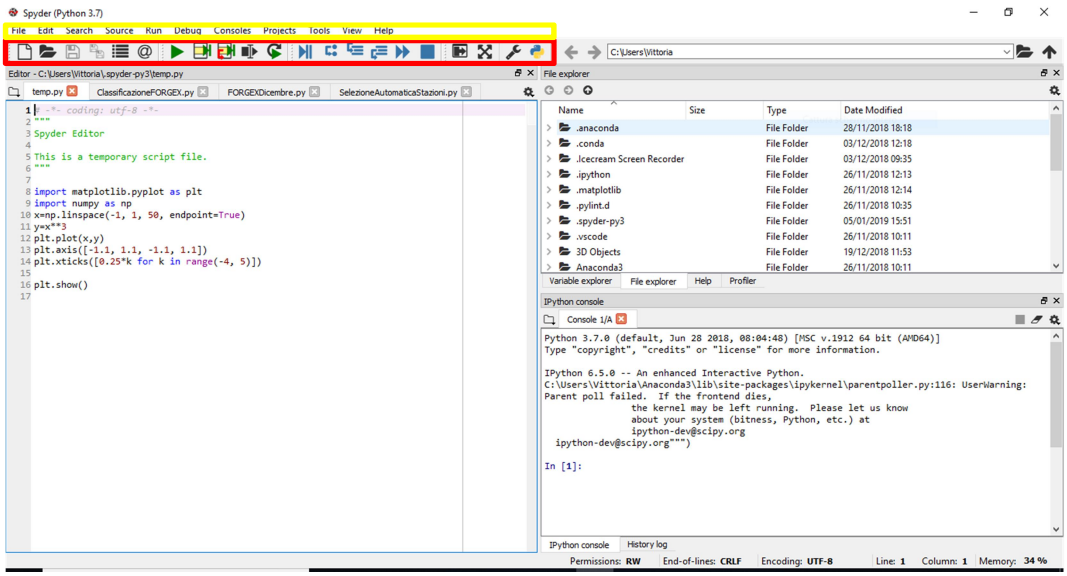


Figura 7. 5: schermata dell'interfaccia grafica del software Spyder utilizzato per progettare il programma.

Esistono numerosi programmi per l'editor di testo, che potrebbero essere utilizzati per implementare un dato applicativo, come per esempio 'Notepad' oppure, maggiormente utilizzato per questo genere di programmi è 'Jupiter'. La scelta di un software dall'interfaccia guidata ha reso la programmazione più leggera e corretta.

Ovviamente software più tecnici, come potrebbe essere 'Jupiter', sono adatti per programmatori più esperti, poiché richiedono una conoscenza maggiore del linguaggio e dell'organizzazione degli ambienti. 'Notepad' a confronto risulta molto simile ad un blocco note di Microsoft: è un editor caratterizzato esclusivamente dalla schermata che garantisce la stesura del programma senza alcun tipo di controllo del linguaggio in cui viene scritto, né tantomeno nella correttezza e fluidità del programma stesso.

7.3.2 *Le principali librerie di Python 3.7*

Python 3.7 è un linguaggio di programmazione, che lavora attraverso pacchetti di librerie standard: queste ultime vengono utilizzate, ognuna con i comandi principali, permettendo la costruzione di diversi applicativi a seconda della richiesta dell'utente.

Nel corso degli anni sono state create numerose librerie, una per ogni campo di applicazione: alcune permettono i calcoli matriciali, altre sono in grado di effettuare calcoli statistici o regressioni lineari. Esistono alcuni pacchetti che permettono di costruire grafici di diverso tipo e modalità, o che vengono utilizzati per la gestione di database fisici, o condivisi in rete.

Per poter utilizzare le diverse funzioni contenute in ogni set, è necessario importare la libreria e indicare al software la posizione di riferimento in cui trovare l'informazione richiesta. Per eseguire questa operazione, prima di iniziare qualunque tipo di calcolo o analisi, è necessario importare l'insieme di applicativi mediante il comando 'import' di Python. A tal proposito, esistono due modalità di richiamo delle librerie:

1) '*import [nome della libreria] as [acronimo libreria]*': è un comando che permetterà di importare tutta la libreria. La sua abbreviazione garantirà il richiamo della funzione, senza dover ogni volta ripetere il nome della libreria;
2) '*from [nome libreria] import [nome funzione]*': è un comando che richiama esclusivamente la funzione dalla libreria di riferimento.

A differenza di altri software per la programmazione, Python, non consente di produrre un ciclo iterativo, semplicemente iterando una variabile come corrispettiva della posizione iterabile. Il ciclo, dovrà essere ben definito dall'utente: si potrebbe desiderare di iterare esclusivamente l'elemento, oppure semplicemente la sua posizione all'interno del vettore, o matrice.

Nel primo caso, non saranno permessi incrementi per i contatori, in quanto si modificherebbe il valore dell'elemento stesso. In tal caso, modificando la serie di valori, si incorrerebbe in una modifica sostanziale dei dati da analizzare, danneggiando l'intero processo.

Nel secondo caso in cui si desidera incrementare un contatore indice della posizione dell'elemento, sarà necessario utilizzare una funzione 'range'. Lo scopo di tale funzione è creare un vettore di affiancamento alla serie di dati iniziali, che abbia la stessa lunghezza di quest'ultimo, ma possieda al suo interno esclusivamente gli indici delle variabili da incrementare.

A seconda dell'operazione da effettuare si dovranno inserire all'interno del programma le librerie e le funzioni ad esse associate, per garantire uno sviluppo più rapido. Nel caso in cui non esista una funzione per l'operazione da svolgere, sarà possibile crearla da zero, e semplicemente richiamandola, applicarla ai dati in esame. Esistono numerosissime librerie diverse, una per ogni tipologia di azione che si desidera svolgere: per il lavoro svolto, sono state utilizzate le librerie descritte nei paragrafi successivi.

7.3.2.1 ***Pandas***

Questa libreria consente l'analisi e lo sviluppo di grandi quantità di dati, raggruppati in database, ordinati e comprensibili. La sintassi di Pandas, segue le linee guida generali del linguaggio Python, con alcuni accorgimenti dovuti alla mole di dati con la quale l'utente si trova a contatto. Ovviamente, lavorare con i database non è semplice come potrebbe essere con delle matrici. Un database potrebbe contenere non solo valori numerici ma anche dei caratteri alfanumerici, identificativi di un dato elemento. Ogni colonna del database ha un nome rappresentativo del suo contenuto, per una più facile identificazione. È possibile modificare queste etichette, in funzione del tipo di dato che si vuole riportare in ogni colonna.

Pandas, al suo interno, raccoglie una serie di funzioni che permettono di leggere, scrivere o modificare, il contenuto di determinati file. A seconda del tipo di file con il quale si desidera lavorare, si osservano diversi comandi:

1) *pd.read_csv('[nome file]')*: permette di richiamare il file con caratteri in virgola mobile, semplicemente inserendo il nome del file.
2) *pd.read_excel('[nome file]'):* è un comando che permette di richiamare qualsiasi tipo di file excel.

È possibile specificare determinate caratteristiche di lettura del file, che variano a seconda delle impostazioni:

1) *sep='/t'* è l'impostazione generale, che permette di mandare a capo alla fine di ogni riga di dati;
2) *sep=';'* indica al programmatore che i diversi valori delle colonne, saranno divisi dalla punteggiatura indicata tra virgolette;
3) *decimal=','* specifica che i valori decimali sono indicati con la punteggiatura descritta. Questa impostazione deve essere tenuta in considerazione specialmente nelle fasi di scrittura del file e di 'import' ad un altro formato.

A differenza di altre librerie 'Pandas', consente di eseguire operazioni, che con altri comandi risulterebbero notevolmente più difficili e lunghe. Un esempio è il comando 'groupby', che viene utilizzato spesso nel corso del'applicativo: tale funzione consente di

raggruppare l'intero database per un carattere specifico. La corretta espressione della funzione è la seguente:

[nome_database].groupby(by='[etichetta della colonna]').[operazione]

Poiché il comando groupby genera un gruppo non utilizzabile a scopo di analisi successive, è bene affiancargli un'operazione come per esempio il conteggio degli elementi totali, oppure il calcolo dei valori medi, o ancora ricavare per quella serie il valore massimo in base al gruppo eseguito.

L'analisi di un database, attraverso la libreria Pandas, non consente l'esecuzione di cicli iterativi semplici: è possibile che all'interno del file non siano contenuti esclusivamente valori numerici, ma anche simboli o caratteri alfanumerici. Eseguendo quindi, una data operazione ciclica su un insieme di elementi, si potrebbero generare dei risultati 'NaN', che dunque potrebbero danneggiare il file iniziale.

7.3.2.2 *Numpy*

La libreria principalmente utilizzata per eseguire operazioni matriciali è 'Numpy', che permette la rappresentazione dei dati attraverso delle matrici multi-dimensionali. La raffigurazione risulta molto simile ad un database, a differenza della definizione a priori delle dimensioni della matrice. Per costruire una matrice con al suo interno un insieme definito di dati, è bene conoscere la numerosità degli stessi. A tal punto si andrà a creare una matrice vuota, all'interno della quale verranno inseriti i dati da analizzare.

Una matrice può contenere valori numerici, letterali o booleani, a seconda del tipo di operazione che si desidera eseguire, e del tipo di rappresentazione richiesta. Solitamente i caratteri booleani, per comodità di visualizzazione, possono essere sostituiti dai valori numerici 0 e 1. La costruzione di matrici differenti, permette una più semplice utilizzazione dei dati: lavorare con le matrici risulta più immediato, poiché è possibile eseguire operazioni tra elementi appartenenti a matrici diverse senza l'ausilio di variabili di appoggio.

Una matrice multi-dimensionale, permette la rappresentazione di diversi dati, e se necessario permette di eseguire su di essa una serie di iterazioni in modo continuativo. I cicli iterativi, vengono costruiti in modo da evidenziare l'operazione che deve essere

eseguita, ma non permettono la modifica della posizione di un dato, senza l'utilizzo di un contatore aggiuntivo.

7.3.2.3 *Matplotlib*

Un'ulteriore libreria molto utile per analisi di grandi quantità di dati, è 'Matplotlib'. È un insieme di funzioni che permettono la costruzione di qualunque tipologia di grafico: si possono costruire dei grafici a barre, dei box plot, semplici grafici a dispersione, oppure anche grafici su più assi cartesiani, confrontabili tra loro.

I grafici possono assumere caratteristiche diverse: esistono diverse impostazioni grafiche, per esempio i colori o i *'marker'* della serie esaminata. Possono, inoltre, presentarsi altri aspetti come lo spessore della linea, continua o tratteggiata, o le caratteristiche del 'marker' identificativo della serie rappresentata sul grafico.

Questa libreria consente, attraverso una serie di comandi particolari, di inserire diverse etichette sul grafico: si osservano le *'label'* per i nomi degli assi e per eventuali punti che si vogliono mettere in evidenza all'interno del grafico. Ad ogni grafico si può associare un dato titolo e una legenda, per semplificarne la lettura.

Attraverso il comando *'subplot'* è possibile raffigurare fino a quattro tipi di grafici diversi all'interno di una stessa figura: in tal modo, per esempio per le diverse durate, potrebbe essere utile visualizzare a colpo d'occhio le differenze che presentano le varie curve di crescita.

7.3.2.4 *Scipy*

Un'altra libreria molto utile per quanto riguarda i calcoli statistici è Scipy. Tale libreria contiene una serie di comandi che garantiscono all'utente di velocizzare operazioni, come le regressioni lineari o le interpolazioni tra differenti dati.

La libreria contiene al suo interno differenti pacchetti legati alle diverse operazioni da svolgere. Per costruire una curva di distribuzione precisa associata alla spezzata in output, si è utilizzato il contenitore 'optimize' di scipy. Questo pacchetto contiene gli algoritmi di ottimizzazione per costruire delle curve il più precise possibile.

Nel corso del programma, la libreria scipy è stata richiamata esclusivamente durante la ricerca dei parametri che stimano correttamente la distribuzione associata alla spezzata in oggetto. L'utilizzo di tale libreria velocizza operazioni, che manualmente risultano molto complesse e con un livello di precisione approssimato.

7.4 *Appendice D: Codice dell'applicativo creato*

Il codice creato per la costruzione delle curve di crescita sulla base del metodo inglese, è stato sviluppato in due parti tra loro correlate. La suddivisione è servita per evitare di costruire un programma molto pesante e cercare di alleggerire un po' l'applicativo. I risultati della parte A vengono rielaborati in modo più veloce nella parte B, così che, nel caso in cui servisse riprodurre altre analisi sullo stesso set di dati, possa essere più facile reperirli.

7.4.1 *Parte A*

7.4.1.1 *Identificazione e selezione del set di dati utilizzabili*

Inizialmente verranno importate le librerie utilizzate nel corso del programma:

```
import numpy as np
import math
import pandas as pd
import matplotlib.pyplot as plt
import copy
from scipy.spatial import distance
```

Successivamente è necessario importare i file utilizzati per l'analisi, e calcolarne la numerosità per evitare serie troppo brevi:

```
percorso=('C:\\Users\Vittoria\Desktop\TesiVittoria\DATI.txt')
percorso1=('C:\\Users\Vittoria\Desktop\TesiVittoria\ANAGRAFICA.xls
x')
info=pd.read_excel(percorso1).iloc[:,[0,2,3]]
dimensione=info.shape
anagrafica=np.full(dimensione, info)
dati=pd.read_csv(percorso, sep='\t', dtype='float')
dim=dati.shape
precipitazione=np.full(dim, dati)
np.set_printoptions(precision=2, suppress=True)

#calcolo della numerosità per ognuna delle stazioni
freq=np.array(dati.groupby('Numero').count())
anagrafica=np.insert(anagrafica,3, freq[:,0], axis=1)

#classificazione secondo la numerosità superiore a 10 elementi
anagrafica=np.insert(anagrafica , 4, list(map(numerosita,
anagrafica[:,3])),axis=1)
anagrafica=np.delete(anagrafica[anagrafica[:,4]==1], 4, axis=0)
anagrafica=np.delete(anagrafica, 4, axis=1)
```

In seguito è necessario definire le coordinate del sito in cui si desidera eseguire l'analisi, e calcolare la distanza per ogni stazione da tale punto:

```
#Inserire le coordinate di un punto specifico
x=int(input('Inserire la coordinata x del sito di interesse: -->
'))
y=int(input('Inserire la coordinata y del sito di interesse: -->
'))
#calcolare la distanza dal punto in esame
utmx=np.array((x-anagrafica[:,1])**2)
utmy=np.array((y-anagrafica[:,2])**2)
radq=((utmx[:]+utmy[:]))
dist=(radq)**(1/2)
anagrafica=np.insert(anagrafica, 4, dist, axis=1)
#ricavare i dati delle sole stazioni accettabili
intersect=np.intersect1d(anagrafica[:,0], precipitazione[:,0])
anagrafica1_matches = anagrafica[np.any(anagrafica[:, 0] ==
intersect[:, None], axis=0)]
valid_station = precipitazione[np.any(precipitazione[:, 0] ==
intersect[:, None], axis=0)]
set_dati=pd.DataFrame(valid_station)
set_dati.to_excel('Stazioni_utili.xlsx')
```

7.4.1.2 *Ricerca della durata caratteristica a cui eseguire l'analisi*

All'utente verrà chiesto di inserire la durata alla quale desidera procedere con l'analisi.

```
#ricerca della durata caratteristica
d=int(input('Per quale durata si vuole eseguire l\'analisi?'))
durate=np.array([1,3,6,12,24])
datid=test(d)
#inserimento di una colonna vuota per valori normalizzati
datid=np.insert(datid, 3, 0, axis=1)
n = np.unique(datid[:,0])
datid=np.array([(datid[datid[:,0]==i]) for i in n] )
```

7.4.1.3 *Calcolo del valore indice e normalizzazione dei dati*

La serie scelta viene classificata per stazione, e per ogni stazione si calcola il valore mediano e si procede con la normalizzazione del set di dati:

```
#calcolo della mediana e normalizzazione
j=np.arange(0, len(datid))
for row in j:
    datid[row][:,3]=(datid[row][:,2])/(np.median(datid[row][:,2]))
    datid[row][:]=np.sort(datid[row][:], axis=0)
norm=copy.deepcopy(datid[:])
norm=np.concatenate(norm)
```

7.4.1.4 *Plotting position*

È necessario calcolare per la serie di dati, raggruppata per stazione, i valori della frequenza e della variabili ridotta ad essa associata:

```
#inserire dei campi vuoti per i, F, y
values=valid_station[:,[0,1]]
values=np.insert(values, 2, 0, axis=1)
values=np.insert(values, 3, 0, axis=1)
values=np.insert(values, 4, 0, axis=1)
#values=np.insert(values, 5, 0, axis=1)
N=np.unique(values[:,0])
values=np.array([(values[values[:,0]==l]) for l in N])
k=np.arange(0, (len(values)))
for item in k:
    values[item][:,2]=np.arange(1, len(values[item])+1)
#azzera l'indice
    values[item][:,3]=(values[item][:,2]-
0.44)/(len(values[item])+0.12)
#valore della frequenza
    values[item][:,4]=(-np.log(-np.log(values[item][:,3])))
#valore della y GUMBEL
#    values[item][:,5]=(1/(1-values[item][:,3])) #valore del T
y=copy.deepcopy(values[:])
y=np.concatenate(y)
```

7.4.1.5 *Classificazione del set di dati con il metodo delle y-slices*

Successivamente si procederà con la costruzione del grafico, raggruppando i dati osservati per le differenti classi di y ottenute:

```
#classificazione e conteggio dei valori di y per slices
y=np.insert(y , 5, list(map(classify, y[:,4])),axis=1)
num1=y[y[:,5]==0]
anag1=num_per_slice(num1, anagrafica)
num2=y[y[:,5]==1]
anag2=num_per_slice(num2, anagrafica)
num3=y[y[:,5]==2]
anag3=num_per_slice(num3, anagrafica)
num4=y[y[:,5]==3]
anag4=num_per_slice(num4, anagrafica)
num5=y[y[:,5]==4]
anag5=num_per_slice(num5, anagrafica)
num6=y[y[:,5]==5]
anag6=num_per_slice(num6, anagrafica)
num7=y[y[:,5]==6]
anag7=num_per_slice(num7, anagrafica)
num8=y[y[:,5]==7]
anag8=num_per_slice(num8, anagrafica)
anag4=diversi(anag1, anag4)
anag5=diversi(anag1, anag5)
anag6=diversi(anag1, anag6)
anag7=diversi(anag1, anag7)
anag8=diversi(anag1, anag8)
```

```
anagrafica=np.delete(anagrafica, [1,2,3], axis=1)
anagrafica=np.insert(anagrafica, 2, anag1[:,2], axis=1)
anagrafica=np.insert(anagrafica, 3, anag2[:,2], axis=1)
anagrafica=np.insert(anagrafica, 4, anag3[:,2], axis=1)
anagrafica=np.insert(anagrafica, 5, anag4[:,2], axis=1)
anagrafica=np.insert(anagrafica, 6, anag5[:,2], axis=1)
anagrafica=np.insert(anagrafica, 7, anag6[:,2], axis=1)
anagrafica=np.insert(anagrafica, 8, anag7[:,2], axis=1)
anagrafica=np.insert(anagrafica, 9, anag8[:,2], axis=1)
anagrafica=anagrafica[anagrafica[:,1].argsort()]
```

7.4.1.6 *Costruzione dei network*

L'applicativo ricostruirà con un processo iterativo, i differenti network in modo da garantire il requisito minimo di dati per slices:

```
#definizione del network1
network1=network(anagrafica,2)
Y1=valori_slice(network1, num1)    #sono i dati da graficare nella
prima slice
norm1=valori_gumbel(network1) #sono tutti i dati del primo network
plot1=np.intersect1d(Y1[:,4], norm1[:,4]) #dati del primo network
per la prima slice
plot1=norm1[np.any(norm1[:,4]==plot1[:,None], axis=0)]
#calcolare il netmax1
netmax1=netmax(norm1)
lnNe1=offset(network1)
netmax1=frequenza(netmax1, lnNe1)
netmax1=net_valid(netmax1)
#definizione del network2
network2=network(anagrafica,3)
Y2=valori_slice(network2, num2) #sono i dati da graficare nella
prima slice
norm2=valori_gumbel(network2)
plot2=np.intersect1d(Y2[:,4], norm2[:,4])
plot2=norm2[np.any(norm2[:,4]==plot2[:,None], axis=0)]
#calcolare il netmax2
netmax2=netmax(norm2)
lnNe2=offset(network2)
netmax2=frequenza(netmax2, lnNe2)
netmax2=net_valid(netmax2)
#definizione del network3
network3=network(anagrafica,4)
Y3=valori_slice(network3, num3) #sono i dati da graficare nella
prima slice
norm3=valori_gumbel(network3)
plot3=np.intersect1d(Y3[:,4], norm3[:,4])
plot3=norm3[np.any(norm3[:,4]==plot3[:,None], axis=0)]
#calcolare il netmax3
netmax3=netmax(norm3)
lnNe3=offset(network3)
netmax3=frequenza(netmax3, lnNe3)
netmax3=net_valid(netmax3)
#definizione del network4
network4=network(anagrafica,5)
```

```
Y4=valori_slice(network4, num4) #sono i dati da graficare nella
prima slice
norm4=valori_gumbel(network4)
plot4=np.intersect1d(Y4[:,4], norm4[:,4])
plot4=norm4[np.any(norm4[:,4]==plot4[:,None], axis=0)]
#calcolare il netmax4
netmax4=netmax(norm4)
lnNe4=offset(network4)
netmax4=frequenza(netmax4, lnNe4)
netmax4=net_valid(netmax4)
#definizione del network5
network5=network(anagrafica,6)
Y5=valori_slice(network5, num5) #sono i dati da graficare nella
prima slice
norm5=valori_gumbel(network5)
plot5=np.intersect1d(Y5[:,4], norm5[:,4])
plot5=norm5[np.any(norm5[:,4]==plot5[:,None], axis=0)]
#calcolare il netmax5
netmax5=netmax(norm5)
lnNe5=offset(network5)
netmax5=frequenza(netmax5, lnNe5)
netmax5=net_valid(netmax5)
#definizione del network6
network6=network(anagrafica,7)
stz=pd.DataFrame(network6[:,(0,1)])
stz.to_excel('StzROI.xlsx')
Y6=valori_slice(network6, num6)
#sono i dati da graficare nella prima slice
norm6=valori_gumbel(network6)
database=pd.DataFrame(norm6)
#si estraggono i dati del network più grande
database.to_excel('Dati_totali_NETWORK.xlsx')
plot6=np.intersect1d(Y6[:,4], norm6[:,4])
plot6=norm6[np.any(norm6[:,4]==plot6[:,None], axis=0)]
plot6=np.intersect1d(Y6[:,4], norm6[:,4])
plot6=norm6[np.any(norm6[:,4]==plot6[:,None], axis=0)]
#calcolare il netmax6
netmax6=netmax(norm6)
lnNe6=(offset(network6))
netmax6=frequenza(netmax6, lnNe6)
netmax6=net_valid(netmax6)
#definizione del network7 non per la numerosità ma per i netmax
network7=network(anagrafica, 8)
Y7=valori_slice(network7, num7) #sono i dati da graficare nella
prima slice
norm7=valori_gumbel(network7)
plot7=np.intersect1d(Y7[:,4], norm7[:,4])
plot7=norm7[np.any(norm7[:,4]==plot7[:,None], axis=0)]
#calcolare il netmax7
netmax7=netmax(norm7)
lnNe7=(offset(network7))
netmax7=frequenza(netmax7, lnNe7)
netmax7=net_valid(netmax7)
```

SI procede con la creazione del grafico rappresentativo dell'intero set di dati:

```
fig=plt.figure(figsize=(6,6))
ax=plt.subplot(111)
ax.plot(plot1[:,4], plot1[:,3], color='r', marker='.',
linestyle='', linewidth=0.3, label='NetSeries1')
ax.plot(netmax1[:,3], netmax1[:,2], marker='x', linestyle='',
color='b', linewidth=0.1, label='Netmax points 1')
ax.plot(plot2[:,4], plot2[:,3], color='g', marker='.',
linestyle='', linewidth=0.3, label='NetSeries2')
ax.plot(netmax2[:,3], netmax2[:,2], marker='x', linestyle='',
color='y', linewidth=0.1, label='Netmax points 2')
ax.plot(plot3[:,4], plot3[:,3], color='orange', marker='.',
linestyle='', linewidth=0.3, label='NetSeries3')
ax.plot(netmax3[:,3], netmax3[:,2], marker='x', linestyle='',
color='c', linewidth=0.1, label='Netmax points 3')
ax.plot(plot4[:,4], plot4[:,3], color='k', marker='.',
linestyle='', linewidth=0.3, label='NetSeries4')
ax.plot(netmax4[:,3], netmax4[:,2], marker='x', linestyle='',
color='w', linewidth=0.1, label='Netmax points 4')
ax.plot(plot5[:,4], plot5[:,3], color='m', marker='.',
linestyle='', linewidth=0.3, label='NetSeries5')
ax.plot(netmax5[:,3], netmax5[:,2], marker='x', linestyle='',
color='r', linewidth=0.1, label='Netmax points 5')
ax.plot(plot6[:,4], plot6[:,3], color='b', marker='.',
linestyle='', linewidth=0.3, label='NetSeries6')
ax.plot(netmax6[:,3], netmax6[:,2], marker='x', linestyle='',
color='g', linewidth=0.1, label='Netmax points 6')
#ax.plot(plot7[:,4], plot7[:,3], color='r', marker='.',
linestyle='', linewidth=0.3, label='NetSeries7')
#ax.plot(netmax7[:,3], netmax7[:,2], marker='x', linestyle='',
color='b', linewidth=0.1, label='Netmax points 7')
#ax.plot(0.3665, 1, marker='o', color='b')
ax.xaxis.set_ticks(np.linspace(-0.6335, 6.3665, 8, endpoint=True))
ax.yaxis.set_ticks([0,4], minor=True)
ax.xaxis.set_ticks_position('bottom')
ax.yaxis.set_ticks_position('left')
ax.patch.set_edgecolor('k')
ax.patch.set_linewidth(1)
ax.xaxis.grid(True, 'major', linewidth=3)
ax.yaxis.grid(True, 'major', linewidth=0.5)
ax.legend(loc='upper left', fontsize=7)
ax.set_title('Rappresentazione dei punti', fontsize=20)
ax.set_xlabel('Reduced variate y')
ax.set_ylabel('Rainfall standardized x')
plt.show()
```

7.4.1.7 *Calcolo dei gradienti per ogni intervallo*

Per ogni intervallo di valori, il codice calcolerà il valore del coefficiente angolare del tratto di retta in esame. Sarà necessario classificare in differenti intervalli i valori dei netmax:

```
#classificazione dei netmax per ogni slice
#prima slice
net11=netmax1[netmax1[:,6]==0]
net21=netmax2[netmax2[:,6]==0]
net31=netmax3[netmax3[:,6]==0]
net41=netmax4[netmax4[:,6]==0]
net51=netmax5[netmax5[:,6]==0]
net61=netmax6[netmax6[:,6]==0]
net71=netmax7[netmax7[:,6]==0]
netslice1=np.concatenate((net11, net21, net31, net41, net51,
net61, net71), axis=0)
netslice1=netslice1[netslice1[:,3].argsort()]
#seconda slice
net12=netmax1[netmax1[:,6]==1]
net22=netmax2[netmax2[:,6]==1]
net32=netmax3[netmax3[:,6]==1]
net42=netmax4[netmax4[:,6]==1]
net52=netmax5[netmax5[:,6]==1]
net62=netmax6[netmax6[:,6]==1]
net72=netmax7[netmax7[:,6]==1]
netslice2=np.concatenate((net12, net22, net32, net42, net52,
net62, net72), axis=0)
netslice2=netslice2[netslice2[:,3].argsort()]
#terza slice
net13=netmax1[netmax1[:,6]==2]
net23=netmax2[netmax2[:,6]==2]
net33=netmax3[netmax3[:,6]==2]
net43=netmax4[netmax4[:,6]==2]
net53=netmax5[netmax5[:,6]==2]
net63=netmax6[netmax6[:,6]==2]
net73=netmax7[netmax7[:,6]==2]
netslice3=np.concatenate((net13, net23, net33, net43, net53,
net63, net73), axis=0)
netslice3=netslice3[netslice3[:,3].argsort()]
#quarta slice
net14=netmax1[netmax1[:,6]==3]
net24=netmax2[netmax2[:,6]==3]
net34=netmax3[netmax3[:,6]==3]
net44=netmax4[netmax4[:,6]==3]
net54=netmax5[netmax5[:,6]==3]
net64=netmax6[netmax6[:,6]==3]
net74=netmax7[netmax7[:,6]==3]
netslice4=np.concatenate((net14, net24, net34, net44, net54,
net64, net74), axis=0)
netslice4=netslice4[netslice4[:,3].argsort()]
#quinta slice
net15=netmax1[netmax1[:,6]==4]
net25=netmax2[netmax2[:,6]==4]
net35=netmax3[netmax3[:,6]==4]
net45=netmax4[netmax4[:,6]==4]
```

```
net55=netmax5[netmax5[:,6]==4]
net65=netmax6[netmax6[:,6]==4]
net75=netmax7[netmax7[:,6]==4]
netslice5=np.concatenate((net15, net25, net35, net45, net55,
net65, net75), axis=0)
netslice5=netslice5[netslice5[:,3].argsort()]
#sesta slice
net16=netmax1[netmax1[:,6]==5]
net26=netmax2[netmax2[:,6]==5]
net36=netmax3[netmax3[:,6]==5]
net46=netmax4[netmax4[:,6]==5]
net56=netmax5[netmax5[:,6]==5]
net66=netmax6[netmax6[:,6]==5]
net76=netmax7[netmax7[:,6]==5]
netslice6=np.concatenate((net16, net26, net36, net46, net56,
net66, net76), axis=0)
netslice6=netslice6[netslice6[:,3].argsort()]
#settima slice
net17=netmax1[netmax1[:,6]==6]
net27=netmax2[netmax2[:,6]==6]
net37=netmax3[netmax3[:,6]==6]
net47=netmax4[netmax4[:,6]==6]
net57=netmax5[netmax5[:,6]==6]
net67=netmax6[netmax6[:,6]==6]
net77=netmax7[netmax7[:,6]==6]
netslice7=np.concatenate((net17, net27, net37, net47, net57,
net67, net77), axis=0)
netslice7=netslice7[netslice7[:,3].argsort()]
#vengono eliminate le colonne superflue
netslice1=np.delete(netslice1, (4,5,6), axis=1)
netslice2=np.delete(netslice2, (4,5,6), axis=1)
netslice3=np.delete(netslice3, (4,5,6), axis=1)
netslice4=np.delete(netslice4, (4,5,6), axis=1)
netslice5=np.delete(netslice5, (4,5,6), axis=1)
netslice6=np.delete(netslice6, (4,5,6), axis=1)
netslice7=np.delete(netslice7, (4,5,6), axis=1)
#costruzione dei set di dati per slices
plot1=np.concatenate((plot1[:,(3,4)], netslice1[:,(2,3)]))
plot1=plot1[plot1[:,1].argsort()]
plot2=np.concatenate((plot2[:,(3,4)], netslice2[:,(2,3)]))
plot2=plot2[plot2[:,1].argsort()]
plot3=np.concatenate((plot3[:,(3,4)], netslice3[:,(2,3)]))
plot3=plot3[plot3[:,1].argsort()]
plot4=np.concatenate((plot4[:,(3,4)], netslice4[:,(2,3)]))
plot4=plot4[plot4[:,1].argsort()]
plot5=np.concatenate((plot5[:,(3,4)], netslice5[:,(2,3)]))
plot5=plot5[plot5[:,1].argsort()]
plot6=np.concatenate((plot6[:,(3,4)], netslice6[:,(2,3)]))
plot6=plot6[plot6[:,1].argsort()]
plot7=np.concatenate((plot7[:,(3,4)], netslice7[:,(2,3)]))
plot7=plot7[plot7[:,1].argsort()]
```

```
#per importare i file e lavorare in maniera separata dal file
origine
df=pd.DataFrame(plot1)
df.to_excel('plot1.xlsx')
df2=pd.DataFrame(plot2)
df2.to_excel('plot2.xlsx')
df3=pd.DataFrame(plot3)
df3.to_excel('plot3.xlsx')
df4=pd.DataFrame(plot4)
df4.to_excel('plot4.xlsx')
df5=pd.DataFrame(plot5)
df5.to_excel('plot5.xlsx')
df6=pd.DataFrame(plot6)
df6.to_excel('plot6.xlsx')
```

7.4.2 *Parte B*

7.4.2.1 *Definizione delle librerie utilizzate*

È necessario definire quali sono le librerie principali che si utilizzano per la costruzione delle differenti curve:

```
import numpy as np
import math
import pandas as pd
import matplotlib.pyplot as plt
import copy
from scipy import stats as stats
from scipy.stats.distributions import genextreme
from scipy.optimize import curve_fit
from scipy.special import gammaln
import seaborn as sns
```

I dati salvati dallo script di selezione verranno importati per essere analizzati:

```
#si importano i differenti dati
percorso=('C:\\Users\Vittoria\Desktop\MATERIALE TESI PROVA\Curva
di crescita MEDIANA\plot1.xlsx')
percorso2=('C:\\Users\Vittoria\Desktop\MATERIALE TESI PROVA\Curva
di crescita MEDIANA\plot2.xlsx')
percorso3=('C:\\Users\Vittoria\Desktop\MATERIALE TESI PROVA\Curva
di crescita MEDIANA\plot3.xlsx')
percorso4=('C:\\Users\Vittoria\Desktop\MATERIALE TESI PROVA\Curva
di crescita MEDIANA\plot4.xlsx')
percorso5=('C:\\Users\Vittoria\Desktop\MATERIALE TESI PROVA\Curva
di crescita MEDIANA\plot5.xlsx')
percorso6=('C:\\Users\Vittoria\Desktop\MATERIALE TESI PROVA\Curva
di crescita MEDIANA\plot6.xlsx')
percorso7=('C:\\Users\Vittoria\Desktop\MATERIALE TESI PROVA\Curva
di crescita MEDIANA\Dati_totali_NETWORK.xlsx')

df=pd.read_excel(percorso)
df2=pd.read_excel(percorso2)
```

```
df3=pd.read_excel(percorso3)
df4=pd.read_excel(percorso4)
df5=pd.read_excel(percorso5)
df6=pd.read_excel(percorso6)
df7=pd.read_excel(percorso7)
#vengono create le matrici multidimensionali
dim=df.shape
plot1=np.full(dim, df)
dim2=df2.shape
plot2=np.full(dim2, df2)
dim3=df3.shape
plot3=np.full(dim3, df3)
dim4=df4.shape
plot4=np.full(dim4, df4)
dim5=df5.shape
plot5=np.full(dim5, df5)
dim6=df6.shape
plot6=np.full(dim6, df6)

dim7=df7.shape
stztot=np.full((dim7), df7)
```

7.4.2.2 *Costruzione della curva station year*

La costruzione della curva station year avviene normalizzando i dati rispetto al valore medio:

```
#normalizzazione station year rispetto alla media
stztot=np.insert(stztot, 3, 0, axis=1)
numeri = np.unique(stztot[:,0])
stztot=np.array([(stztot[stztot[:,0]==i]) for i in numeri] )
#calcolo mediana e normalizzazione
j=np.arange(0, len(stztot))
for row in j:
   stztot[row][:,3]=(stztot[row][:,2])/(np.mean(stztot[row][:,2]))
norm1=copy.deepcopy(stztot[:])
norm1=np.concatenate(norm1)
#si crea la matrice contentente i dati normalizzati del network
emp=np.full((dim7[0], 3), -999, float)
emp[:,0]=norm1[:,3]
emp=emp[emp[:,0].argsort()]
i=np.arange(1, len(norm1)+1)
Freq=(i-0.44)/(len(norm1)+0.12)
ysty=-np.log(-np.log(Freq))
emp[:,1]=ysty
emp[:,2]=Freq
per_rit=1/(1-Freq)
emp=np.insert(emp, 3, per_rit, axis=1)
emp=np.delete(emp, np.where(emp[:,1]<=-0.655), axis=0)
#emp=np.delete(emp, np.where(emp[:,1]>=6), axis=0)
```

7.4.2.3 *Costruzione della spezzata interpolante i punti*

I valori dei differenti network verranno raggruppati in un'unica matrice:

```
ploty=()
ploty=np.concatenate((plot1[:,1], plot2[:,1], plot3[:,1],
plot4[:,1], plot5[:,1], plot6[:,1]))
np.set_printoptions(precision=6, suppress=True)
plotxi=()
plotxi=np.concatenate((plot1[:,0], plot2[:,0], plot3[:,0],
plot4[:,0], plot5[:,0], plot6[:,0]))
#viene identificata la prima spezzata
dim=plotxi.shape[0]
matrice=np.empty((dim, 2), dtype=float)
matrice=np.insert(matrice, 0, ploty, axis=1)
matrice=np.insert(matrice, 1, plotxi, axis=1)
matrice=np.delete(matrice, (2,3), axis=1)
```

E' necessario impostare i parametri per la costruzione della prima interpolante con un numero di intervalli pari al numero di network complessivi:

```
#si definisce il numero di slices pari al numero di network
l=6
#si imposta il valore mediano ridotto
yf=0.3665
xf=1
#imposto le percentuali delle slices
ymin=min(matrice[:,0])
ymax=max(matrice[:,0])
pcp=20/100
pcd=80/100
#in tal modo arrontonda al valore esatto più vicino
lp=int(round((l*(pcp)), 0))
ld=int(round((l*pcd), 0))
#calcolo i valori delle y prima di yf
deltaip=((yf-ymin)/lp)
deltaid=((ymax-yf)/ld)
yp=np.empty(lp+1)
yp[0]=ymin
yp[lp]=yf
for i in np.arange(1,lp):
    yp[i]=yp[i-1]+deltaip
#calcolo i valori delle y dopo yf
yd=np.empty(ld+1)
yd[0]=yf
yd[ld]=ymax
for item in np.arange(0, ld):
    yd[item+1]=yd[item]+deltaid
y=np.concatenate((yp,yd))
y=np.unique(y)
#si creano le matrici contententi i dati per slices
psy=np.full((dim, l), '-999', dtype=float)
psx=np.full((dim, l), '-999', dtype=float)
```

I dati saranno classificati in modo da inserire i differenti valori nelle rispettive colonne delle matrici, identificative della slices di riferimento.

```
#i dati vengono classificati in base ai diversi intervalli
for ik in np.arange(1, l+1):
    for im in np.arange(1, len(ploty)+1):
        if ploty[im-1]>y[ik-1] and ploty[im-1]<y[ik]:
            psy[im-1, ik-1]=ploty[im-1]
            psx[im-1, ik-1]=plotxi[im-1]

psy[0,0]=matrice[0, 0]
psy[-1,-1]=matrice[-1,0]
psx[0,0]=matrice[0, 1]
psx[-1,-1]=matrice[-1,1]
#si filtra le matrici in modo da oscurare i valori non necessari
psx=np.ma.masked_inside(psx, -999,-1)
psy=np.ma.masked_inside(psy, -999, -1)
```

In questo modo per ogni colonna si calcoleranno i differenti gradienti:

```
#si calcolano i gradienti per ogni colonna intesa come una slice
m1=list(gradiente(psy[:,i], psx[:, i]) for i in np.arange(0, l))
```

Vengono definite le ordinate di modello associate agli estremi delle y:

```
#calcolo i valori delle x modello
xp=np.zeros(lp+1)
xp[-1]=xf

xd=np.zeros(ld+1)
xd[0]=xf
#definizione valori prima di xf
row=lp
it=lp
while row>0 and it>0:
    xp=np.insert(xp, lp, (xp[lp]-m1[it-1]*(y[row]-y[it-1])),
axis=0)
    row=row-1
    it=it-1
#definizione valori dopo xf
rows=0
il=lp
while row<=ld and il<l:
    xd=np.insert(xd, (rows+1), (xd[rows]+m1[il]*(y[il+1]-y[il])),
axis=0)
    rows=rows+1
    il=il+1
#
xp=np.delete(xp, np.where(xp[:]==0))
xd=np.delete(xd, np.where(xd[:]==0))
x=np.concatenate((xp,xd))
x=np.unique(x)
```

Si costruisce la matrice identificativa dei valori di modello stimati per l'intero set di dati:

```
xmodellp=np.full((dim, lp), -999, dtype=float)
```

```
xmodelld=np.full((dim, ld), -999, dtype=float)
#definizione dei valori di modello precedenti a xf
rowp=lp
itp=lp-1
while rowp>0 and itp>=0:
    xmodellp=np.insert(xmodellp,lp+1, (x[rowp]+m1[rowp-
1]*(psy[:,itp]-y[rowp])))
    rowp=rowp-1
    itp=itp-1
xmodellp=np.delete(xmodellp, np.where(xmodellp[:]==-999))
xmodellp=np.delete(xmodellp, np.where(xmodellp[:]<=-1))
#definizione dei valori di modello successivi a xf
rowd=lp
itd=lp
while rowd<=l and itd<l:
    xmodelld=np.append(xmodelld,(x[rowd] + m1[rowd]*(psy[:,itd] -
y[rowd])))
#    xmodelld=np.delete(xmodelld, np.where(xmodelld[:]==-999))
#    xmodelld=np.delete(xmodelld, np.where(xmodelld[:]<-1))
#    xmodelld=np.delete(xmodelld, np.where(xmodelld[:]>100))
    rowd=rowd+1
    itd=itd+1
#xmodelld=np.append(xmodelld, (x[rowd]+m1[rowd]*(psy[:,itd]-
y[rowd])))
xmodelld=np.delete(xmodelld, np.where(xmodelld[:]==-999))
xmodelld=np.delete(xmodelld, np.where(xmodelld[:]<-1))
xmodelld=np.delete(xmodelld, np.where(xmodelld[:]>100))
xmodell=np.concatenate((xmodellp, xmodelld))
if xmodell.shape!=plotxi.shape:
    xmodelld=np.insert(xmodelld, 0, 1)
    xmodell=np.concatenate((xmodellp, xmodelld))
```

7.4.2.4 *Discretizzazione dell'intervallo per ottenere la migliore spezzata*

Si applica la funzione riportata in [Appendice E: Codice delle funzioni costruite per la realizzazione dell'applicativo] relativa alla discretizzazione delle differenti slices:

```
xmodell8,x8, y8, m8=model(8, matrice, ploty, plotxi)
xmodell9, x9, y9, m9=model(9, matrice, ploty, plotxi)
xmodell10, x10, y10, m10=model(10, matrice, ploty, plotxi)
xmodell11, x11, y11, m11=model(11, matrice, ploty, plotxi)
xmodell12,x12, y12, m12=model(12, matrice, ploty, plotxi)
xmodell13, x13, y13, m13=model(13, matrice, ploty, plotxi)
xmodell14, x14, y14, m14=model(14, matrice, ploty, plotxi)
xmodell15,x15, y15, m15=model(15, matrice, ploty, plotxi)
xmodell16, x16, y16, m16=model(16, matrice, ploty, plotxi)
xmodell17, x17, y17, m17=model(17, matrice, ploty, plotxi)
xmodell18, x18, y18, m18=model(18, matrice, ploty, plotxi)
```

Si calcola il valore dell'equazione degli scarti ottenuta per le differenti curve di modello:

```
eq6=EQ1(xmodell, plotxi, 0.05, m1)
eq8=EQ1(xmodell8, plotxi, 0.05, m8)
eq9=EQ1(xmodell9, plotxi, 0.05, m9)
eq10=EQ1(xmodell10, plotxi, 0.05, m10)
eq11=EQ1(xmodell11, plotxi, 0.05, m11)
eq12=EQ1(xmodell12, plotxi, 0.05, m12)
eq13=EQ1(xmodell13, plotxi, 0.05, m13)
eq14=EQ1(xmodell14, plotxi, 0.05, m14)
eq15=EQ1(xmodell15, plotxi, 0.05, m15)
eq16=EQ1(xmodell16, plotxi, 0.05, m16)
eq17=EQ1(xmodell17, plotxi, 0.05, m17)
eq18=EQ1(xmodell18, plotxi, 0.05, m18)

#si costruisce il grafico delle differenti spezzate ottenute con
la nuvola di punti
fig=plt.figure(figsize=(6,6))
ax=plt.subplot(111)
ax.plot(matrice[:,0], matrice[:,1], color='orange', marker='.',
linestyle='', linewidth=0.05, label='NetSeries1')
ax.plot(y, x, color='c', marker='', linestyle='-', linewidth=2.5,
label='Retta spezzata con l=6')
ax.plot(y8, x8, color='m', marker='', linestyle='-',
linewidth=2.5, label='Retta spezzata con l=8')
ax.plot(y10, x10, color='b', marker='', linestyle='-',
linewidth=2.5, label='Retta spezzata con l=10')
ax.plot(y13, x13, color='r', marker='', linestyle='-',
linewidth=2.5, label='Retta spezzata con l=12')
ax.plot(y15, x15, color='y', marker='', linestyle='-',
linewidth=2.5, label='Retta spezzata con l=15')
ax.plot(y16, x16, color='c', marker='', linestyle='-',
linewidth=2.5, label='Retta spezzata con l=14')
ax.xaxis.set_ticks(y)
ax.yaxis.set_ticks([0,4], minor=True)
ax.xaxis.set_ticks_position('bottom')
ax.yaxis.set_ticks_position('left')
ax.patch.set_edgecolor('k')
ax.patch.set_linewidth(1)
ax.xaxis.grid(True, 'major', linewidth=1)
ax.yaxis.grid(True, 'major', linewidth=0.5)
ax.legend(loc='center left', bbox_to_anchor=(1, 0.5))
#ax.set_title('Rette spezzate', fontsize=20)
ax.set_xlabel('Reduced variate y')
ax.set_ylabel('Rainfall standardized x')

#il numero di slices è scelto in base al valore dell'eq1 risultato
#come più basso possibile
#si costruisce il grafico della spezzata a confronto con il set di
dati
fig=plt.figure(figsize=(8,8))
ax=plt.subplot(111)
#ax.plot(matrice[:,0], matrice[:,1], color='orange', marker='.',
linestyle='', linewidth=0.1, label='NetSeries1')
ax.plot(y14, x14, color='m', marker='', linestyle='-',
linewidth=2, label='Retta spezzata con L=14')
```

```
ax.plot(y14, x14, color='m', marker='o', linestyle='',
linewidth=2)
ax.xaxis.set_ticks(y11, minor=True)
ax.yaxis.set_ticks([0,4], minor=True)
ax.xaxis.set_ticks_position('bottom')
ax.yaxis.set_ticks_position('left')
ax.patch.set_edgecolor('k')
ax.patch.set_linewidth(1)
ax.xaxis.grid(True, 'major', linewidth=1)
ax.yaxis.grid(True, 'major', linewidth=0.5)
ax.legend(loc='center left', bbox_to_anchor=(1, 0.5))
#ax.set_title('Growth curves con P=0.05', fontsize=20)
ax.set_xlabel('Reduced variate y')
ax.set_ylabel('Rainfall standardized x')
```

7.4.2.5 *Ricerca dei parametri che garantiscono il best fitting*

Si applica la funzione 'curve_fit', note le funzioni per le quali ricavare i parametri che garantiscono un migliore adattamento

```
#curve fitting con la funzione nota:
#si ricavano le nuove x in funzione delle y
if xmodell14.shape==plotxi.shape:
    a, b=curve_fit(GEV, ploty, xmodell14, bounds=(-0.7, 6))
    print('I parametri stimata con la funzione sono:', a)
    alfa=a[0]
    epsilon=a[1]
    k=a[2]
    xGev=GEV(ploty, alfa, epsilon, k)
    plt.plot(ploty, xGev, linewidth=3, color='g', label='GEV,
FORGEX')
    c,d=curve_fit(GUMBEL, ploty, xmodell14, bounds=(-0.7, 6))
    print('I parametri stimata con la funzione sono: ' , c)
    alfa1=c[0]
    epsilon1=c[1]
    xGumbel=GUMBEL(ploty, alfa1, epsilon1)
    plt.plot(ploty, xGumbel, linewidth=3, color='orange',
label='GUMBEL')
    plt.legend()
else:
    a, b=curve_fit(GEV, y12, x12, bounds=(-0.7, 6))
    print('I parametri stimata con la funzione sono:', a)
    alfa=a[0]
    epsilon=a[1]
    k=a[2]
    xGev=GEV(y12, alfa, epsilon, k)
    plt.plot(y12, xGev, linewidth=3, color='g', label='GEV')
    c,d=curve_fit(GUMBEL, y12, x12, bounds=(-0.7, 6))
    print('I parametri stimata con la funzione sono: ' , c)
    alfa1=c[0]
    epsilon1=c[1]
    xGumbel=GUMBEL(y12, alfa1, epsilon1)
    plt.plot(y12, xGumbel, linewidth=3, color='orange',
label='GUMBEL')
    plt.legend(loc='center left', bbox_to_anchor=(1, 0.5))
    diffGEV=sum((xGev-x12)**2)
```

```
    if diffGEV<1:
        print('La distribuzione scelta Ã¨ adatta al numero di
slice scelta')
    else:
        print('Modifica gli intervalli L')
```

7.4.2.6 *Costruzione della curva empirica di stazione*

Si ricava la serie di dati della stazione strumentata per la costruzione della curva empirica di stazione.

```
percorso3=('C:\\Users\Vittoria\Desktop\Tesi
Vittoria\ANAGRAFICA.xlsx')
percorso4=('C:\\Users\Vittoria\Desktop\Tesi Vittoria\DATI.txt')
info=pd.read_excel(percorso3)
stz=pd.read_csv(percorso4,sep='\t', dtype='float')
n=np.array(info[info.iloc[:,1]=='Spigno_Monferrato'].iloc[0,0])
dati=stz[stz.iloc[:,0]==n]
#procedura di identificazione del valore mediano per il sito noto
mediana=np.array(dati.groupby(by='Numero').median())
MEDIAcal=np.array(dati.groupby(by='Numero').mean())
N=np.array(dati.groupby(by='Numero').count().iloc[:,0])
#verifica della numerosità del set di dati
if N>10:
    print('L\'analisi procede con un grado di precisione elevato')
else:
    print('Il numero di dati nel sito non è¨ sufficiente a
garantire una precisione elevata, N=', N)
```

Verrà richiesto all'utente di selezionare la durata per quale si esegue l'analisi. In questo modo, a seconda del valore selezionato, si estrarranno i dati corrispondenti e il valore medio utilizzato per la normalizzazione della serie.

```
d=int(input('Inserisci la durata di analisi :'))
if d==1:
    index=mediana[:,1]
    mediacalc=MEDIAcal[:,1]
    norm=dati.iloc[:,2]/MEDIAcal[:,1]
else:
    if d==3:
        index=mediana[:,2]
        mediacalc=MEDIAcal[:,2]
        norm=dati.iloc[:,3]/MEDIAcal[:,2]
    else:
        if d==6:
            index=mediana[:,3]
            mediacalc=MEDIAcal[:,3]
            norm=dati.iloc[:,4]/MEDIAcal[:,3]
        else:
            if d==12:
                index=mediana[:,4]
                mediacalc=MEDIAcal[:,4]
                norm=dati.iloc[:,5]/MEDIAcal[:,4]
            else:
```

```
if d==24:
    index=mediana[:,5]
    mediacalc=MEDIAcal[:,5]
    norm=dati.iloc[:,6]/MEDIAcal[:,5]
```

Si calcolerà della serie in esame la frequenza e i valori della variabile ridotta:

```
i=np.arange(1, len(norm)+1)
frequenza=(i-0.44)/(N+0.12)
Y=-np.log(-np.log(frequenza))
Tempo=1/(1-frequenza)
mediastimata=float(input('Inserire il valore della media del sito
scelto'))
val=max(emp[:,1])
m=max(ploty)
#si allungano le serie di dati sulla base del range di dati del
station year
yagg=np.linspace(m+0.1, val)
ploty=np.append(ploty, yagg)
xGev=GEV(ploty, alfa, epsilon, k)
xF=(xGev*index)/mediastimata
norm=np.array(norm.sort_values())
dim1=Y.shape[0]
#i valori sono inseriti in una matrici per essere visualizzati e
richiamati più velocemente
valori=np.full((dim1, 2), 0, dtype=float)
valori=np.insert(valori, 0, Y, axis=1)
valori=np.insert(valori, 1, norm, axis=1)
valori=np.insert(valori, 2, Tempo, axis=1)
valori=np.delete(valori, np.where(valori[:,0]<=-0.655), axis=0)
#viene costruito il grafico delle distribuzioni FORGEX, ROSSO e
KRIGING in grafici differenti
fig=plt.figure(figsize=(8,8))
ax=plt.subplot(111)
#ax.plot(ploty, xFcal, color='orange', marker='', linestyle='-',
label='GEV, FORGEX, corretta (media calcolata)', linewidth=3)
#ax.plot(Y, frequenza, color='y', marker='', linestyle='-',
label='Curva di frequenza empirica di stazione')
ax.plot(ploty, xF, color='g', marker='', linestyle='-',
label='GEV, FORGEX, corretta', linewidth=3)
ax.plot(emp[:,1], emp[:,0], color='y', marker='o', linestyle='',
label='Curva Station Year')
ax.plot(valori[:,0], valori[:,1], color='c', marker='o',
linestyle='', label='Curva di frequenza empirica di stazione')
ax.xaxis.set_ticks_position('bottom')
ax.yaxis.set_ticks_position('left')
ax.patch.set_edgecolor('k')
ax.patch.set_linewidth(1)
ax.xaxis.grid(False, 'major', linewidth=1)
ax.yaxis.grid(False, 'major', linewidth=0.5)
ax.set_xlabel('Reduced variate y')
ax.set_ylabel('Rainfall standardized x')
ax.legend()
fig=plt.figure(figsize=(8,8))
ax=plt.subplot(111)
```

```
#ax.plot(ploty, xFcal, color='orange', marker='', linestyle='-',
label='GEV, FORGEX, corretta (media calcolata)', linewidth=3)
#ax.plot(Y, frequenza, color='y', marker='', linestyle='-',
label='Curva di frequenza empirica di stazione')
#ax.plot(emp[:,1], emp[:,0], color='y', marker='o', linestyle='',
label='Curva Station Year')
#ax.plot(valori[:,0], valori[:,1], color='c', marker='o',
linestyle='', label='Curva di frequenza empirica di stazione')
ax.plot(ploty, xF, color='g', marker='', linestyle='-',
label='GEV, FORGEX, corretta', linewidth=3)
ax.xaxis.grid(True, 'major', linewidth=1)
ax.yaxis.grid(True, 'major', linewidth=0.5)
ax.set_xlabel('Reduced variate y')
ax.set_ylabel('Rainfall standardized x')
#ax.legend()
```

7.4.2.7 *Costruzione delle curve di confronto*

Si inseriscono i parametri appartenenti ai differenti metodi:

```
#i parametri stimati con il rosso sono i seguenti:
ALFA=float(input('Inserire il valore di alfa per il sito in esame:
'))
EPSILON=float(input('Inserire il valore di epsilon per il sito in
esame:'))
KAPPA=float(input('Inserire il valore di kappa per il sito in
esame: '))
valuemaxR=GEV(val, ALFA, EPSILON, KAPPA)
#si calcolano le ordinate
xR=GEV(ploty, ALFA, EPSILON, KAPPA)
plt.plot(ploty, xR, color='b', label='GEV, Rosso-DeMichele',
linewidth=3)
#plt.legend()
#i parametri del kriging sono i seguenti:
theta2=float(input('Inserire il valore del parametro theta2 per il
sito in esame: '))
theta3=-(float(input('Inserire il valore del parametro theta3 per
il sito in esame: ')))
#si calcolano le ordinate
F=np.exp(-np.exp(-ploty))
xK=1-(theta2/theta3)*((np.exp(gammaln(1-theta3)))-(-np.log(F))**(-
theta3))
plt.plot(ploty, xK, color='r', label='GEV, Kriging ARPA',
linewidth=3)
ax.plot(emp[:,1], emp[:,0], color='y', marker='o', linestyle='',
label='Curva Station Year')
ax.plot(valori[:,0], valori[:,1], color='c', marker='o',
linestyle='', label='Curva di frequenza empirica di stazione')
plt.legend()
#si costruiscono gli stessi grafici in funzione del T
T=1/(1-np.exp(-((1-(k/alfa)*(xF-epsilon))**(1/k))))
fig=plt.figure(figsize=(8,8))
ax=plt.subplot(111)
ax.plot(T, xF, color='g', marker='x', label='GEV, FORGEX
corretta')
ax.set_xscale('log')
T1=(1/(1-np.exp(-(1-(KAPPA/ALFA)*(xR-EPSILON))**(1/KAPPA))))
```

```
ax.plot(T1, xR, color='b', marker='x', label='GEV, Rosso-
DeMichele')
#ax.set_xscale('log')
#ax.xaxis.grid(True, 'minor', linewidth=1)
#ax.yaxis.grid(True, 'major', linewidth=0.5)
#ax.set_xlabel('Tempo di ritorno (anni) T')
#ax.set_ylabel('Rainfall standardized x')
#ax.legend()
T2=1/(1-F)
ax.plot(T2, xK, color='r', marker='x', label='GEV, Kriging ARPA')
ax.plot(valori[:,2],  valori[:,1], color='c',
marker='o',linestyle='', label='Curva di frequenza empirica di
stazione')
ax.set_xscale('log')
ax.xaxis.grid(True, 'minor', linewidth=1)
ax.yaxis.grid(True, 'major', linewidth=0.5)
ax.set_xlabel('Tempo di ritorno (anni) T')
ax.set_ylabel('Rainfall standardized x')
ax.legend()
```

La procedura implementa il calcolo della grandezza di progetto per le differenti distribuzioni:

```
#FORGEX
hF=mediastimata*xF
#hFcal=mediastimata*xFcal
fig=plt.figure(figsize=(8,8))
ax=plt.subplot(111)
ax.plot(ploty, hF, marker='o', color='g', label='CPP, FORGEX,
corretta')
#ax.plot(T, hF, marker='o', color='g', label='CPP, FORGEX,
corretta')
#ax.plot(ploty, hFcal, color='orange', marker='o', linestyle='-',
label='CPP, FORGEX, corretta (media calcolata)', linewidth=3)
ax.legend()
#ROSSO
hR=mediastimata*xR
ax.plot(ploty, hR, marker='o', color='b', label='CPP, Rosso-
DeMichele')
ax.legend()
#KRIGING
hK=mediastimata*xK
ax.plot(ploty, hK, marker='o', color='r', label='CPP, Kriging
ARPA')
ax.legend()
ax.xaxis.grid()
ax.yaxis.grid()
ax.set_xlabel('Variabile ridotta y')
ax.set_ylabel('Altezza di precipitazione h')
```

7.5 *Appendice E: Codice delle funzioni costruite per la realizzazione dell'applicativo*

7.5.1 *Parte A*

In questo capitolo sono definite le funzione in ordine di utilizzo, per il primo script utilizzato per la stima delle diverse curve di crescita.

```
#funzione per la verifica della numerosità della serie di dati
def numerosita(n):
    if n>=10:
        return 1
    else:
        return -1

#funzione per la selezione della durata d'esame.
def test(d):
    while True:
        if d not in durate:
            print('Le durate accettabili sono 1,3,6,12,24 ore')
            d=int(input('Per quale durata si vuole eseguire
l\'analisi?'))
        else:
            if d==1:
                datid=valid_station[:,[0,1,2]]
                return datid
            else:
                if d==3:
                    datid=valid_station[:,[0,1,3]]
                    return datid
                else:
                    if d==6:
                        datid=valid_station[:,[0,1,4]]
                        return datid
                    else:
                        if d==12:
                            datid=valid_station[:,[0,1,5]]
                            return datid
                        else:
                            if d==24:
                                datid=valid_station[:,[0,1,6]]
                                return datid

#funzione utilizzata per la classificazione dei differenti
#valori delle variabili ridotte y
def classify(n):
    if n<-0.6335:
        return 'nan'
    else:
        if n<=0.3665 and n>=-0.6335:
```

```
            return 0
        else:
            if n<=1.3665 and n>0.3665:
                return 1
            else:
                if n<=2.3665 and n>1.3665:
                    return 2
                else:
                    if n<=3.3665 and n>2.3665:
                        return 3
                    else:
                        if n<=4.3665 and n>3.3665:
                            return 4
                        else:
                            if n<=5.3665 and n>4.3665:
                                return 5
                            else:
                                if n<=6.3665 and n>5.3665:
                                    return 6
                                else:
                                    if n<=7.3665 and n>6.3665:
                                        return 7
                                    else:
                                        if n>7.3665:
                                            return 8


#calcolo della numerosità dei dati per ogni slices
def num_per_slice(elemento, infoGenerali):
    df=pd.DataFrame(elemento)
    Int=np.array((df.groupby(by=0).count()).iloc[:,0])
    inters=np.intersect1d(elemento[:,0], infoGenerali[:,0])
    anag=infoGenerali[np.any(infoGenerali[:,0]==inters[:,None],
axis=0)]
    anag=np.insert(anag, 5, Int, axis=1)
    anag=np.delete(anag, (1,2,3), axis=1)
    return anag

#funzione che identifica i valori differenti tra i set di dati
#ottenuti per ogni slices.
def diversi(valori1, valori2):
    intersezione4=np.setxor1d(valori1[:,0], valori2[:,0])
    dim=intersezione4.shape[0]
    matvuota=np.zeros((dim,1))
    matvuota=np.insert(matvuota, 0, intersezione4, axis=1)
    matvuota=np.insert(matvuota, 2, 0, axis=1)
    valori2=np.append(valori2, matvuota, axis=0)
    valori2=valori2.reshape(-1, 3)
    return valori2

#funzione per la classificazione dei network
def network(insieme, numslice):
    valide=insieme[insieme[:,1]<=200000]
    valide=np.insert(valide, 10, 0, axis=1)
    n=valide.shape[0]
    num=0
    network=()
    cont=0
```

```
    somma=0
    while somma<=20 and num<n:
        if valide[cont,10]==0:
            network=np.append(network, valide[cont,:])
            somma=somma+valide[cont,numslice]
            valide[cont,10]=1
            cont=cont+1
        num=num+1
    network=network.reshape(-1,11)
    return network

#funzione che identifica i valori delle slices dei diversi network
def valori_slice(NETWORK, dati_slice):
    Y3=np.intersect1d(NETWORK[:,0], dati_slice[:,0])
    Y3=dati_slice[np.any(dati_slice[:,0]==Y3[:,None], axis=0)]
    return Y3

#definizione dei valori di y da utilizzare per il grafico
def valori_gumbel(network):
    intersect1=np.intersect1d(network[:,0], norm[:,0])
    norm1=norm[np.any(norm[:, 0] == intersect1[:, None], axis=0)]
 #tutti i valori delle ordinate
    intersect2=np.intersect1d(network[:,0], y[:,0])
    gumbel1=y[np.any(y[:,0]==intersect2[:,None], axis=0)]
#tutti i valori delle ascisse
    norm1=np.insert(norm1, 4, gumbel1[:,4], axis=1)
    norm1=np.insert(norm1, 5, gumbel1[:,5], axis=1)
    df=pd.DataFrame(norm1).dropna()
    dimensione=df.shape
    norm1=np.full(dimensione, df)
    return norm1

#calcolo della serie di netmax
def netmax(NORM):
    max1=pd.DataFrame(NORM)
    max1=max1.groupby(by=1).max()
    d=max1.shape
    netmax1=np.full(d, max1)
    netmax1=netmax1[netmax1[:,3].argsort()]
    return netmax1

#calcolo della distanza di offset della serie di netmax
def offset(NETWORK):
    N=NETWORK.shape[0]
    if N==1:
        N=0.01
        lnNe=0
        return lnNe
    else:
        distanza=(np.mean(np.ediff1d(NETWORK[:,1])))/1000
        area=((2.5)*(distanza**2))
        print(area)
        if d>24:
            D=d
        else:
            if d<15:
                D=d/18
            else:
```

```
                D=d/24
        lnNe=np.log(N)*(0.081+0.085*np.log(area)-0.051*np.log(N)-
0.027*np.log(D))
        print('La distanza di offset Ã¨: ', lnNe)
        Ne=np.exp(lnNe)
        print('Il numero di stazioni indipendenti Ã¨: ', Ne)
        return lnNe
#calcolo delle variabili della F e delle y per la serie di netmax
def frequenza(insieme, lnNe_n):
    N=insieme.shape[0]
    insieme=np.insert(insieme, 4, 0, axis=1)
    insieme=np.insert(insieme, 5, 0, axis=1)
    insieme[:,5]=np.arange(1, len(insieme)+1)
#azzera l'indice
#    lnNe_n=offset(insieme) # calcola la distanza di offset
    insieme[:,4]=((insieme[:,5]-0.44)/(N+0.12)) #valore della
frequenza
    insieme[:,3]=(-np.log(-np.log(insieme[:,4])))+np.abs(lnNe_n)
    return insieme

#classifica i valori dei netmax per ogni slices.
def net_valid(insieme_netmax):
    insieme_netmax=np.delete(insieme_netmax, 6, axis=1)
    insieme_netmax=np.insert(insieme_netmax, 6, list(map(classify,
insieme_netmax[:,3])),axis=1)
    df=pd.DataFrame(insieme_netmax).dropna()
    dimensione=df.shape
    valido=np.full(dimensione, df)
    return valido

#classificazione dei valori per le singole slice
def single_slice(slice_num, massimi_net):
    slice1=np.intersect1d(slice_num, massimi_net)
    slice1=massimi_net[np.any(massimi_net[:,2]==slice1[:,None],
axis=0)]
    X=np.concatenate((slice_num[:,4], slice1[:,3]), axis=0)
    X=np.sort(X)
    Y=np.concatenate((slice_num[:,3], slice1[:,2]), axis=0)
    Y=np.sort(Y)
    return X, Y

#definizione del calcolo del gradiente
def gradiente(x,y):
    if x.shape[0]<=2 and y.shape[0]<=2:
        grad=None
    else:
        grad=(((x.mean())*(y.mean()))-
((x*y).mean()))/(((x.mean())**2)-(x**2).mean())
    return grad
```

7.5.2 *Parte B*

In questo capitolo sono definite le funzione in ordine di utilizzo, per il secondo script utilizzato per la stima delle diverse curve di crescita.

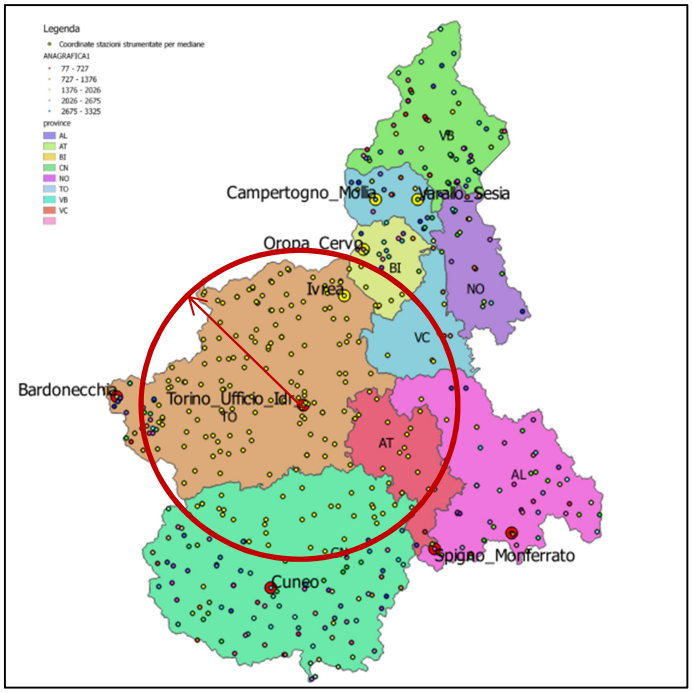
```
#funzione identificativa dei valori di modello per la costruzione
della spezzata
def model(L, matrix, ploty,plotxi):
    yf=0.3665
    xf=1
    #imposta le percentuali
    ymin=min(matrix[:,0])
    ymax=max(matrix[:,0])
    pcp=20/100
    pcd=80/100
    #in tal modo arrontonda al valore esatto più vicino
    lp=int(round((L*(pcp)), 0))
    ld=int(round((L*pcd), 0))
    #calcola i valori delle y prima di yf
    deltaip=((yf-ymin)/lp)
    deltaid=((ymax-yf)/ld)
    yp=np.empty(lp+1)
    yp[0]=ymin
    yp[lp]=yf
    for i in np.arange(1,lp):
        yp[i]=yp[i-1]+deltaip
    #calcola i valori delle y dopo yf
    yd=np.empty(ld+1)
    yd[0]=yf
    yd[ld]=ymax

    for item in np.arange(0, ld):
        yd[item+1]=yd[item]+deltaid

    y=np.concatenate((yp,yd))
    y=np.unique(y)

    psy=np.full((dim, L), '-999', dtype=float)
    psx=np.full((dim, L), '-999', dtype=float)

    for ik in np.arange(1, L+1):
        for im in np.arange(1, len(ploty)+1):
            if ploty[im-1]>y[ik-1] and ploty[im-1]<y[ik]:
                psy[im-1, ik-1]=ploty[im-1]
                psx[im-1, ik-1]=plotxi[im-1]

    psy[0,0]=matrix[0, 0]
    psy[-1,-1]=matrix[-1,0]
    psx[0,0]=matrix[0, 1]
    psx[-1,-1]=matrix[-1,1]
    #filtra le matrici in modo da oscurare i valori non necessari
    psx=np.ma.masked_inside(psx, -999,-1)
    psy=np.ma.masked_inside(psy, -999, -1)
    #calcola i gradienti per ogni colonna intesa come una slice
    m1=list(gradiente(psy[:,i], psx[:, i]) for i in np.arange(0,
L))
```

```
    #calcola i valori delle x modello
    xp=np.zeros(lp+1)
    xp[-1]=xf

    xd=np.zeros(ld+1)
    xd[0]=xf
    #definizione valori prima di xf
    row=lp
    it=lp
    while row>0 and it>0:
        xp=np.insert(xp, lp, (xp[lp]-m1[it-1]*(y[row]-y[it-1])),
axis=0)
        row=row-1
        it=it-1
    #definizione valori dopo xf
    rows=0
    il=lp
    while row<=ld and il<L:
        xd=np.insert(xd, (rows+1), (xd[rows]+m1[il]*(y[il+1]-
y[il])), axis=0)
        rows=rows+1
        il=il+1
    xp=np.delete(xp, np.where(xp[:]==0))
    xd=np.delete(xd, np.where(xd[:]==0))
    x=np.concatenate((xp,xd))
    x=np.unique(x)

    #definizione dei valori di modello
    xmodellp=np.full((dim, lp), -999, dtype=float)
    xmodelld=np.full((dim, ld), -999, dtype=float)

    rowp=lp
    itp=lp-1
    while rowp>0 and itp>=0:
        xmodellp=np.insert(xmodellp,lp+1, (x[rowp]+m1[rowp-
1]*(psy[:,itp]-y[rowp])))
        rowp=rowp-1
        itp=itp-1

    xmodellp=np.delete(xmodellp, np.where(xmodellp[:]==-999))
    xmodellp=np.delete(xmodellp, np.where(xmodellp[:]<=-1))

    rowd=lp
    itd=lp

    while rowd<=L and itd<L:
        xmodelld=np.append(xmodelld,(x[rowd] +
m1[rowd]*(psy[:,itd] -y[rowd])))
    #    xmodelld=np.delete(xmodelld, np.where(xmodelld[:]==-999))
    #    xmodelld=np.delete(xmodelld, np.where(xmodelld[:]<-1))
    #    xmodelld=np.delete(xmodelld, np.where(xmodelld[:]>100))
        rowd=rowd+1
        itd=itd+1

    #xmodelld=np.append(xmodelld, (x[rowd]+m1[rowd]*(psy[:,itd]-
y[rowd])))
    xmodelld=np.delete(xmodelld, np.where(xmodelld[:]==-999))
    xmodelld=np.delete(xmodelld, np.where(xmodelld[:]<0))
```

```
    xmodelld=np.delete(xmodelld, np.where(xmodelld[:]>100))
    xmodell=np.concatenate((xmodellp, xmodelld))
    if xmodell.shape==plotxi.shape:
        return xmodell, x, y, m1
    else:
        xmodelld=np.insert(xmodelld, 0, 1)
        xmodell=np.concatenate((xmodellp, xmodelld))
        return xmodell, x, y, m1

#funzione per il calcolo dell'equazione agli scarti minimi
def EQ1(XMODELL, Xi, parametro, M):
    if XMODELL.shape==Xi.shape:
        k=2
        vett=()
        while k<len(M):
            diff=(M[k]-M[k-1])**2
            vett=np.append(vett, diff)
            k=k+1
        somma=sum(vett)
        mat=()
        dif=(Xi-XMODELL)**2
        mat=np.append(mat, dif)
        som=sum(mat)
        Mt=len(Xi)
        var=Mt/(l-1)
        eq1=som+parametro*var*somma
        return eq1
    else:
        return 'I valori ottenuti non sono idonei per la scelta'

#funzione per ricavare la distribuzione GEV
def GEV(Y, ALFA, EPSILON, K):
    return EPSILON-(ALFA/K)*(np.exp(-K*Y)-1)

#funzione per ricavare la distribuzione GUMBEL
def GUMBEL(Y, ALFA, EPSILON):
    return EPSILON+(ALFA*Y)
```

7.6 *Appendice F: Schede delle località analizzate*

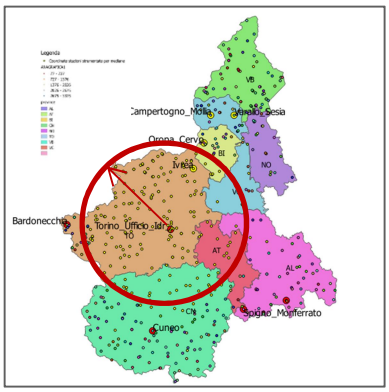
7.6.1 *Torino Ufficio Idrografico*



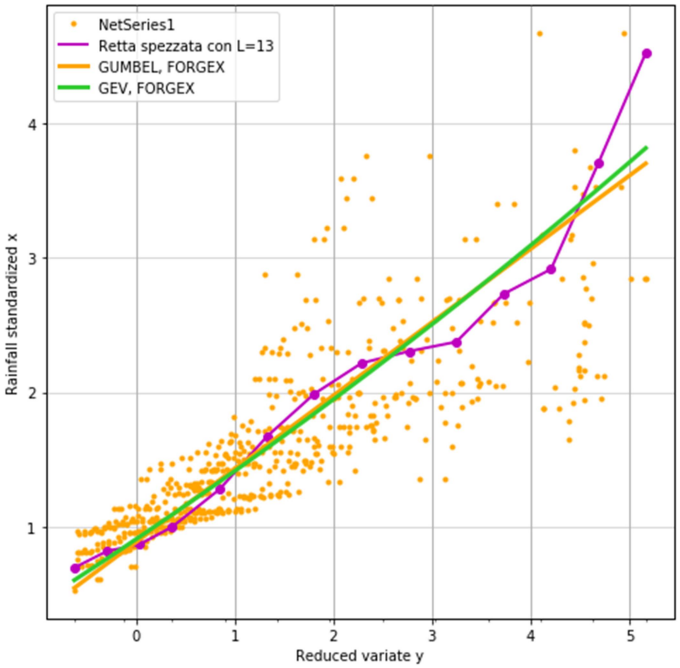
Localizzazione	*X=395177*
	Y=4991310
Raggio (km)	*76*
Numero stazioni	*191*
Numerosità dati stazione strumentata	*60*

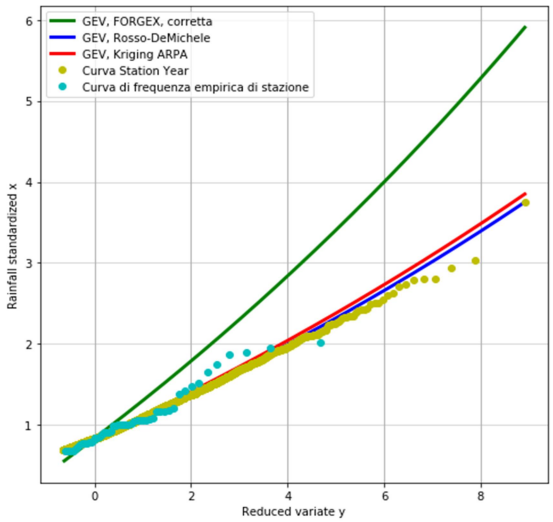
Durata (h)	N° segmenti	Valore Scarti minimo
1	*13*	*127*
3	*11*	*120*
6	*10*	*121*
12	*11*	*190*
24	*14*	*120*

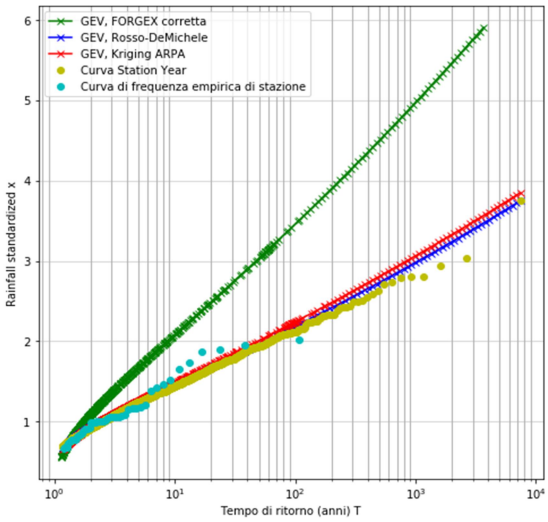
7.6.1.1 *Torino Ufficio Idrografico: durata 1h*

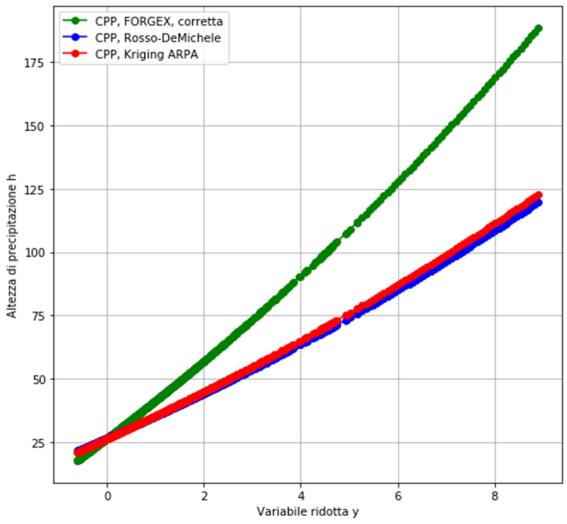


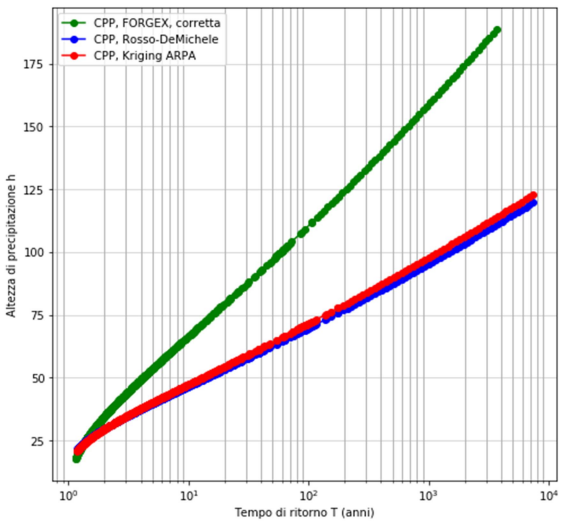
FORGEX		
α	ε	k
0.494	0.907	-0.049
VAPI		
α	ε	k
0.262	0.836	-0.048
Kriging		
θ2*	θ3	-
0.27755	-0.04354	-



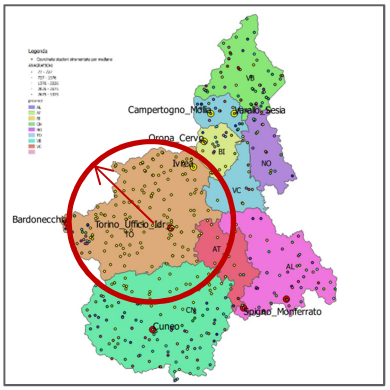




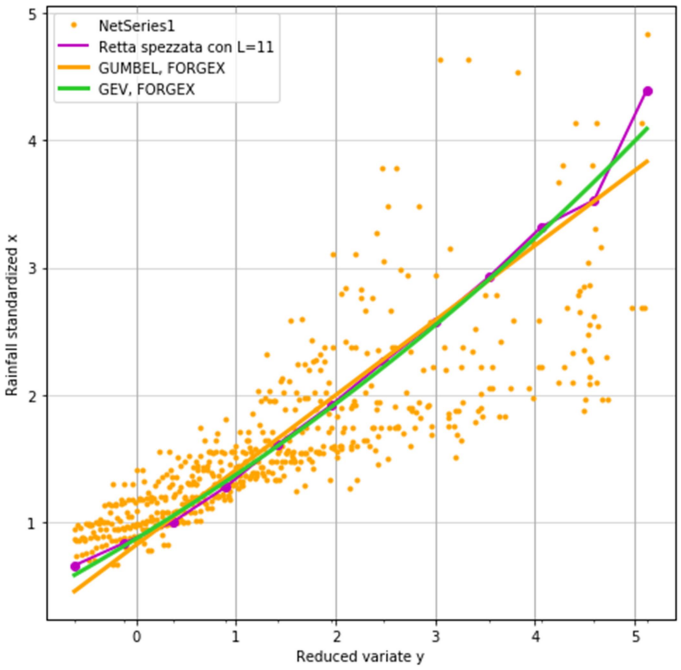


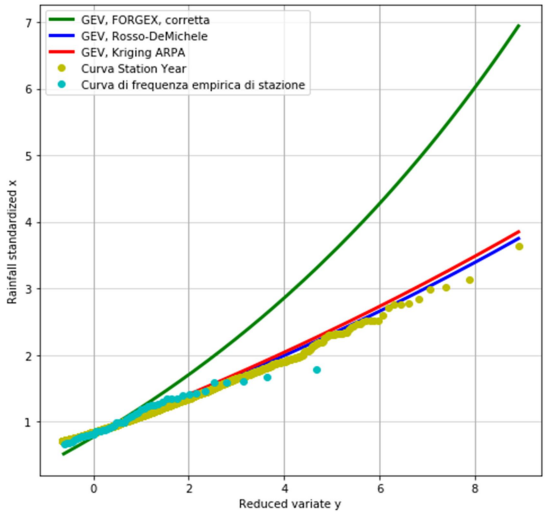


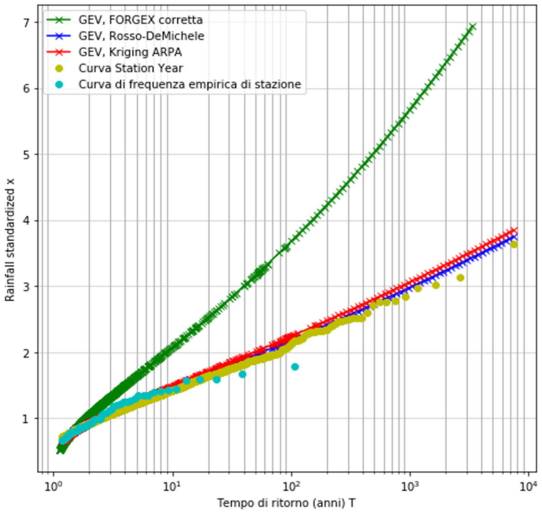
7.6.1.2 *Torino Ufficio Idrografico: durata 3h*

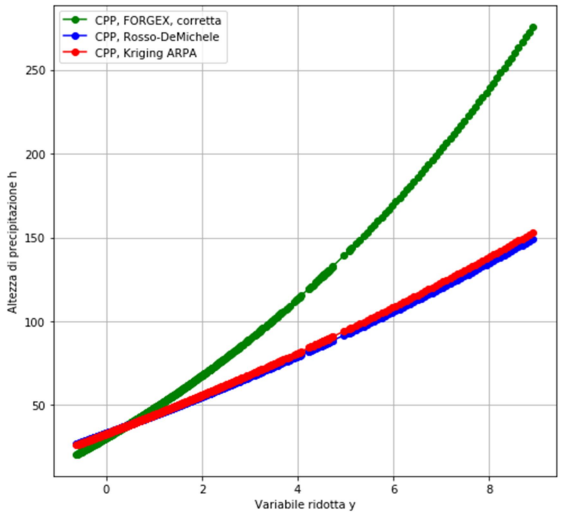


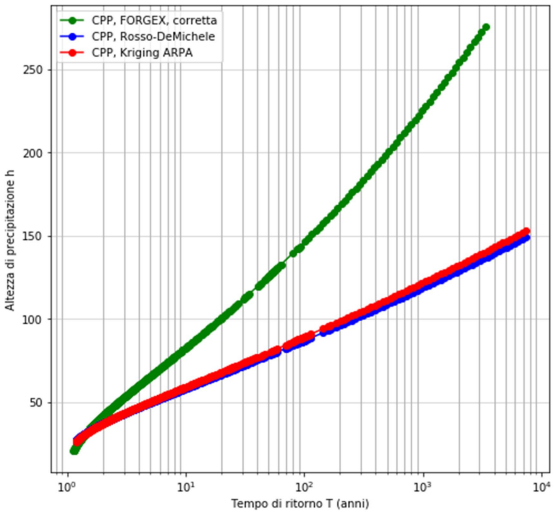
FORGEX		
α	ε	k
0.476	0.869	-0.103
VAPI		
α	ε	k
0.262	0.836	-0.048
Kriging		
θ2*	θ3	-
0.27755	-0.04354	-



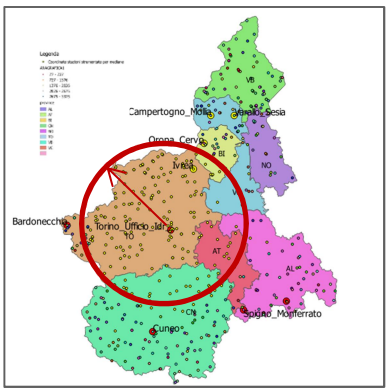




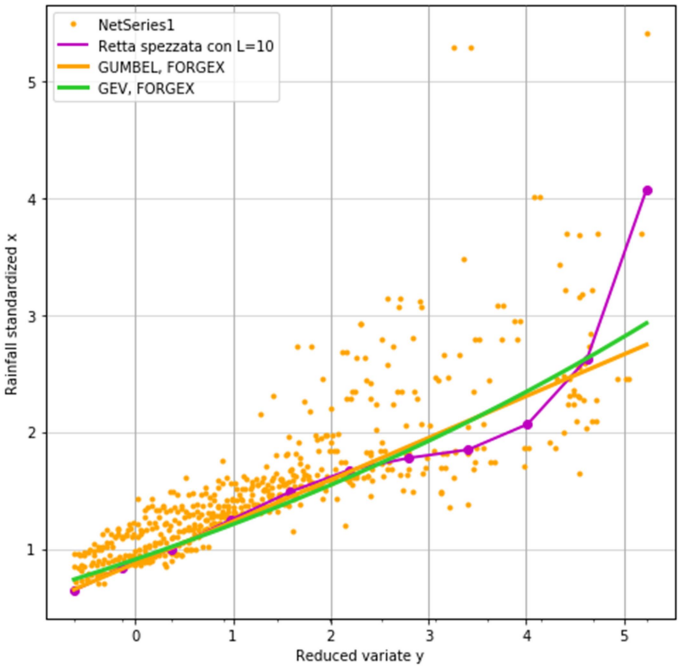


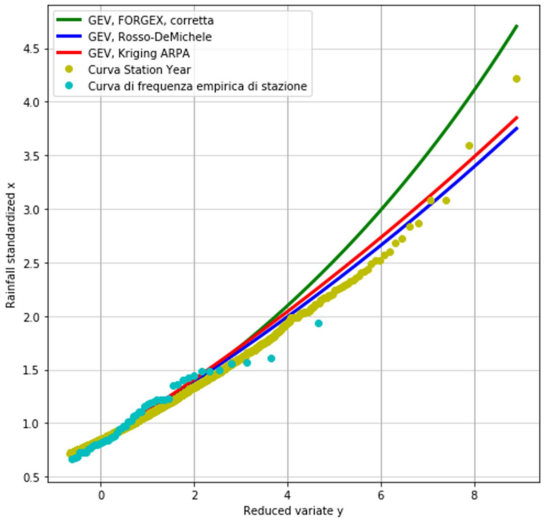


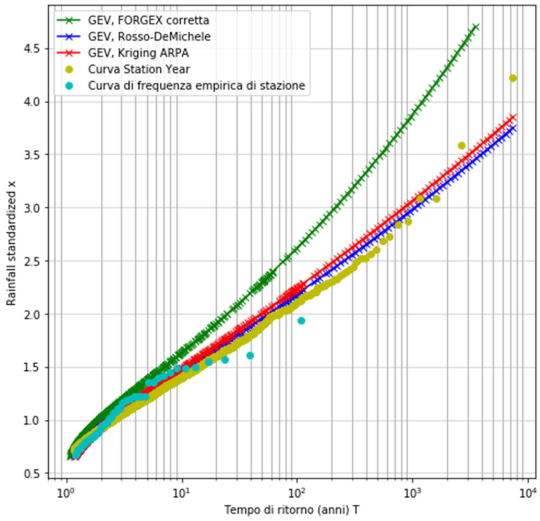
7.6.1.3 *Torino Ufficio Idrografico: durata 6h*

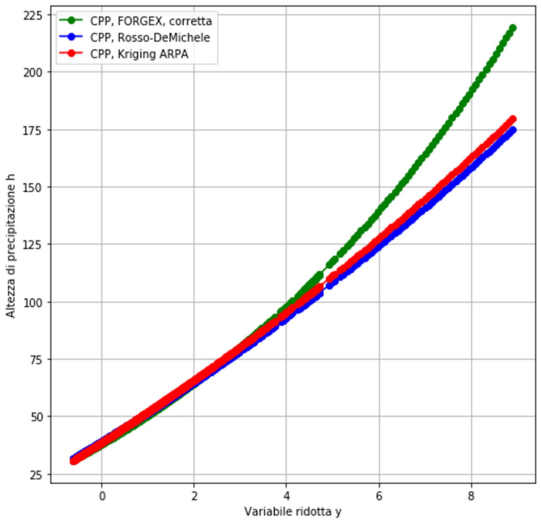


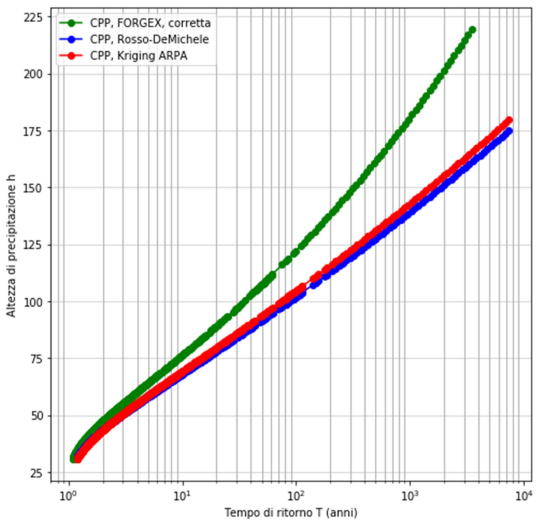
FORGEX		
α	ε	k
0.283	0.910	-0.112
VAPI		
α	ε	k
0.262	0.836	-0.048
Kriging		
θ2*	θ3	-
0.27755	-0.04354	-



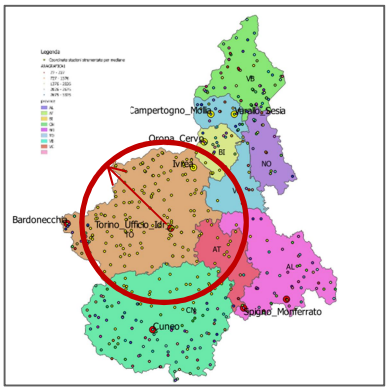


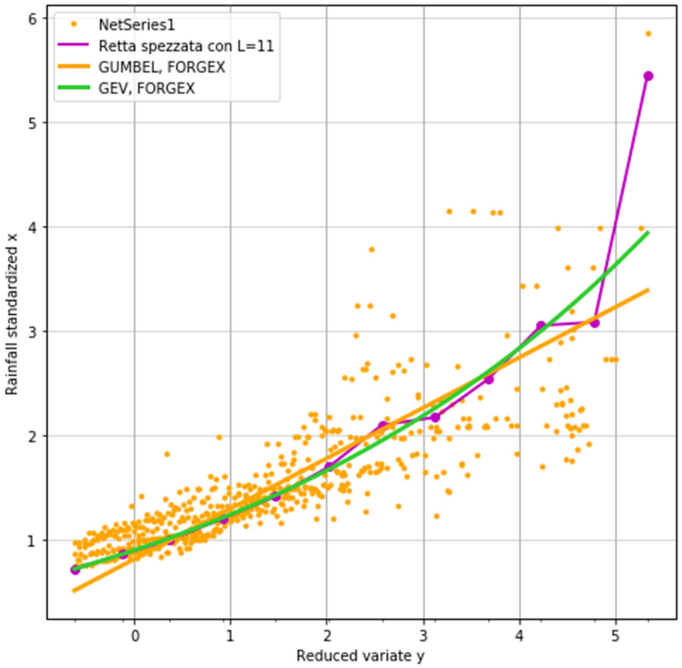




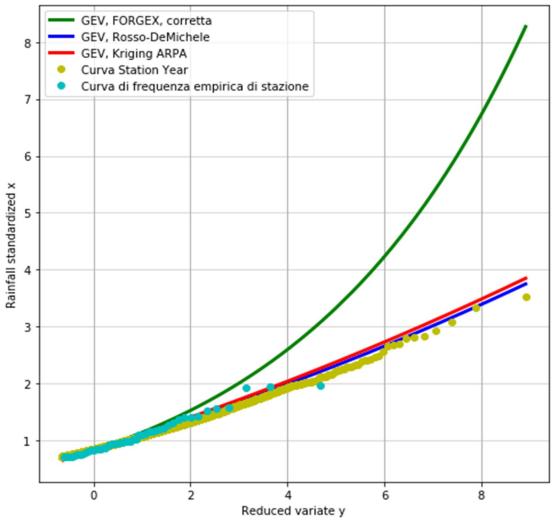


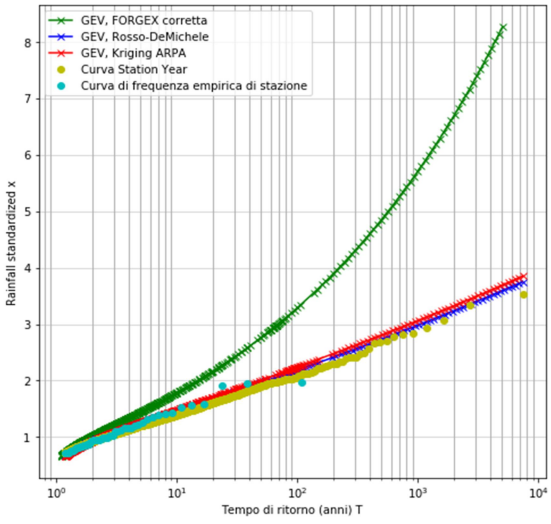
7.6.1.4 *Torino Ufficio Idrografico: durata 12h*

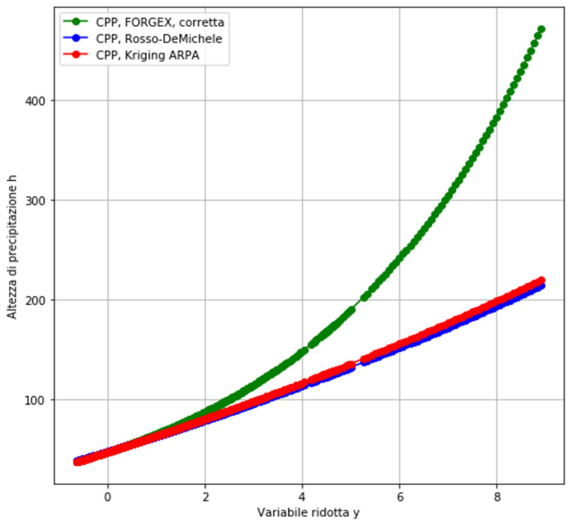


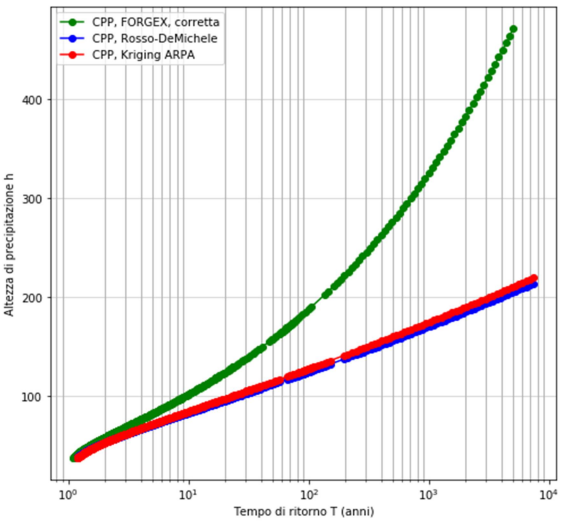


FORGEX		
α	ε	k
0.306	0.904	-0.212
VAPI		
α	ε	k
0.262	0.836	-0.048
Kriging		
θ2*	θ3	-
0.27755	-0.04354	-

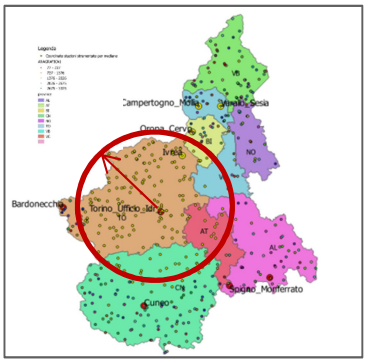




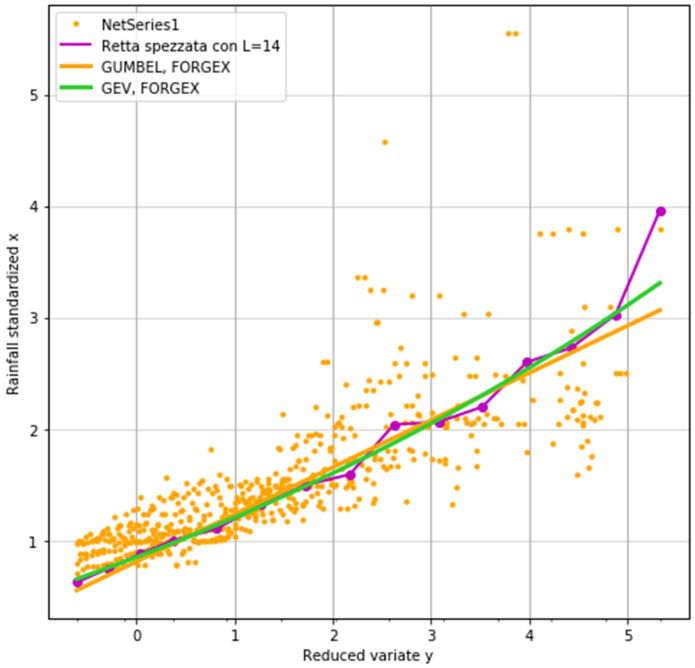


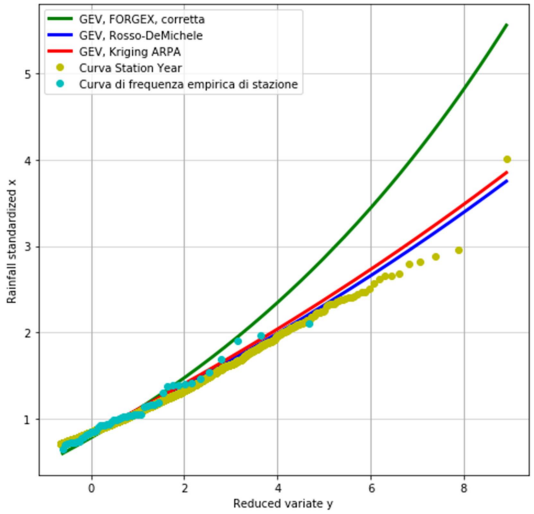


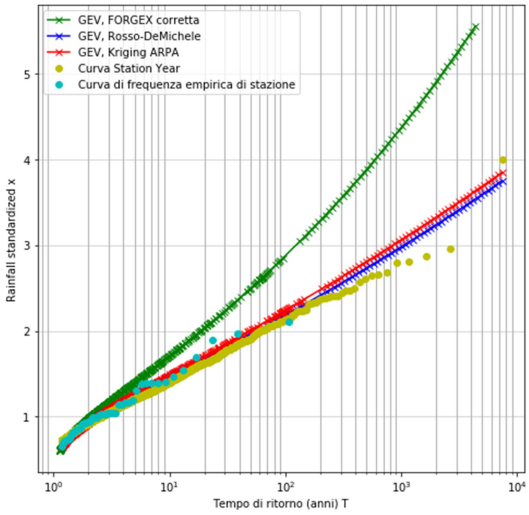
7.6.1.5 *Torino Ufficio Idrografico: durata 24h*

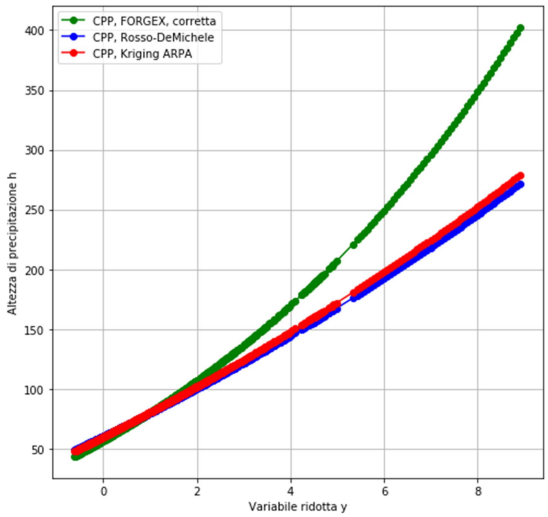


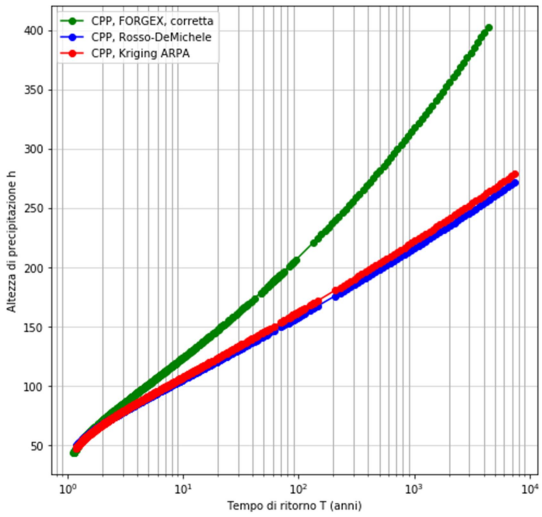
FORGEX		
α	ε	k
0.334	0.851	-0.114
VAPI		
α	ε	k
0.262	0.836	-0.048
Kriging		
θ2*	θ3	-
0.27755	-0.04354	-



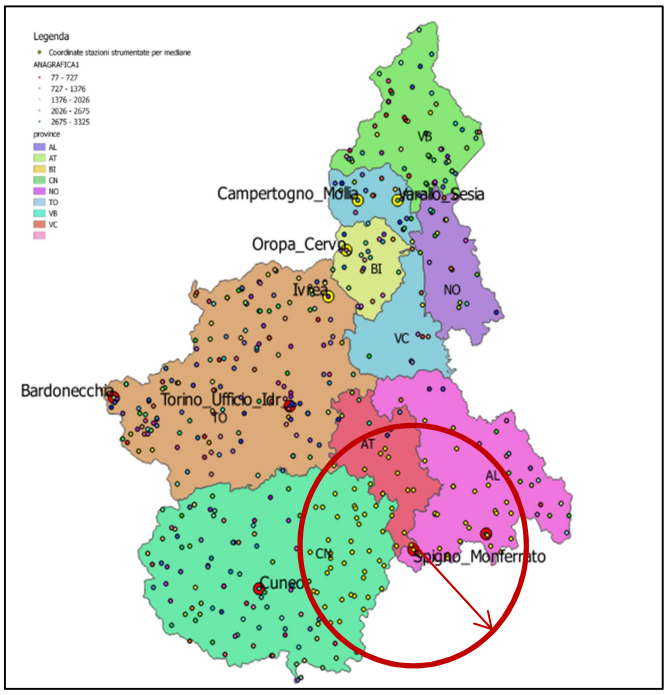








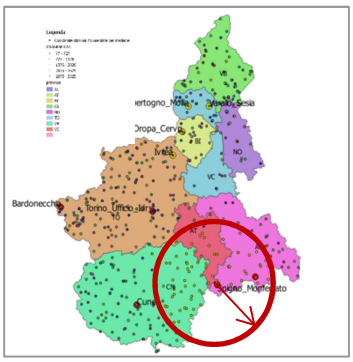
7.6.2 *Spigno Monferrato*

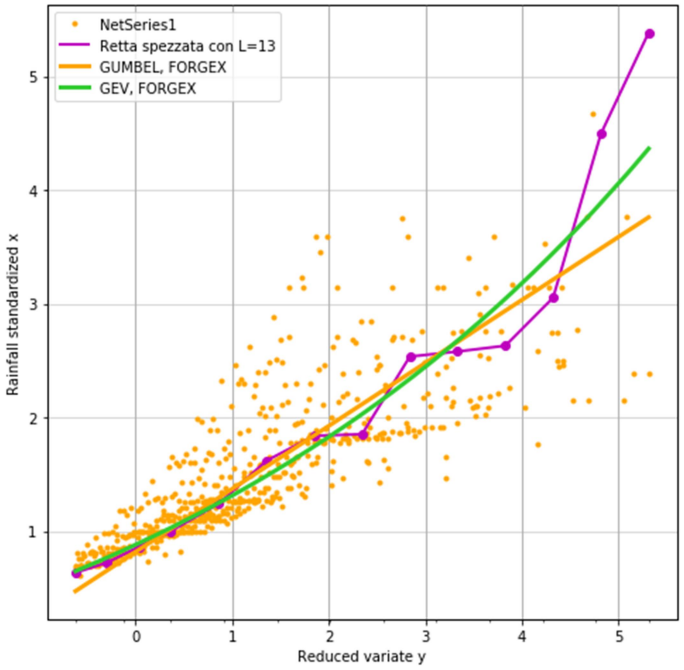


Localizzazione	*X=448514*
	Y=4933260
Raggio (km)	*56.3*
Numero stazioni	*79*
Numerosità dati stazione strumentata	*61*

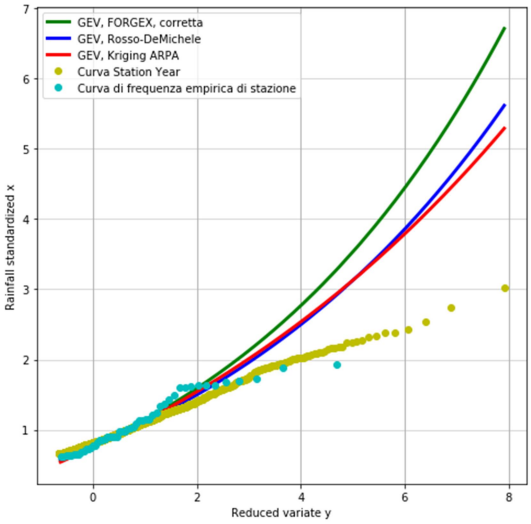
Durata (h)	N° segmenti	Valore Scarti minimo
1	*13*	*178*
3	*10*	*147*
6	*10*	*130*
12	*10*	*128*
24	*14*	*144*

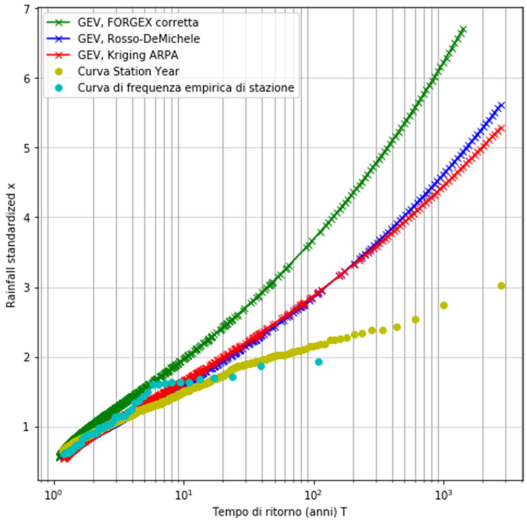
7.6.2.1 *Spigno Monferrato: durata 1h*

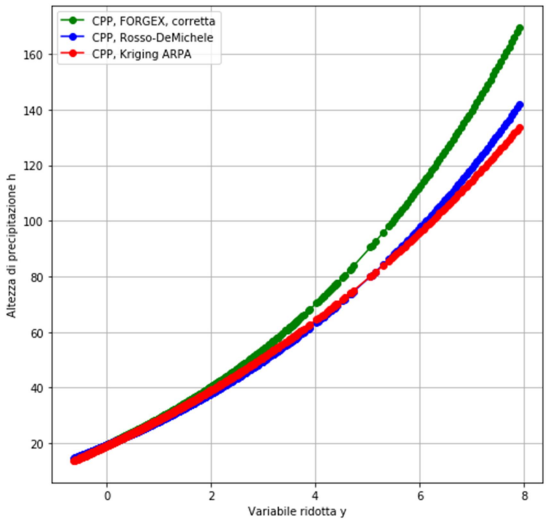


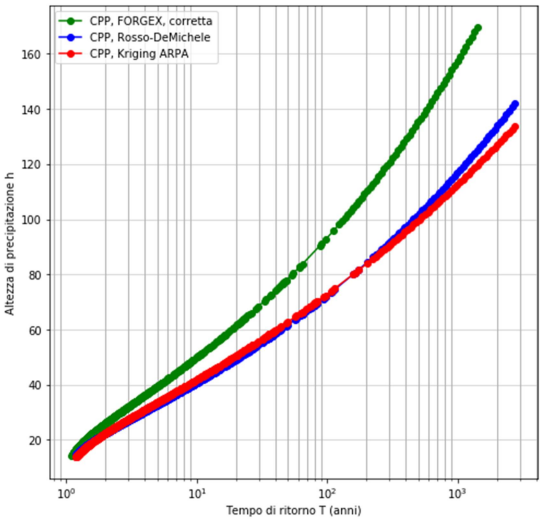


FORGEX		
α	ε	k
0.395	0.880	-0.176
VAPI		
α	ε	k
0.313	0.764	-0.154
Kriging		
θ2*	θ3	-
0.35067	-0.11512	-

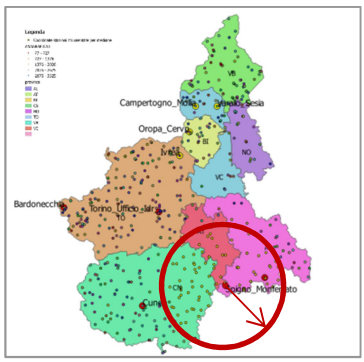




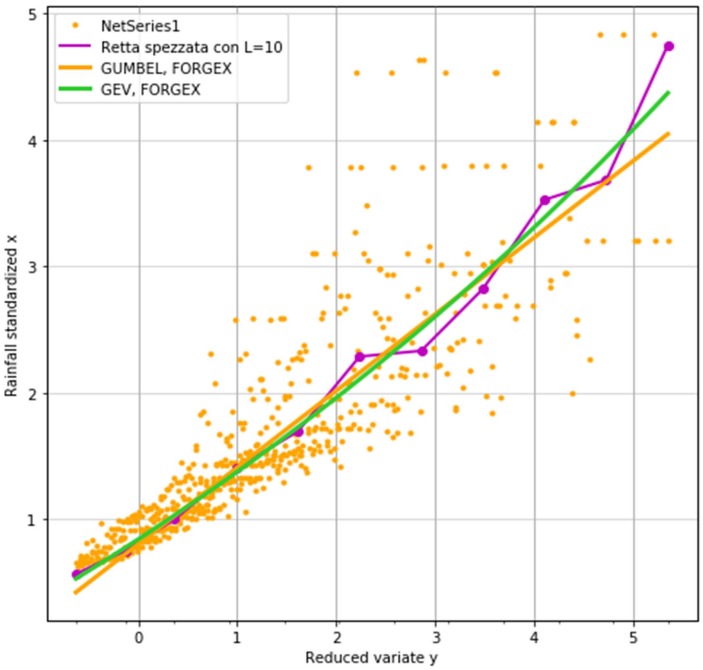


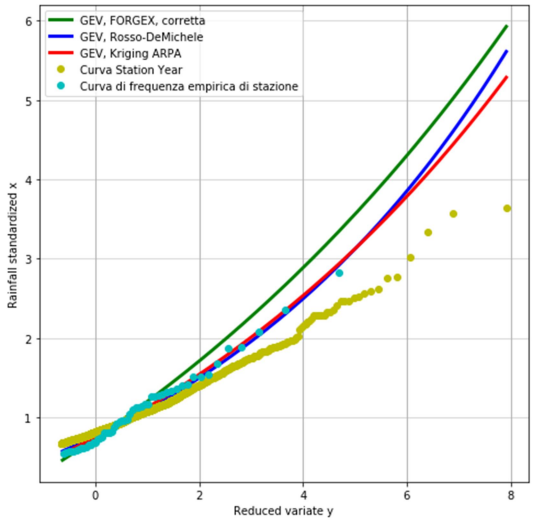


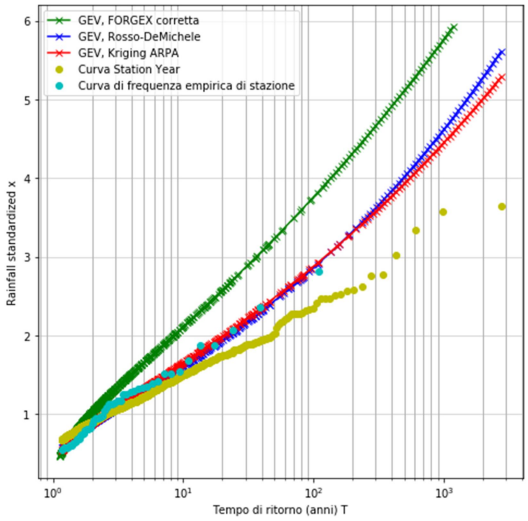
7.6.2.2 *Spigno Monferrato: durata 3h*

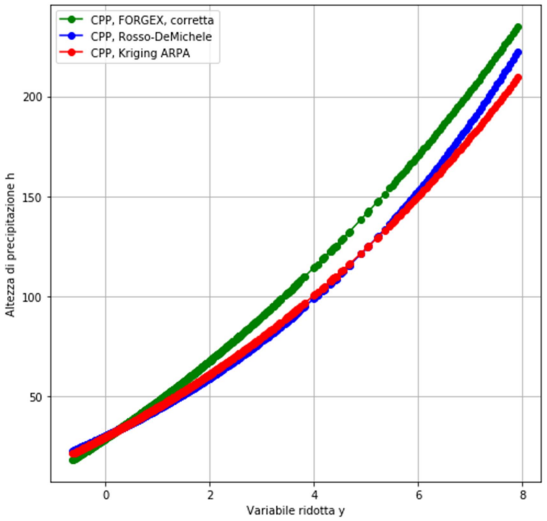


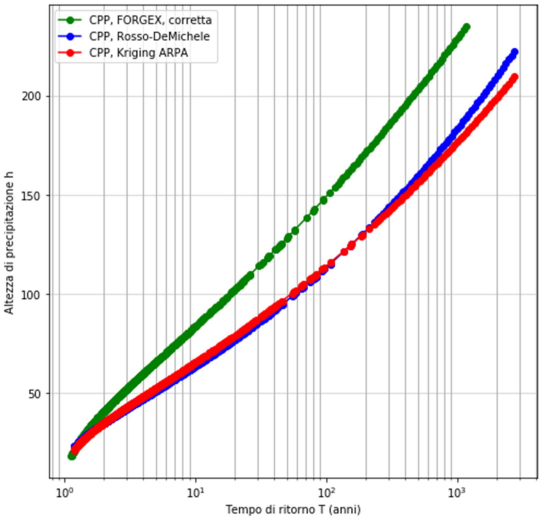
FORGEX		
α	ε	k
0.508	0.841	-0.092
VAPI		
α	ε	k
0.313	0.764	-0.154
Kriging		
θ2*	θ3	-
0.35067	-0.11512	-



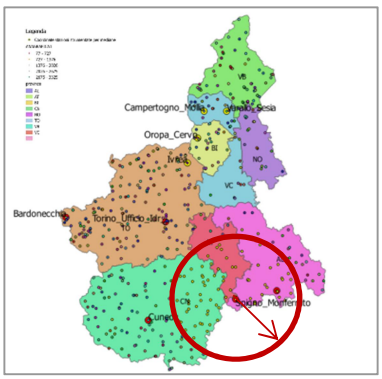




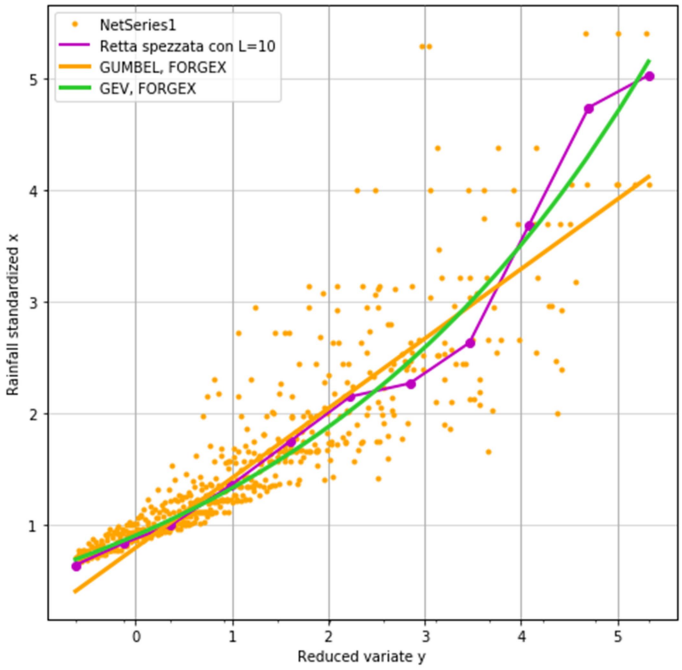


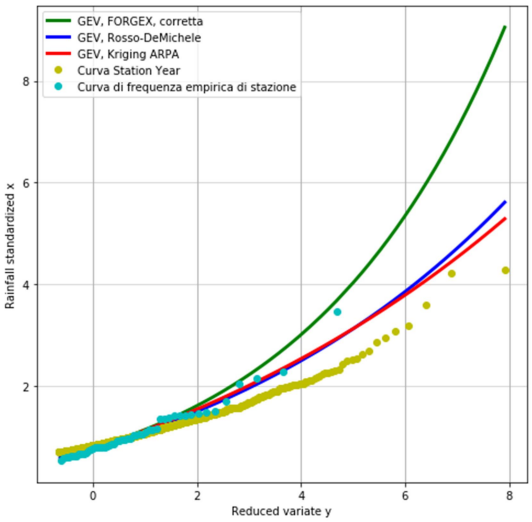


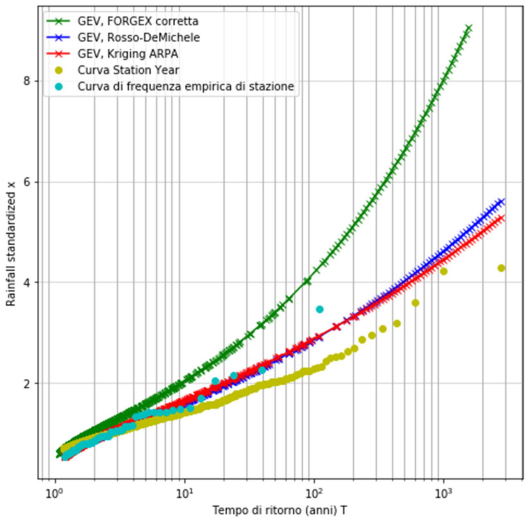
7.6.2.3 *Spigno Monferrato: durata 6h*

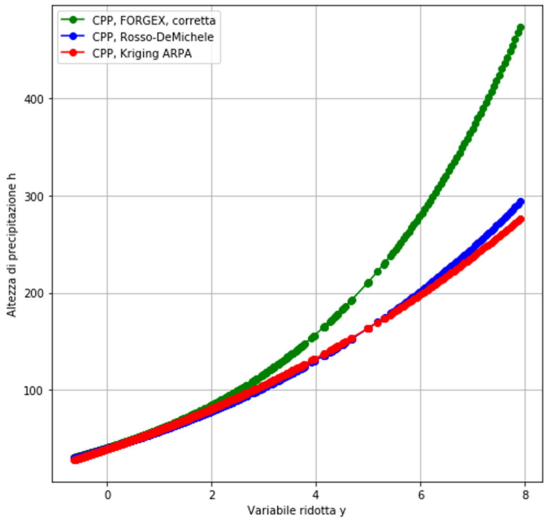


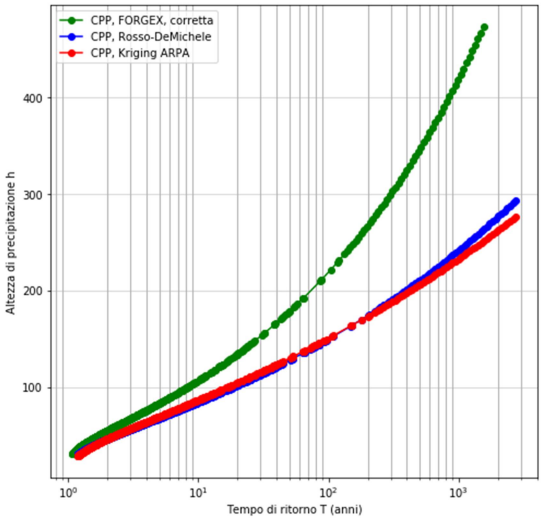
FORGEX		
α	ε	k
0.372	0.906	-0.257
VAPI		
α	ε	k
0.313	0.764	-0.154
Kriging		
θ2*	θ3	-
0.35067	-0.11512	-



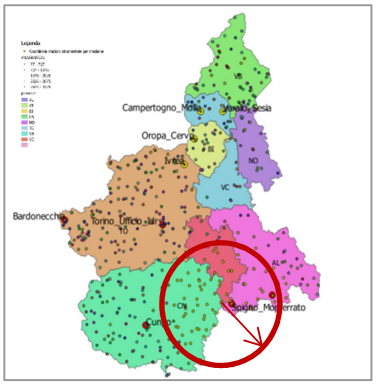


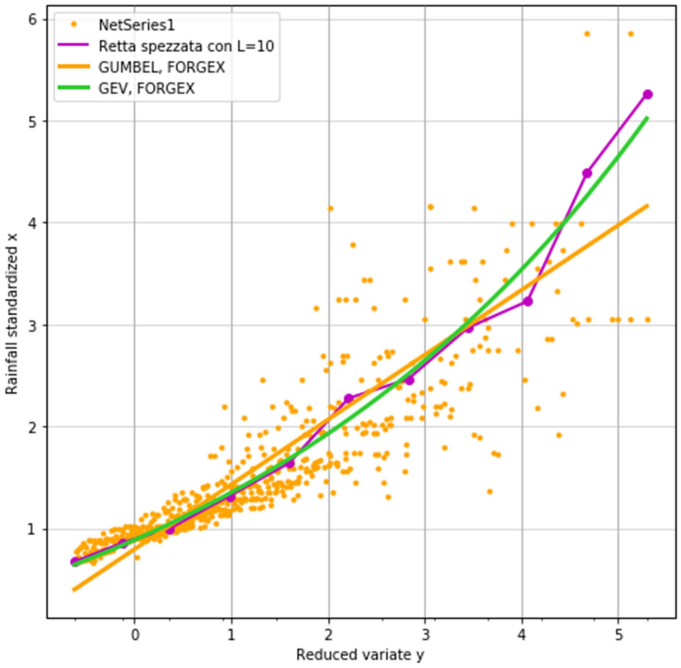




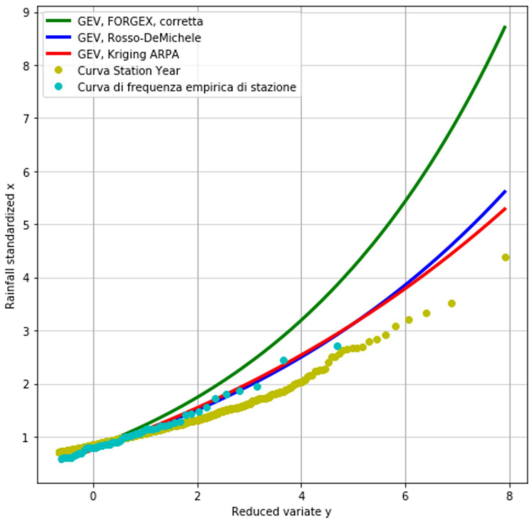


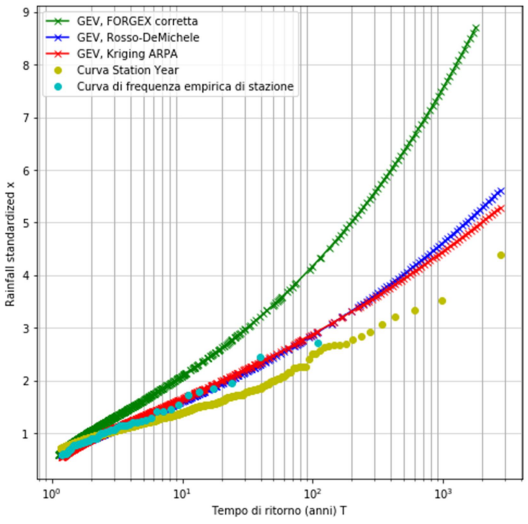
7.6.2.4 *Spigno Monferrato: durata 12h*

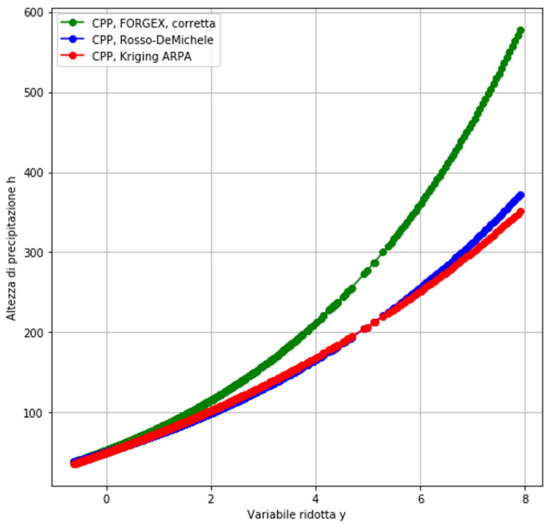


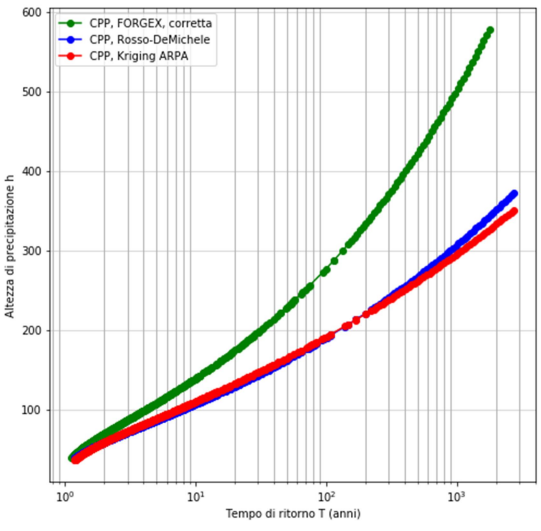


FORGEX		
α	ε	k
0.416	0.885	-0.217
VAPI		
α	ε	k
0.313	0.764	-0.154
Kriging		
θ2*	θ3	-
0.35067	-0.11512	-

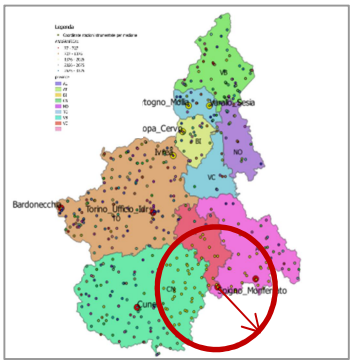




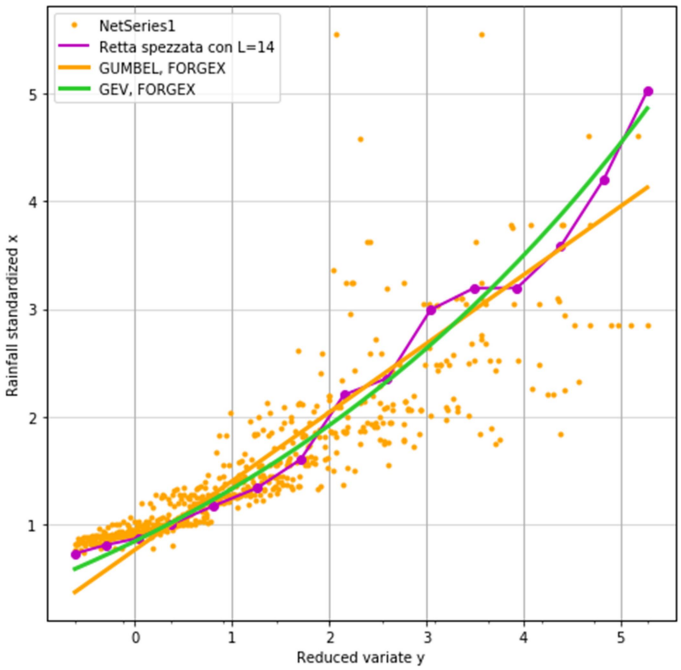


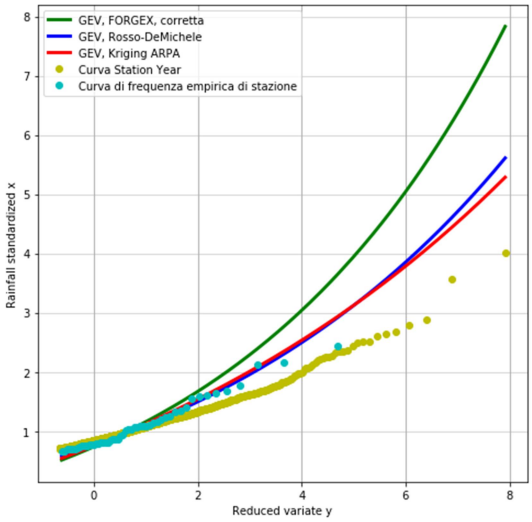


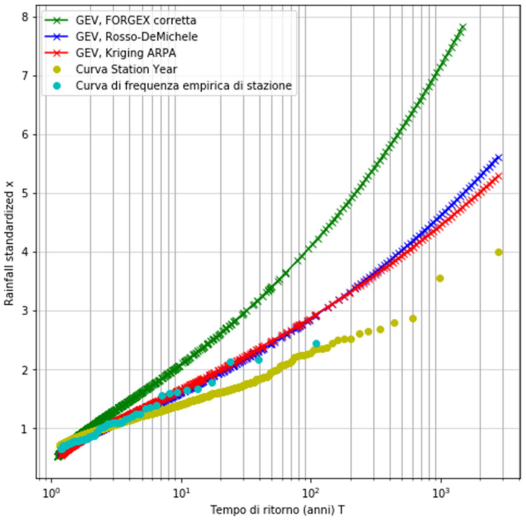
7.6.2.5 *Spigno Monferrato: durata 24h*

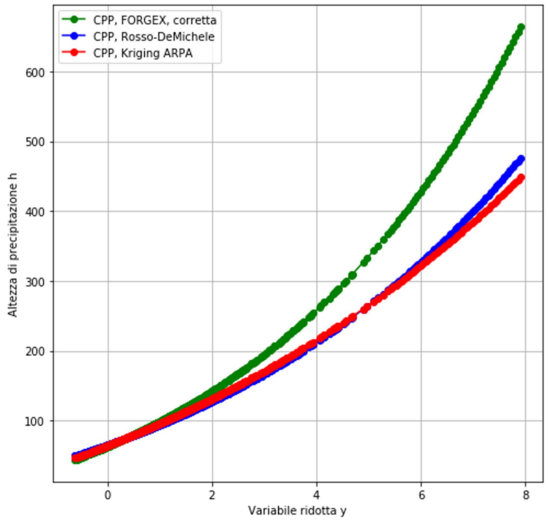


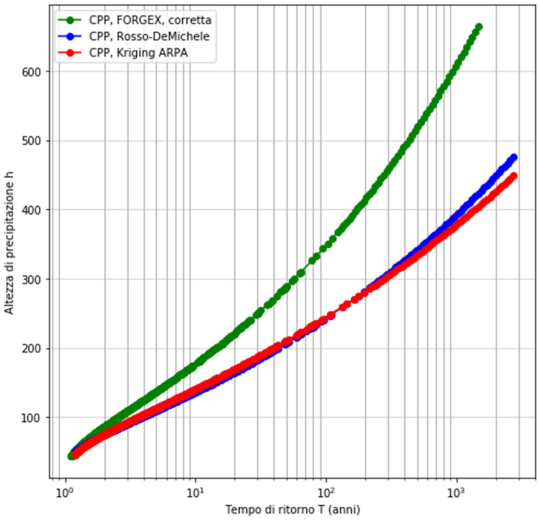
FORGEX		
α	ε	k
0.441	0.847	-0.190
VAPI		
α	ε	k
0.313	0.764	-0.154
Kriging		
θ2*	θ3	-
0.35067	-0.11512	-



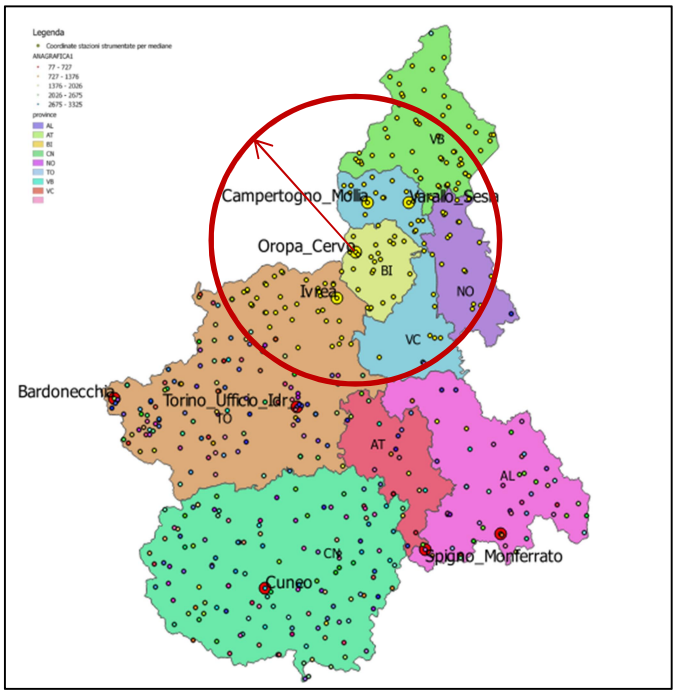








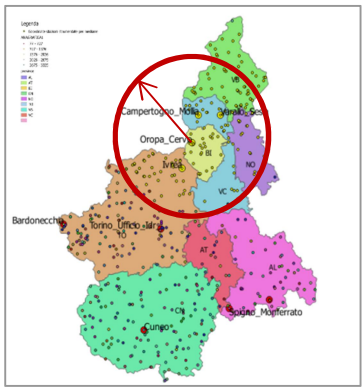
7.6.3 *Oropa Cervo*



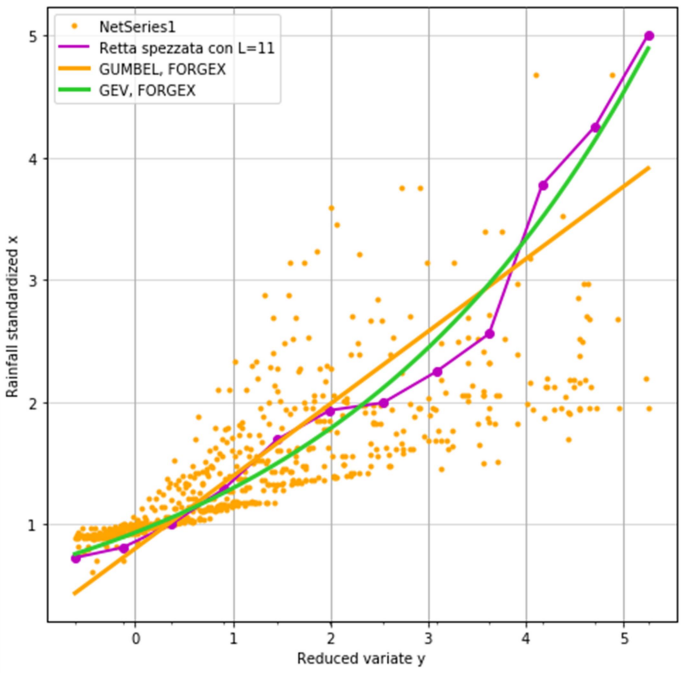
Localizzazione	*X=419601*
	Y=5053920
Raggio (km)	*85*
Numero stazioni	*161*
Numerosità dati stazione strumentata	*63*

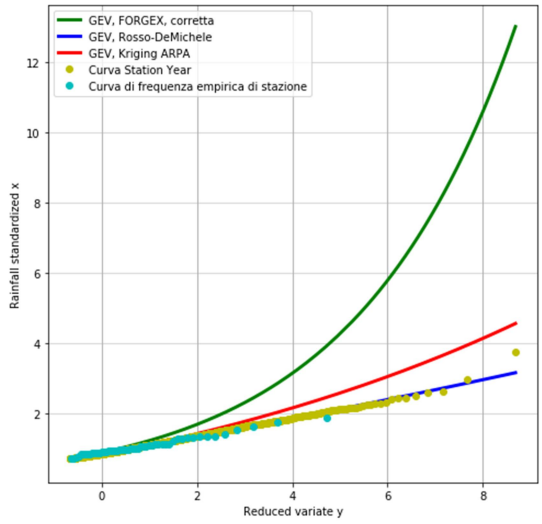
Durata (h)	N° segmenti	Valore Scarti minimo
1	*11*	*192*
3	*14*	*182*
6	*10*	*138*
12	*11*	*124*
24	*11*	*166*

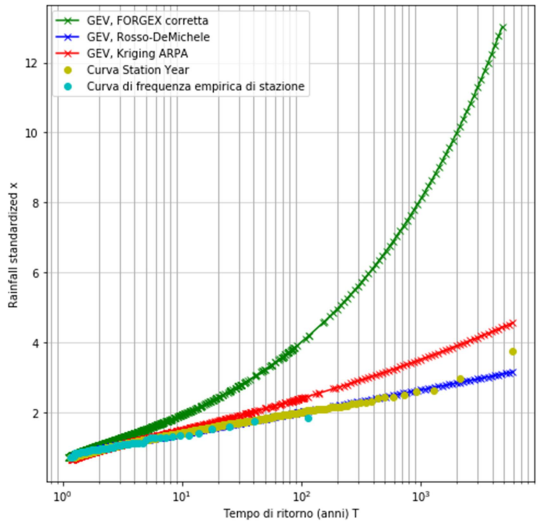
7.6.3.1 *Oropa Cervo: durata 1h*

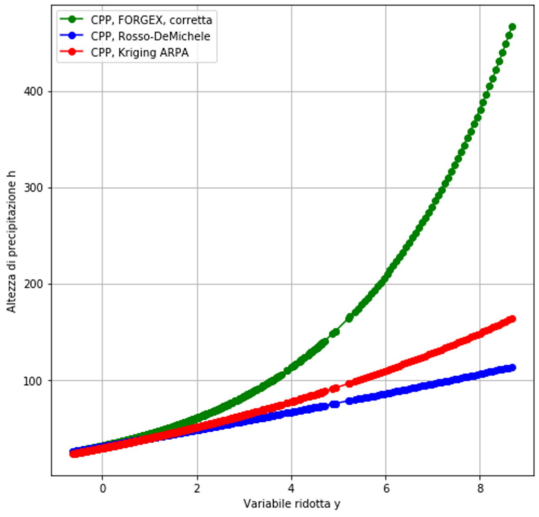


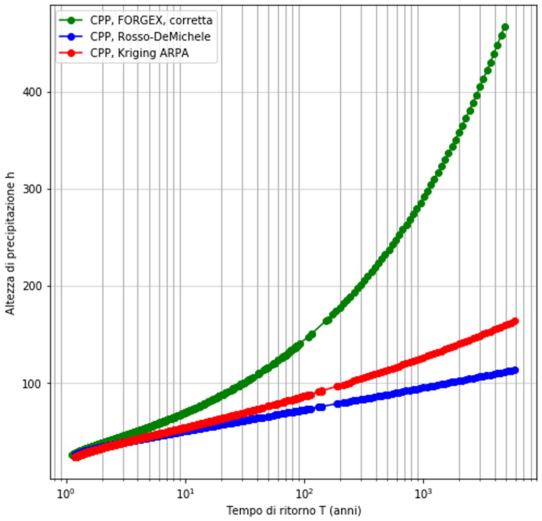
FORGEX		
α	ε	k
0.311	0.933	-0.298
VAPI		
α	ε	k
0.236	0.858	-0.026
Kriging		
θ2*	θ3	-
0.27308	-0.09838	-



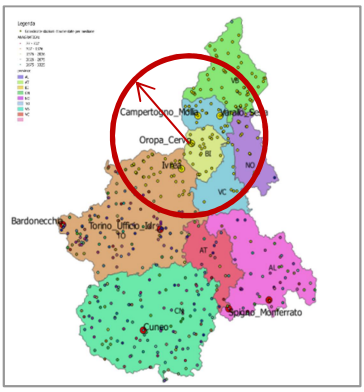




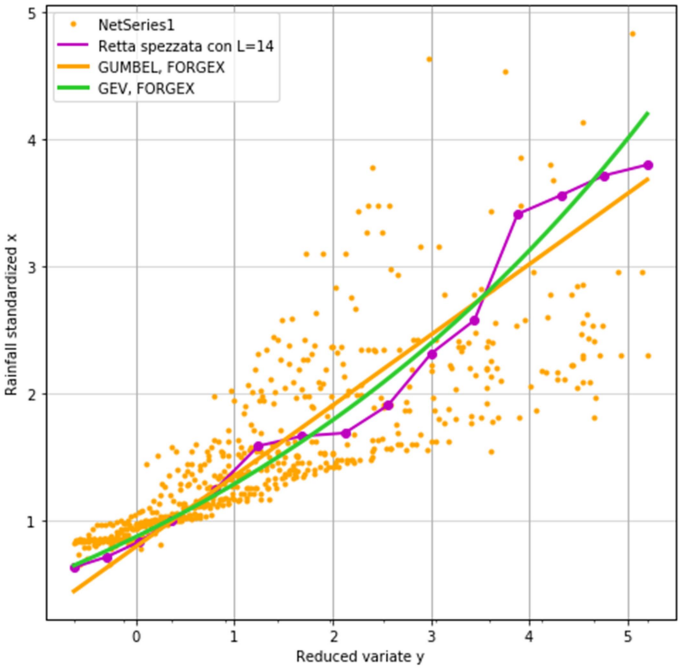


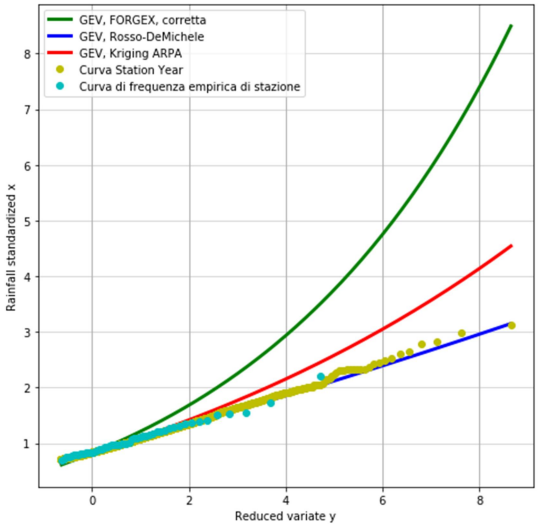


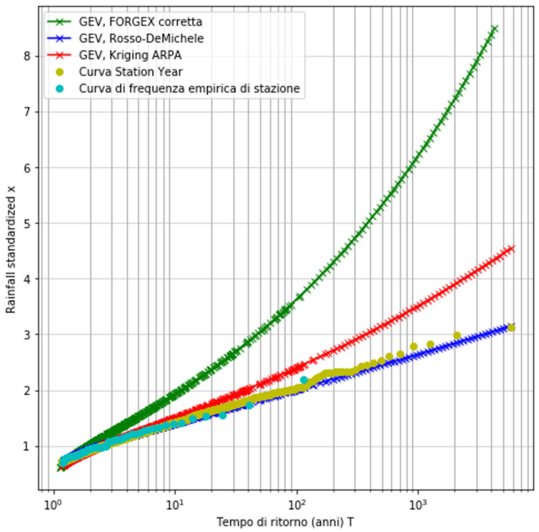
7.6.3.2 *Oropa Cervo: durata 3h*

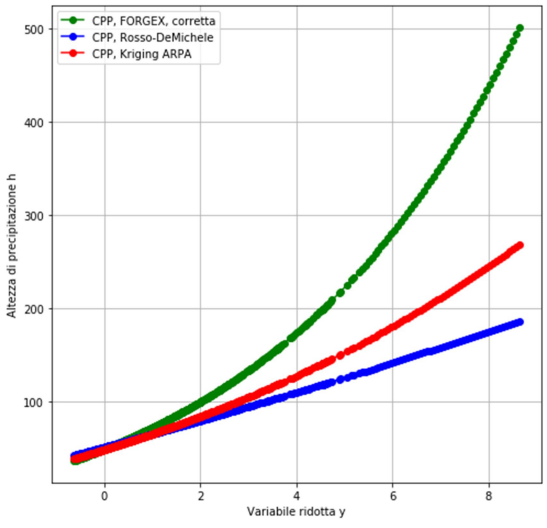


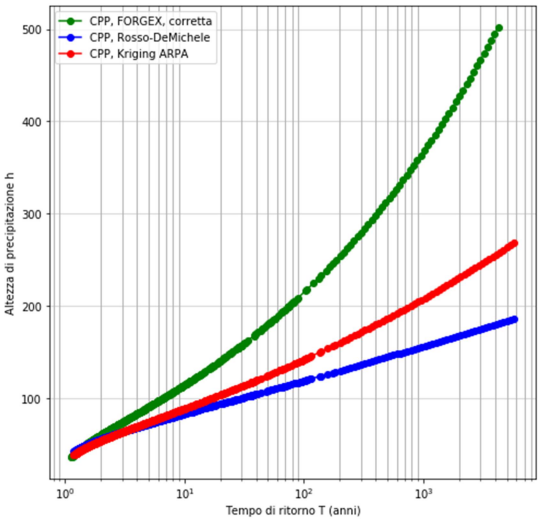
FORGEX		
α	ε	k
0.377	0.877	-0.187
VAPI		
α	ε	k
0.236	0.858	-0.026
Kriging		
θ2*	θ3	-
0.27308	-0.09838	-



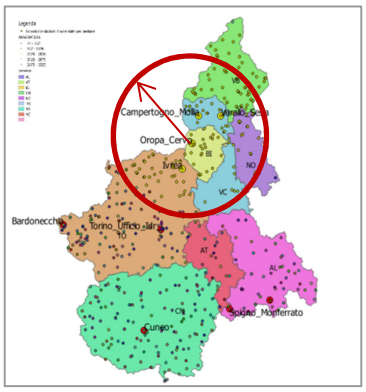




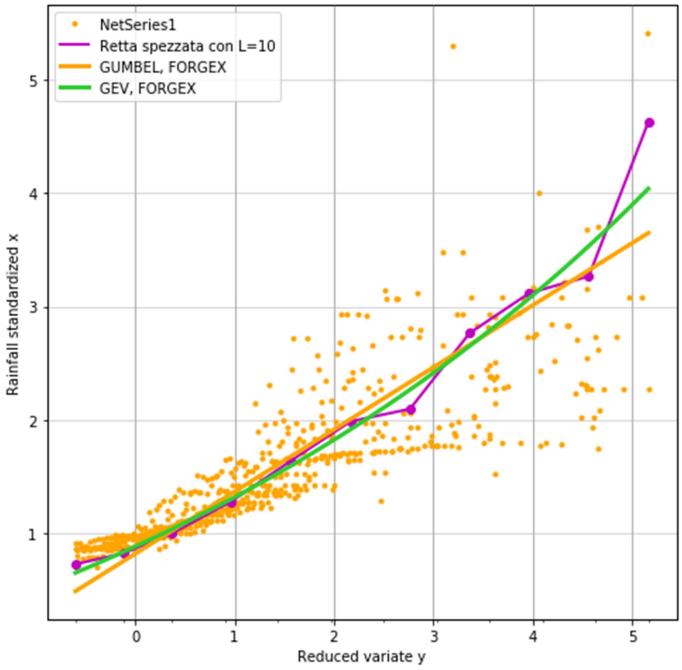


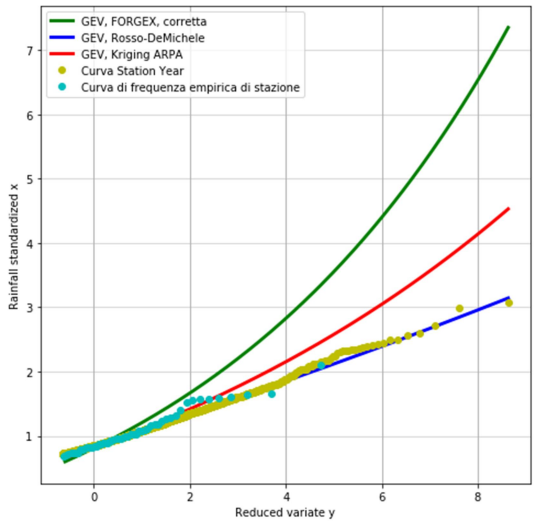


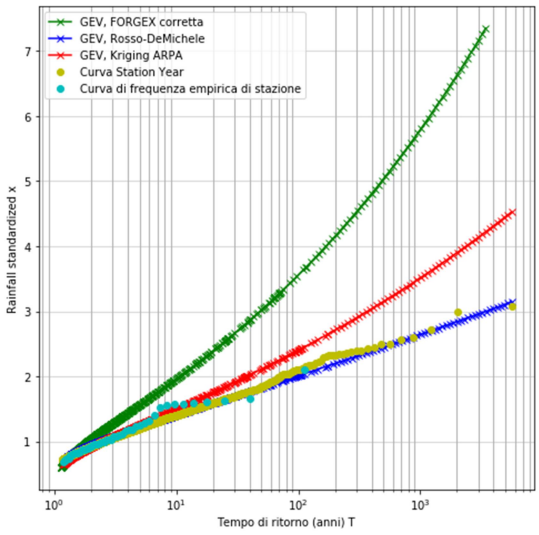
7.6.3.3 *Oropa Cervo: durata 6h*

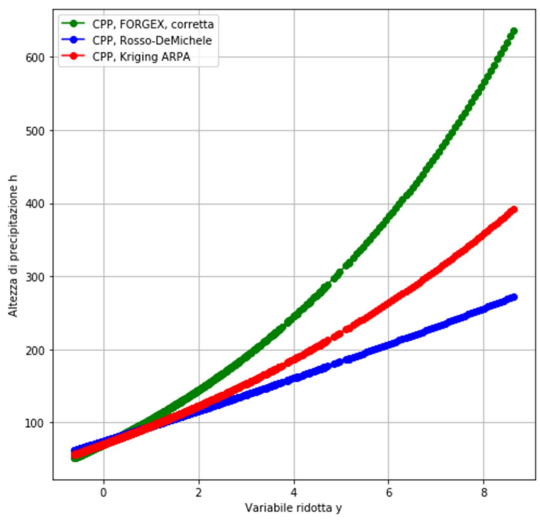


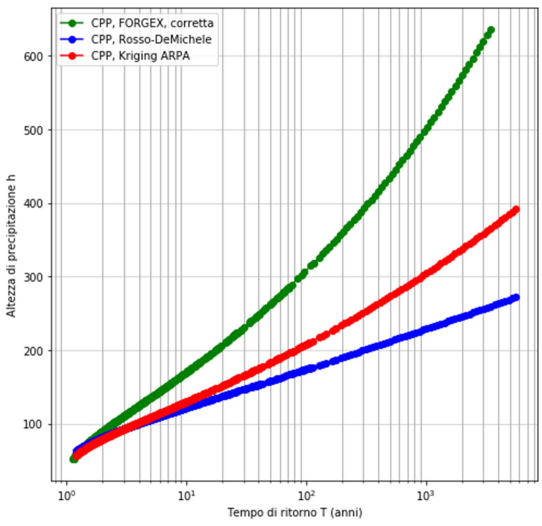
FORGEX		
α	ε	k
0.401	0.886	-0.151
VAPI		
α	ε	k
0.236	0.858	-0.026
Kriging		
θ2*	θ3	-
0.27308	-0.09838	-



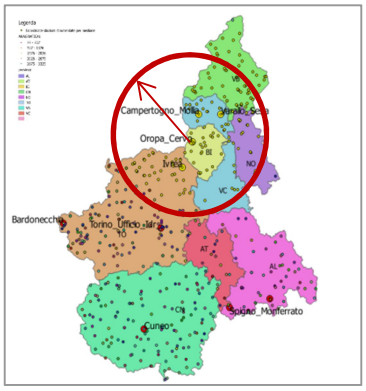




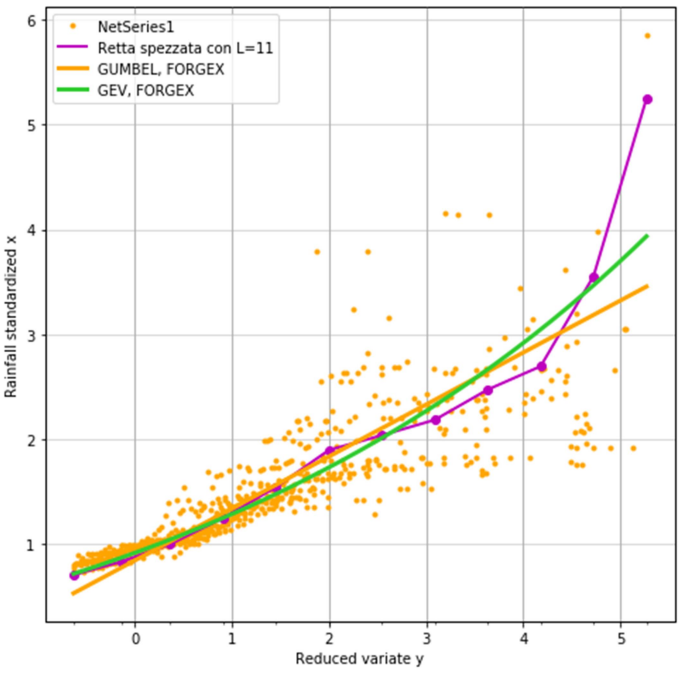


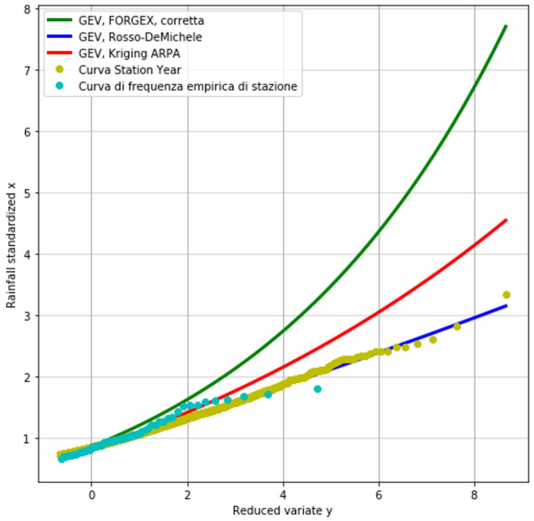


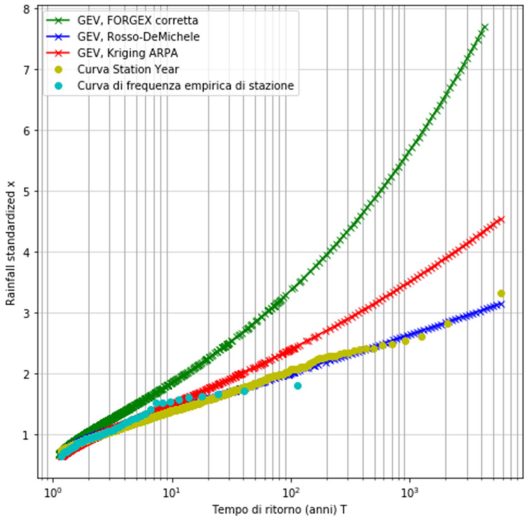
7.6.3.4 *Oropa Cervo: durata 12h*

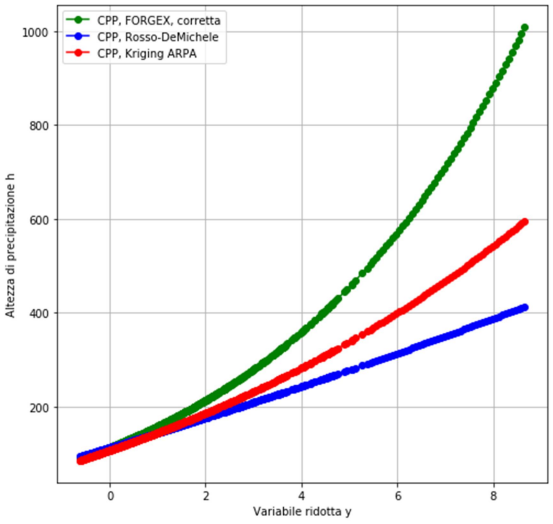


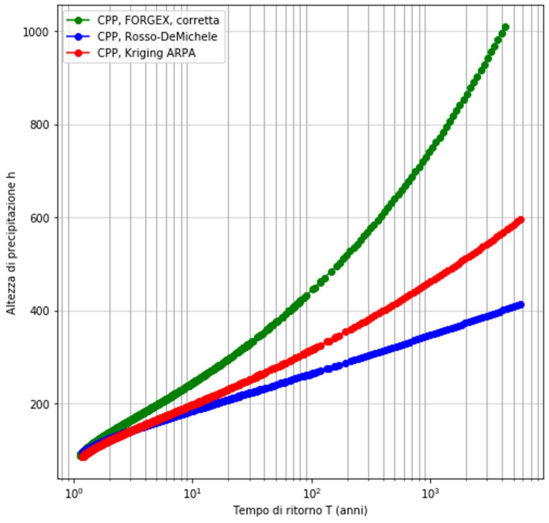
FORGEX		
α	ε	k
0.336	0.917	-0.187
VAPI		
α	ε	k
0.236	0.858	-0.026
Kriging		
θ2*	θ3	-
0.27308	-0.09838	-



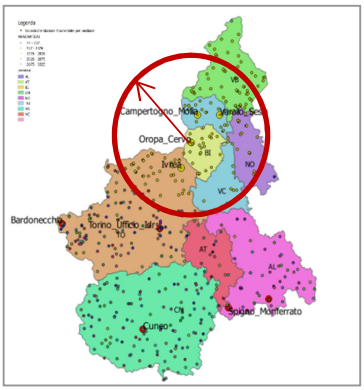




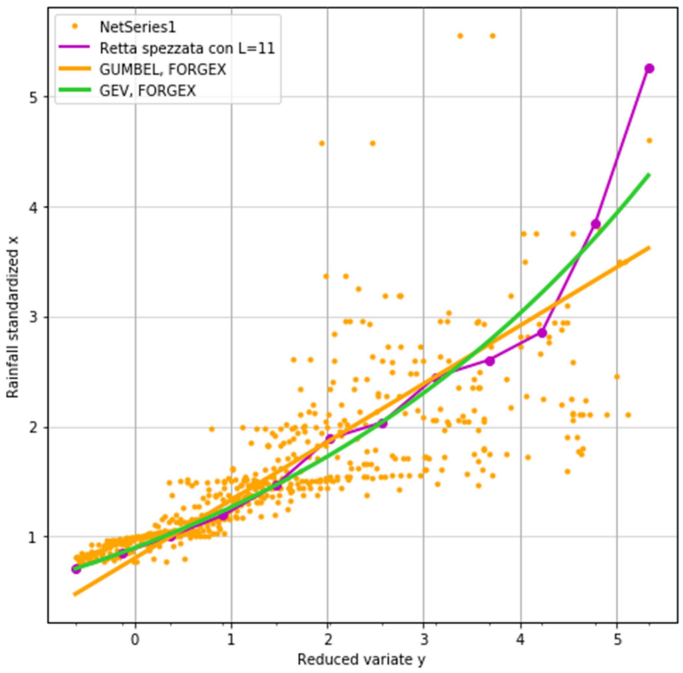


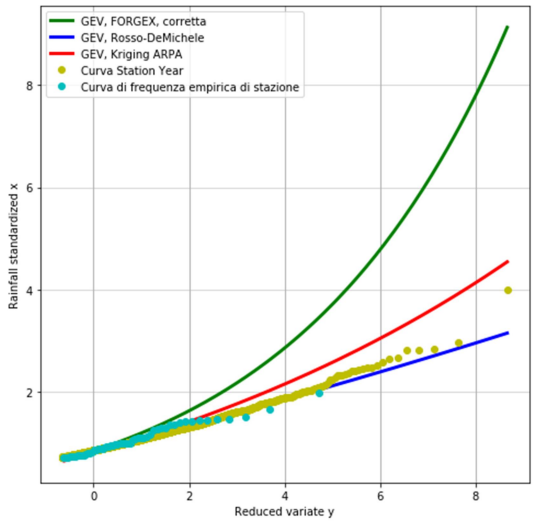


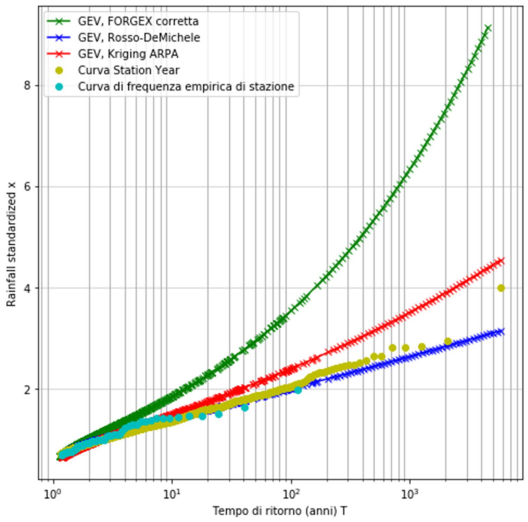
7.6.3.5 *Oropa Cervo: durata 24h*

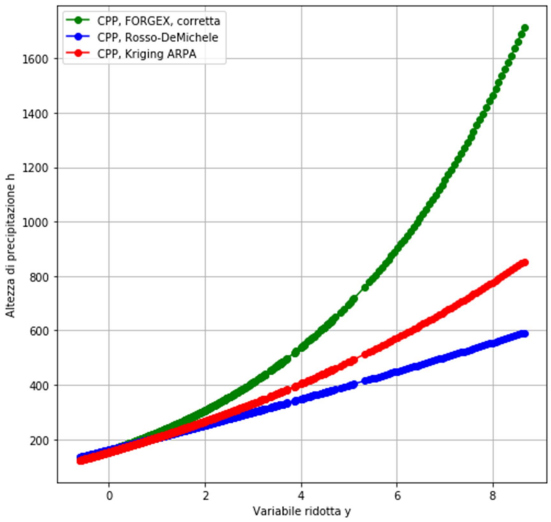


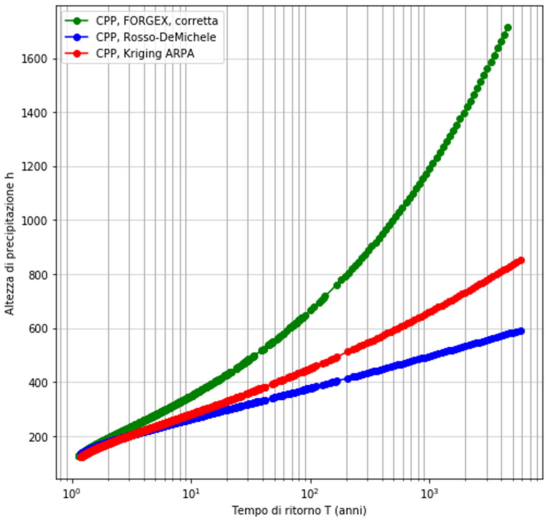
FORGEX		
α	ε	k
0.328	0.895	-0.224
VAPI		
α	ε	k
0.236	0.858	-0.026
Kriging		
θ2*	θ3	-
0.27308	-0.09838	-



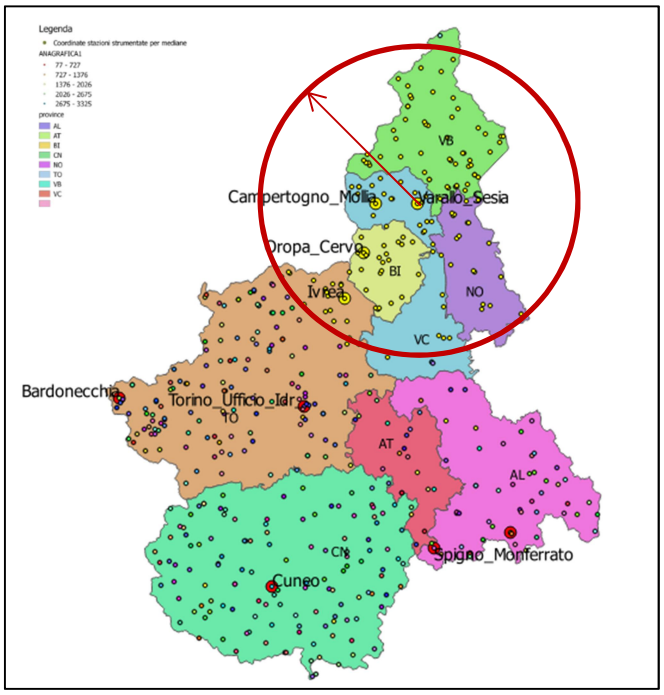








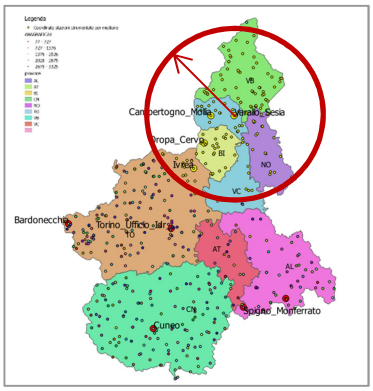
7.6.4 *Varallo Sesia*



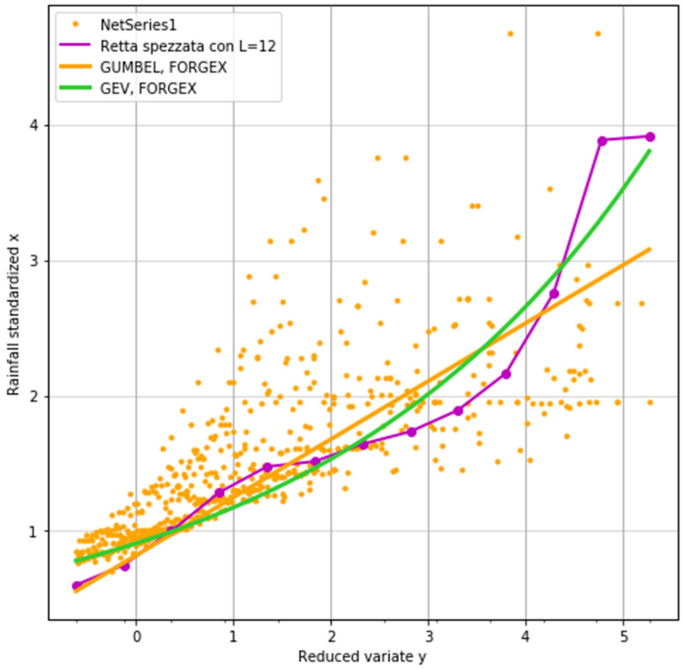
Localizzazione	*X=441874*
	Y=5074050
Raggio (km)	*102,6*
Numero stazioni	*149*
Numerosità dati stazione strumentata	*59*

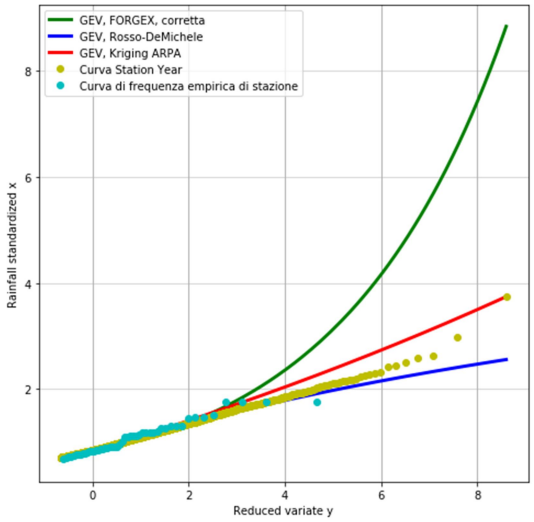
Durata (h)	N° segmenti	Valore Scarti minimo
1	*12*	*197*
3	*11*	*154*
6	*15*	*189*
12	*11*	*147*
24	*11*	*146*

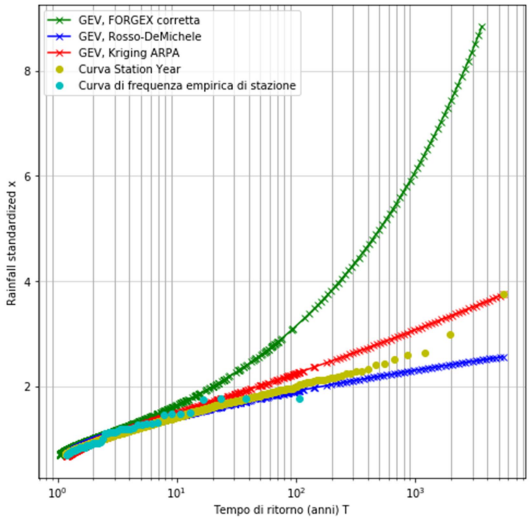
7.6.4.1 *Varallo Sesia: durata 1h*

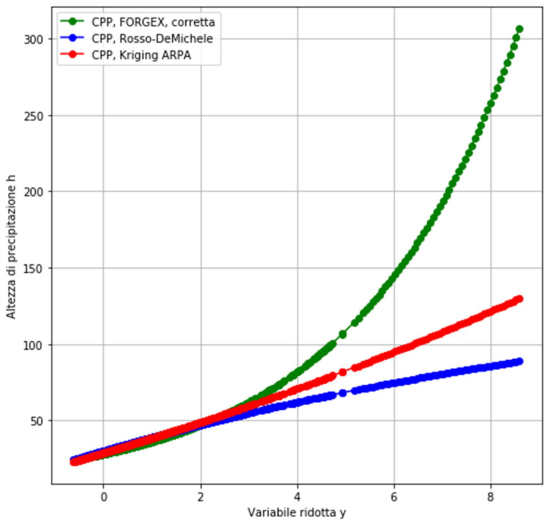


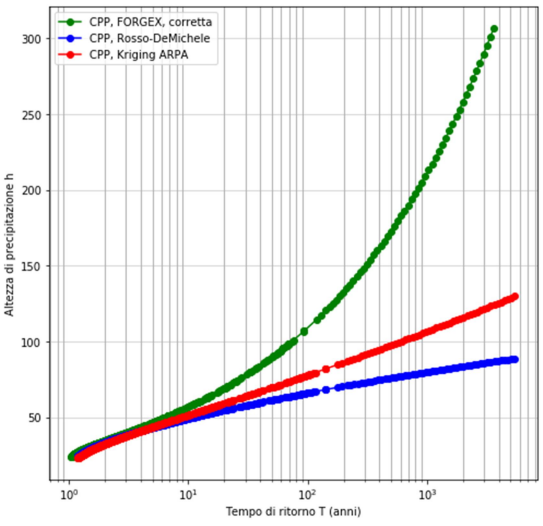
FORGEX		
α	ε	k
0.227	0.907	-0.296
VAPI		
α	ε	k
0.266	0.865	0.074
Kriging		
θ2*	θ3	-
0.27475	-0.04735	-



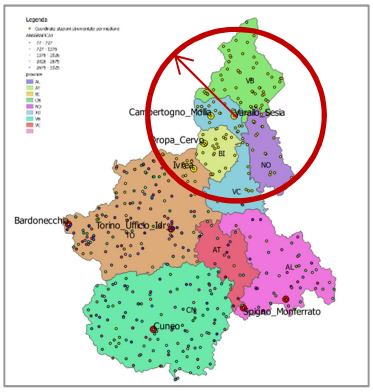




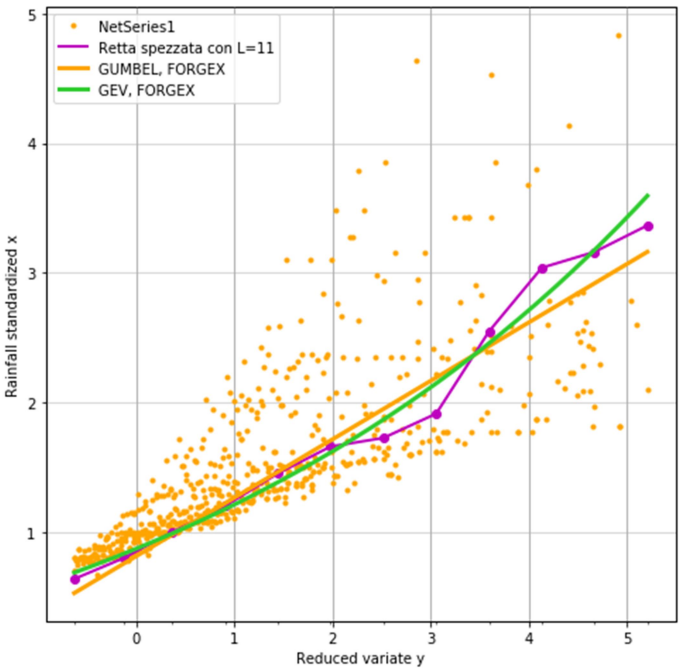


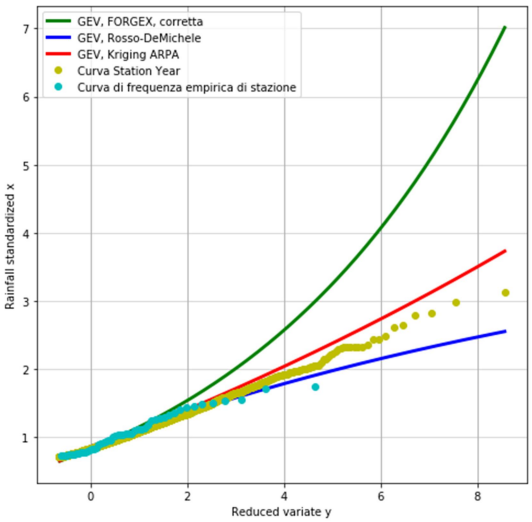


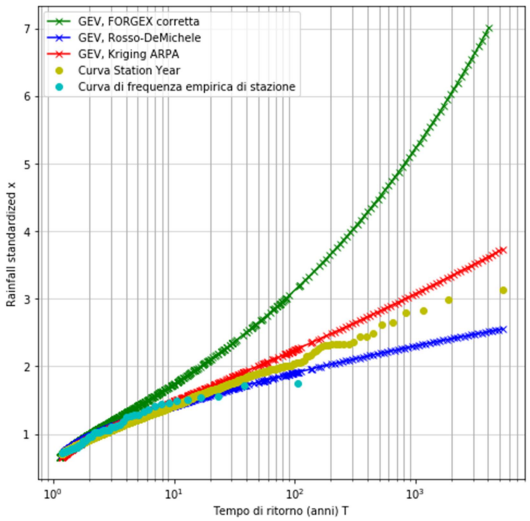
7.6.4.2 *Varallo Sesia: durata 3h*

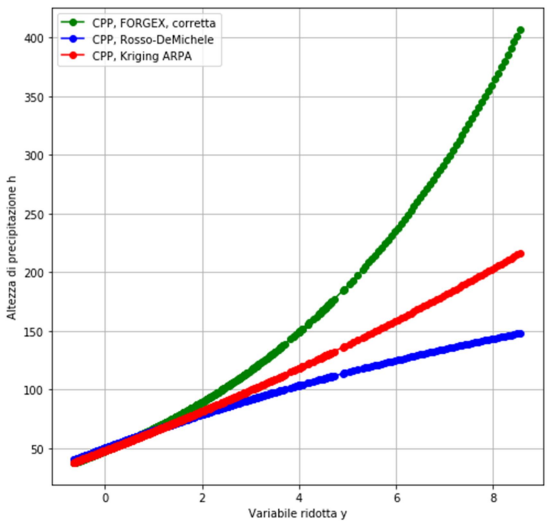


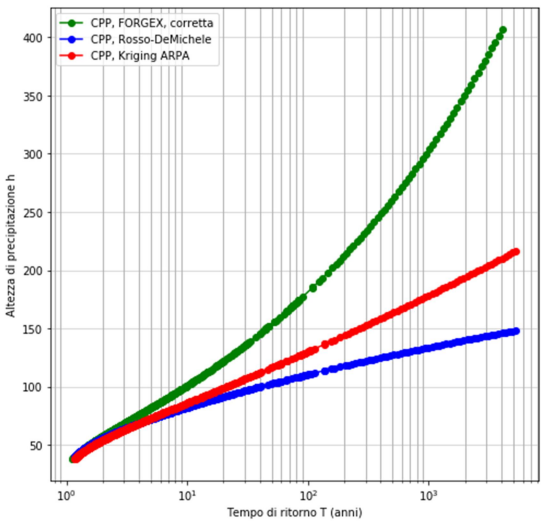
FORGEX		
α	ε	k
0.309	0.872	-0.185
VAPI		
α	ε	k
0.266	0.865	0.074
Kriging		
θ2*	θ3	-
0.27475	-0.04735	-



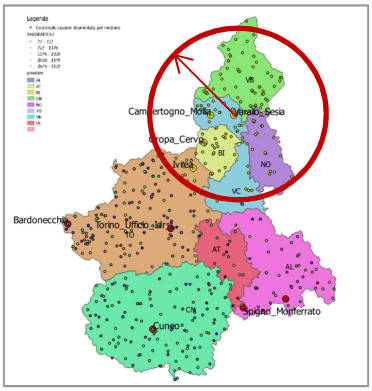




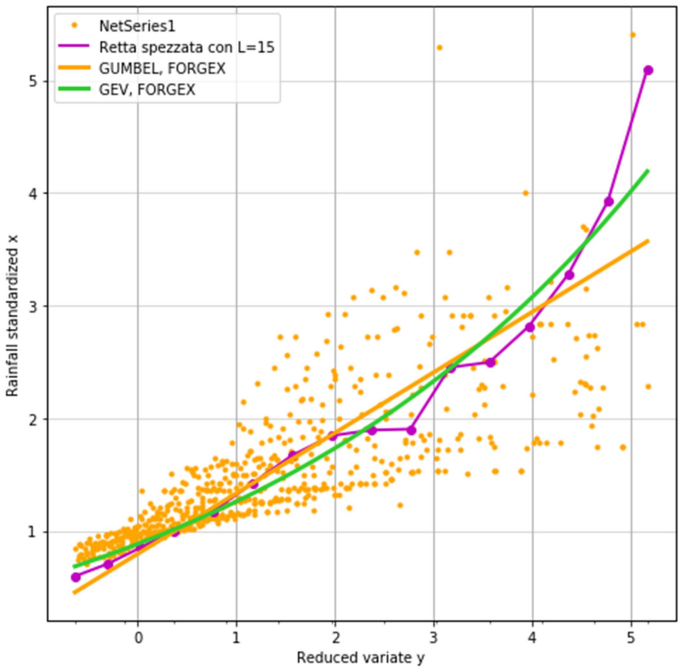


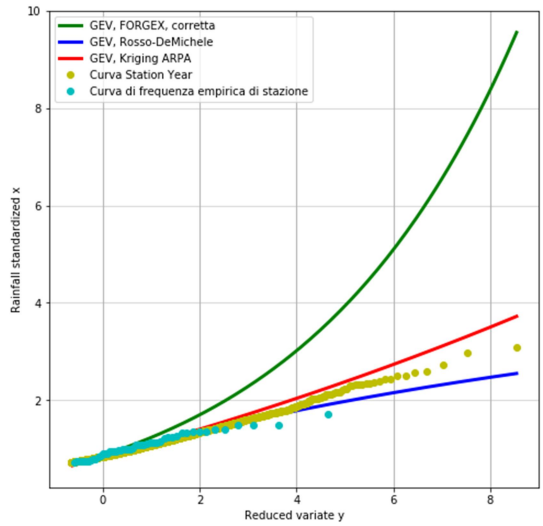


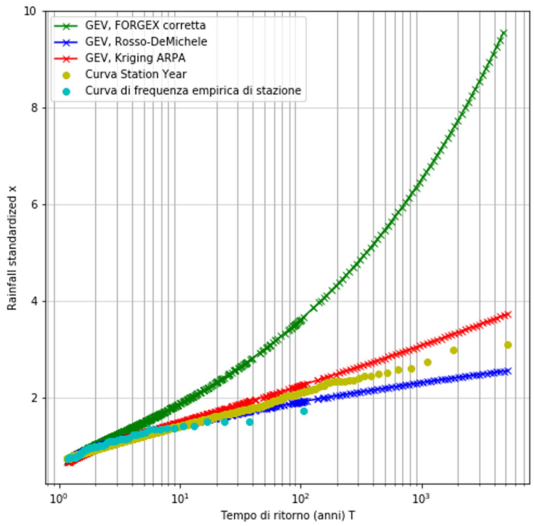
7.6.4.3 *Varallo Sesia: durata 6h*

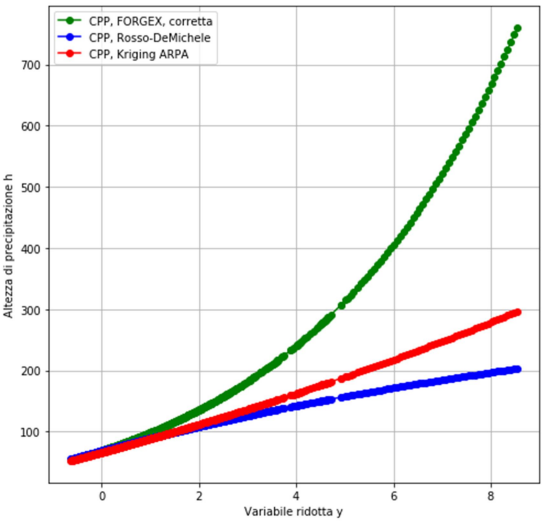


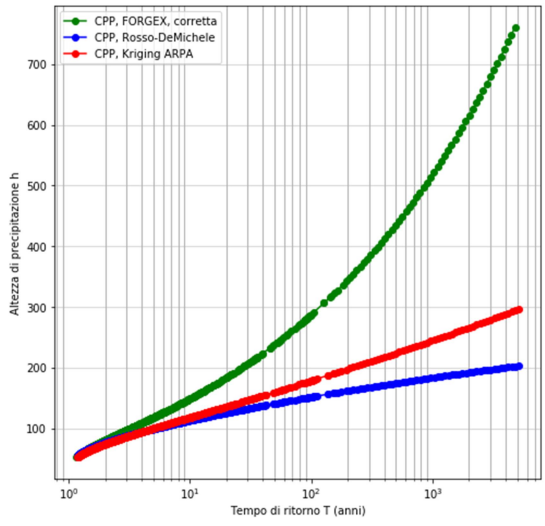
FORGEX		
α	ε	k
0.335	0.882	-0.227
VAPI		
α	ε	k
0.266	0.865	0.074
Kriging		
θ2*	θ3	-
0.27475	-0.04735	-



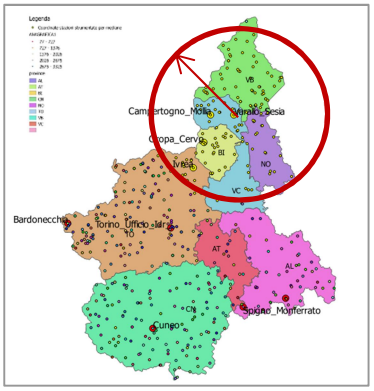




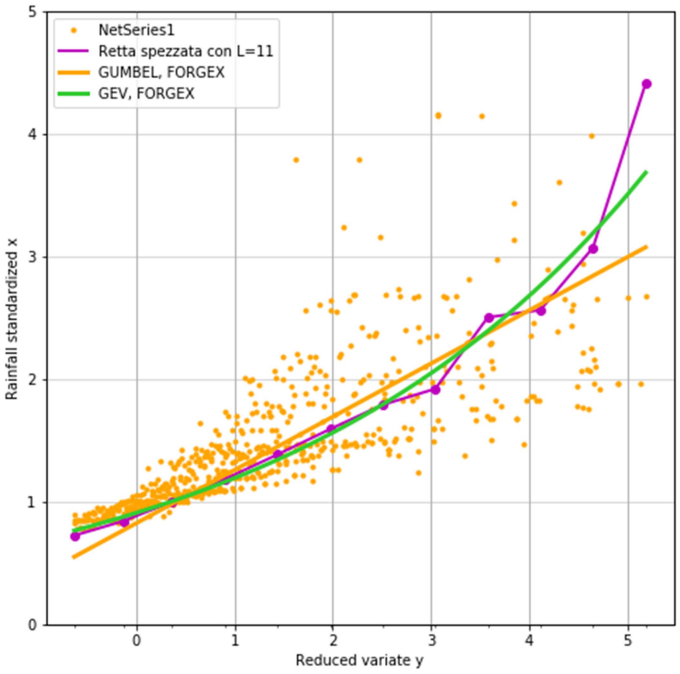


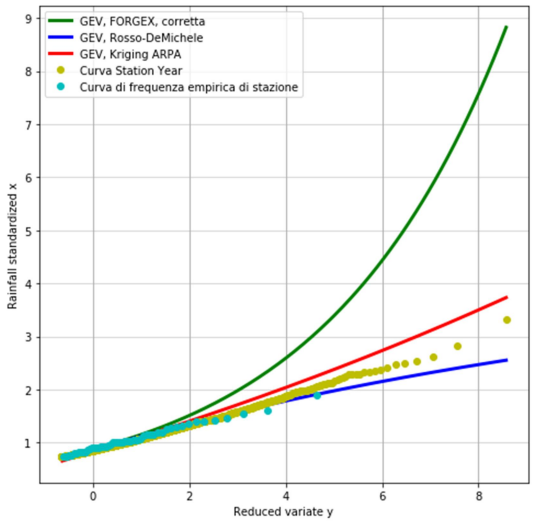


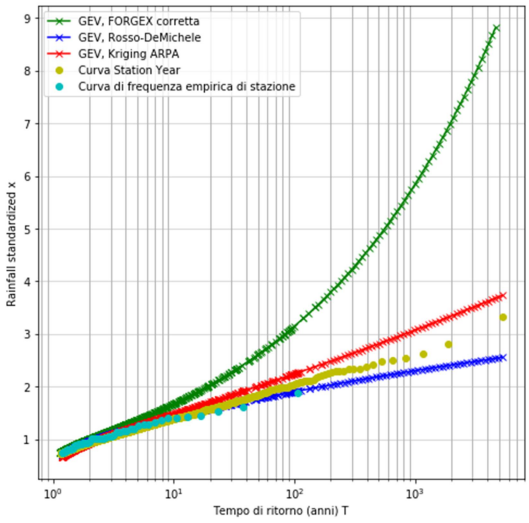
7.6.4.4 *Varallo Sesia: durata 12h*

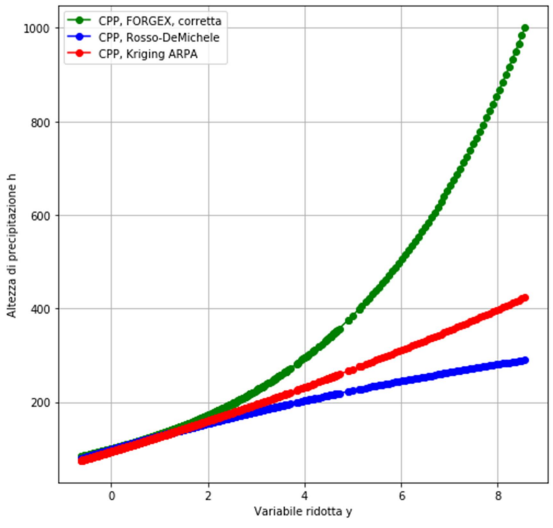


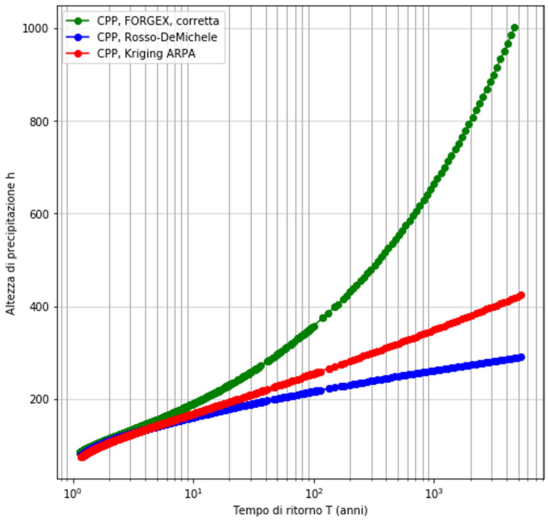
FORGEX		
α	ε	k
0.247	0.908	-0.226
VAPI		
α	ε	k
0.266	0.865	0.074
Kriging		
θ2*	θ3	-
0.27475	-0.04735	-



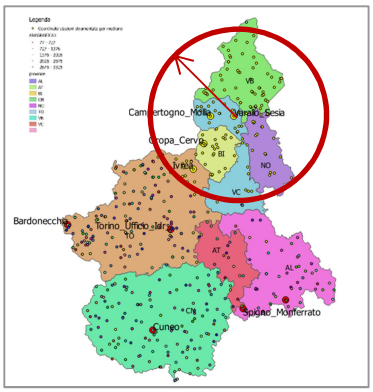




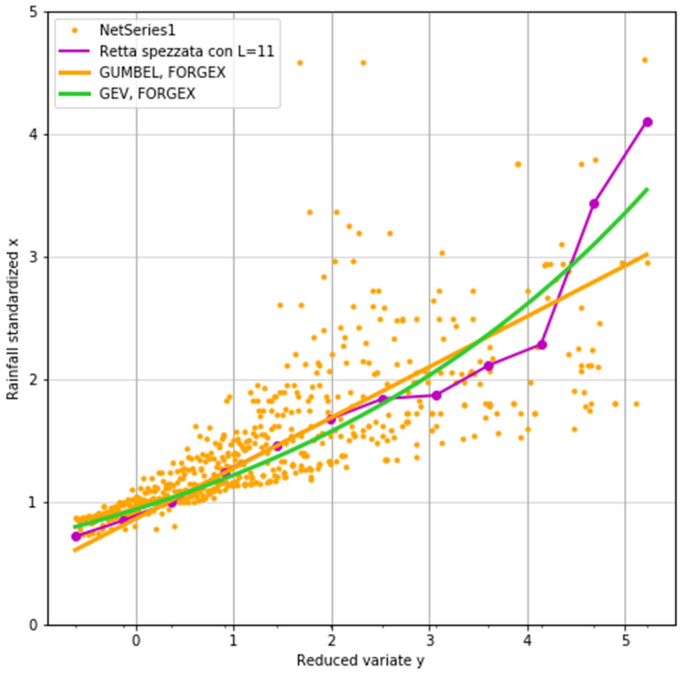


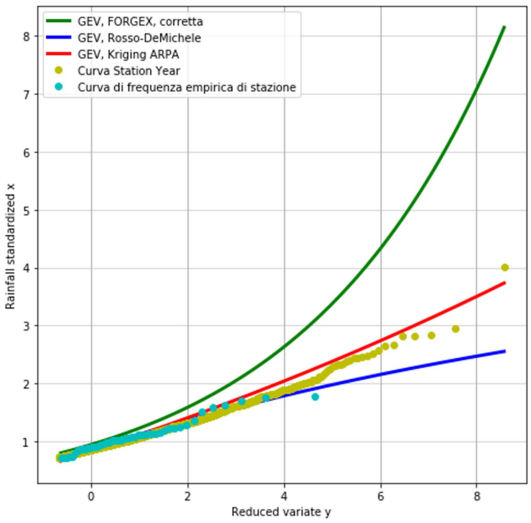


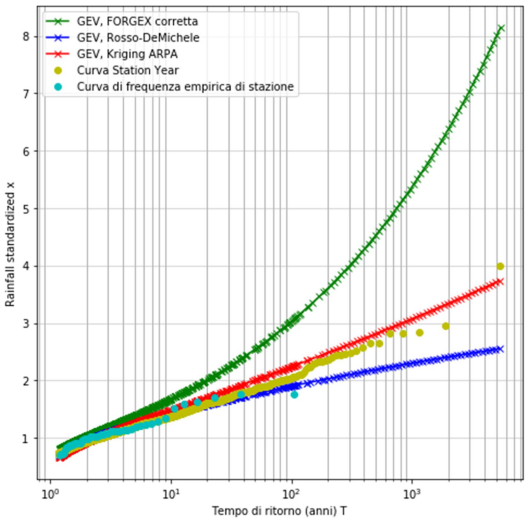
7.6.4.5 *Varallo Sesia: durata 24h*

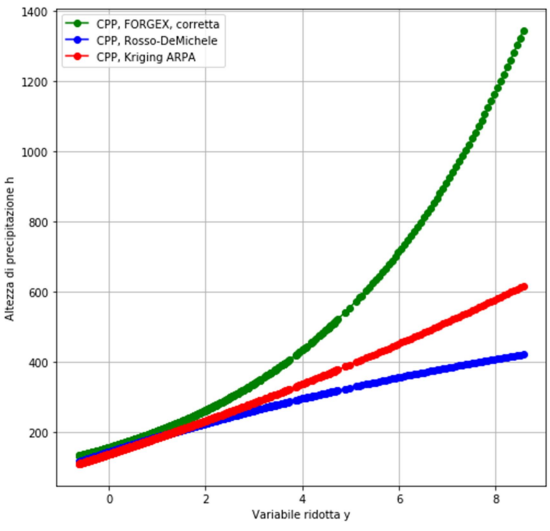


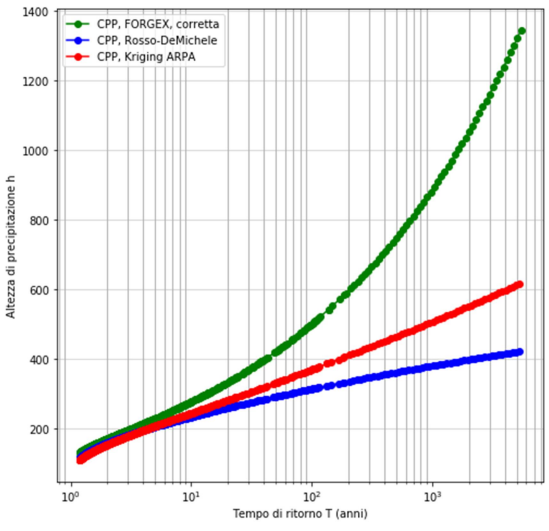
FORGEX		
α	ε	k
0.247	0.937	-0.243
VAPI		
α	ε	k
0.266	0.865	0.074
Kriging		
θ2*	θ3	-
0.27475	-0.04735	-



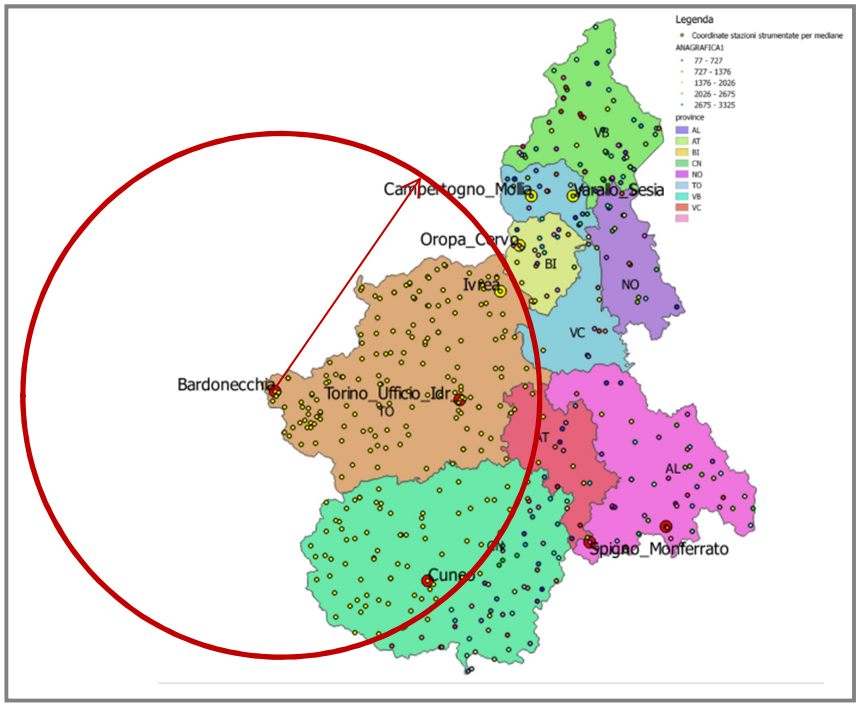








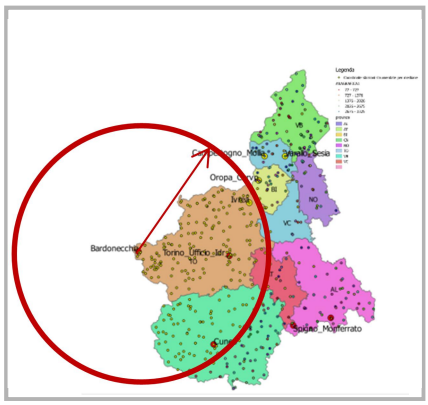
7.6.5 *Bardonecchia*



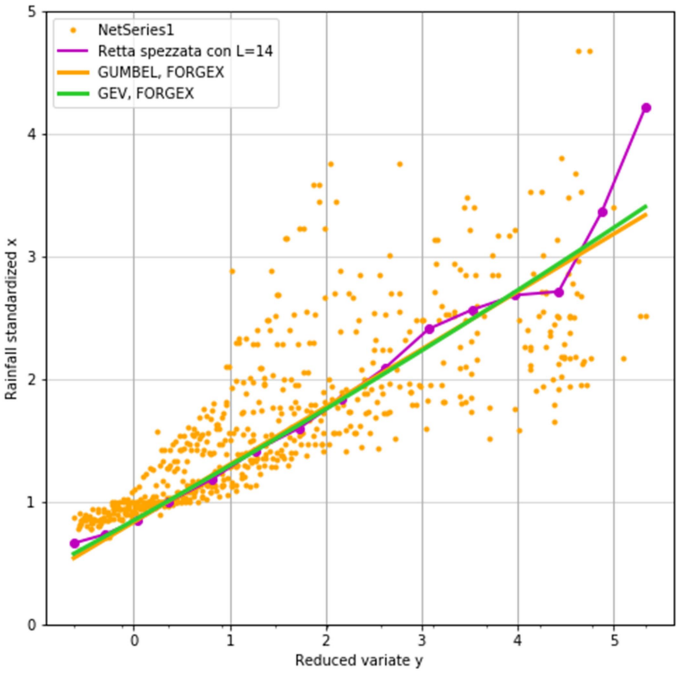
Localizzazione	*X=319124*
	Y=4994860
Raggio (km)	*146*
Numero stazioni	*241*
Numerosità dati stazione strumentata	*50*

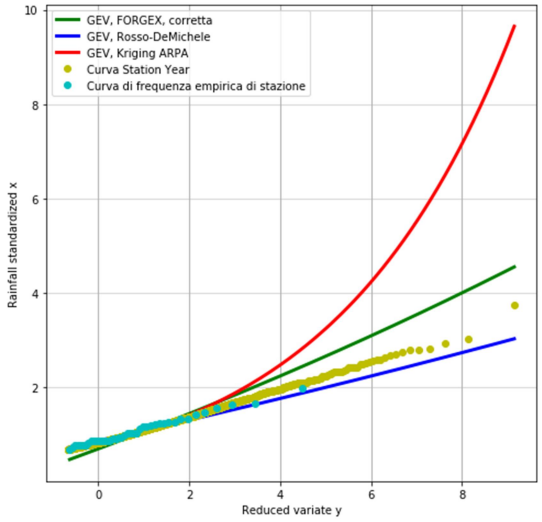
Durata (h)	N° segmenti	Valore Scarti minimo
1	*14*	*154*
3	*10*	*170*
6	*10*	*176*
12	*15*	*139*
24	*15*	*163*

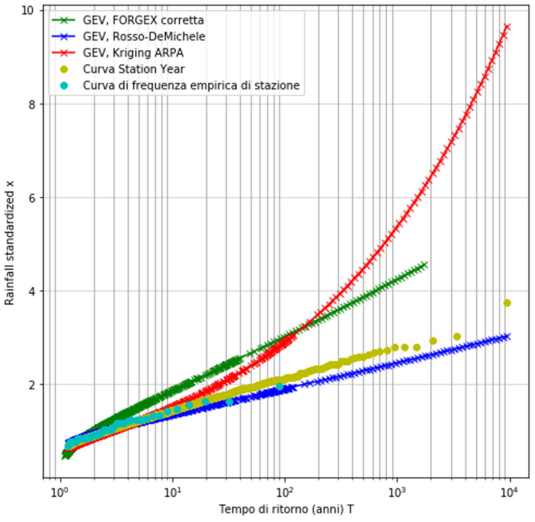
7.6.5.1 *Bardonecchia: durata 1h*

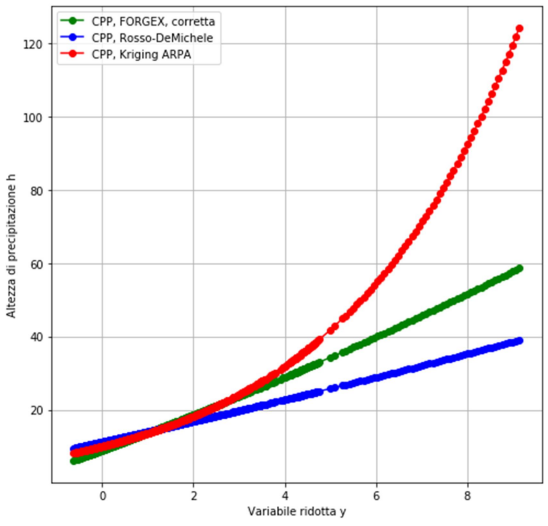


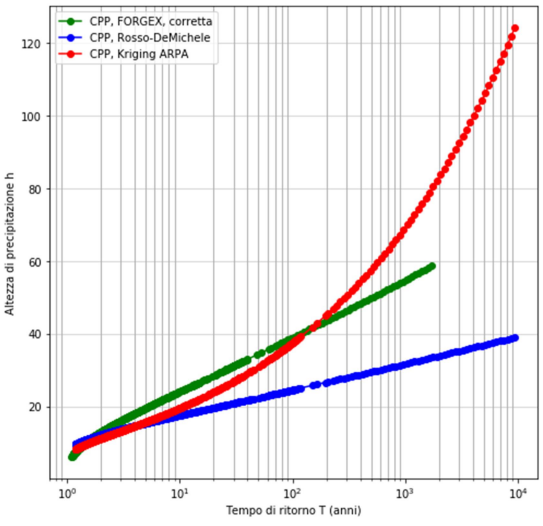
FORGEX		
α	ε	k
0.438	0.847	-0.032
VAPI		
α	ε	k
0.214	0.872	-0.021
Kriging		
θ2*	θ3	-
0.24424	-0.25417	-



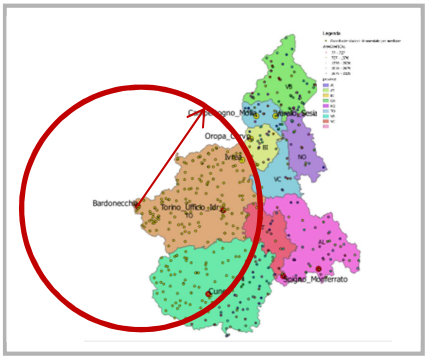




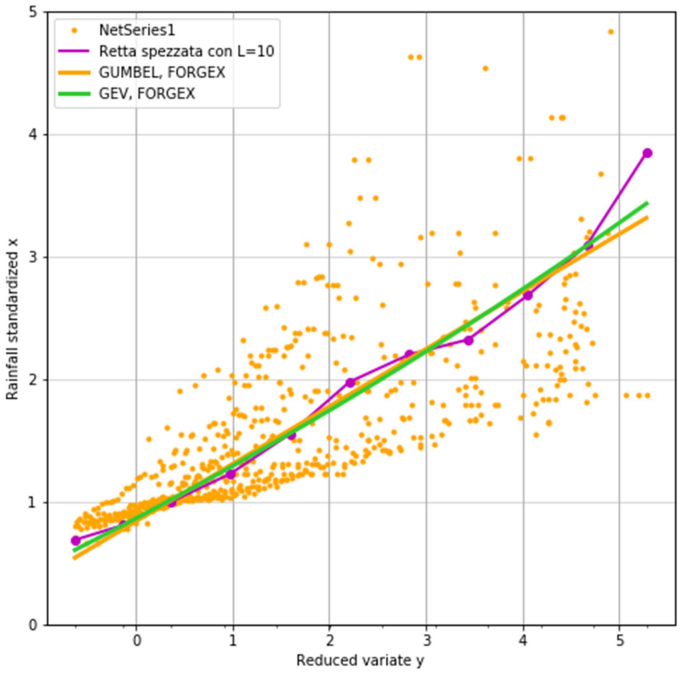


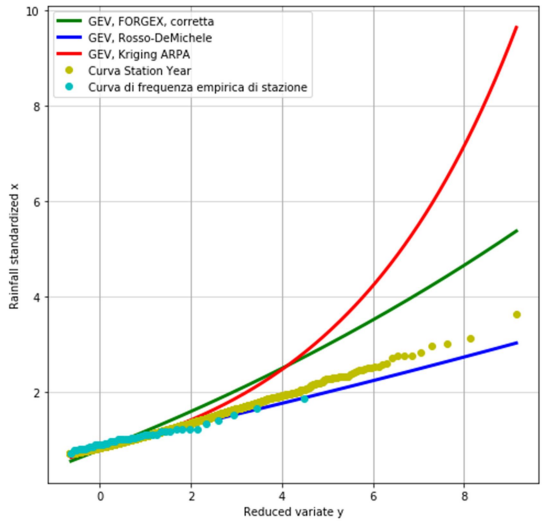


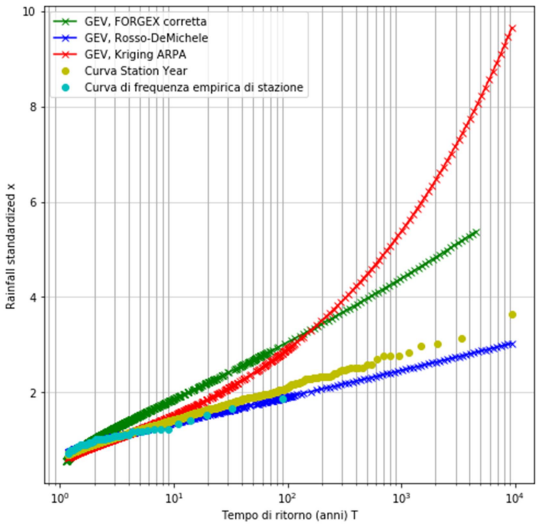
7.6.5.2 *Bardonecchia: durata 3h*

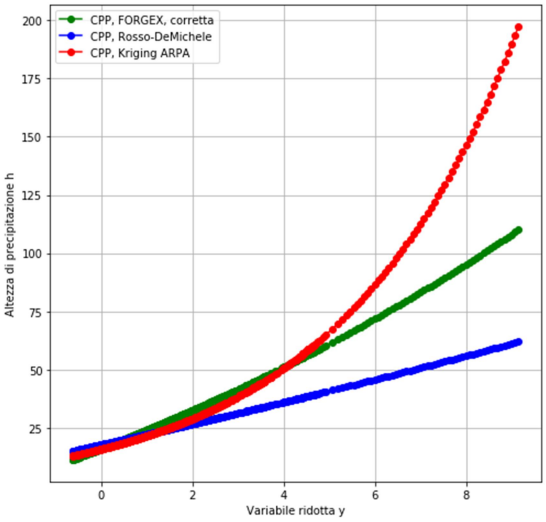


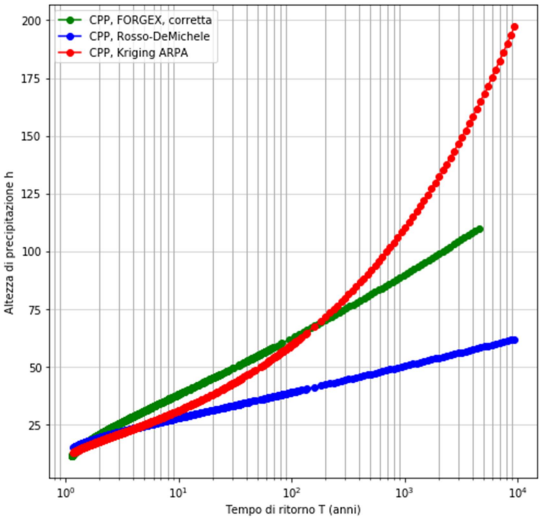
FORGEX		
α	ε	k
0.414	0.863	-0.058
VAPI		
α	ε	k
0.214	0.872	-0.021
Kriging		
θ2*	θ3	-
0.24424	-0.25417	-



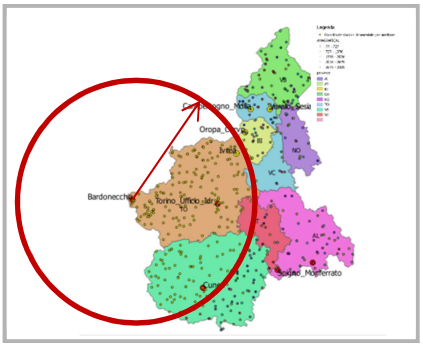




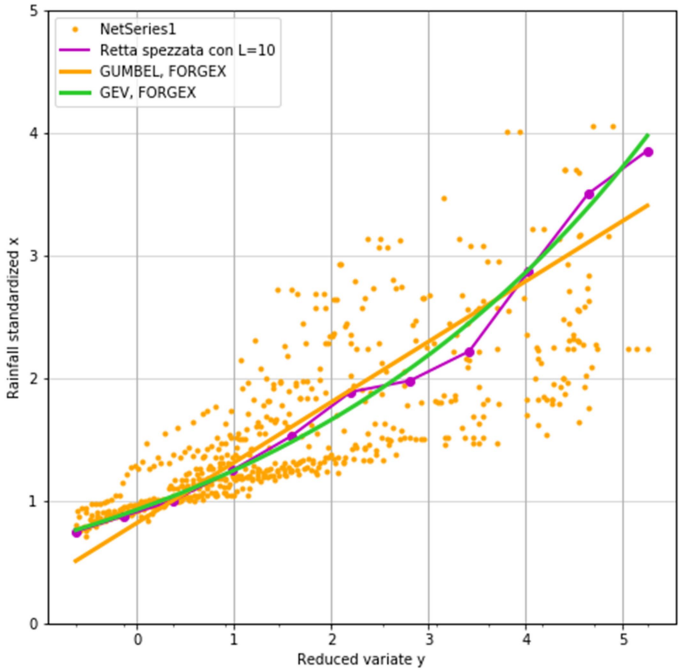


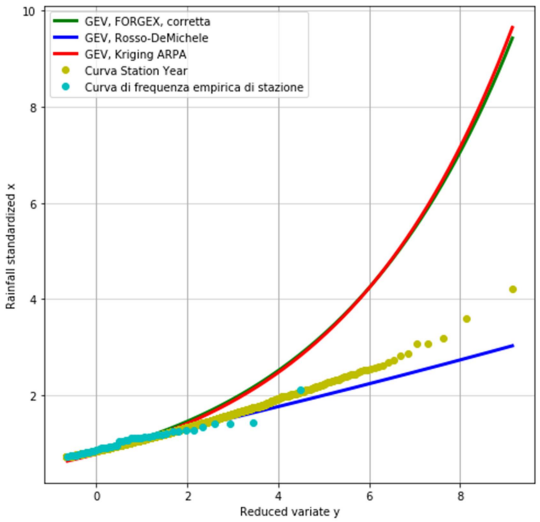


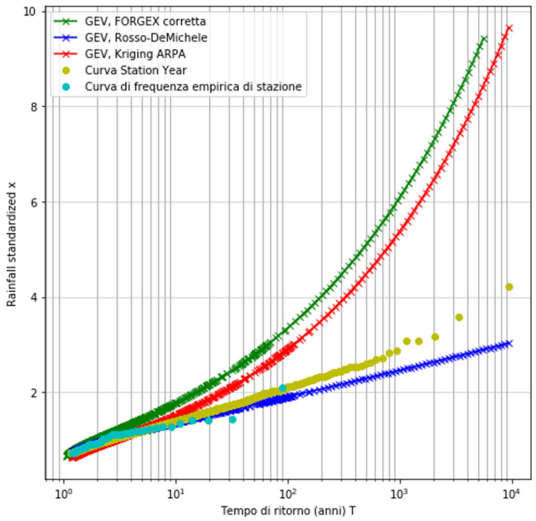
7.6.5.3 *Bardonecchia: durata 6h*

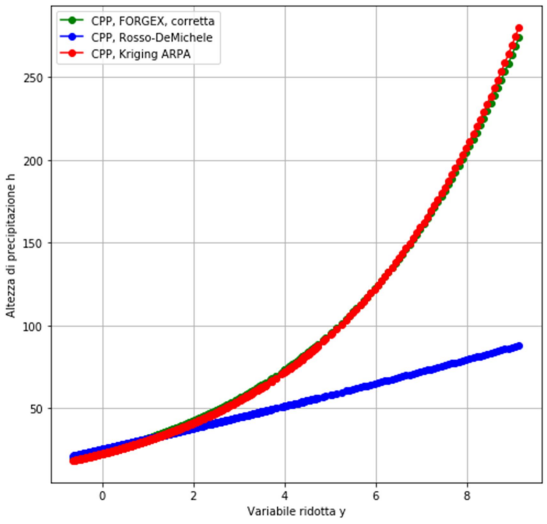


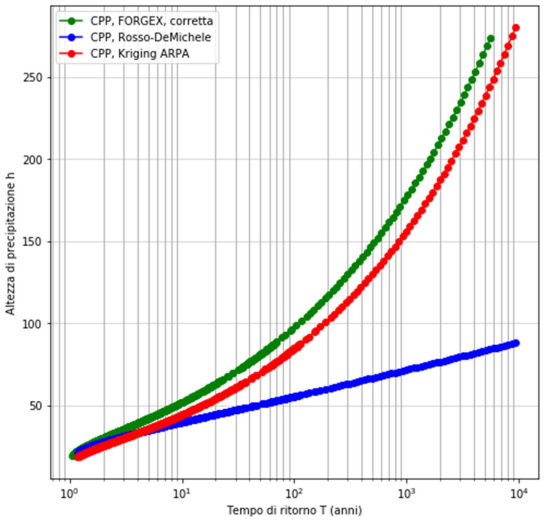
FORGEX		
α	ε	k
0.283	0.927	-0.246
VAPI		
α	ε	k
0.214	0.872	-0.021
Kriging		
θ2*	θ3	-
0.24424	-0.25417	-



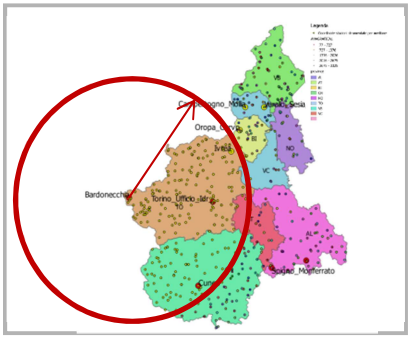




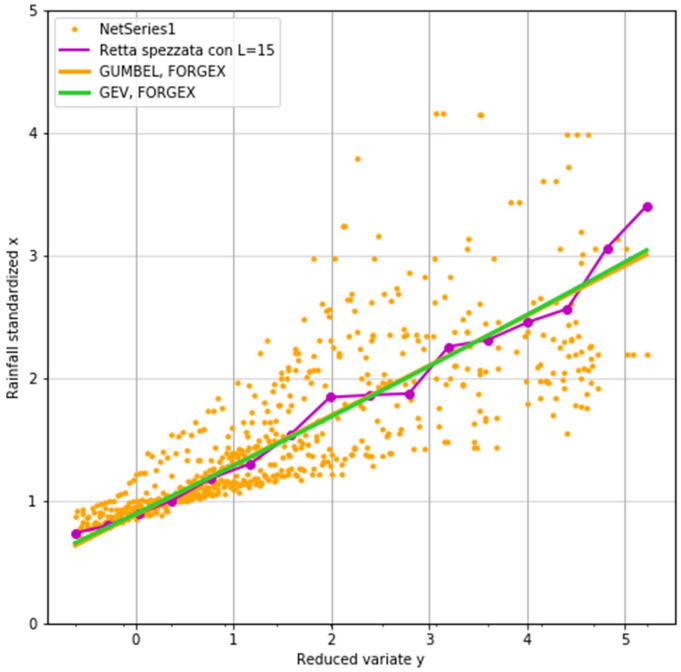


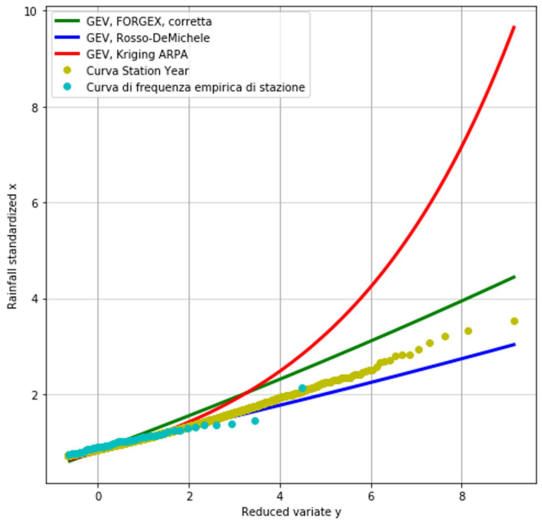


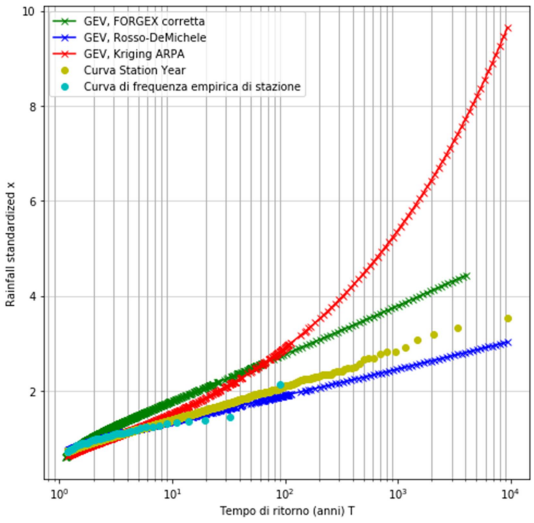
7.6.5.4 *Bardonecchia: durata 12h*

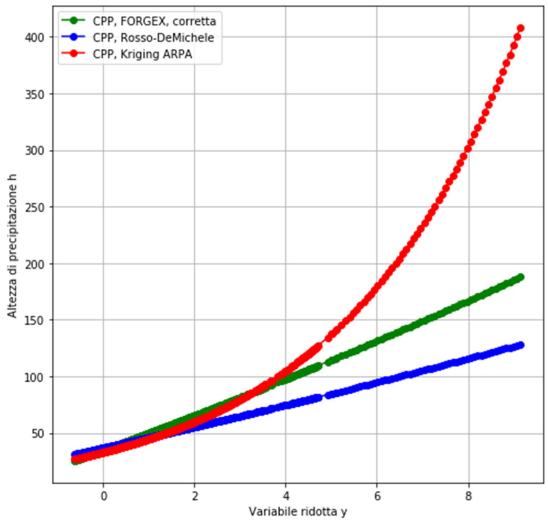


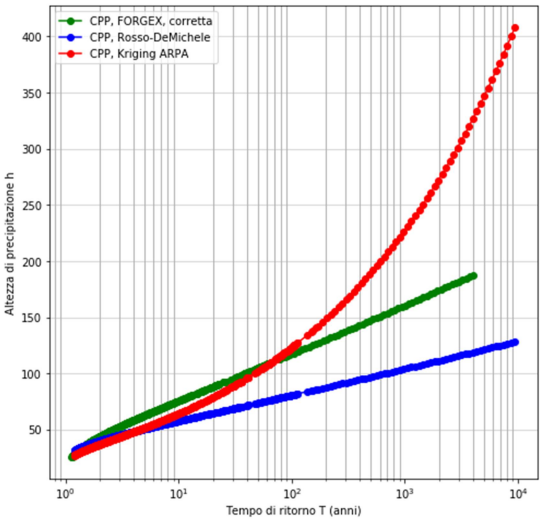
FORGEX		
α	ε	k
0.387	0.892	-0.023
VAPI		
α	ε	k
0.214	0.872	-0.021
Kriging		
θ2*	θ3	-
0.24424	-0.25417	-



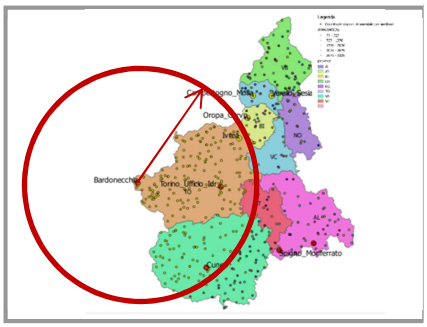




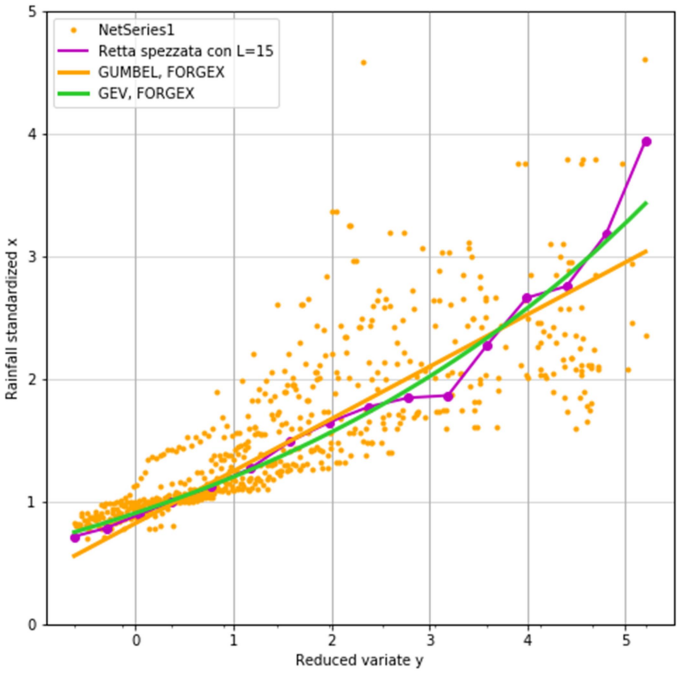


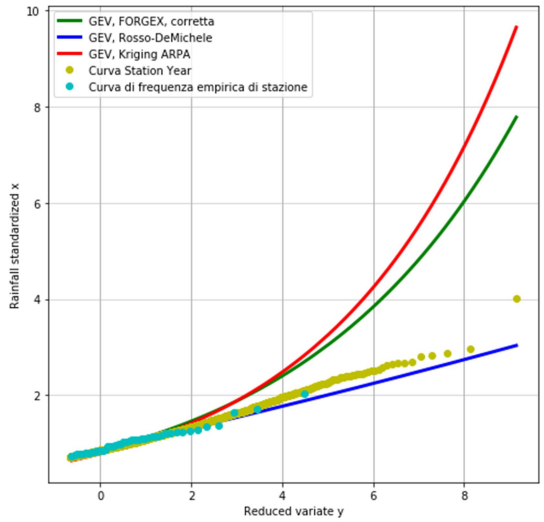


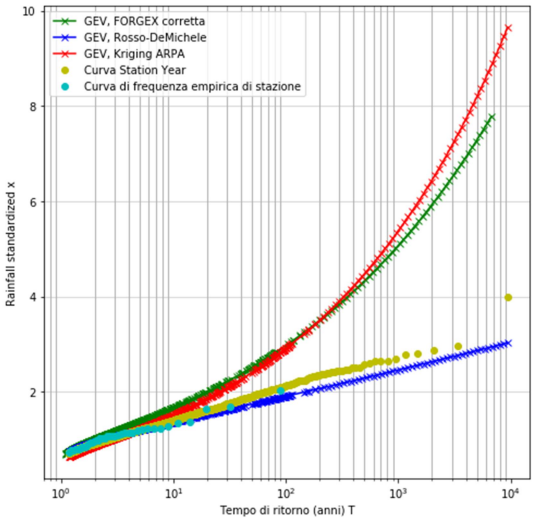
7.6.5.5 *Bardonecchia: durata 24h*

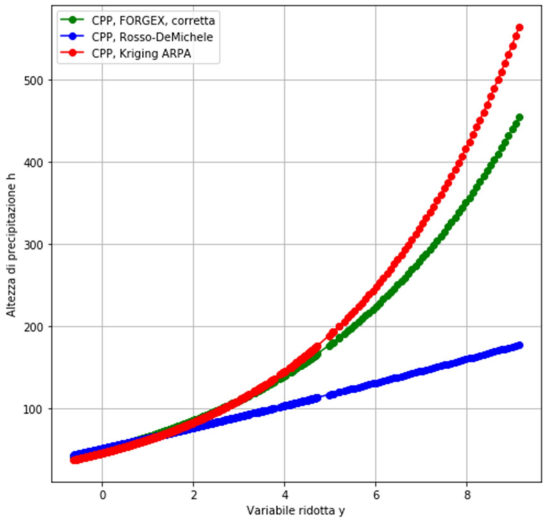


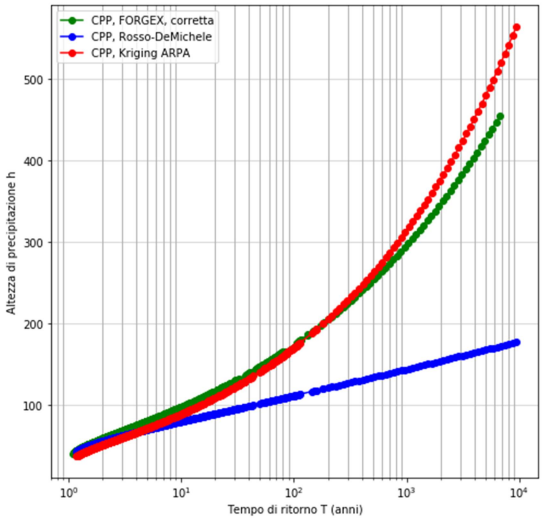
FORGEX		
α	ε	k
0.265	0.905	-0.212
VAPI		
α	ε	k
0.214	0.872	-0.021
Kriging		
θ2*	θ3	-
0.24424	-0.25417	-











7.7 *Appendice G: mappe regionalizzate per il Kriging*

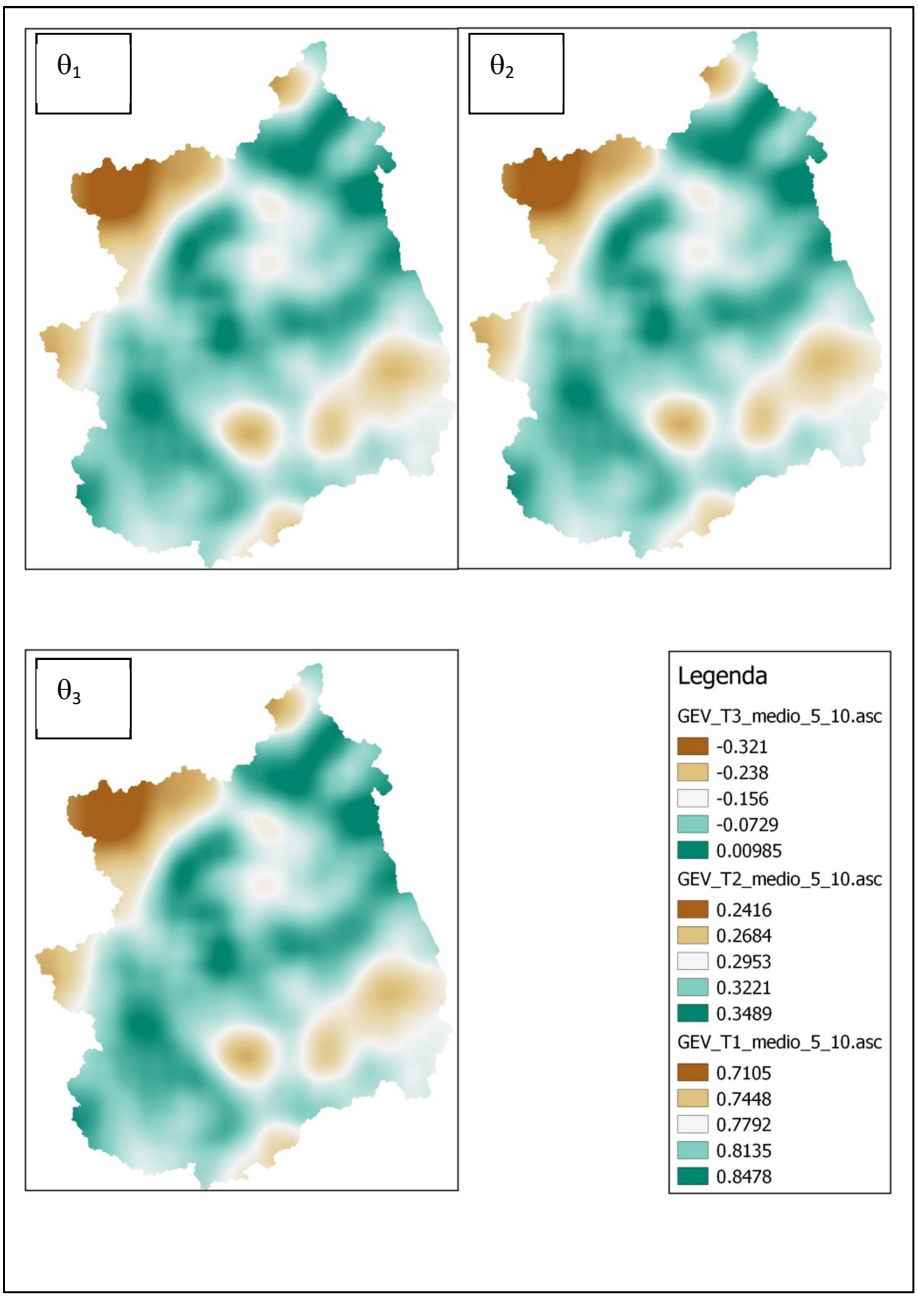


Figura 7. 6: Rappresentazione delle mappe regionalizzate dei parametri θ_1 θ_2 θ_3 utilizzati per il calcolo della distribuzione con il metodo Kriging

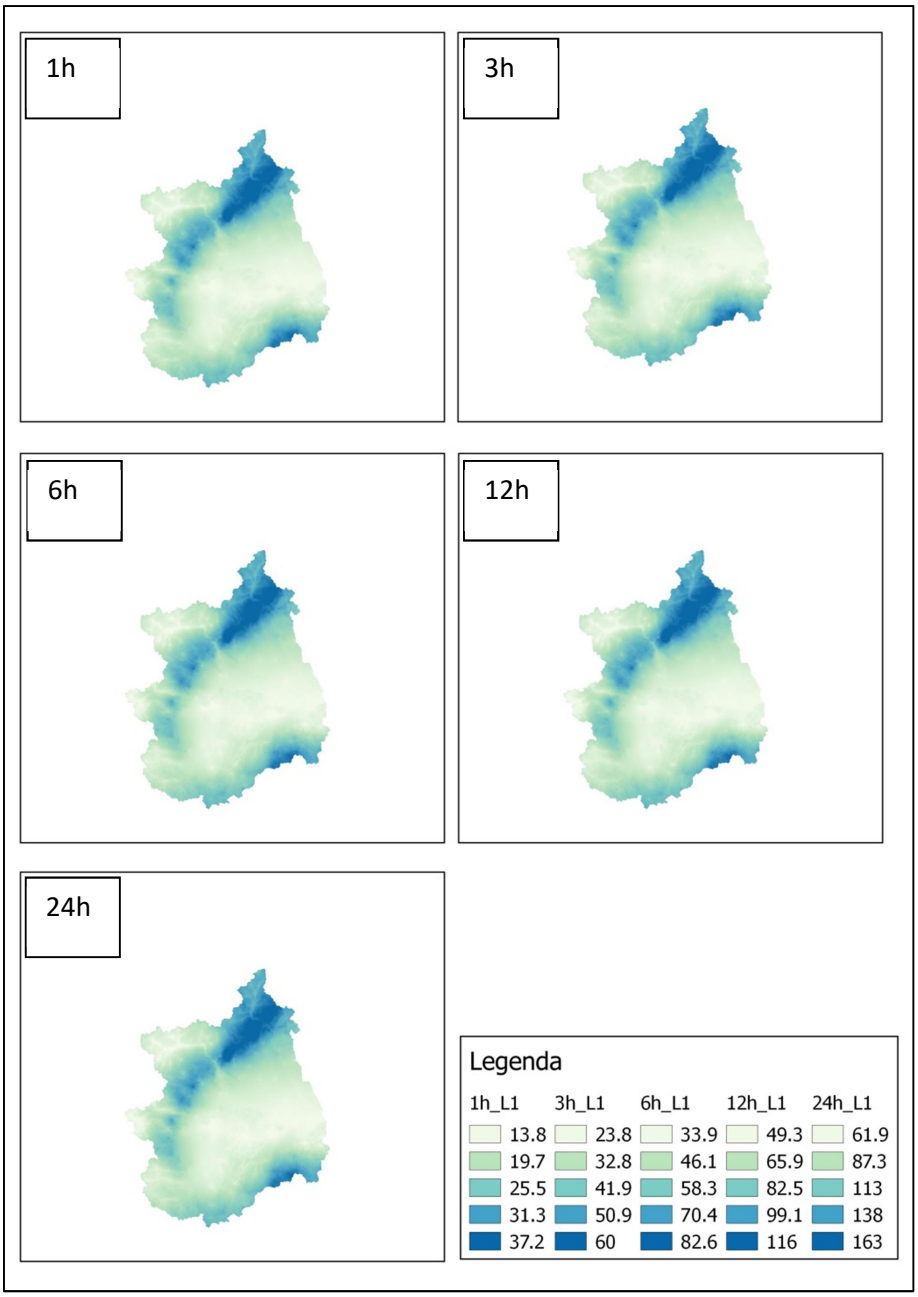


Figura 7. 7: rappresentazione delle mappe contenenti la regionalizzazione del valore medio, utilizzato per la costruzione della curva CPP.

Bibliografia

Brunner, M. I., Furrer, R., Sikorska, A. E., Viviroli, D., Seibert, J., & Favre, A. (2018, Febbraio 24). Synthetic design hydrographs for ungauged catchments: a comparison of regionalization methods. *Stochastic environmental Research and Risk Assestment* , p. 1993-2023.

Carraro, M. (2008). *Analisi regionalizzata delle precipitazioni per l'individuazione di curve segnalatrici di possibilità pluviometrica di riferimento.* Padova: Nordest Ingegneria s.r.l.

Claps , P., Laio, F., & Viglione , A. (2007). *Valutazione delle risorse idriche utilizzabili per obiettivi multipli attraverso la realizzazione di alcuni grandi invasi artificiali in Piemonte.* Regione Piemonte.

Dales, M. Y., & Reed , D. W. (1989). *Regional Flood and Storm hazard assesment.* Wallingford, UK: Institute of Hydrology.

Das, S. (2018, Luglio 24). Extreme rainfall estimation at ungauged sites: Comparison between region-of-influence approach of regional analysis and spatial interpolation technique. *International Journal of Climatology* , p. 1-17.

De Michele, C., & Rosso, R. (2001). *Rapporto sulla Valutazione delle Piene in Italia Nord Occidentale.* Milano: Consiglio Nazionale delle Ricerche.

Faulker, D. (1999). *Flood Estimation Handbook: Volume 2. Rainfall frequency Estimation.* Wallingford, UK: Istitute of Hydrology.

Faulkner, D. S., & Prudhomme, C. (1998). Mapping an index of extreme rainfall across the UK. *Hydrology and Earth system Sciences* , p. 183-194.

Ganora, D., Laio , F., & Claps , P. (s.d.). *Sintesi dei metodi di stima regionali delle piene di progetto.* Politecnico di Torino.

Houghton-Carr, H. (1999). *Flood Estimation Handbook: Volume 4. Restatement and application of the Flood Studies Report rainfall-runoff method .* Wallingford, UK: Istitute of Hydrology.

Kim, Jungwon & Nam, Woosung & Shin , Ju-young & Heo, & Jun-Haeng. (2019). *Application of the FORGEX method to Annual Maximum Rainfall in Korea.*

Met Eireann . (Ottobre 2007). *Estimation of Point Rainfall Frequencies.* Office of Public Works (OPW).

Ouarda, T. (2001). *Hydrological Frequency Analysis, Regional.* John Wiley and Sons.

Products, E. S.-E. (s.d.). *Valori estremi della precipitazione e loro ragguaglio areale nel bacino del fiume Reno.* Bologna: ET&P Srl.

Reed , D. W., Faulkner, D. S., & Stewart, E. J. (1999, 3 (2)). The FORGEX method of rainfall growth estimation II: Description. *Hydrology and Earth System Sciences Discussions, European Geosciences Union*, p. 197-203.

Reed, D. (1999). *Flood Estimation Handbook: Volume 1. Overview.* Wallingford, UK: Istitute of Hydrology.

Reed, D. W., & Stewart , E. J. (1994). Iuter-site and inter-duration dependencein rainfall extremes. *Statistics for the Environment 2: water-related issues* , p. 125-143.

Robson , A., & Reed, D. (1999). *Flood Estimation Handbook: Volume 3. Statistical procedures for flood frequency estimation.* Wllingford: Istitute of Hydrology.

Stewart, E. J., & Reynard, N. S. (1991). *Rainfall Frequency estimation in England and Wales.* Wallingford: Institute of Hydrology .

Svensson , C., & Jones , D. A. (2010, 3). Review of rainfall frequency estimation methods . *Journal of Flood Risk Management* , p. 296-313.

Viglione, A. (2007). *Metodi statistici non-supervised per la stima di grandezze idrologiche in siti non strumentati.* Tesi di dottorato presso il Politecnico di Torino.

Sitografia

https://elearning.unipd.it/scuolaamv/pluginfile.php/17950/mod_resource/content/1/II_L7_31032014.pdf, 'Idraulica e Idrologia, Lezione 7'.

http://profpozzi-classe004.blogspot.com/2017/11/le-coordinate-geografiche.html

http://www.biostatistica.unich.it/mat_didattica/MENNA/A_Menna_UniCH_Lez2_2008_DistribuzioneNormale.pdf

http://geomatica.como.polimi.it/corsi/cartografialaurea/cartografiaufficialeitaliana/cartografiaufficialeitalianaigmi.htm

https://geomappando.com/2016/07/07/sistemi-riferimento-coordinate-2-parte/

http://dispense.dmsa.unipd.it/putti/idrologia/stat/node13.html

http://www.idrologia.polito.it/didattica/PIT/2013/2_AnalisiRegionale/RegionalFloodFrequencyAnalysis.pdf

https://it.wikipedia.org/wiki/Kriging

Ringraziamenti

''L'esperienza mi ha insegnato che l'uomo, quando desidera raggiungere un traguardo, trova dentro di sé risorse impensabili.''

Umberto Veronesi

Questo capitolo conclude ufficialmente il mio lungo percorso universitario, che tra alti e bassi è giunto alla fine. Durante la mia carriera universitaria, ho avuto modo di incontrare molte persone, alcune delle quali hanno contribuito a questo traguardo, altre invece che nel cammino si sono allontanate per interessi differenti.

Primo fra tutti vorrei ringraziare il Prof. Pierluigi Claps, che mi ha permesso di intraprendere e approfondire l'argomento che maggiormente mi ha incuriosito nel mio percorso di studi. Una parola d'obbligo vorrei rivolgerla all'Ing. Andrea Libertino, che è stato in grado di aiutarmi durante la conclusione dell'elaborato, sopportando i mille dubbi che ogni giorno riproponevo. Ultimo ma non meno importante, vorrei ringraziare l'Ing. Gianluca Vezzù per l'aiuto indispensabile che mi ha fornito durante la costruzione dell'applicativo in Python.

Un ringraziamento speciale è da rivolgere alla mia famiglia e ai miei nonni, che in questi lunghi anni non hanno mai smesso un momento di credere in me, nonostante a volte la situazione sembrava sfuggirmi dalle mani. In particolare vorrei ringraziare mia Mamma e mio Fratello, che soprattutto negli ultimi mesi, hanno sopportato qualunque tipologia di crisi nervosa che mi ha sopraffatto. E infine vorrei che questi ringraziamenti arrivassero

dal mio Papà, sempre in grado di sorreggermi e aiutarmi in qualunque momento, senza che io me ne accorga.

Durante questo lungo percorso ho avuto modo di incontrare due persone, che sono diventate per me un punto di riferimento. Moreno, il mio fidanzato, che negli ultimi sei anni ha combattuto nelle retrovie la mia lunga battaglia, sempre pronto ad offrirmi una parola di conforto e una spalla su cui appoggiarmi.

In secondo luogo vorrei ringraziare Elena, sempre presente nei momenti migliori, ma anche in quelli più difficili, senza la quale probabilmente oggi non sarei la persona che sono diventata.

Durante il mio cammino, ho incontrato molte persone che hanno reso le giornate al poli molto più leggere e spensierate: vorrei ringraziare Balan, Wonder, Daniele e Antonio, ma anche le persone conosciute nel corso degli ultimi anni della Magistrale, come Stella, Gioacchino e Sarah.

Un ringraziamento particolare vorrei dedicarlo a Mara, con la quale ho condiviso fianco a fianco questo ultimo anno, costruendo in questo modo un'amicizia speciale.

Grazie!