

Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA

**MODELLISTICA FISICA PER LA PROTEZIONE
DELL'AMBIENTE**

Ciclo XX

Settore scientifico disciplinare di afferenza: GEO/12

TITOLO TESI

**MODELLO BAYESIANO PER LA RIDUZIONE
DELL'INCERTEZZA NELLA PREVISIONE DELLE PIENE
DEL FIUME RENO**

Presentata da Fabrizio Nerozzi

**Coordinatore Dottorato:
Prof. Ezio Todini**

**Relatore:
Prof. Stefano Tibaldi**

Esame finale anno 2008

*Anything goes,
Rodica.*

Indice

Abstract	4
Introduzione	7
1 Il problema della previsione di piena	15
1.1 Premessa	15
1.2 Scopo del lavoro	20
1.3 Il modello stocastico autoregressivo	21
1.4 Il modello idrologico TOPKAPI	27
1.4.1 La componente del suolo.	29
1.4.2 La componente del deflusso superficiale.	31
1.4.3 La componente del deflusso in rete drenante.	32
1.4.4 La componente dell'accumulo e scioglimento della neve.	33
1.4.5 La curva aree sature–volumi immagazzinati.	34
1.4.6 L'uso del modello digitale del terreno.	35
1.4.7 I parametri del modello.	37
1.4.8 I dati idro–meteorologici in ingresso.	40
1.5 L'accoppiamento dei modelli idro–meteorologici	41
1.5.1 Modellistica dei processi atmosferici alla mesoscala	41
1.5.2 Le previsioni di “ensemble”	44
1.6 Il bacino del fiume Reno	45
2 Il modello bayesiano di previsione	51
2.1 La teoria bayesiana della stima	51
2.2 Filtro di Kalman (H^2)	56
2.3 Filtro di Kalman adattativo	63
2.4 Modellazione dell'errore	65
2.5 Filtro minimax (H^∞)	69

3	Risultati	81
3.1	Riepilogo meteorologico	81
3.2	Analisi statistica	89
3.3	Analisi degli eventi di piena	96
3.3.1	7–8 novembre 2003	96
3.3.2	27–28 novembre 2003	107
3.3.3	31 ottobre – 1 novembre 2004	115
4	Conclusioni e prospettive future	127
	Bibliografia	133
	Ringraziamenti	138

Abstract

Many efforts have been devoting since last years to reduce uncertainty in hydrological modeling predictions. The principal sources of uncertainty are provided by input errors, for inaccurate rainfall prediction, and model errors, given by the approximation with which the water flow processes in the soil and river discharges are described.

The aim of the present work is to develop a bayesian model in order to reduce the uncertainty in the discharge predictions for the Reno river. The 'a priori' distribution function is given by an autoregressive model, while the likelihood function is provided by a linear equation which relates observed values of discharge in the past and hydrological TOPKAPI model predictions obtained by the rainfall predictions of the limited-area model COSMO-LAMI. The 'a posteriori' estimations are provided throw a H^∞ filter, because the statistical properties of estimation errors are not known. In this work a stationary and a dual adaptive filter are implemented and compared. Statistical analysis of estimation errors and the description of three case studies of flood events occurred during the fall seasons from 2003 to 2005 are reported. Results have also revealed that errors can be described as a markovian process only at a first approximation.

For the same period, an ensemble of 'a posteriori' estimations is obtained

throw the COSMO-LEPS rainfall predictions, but the spread of this 'a posteriori' ensemble is not able to encompass observation variability. This fact is related to the building of the meteorological ensemble, whose spread reaches its maximum after 5 days.

In the future the use of a new ensemble, COSMO-SREPS, focused on the first 3 days, could be helpful to enlarge the meteorological and, consequently, the hydrological variability.

Introduzione

La scienza cerca di spiegare ogni fenomeno mediante un insieme di leggi deterministiche, modelli, in cui sono stabilite le relazioni matematiche tra opportune grandezze in grado di descrivere gli effetti osservabili. Tuttavia, l'esperienza dimostra che in natura non esistono fenomeni per i quali sia possibile, tramite queste relazioni, determinare con esattezza gli effetti, sia perché le misure degli osservabili non sono esenti da errore, sia perché di solito i fenomeni sono talmente complessi che difficilmente possono essere compresi nella loro interezza.

La conoscenza, o meglio, la pretesa della conoscenza esatta dei fenomeni naturali appare perciò più come un'astrazione concettuale, piuttosto che una concreta aspirazione. Si potrebbe tentare un'analogia tra l'idea che soggiace alla base della scienza deterministica di conoscere esattamente la realtà che determina un dato fenomeno e la conoscenza della cosa in sé kantiana. Il pensiero di Kant si pone infatti al vertice di una lunga riflessione sulla vera natura dei fenomeni iniziata già nel V sec. a.C. con l'atomismo di Democrito; da Cartesio a Kant i due principali indirizzi filosofici, razionalismo ed empirismo, pur divergendo su quasi tutto, concordano sulla natura puramente fenomenica del mondo sensibile e quotidiano. Nell'accezione di Kant il "fenomeno" indica "la realtà così come possiamo farne esperienza", mentre al di là dell'apparenza, ossia di come cioè le cose ci appaiono,

si colloca il noumeno (da *nooúmenon* in greco, “ciò che viene pensato”), che è il modo in cui il pensiero cerca di rappresentarsi ciò che va oltre la sua capacità di conoscere. Kant spiega che la ragione non può accedere al noumeno se non come un “concetto-limite”, del quale possiamo solo dire che è a fondamento di tutto ciò di cui facciamo esperienza. In Kant i termini “noumeno” e cosa in sé non sono sovrapponibili. Il noumeno è una rappresentazione o idea della ragione, e come tale risiede nella mente umana, mentre la cosa in sé è invece ciò a cui il noumeno si riferisce. È la realtà esterna al soggetto, ciò con cui per definizione non si può entrare in alcun rapporto se non tramite il pensiero filosofico, poiché si pone al di là di ogni esperienza possibile.

Il luogo deputato per continuare una discussione sul piano filosofico non può essere certamente questo, ma l’invasione di campo è stata utile per definire in termini generali ciò che è stato poi l’oggetto del lavoro svolto in questa tesi: l’incertezza associata alla conoscenza della realtà e, quindi, alla previsione di un determinato fenomeno. L’incertezza è un’esperienza che si riscontra quotidianamente in tutte le attività umane ma soltanto di recente ha cominciato ad essere oggetto di un’ampia discussione nella comunità scientifica idrologica [23, 26]. Può essere considerata come un attributo informativo alla stima od alla previsione del fenomeno, ed è legata alla teoria della probabilità, nonostante molti autori siano convinti che l’incertezza possa essere efficacemente trattata in un ambito molto più ampio, all’interno del quale il contenuto statistico è soltanto una, seppure importante, delle componenti informative [22, 23].

Il significato di incertezza in ambito idrologico può essere compreso in due categorie principali: incertezza nella descrizione del processo da parte del modello ed incertezza nella previsione di piena, legata a sua volta all’incertezza nella

previsione della precipitazione. Per entrambi i casi, comunque, si tratta dell'incertezza relativa alla previsione deterministica di un modello, che è il problema che interessa maggiormente agli utenti finali. Infatti, la maggior parte dei metodi diretti a valutare l'incertezza di un sistema è di derivarla dalla conoscenza della proprietà statistiche del modello e dei dati in ingresso. Questo approccio, però, può risultare limitativo per due ragioni principali: se da un lato la determinazione delle proprietà statistiche delle stime può risultare di difficile realizzazione, dall'altro lato queste proprietà possono non essere conosciute dettagliatamente. Si è cercato di superare le difficoltà del primo tipo sviluppando dei metodi analitici approssimati, quali ad esempio il "first-order reliability method" (FORM) ed il "second-order reliability method" (SORM), mentre per il secondo problema soltanto recentemente sono stati proposti dei metodi di valutazione dell'incertezza nell'ambito della previsione idrologica, basati sullo sviluppo di tecniche di analisi statistica degli errori dei modelli, quale ad esempio il "Bayesian Forecasting System", e sull'approccio meta-gaussiano [14, 22]. Tuttavia, il punto debole di questi metodi è la correttezza delle assunzioni di stazionarietà e di ergodicità attribuite agli errori dei modelli, necessarie per potere derivare le loro proprietà statistiche sulla base di una simulazione [23].

Un'altra possibilità di valutare l'incertezza è quella di usare le proprietà statistiche ottenute da una simulazione per generare con il metodo Monte Carlo altre simulazioni. Questo è l'approccio su cui si basa il "Generalized Likelihood Uncertainty Estimation" (GLUE) [3]. Le serie relative ad ognuna delle simulazioni vengono pesate mediante un sistema di verosimiglianza. Un problema legato a questo tipo di approccio è quello di scegliere un'appropriata procedura di campionamento. Infatti, se fare diverse simulazioni variando i parametri è relativamente

semplice, non appare invece ancora chiaro come campionare lo spazio della struttura del modello. Un secondo problema è relativo al fatto che la misura di verosimiglianza normalizzata, impiegata dalla GLUE, non è una stima consistente della densità di probabilità delle simulazioni dei modelli e, conseguentemente, i limiti degli intervalli di confidenza non sono necessariamente in grado di comprendere le osservazioni. A meno che questo problema non venga risolto, GLUE non dovrebbe essere considerato come un metodo probabilistico, quanto piuttosto un'analisi di sensibilità pesata [16, 23]. Altri metodi non probabilistici nati negli anni '50 sono una generalizzazione della teoria delle probabilità, tra i quali la teoria degli insiemi casuali, la teoria dell'evidenza, la teoria degli insiemi fuzzy e la teoria delle possibilità.

In questa breve panoramica sono state individuate quattro categorie di metodi di analisi per determinare l'incertezza della stima di un modello e possono essere così denominate: metodi analitici approssimati, tecniche basate sull'analisi statistica degli errori del modello, metodi numerici approssimati o analisi di sensibilità ed infine metodi non probabilistici.

La scelta di quale metodo usare per determinare l'incertezza è piuttosto complessa e dovrebbe essere specifica del tipo di modello, se si tratta di simulazioni o di previsioni, del tipo di informazione disponibile e delle prerogative dei diversi metodi. La prima indicazione è che il metodo debba essere in grado di contenere la maggior parte dell'informazione. Ad esempio, un'informazione a priori molto importante è data dall'ultimo errore di previsione, poiché gli errori dei modelli idrologici sono fortemente correlati. Di questa informazione non si tiene conto nella formulazione classica della GLUE e nell'approccio meta-Gaussiano, mentre viene considerata nel metodo SODA e nel Bayesian Forecasting System [23].

Quando sono disponibili serie temporali piuttosto lunghe di variabili osservate, informazioni importanti possono essere dedotte dall'analisi statistica degli errori. Questa tecnica può essere infatti utile per verificare l'ipotesi di stazionarietà. Tuttavia, i metodi probabilistici per la stima dell'incertezza hanno un'utilità limitata nel caso di bacini che non dispongono di reti osservative e dovrebbero essere utilizzati con attenzione per stimare l'incertezza di eventi eccezionali per i quali è sempre scarsa l'informazione storica.

Gli approcci non probabilistici, includendo anche le analisi di sensibilità, appaiono probabilmente quelli più promettenti. La GLUE è in grado di esplorare la variabilità delle stime di un modello campionando lo spazio delle possibili soluzioni, ma l'incertezza è sempre rilevante e le indicazioni che otteniamo da questi metodi non possono essere espresse in termini di probabilità.

Come già affermato, le sorgenti di incertezza possono essere divise in due categorie: la prima, denominata idrologica, è dovuta agli errori con cui vengono stimati i parametri del modello. Quando la previsione di piena è formulata con un modello non strutturale, cosiddetto "data driven", per il quale cioè la struttura di relazioni non è imposta "a priori" ma individuata dalle osservazioni, possono essere usati degli algoritmi ricorsivi, generalmente basati sul filtro di Kalman o altre tecniche di stima simili, per definire i parametri del modello.

Il problema diventa notevolmente più complesso quando sono usati invece modelli strutturali. In genere, le procedure per la definizione dei parametri sono divise in tre categorie. La prima si basa sulla correzione dei dati forniti in ingresso al modello, ad esempio la precipitazione, finché non si raggiunge un ragionevole accordo tra i dati osservati di piena e le stime del modello. Purtroppo questa procedura ha una validità limitata soltanto ai modelli di tipo serbatoio. Nella se-

condotta procedura i parametri del modello vengono calibrati in tempo reale. I limiti di questo approccio sono due: il notevole costo computazionale degli algoritmi utilizzati, quali ad esempio i filtri di Kalman non-lineari, ed il fatto che le stime dei parametri possono essere altamente instabili, con il conseguente aumento dell'incertezza nella previsione della piena. (Todini, c.p.).

La terza procedura si basa sullo sviluppo di un modello per l'errore, condizionatamente al modello strutturale. Questo modello di errore può essere realizzato in diversi modi, ad esempio con delle tecniche auto-regressive, ARMA, o con modelli ANN basati su metodi di "Nearest-Neighbours". Più recentemente sono stati tentati alcuni approcci bayesiani basati sull'applicazione di filtri di Kalman contenenti nel vettore di stato sia i valori previsti che quelli osservati (MISP) [EFFORTS, <http://www.progea.net>], oppure su una regressione tra i valori previsti e quelli osservati [14]. È importante sottolineare che, se per assimilare le osservazioni e calibrare i valori dei parametri il filtro di Kalman è lo strumento ottimale, perde molto rapidamente la sua efficacia quando viene usato per la previsione e non può essere utilizzato per valutarne l'incertezza.

I metodi per la valutazione dell'incertezza dei parametri del modello non possono rappresentare completamente l'incertezza della previsione, poiché non tengono conto di altri fattori, ugualmente, se non più importanti, quali ad esempio la precipitazione in ingresso al modello. Alla scala dei bacini di piccola e media dimensione, il problema di quantificare la precipitazione che può accadere è ancora un problema aperto. Nonostante l'utilizzo di modelli meteorologici ad area limitata e ad alta risoluzione abbia introdotto un miglioramento quantitativo nella previsione a breve termine di eventi intensi a scala locale, come i fenomeni temporaleschi, la previsione della precipitazione è tuttora affetta da errori che

possono essere considerevoli alle scale spazio-temporali tipiche delle applicazioni idrologiche.

Come è emerso dai risultati del progetto EFFS la previsione quantitativa della pioggia da parte dei modelli numerici è fortemente non accurata e non può essere usata direttamente per determinare conseguentemente l'incertezza della previsione idrologica. Come alternativa all'accoppiamento diretto è stato proposto all'interno del progetto "A Comprehensive Forecasting System for Flood Risk Mitigation and Control" (AFORISM) un approccio basato sulla combinazione bayesiana tra un ensemble di serie storiche di precipitazioni, scelte condizionatamente alle ultime misure osservate, e le precipitazioni previste dal modello ad area limitata (Todini, c.p.).

In campo meteorologico, uno dei metodi perseguiti con maggiore successo per determinare il grado di incertezza della previsione è quello di associarlo alla dispersione degli elementi di un ensemble. La grande diffusione di questi metodi ha portato sempre più ad affrontare il problema della previsione, in particolare quella degli scenari di rischio, mediante un approccio probabilistico. Alla luce dei recenti sviluppi, i sistemi di ensemble meteorologici disponibili hanno cominciato ad essere considerati strumenti in grado di fornire una previsione quantitativa di precipitazione adeguata per alimentare i modelli idrologici, permettendo, da un lato, di migliorare la qualità delle previsioni dei deflussi e, dall'altro, di giungere ad una quantificazione dell'incertezza associata a tali previsioni [2].

Nel primo capitolo, oltre ad essere delineato lo scopo del lavoro, è data anche una descrizione generale dei principali metodi numerici utilizzati in idrologia per il preannuncio della piena. Sono descritti, in maniera non formale, i processi stocastici, ed in particolare una classe specifica dei processi stocastici, rappresentabili

mediante un modello autoregressivo. Viene poi fornita una descrizione dettagliata del modello ad afflussi–deflussi TOPKAPI, effettivamente utilizzato. Nel capitolo 2 si introduce il modello bayesiano per la riduzione dell’incertezza nella previsione della piena, che combina le stime del modello stocastico autoregressivo con quelle del modello TOPKAPI. Dopo una prima parte in cui si riprende la teoria del “Bayesian Forecasting System”, viene presentata la formulazione del filtro di Kalman (H^2), sia stazionario che adattativo, e sia per errori non autocorrelati che autocorrelati. Quindi, viene descritto il filtro H^∞ come generalizzazione del filtro H^2 , e la possibilità di un suo utilizzo nella previsione di portata. I risultati del lavoro sono presentati nel capitolo 3. Nella prima parte del capitolo si possono trovare quelli generali, dati dalle elaborazioni statistiche compiute per le stagioni autunnali dal 2003 al 2005, mentre nella seconda parte l’analisi di tre eventi di piena a seguito di fenomeni temporaleschi avvenuti sul sottobacino montano. Infine, nel capitolo dedicato alle conclusioni vengono discussi i risultati, anche alla luce di un possibile utilizzo di questo sistema ai fini di Protezione Civile, e si danno alcune indicazioni per un possibile sviluppo del lavoro.

Capitolo 1

Il problema della previsione di piena

1.1 Premessa

La piena viene intesa come il rapido sopraelevamento della superficie libera di un corso d'acqua dovuto all'aumento di portata. Il fenomeno è determinato da precipitazioni di forte intensità o, più raramente, dalla rapida fusione di nevai e ghiacciai [9].

La previsione della piena viene generalmente effettuata attraverso l'utilizzo di modelli idrologici, classificabili secondo diversi criteri. Il primo criterio è sulla base dei dati forniti in ingresso, distinguendo tra i modelli idraulici, basati sui dati idrometrici, in grado di descrivere soltanto i processi interessanti la rete drenante, come ad esempio il trasferimento dei flussi lungo l'alveo del fiume, da quelli basati sui dati idro-meteorologici ed idrometrici, capaci invece di rappresentare anche i processi lungo i versanti, come i modelli di trasformazione afflussi-deflussi. Le condizioni di deflusso possono rappresentare un secondo criterio. Si possono distinguere i modelli completi, o di simulazione continua, in grado di rappresentare

il comportamento del bacino idrografico in diverse situazioni e modelli ad equivalenza limitata, o modelli ad evento, in grado di descrivere la piena soltanto in occasioni di eventi particolari. Si possono distinguere i modelli secondo lo scopo cui sono destinati, per la ricerca, per la progettazione o per la gestione operativa, per la scala spaziale, monodimensionali, bidimensionali e tridimensionali.

In questa sede ci è sembrato opportuno attuare una classificazione basata sul grado di conoscenza dei processi idrologici, distinguendo quindi tra i modelli stocastici, per i quali non viene presupposto nessun tipo di conoscenza, i modelli integrali ed infine i modelli differenziali distribuiti, nei quali invece il comportamento delle diverse parti del bacino è descritto mediante una formulazione matematica complessa ed onerosa.

I modelli stocastici hanno una natura empirica e si basano su un operatore matematico che mette in relazione le variabili di stato ad istanti successivi. Il modello autoregressivo, che verrà descritto in seguito, ne è un esempio. Esistono altri modelli stocastici che mettono in relazione la portata defluente all'afflusso meteorico. È bene sottolineare che il sistema fisico da modellare viene trattato come una scatola nera, *black box*, e le relazioni di tipo statistico che sostanziano il modello stocastico non hanno nessuna pretesa esplicativa di come i processi idrologici realmente avvengono. I parametri del modello vengono stimati, come descritto in seguito per il modello autoregressivo, mediante delle tecniche statistiche di minimizzazione di una funzione di rischio.

I modelli integrali, o concettuali, occupano una posizione intermedia tra quelli stocastici e quelli che perseguono un approccio puramente deterministico. Possono essere ulteriormente divisi tra quelli di tipo concentrato e quelli di tipo semi-distribuito.

I modelli integrali di tipo concentrato sono in grado di descrivere la dinamica globale del bacino, o del sottobacino, a seguito di una sollecitazione. Ciò è consentito attraverso un'integrazione spaziale delle precipitazioni, della caratteristica del suolo e della topografia [5]. La determinazione dei parametri è in genere affidata alla realizzazione di tecniche statistiche, ma le complesse relazioni tra le diverse componenti del modello rendono difficile una loro definizione da un punto di vista fisico o idrologico. Soltanto nei casi in cui la risposta impulsiva derivi dall'integrazione analitica delle equazioni differenziali del moto, come ad esempio per la propagazione mono-dimensionale cinematica o parabolica, è possibile trovare i valori dei parametri sulla base di considerazioni di natura fisica [31]. In questa categoria si collocano tutti i modelli formulati sotto forma di risposta impulsiva, come il canale lineare, il serbatoio lineare, la cascata di serbatoi lineari e non, il modello di Nash, nel quale la conoscenza del sistema fisico viene introdotta attraverso vincoli matematici che esprimono la continuità della massa o le condizioni di regolarità delle soluzioni di un sistema dinamico retto da equazioni differenziali lineari [25]

I modelli integrali semidistribuiti ipotizzano la conoscenza dei singoli fenomeni che interagendo tra loro danno luogo al comportamento complessivo del bacino, o del sottobacino. La stima dei parametri avviene abbinando considerazioni di natura fisica a funzioni di rischio da minimizzare. In questa categoria si collocano lo Stanford Watershed Model IV, il SSARR, il Sacramento, il Tank, nei quali i singoli fenomeni vengono rappresentati a livello di sottobacino mediante relazioni empiriche o risposte impulsive, e quindi integrati a livello di bacino, imponendo l'uguaglianza delle condizioni al contorno [6, 7, 27, 28, 31]. Un'evoluzione di questa categoria è rappresentata dai modelli, quali lo Xinanjiang, il

Probability Distribution Moisture Model, il Synthesized Constrained Linear Systems, nei quali il deflusso superficiale segue un meccanismo di saturazione dal basso, anche detto per eccesso di saturazione, o di Dunne [24, 33, 38]. Il suolo raggiunge la saturazione in superficie a partire dal basso a causa della precipitazione che vi cade sopra e del deflusso ipodermico verso valle. Una componente tipica di tali modelli è una curva a due parametri che rappresenta la relazione tra il volume totale di acqua immagazzinato nel suolo e l'estensione delle aree sature. La parametrizzazione della curva è data da relazioni empiriche, di natura statistica e non da una rappresentazione fisica del processo. Al contrario, il modello TOPMODEL ipotizza che la dinamica della zona satura, ovvero l'accumulo di umidità nel suolo, possa essere rappresentata mediante la successione di stati stazionari della tavola d'acqua che si origina nello strato di suolo più superficiale, definendo così una relazione fisica tra il volume totale di acqua immagazzinata nel suolo e l'estensione delle aree sature [4]. Ciò rende ragione di come talvolta i modelli semi-distribuiti possano essere considerati come modelli fisicamente basati. Questo genere di modelli sono frequentemente usati per il controllo delle piene, in cui i processi sono rappresentati da semplici relazioni fisiche, ma il loro limite è legato alla calibrazione, la cui validità è legata alle condizioni del suolo in atto. Ciò sconsiglia l'utilizzo in tempo reale dei modelli semi-distribuiti.

I modelli differenziali distribuiti, o deterministici, costituiscono la categoria dei modelli più complessa e completa. La formulazione matematica del moto dei flussi in superficie, in falda e di scambio tra suolo ed atmosfera è nota. La simulazione di questi processi viene effettuata tramite la soluzione di un sistema di equazioni differenziali che, nei casi generali, non sono risolvibili analiticamente. L'integrazione delle equazioni può essere effettuata mediante metodi numerici

approssimati, basati sulla schematizzazione sopra un reticolo spazio-temporale. Le tecniche numeriche maggiormente utilizzate in idrologia sono il metodo delle differenze finite, il metodo degli elementi finiti ed il metodo delle differenze finite integrate. Per alcuni processi idrologici si deve ricorrere invece a leggi empiriche, come avviene ad esempio per la legge di Darcy che governa il flusso attraverso un mezzo poroso o per la legge di Manning relativa al flusso nei canali, che devono essere validate mediante appositi esperimenti di calibrazione.

Il valore aggiunto di questi modelli è che, oltre ai deflussi ed il volume di acqua immagazzinato nel suolo, possono essere utilizzati per prevedere gli effetti di eventuali cambiamenti intervenuti nel bacino, di eventuali cambiamenti nella destinazione d'uso del suolo, o la possibilità di rappresentare ed analizzare fenomeni di trasporto e di diffusione di sostanze inquinanti [31]. Ciò che rende invece questi modelli difficilmente realizzabili è da un lato l'integrazione dei vari fenomeni idrologici, caratterizzati da scale di tempo assai diverse, che variano da minuti e ore per il deflusso superficiale, a giorni per il deflusso ipodermico, a mesi per i deflussi in falda. Dall'altro lato, nonostante le ipotesi semplificative, la risoluzione delle equazioni che governano i processi idrologici richiede una cospicua mole di dati in ingresso e di tempo di calcolo.

Tra i modelli differenziali distribuiti più noti e completi va annoverato lo *Système Hydrologique Européen* (SHE), sviluppato tra il *Danish Hydraulic Institute*, l'*Institute of Hydrology of Wallingford* e la *SOGREAH*. Questo modello si basa su un sistema di equazioni differenziali che descrivono l'equazioni di continuità della massa e della quantità di moto sulla base delle condizioni al contorno [1]. In seguito, con lo sviluppo dei sistemi di calcolo, dei sistemi informativi geografici (GIS) e dell'evoluzione della modellistica idrologica sono stati sviluppati altri

modelli, caratterizzati da una parametrizzazione più semplice, quali, ad esempio, DHSVM, LISFLOOD, WATFLOOD, TOPKAPI e FEWS NET Stream flow model, che possono essere utilizzati per la gestione operativa nel preannuncio delle piene [8, 13, 32, 34, 35].

1.2 Scopo del lavoro

Un problema molto importante che sta nelle motivazioni del presente lavoro è la comunicazione di una previsione incerta a coloro che devono prendere decisioni in caso di allerta. Generalmente l'efficacia di un sistema di allerta dipende anche dalla semplicità dei messaggi diffusi che derivano dal corretto utilizzo di procedure probabilistiche, tali da rendere ottimale la prevenzione dal rischio di alluvione. Krzysztofowics afferma che *“Rational decision making (for flood warning, navigation, or reservoir systems) requires that uncertainty about a hydrological predictand (such as river stage, discharge, or runoff volume) be quantified in terms of a probability distribution, conditional on all available information and knowledge”* [14].

Si possono individuare due punti importanti: il primo è che l'obiettivo della previsione è di dare una descrizione dell'incertezza dei valori futuri della portata di piena e non dell'incertezza della stima del modello idrologico. Il secondo punto riguarda invece come esprimere l'incertezza in termini di densità di probabilità. Attorno a questa duplice considerazione si è svolto il lavoro di ricerca, che ha avuto come obiettivo la formulazione di un modello bayesiano per la riduzione dell'incertezza nella previsione idrologica, mediante la combinazione delle previsioni di piena, date dal modello TOPKAPI, accoppiato al modello meteorologico

COSMO–LAMI, e dal modello statistico autoregressivo, per ottenere un'unica stima, associata al valore minimo di una determinata funzione di rischio. Il metodo bayesiano si svolge in tre momenti: scelta della funzione di distribuzione a priori, modellazione della funzione di verosimiglianza e determinazione della funzione di distribuzione a posteriori, ed ha come vantaggi il fatto di incorporare l'informazione contenuta nei dati storici, di correggere eventuali errori sistematici mediante la funzione di verosimiglianza, affinché la densità di probabilità a posteriori sia ben calibrata.

Il sistema di previsioni di ensemble a mesoscala COSMO–LEPS è stato utilizzato per fornire una previsione quantitativa di precipitazione in ingresso al modello TOPKAPI, al fine di costituire una catena modellistica meteo–idrologica operativa che elabori, in tempo reale, una previsione probabilistica di portata. L'incertezza della previsione meteorologica viene così propagata all'incertezza della previsione di portata effettuata con il modello idrologico, affinché se ne possa tenere conto nel successivo processo decisionale a carico delle Autorità responsabili.

In questo lavoro le simulazioni sono state effettuate nel sottobacino montano del fiume Reno, localizzato nel versante Nord dell'Appennino Tosco–Emiliano. Nel punto di osservazione scelto, la Chiusa di Casalecchio, la previsione delle piene di portata è relativamente complessa, poiché legata al carattere prevalentemente torrentizio del fiume, con tempi di corrivazione dell'ordine di 6–8 ore.

1.3 Il modello stocastico autoregressivo

Una successione ordinata e finita dei livelli idrologici di un fiume può essere rappresentata come parte di un'unica realizzazione di un processo stocastico. In

modo non rigoroso un processo stocastico può essere definito come una famiglia di variabili casuali, tale per cui è data una variabile casuale x_t per ogni istante t di un intervallo temporale T [37]. È possibile, in linea di principio, definire una funzione di densità $f(\dots, x_{t-1}, x_t, x_{t+1}, \dots)$ per il processo stocastico. Per ogni insieme finito $t \in T$, la funzione di distribuzione del processo stocastico può essere inferita dalla funzione di distribuzione di probabilità congiunta dell'insieme delle variabili casuali x_{t_1}, \dots, x_{t_n} . È possibile marginalizzare la funzione di densità per ogni sottoinsieme delle sue componenti, ovvero sono date le funzioni di distribuzione marginale per ognuna delle x_t .

Tuttavia, l'utilizzo del processo stocastico come base per una procedura inferenziale apre due ordini di problemi. Il primo è che se la serie osservata è l'unica realizzazione tra tutte quelle possibili, allora la possibilità di inferire sul processo non può essere data per scontata. Infatti, non c'è modo di dire quali caratteristiche sono specifiche della serie osservata e quali invece del processo stocastico. Se anche fosse possibile usare una sola realizzazione per fare inferenza sulle caratteristiche del processo, è necessario, e questo è il secondo tipo di problema che si può presentare, che queste caratteristiche campionate dalla serie siano stabili nel tempo. Si giunge pertanto alla definizione di due proprietà che i processi stocastici possono avere o non avere: la stazionarietà e l'ergodicità.

Si può parlare di stazionarietà forte (o stretta) e stazionarietà debole. Un processo stocastico $\{x_t; t \in T\}$ è detto strettamente stazionario se, per ogni h , ogni insieme finito $(x_{t_1+h}, \dots, x_{t_n+h})$ gode della proprietà di avere la medesima funzione di distribuzione congiunta dell'insieme $(x_{t_1}, \dots, x_{t_n})$, ovvero che le funzioni di distribuzione marginali sono invarianti.

La stazionarietà debole, invece, riguarda quei processi stocastici $\{x_t; t \in T\}$

che:

- (i) per ogni t il valore atteso della variabile casuale $\mathbf{E}[x_t]$ è una costante,
- (ii) la funzione di distribuzione di $(x_{t_1+h}, \dots, x_{t_n+h})$ ha la stessa matrice di covarianza per ogni h , ovvero la covarianza è una funzione di h ed è chiamata funzione di covarianza, o autocovarianza.

La stazionarietà forte non implica la stazionarietà debole. Infatti, un processo stocastico può essere stazionario in senso forte ma non possedere momenti; viceversa, l'invarianza nel tempo della matrice di covarianza non implica l'invarianza delle funzioni di distribuzione marginali. Soltanto in un caso le due proprietà coincidono, quando il processo è gaussiano, ossia quando la distribuzione congiunta di un qualunque sottoinsieme di elementi del processo è una normale multivariata [37].

L'ergodicità è una condizione che limita la memoria del processo. Un processo non ergodico è tale per cui le caratteristiche di persistenza sono così accentuate da non potere fare nessuna inferenza significativa sulle proprietà statistiche del processo sulla base di una sua realizzazione finita. In un processo ergodico, al contrario, la memoria del processo è debole e, anche per realizzazioni finite del processo, è possibile desumere le sue proprietà statistiche. In questo caso, la significatività delle inferenze aumenta con l'aumentare della lunghezza del segmento del processo preso in esame. Un processo è ergodico quando eventi lontani tra loro nel tempo possono essere considerati praticamente indipendenti, ossia quando
$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbf{E}[x_t x_{t-k}^T] = 0$$

I processi autoregressivi costituiscono una classe particolare di processi stocastici e forniscono una rappresentazione intuitiva di una serie temporale persistente.

La determinazione della variabile casuale x_t è data da una funzione lineare delle determinazioni nel passato delle variabili casuali del processo stocastico, x_{t-1} , x_{t-2} , ..., e da un'ulteriore variabile casuale, relativa ad un processo stocastico "white noise", ossia un processo stocastico stazionario che non ha alcuna memoria di sé. Ad esempio, uno schema autoregressivo di ordine p per una data serie temporale $\{x_t\}$ stabilisce che per coefficienti costanti $\alpha_1, \dots, \alpha_p$, la serie temporale $u_t = x_t - \alpha_1 x_{t-1} - \dots - \alpha_p x_{t-p}$ sia un rumore bianco con le seguenti proprietà statistiche:

$$\begin{cases} \mathbf{E}[u_t] = 0 \\ \mathbf{E}[u_s u_t^T] = \delta_{st} \mathbf{Q} \end{cases} \quad (1.1)$$

Un processo stocastico autoregressivo di ordine p , $\text{AR}(p)$, può essere rappresentato come un processo multivariato di primo ordine, $\text{VAR}(1)$:

$$\begin{bmatrix} x_t \\ x_{t-1} \\ \dots \\ x_{t-p-2} \\ x_{t-p-1} \end{bmatrix} = \begin{bmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_{p-1} & \alpha_p \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ & & \dots & & \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ x_{t-2} \\ \dots \\ x_{t-p-1} \\ x_{t-p} \end{bmatrix} + \begin{bmatrix} u_t \\ 0 \\ \dots \\ 0 \\ 0 \end{bmatrix} \quad (1.2)$$

e quindi in forma vettoriale:

$$\begin{aligned} \mathbf{x}_t &= \mathbf{F} \mathbf{x}_{t-1} + \mathbf{u}_t \\ &= \mathbf{F}^t \mathbf{x}_0 + \mathbf{F}^{t-1} \mathbf{u}_0 + \dots + \mathbf{F} \mathbf{u}_{t-1} + \mathbf{u}_t \end{aligned} \quad (1.3)$$

La crescita della varianza dell'errore della stima è descritta da:

$$\begin{aligned}
\mathbf{Q}_t &= \mathbf{E}[(x_t - \mathbf{F}\hat{x}_{t-1})(x_t - \mathbf{F}\hat{x}_{t-1})^T] = \mathbf{E}[(x_t - \mathbf{F}^t x_0)(x_t - \mathbf{F}^t x_0)^T] \\
&= \mathbf{E}[(\mathbf{F}^{t-1}u_0 + \dots + \mathbf{F}u_{t-1} + u_t)(\mathbf{F}^{t-1}u_0 + \dots + \mathbf{F}u_{t-1} + u_t)^T] \\
&= \mathbf{Q} + \mathbf{FQF}^T + \dots + \mathbf{F}^{t-1}\mathbf{QF}^{t-1T} \\
&= \mathbf{Q} + \mathbf{FQ}_{t-1}\mathbf{F}^T
\end{aligned}$$

I coefficienti $\alpha_1, \dots, \alpha_p$ possono essere stimati da una realizzazione finita del processo stocastico.

In generale, uno stimatore è una variabile casuale osservabile, determinata dagli elementi dello spazio campionario, che può essere usato al posto del valore vero, incognito, del parametro relativo alla popolazione. È importante che lo stimatore sia costruito in modo tale che la funzione di distribuzione dei valori possibili che esso può assumere sia massimamente concentrata in un intorno del valore vero del parametro della popolazione. Gli *stimatori lineari* sono una classe di stimatori relativamente semplici, che possono essere ottenuti come funzioni lineari degli elementi del campione, senza peraltro formulare nessuna ipotesi sulla funzione di distribuzione della popolazione, se non il fatto di assumere finiti i momenti del primo e del secondo ordine. Ad esempio, se $\mathbf{x} = (x_1, x_2, \dots, x_n)$ è un campione estratto da una popolazione con media μ e varianza σ^2 , allora $\hat{x} = a_1x_1 + a_2x_2 + \dots + a_nx_n$ rappresenta uno stimatore di μ , qualora a_1, a_2, \dots, a_n siano costanti note. Inoltre, se $a_1 + a_2 + \dots + a_n = 1$, ne consegue che il valore atteso dello stimatore è identico al valore vero del parametro della popolazione: $\mathbf{E}[\hat{x}] = \mu$. Stimatori di questo tipo sono detti *stimatori lineari non devianti* o *corretti*. Occorre quindi stabilire un criterio di scelta per gli infiniti stimatori lineari

non devianti. Un criterio largamente condiviso è quello di scegliere lo stimatore lineare corretto con la varianza minima. Tale stimatore viene perciò detto *stimatore corretto migliore o più efficiente*. Nell'esempio precedente la varianza dello stimatore \hat{x} è data da $\sigma^2(\hat{x}) = (a_1^2 + a_2^2 + \dots + a_n^2)\sigma^2$. Si dimostra che la funzione $\sigma^2(\hat{x})$ ha un unico minimo in corrispondenza di $a_1 = a_2 = \dots = a_n = \frac{1}{n}$ e per questo motivo la media campionaria $\bar{x} = (x_1 + x_2 + \dots + x_n)/n$ rappresenta l'esempio più semplice di uno stimatore lineare non deviato ed a varianza minima. La stima dei coefficienti di regressione $\beta = (\beta_1, \beta_2, \dots, \beta_k)$ di un modello lineare può essere considerata come una generalizzazione di quanto detto per la stima della media di una popolazione [21, 37].

Il teorema di *Gauss-Markov* afferma che per un modello lineare, sotto determinate condizioni, gli stimatori lineari corretti più efficienti sono identici agli stimatori ottenuti con il metodo dei minimi quadrati.

Teorema 1.1 *Siano $\mathbf{X}' = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ una successione di n vettori di dimensione $k < n$ linearmente indipendenti e $\mathbf{y}' = (y_1, y_2, \dots, y_n)$ un vettore di n variabili casuali indipendenti, con valore atteso uguale a $\mathbf{X}\beta$ e tutte con la medesima varianza σ^2 . Posto $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ essere lo stimatore dei minimi quadrati e $\mathbf{b} = \mathbf{L}\mathbf{y}$ un qualunque altro stimatore lineare non deviato, allora $\Sigma(\mathbf{b}) - \Sigma(\hat{\beta})$ è una matrice definita positiva o semipositiva.*

Dimostrazione. Poichè lo stimatore $\mathbf{b} = \mathbf{L}\mathbf{y}$ non è deviato, per definizione $\mathbf{E}(\mathbf{b}) = \mathbf{L}\mathbf{X}\beta = \beta$, da cui risulta $\mathbf{L}\mathbf{X} = \mathbf{I}$. La matrice di varianza-covarianza di \mathbf{b} è data da $\Sigma(\mathbf{b}) = (\mathbf{L}\mathbf{y} - \beta)(\mathbf{L}\mathbf{y} - \beta)' = \mathbf{L}(\mathbf{y} - \mathbf{X}\beta)(\mathbf{y} - \mathbf{X}\beta)'\mathbf{L}' = \sigma^2\mathbf{L}\mathbf{L}'$. Posto $\mathbf{M} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, $\mathbf{M}\mathbf{X} = \mathbf{I}$. Scomponendo \mathbf{L} come $\mathbf{L} = \mathbf{M} + \mathbf{D}$, ne consegue $\mathbf{D}\mathbf{X} = 0$. Allora:

$\Sigma(\mathbf{b}) - \Sigma(\hat{\beta}) = \sigma^2((\mathbf{M} + \mathbf{D})(\mathbf{M} + \mathbf{D})' - \mathbf{M}\mathbf{M}') = \sigma^2(\mathbf{M}\mathbf{D}' + \mathbf{D}\mathbf{M}' + \mathbf{D}\mathbf{D}')$.
 Siccome $\mathbf{M}\mathbf{D}' = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{D}\mathbf{X})' = 0$, $\mathbf{D}\mathbf{M}' = \mathbf{D}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = 0$ e $\mathbf{D}\mathbf{D}'$ è una matrice definita positiva o semipositiva, la tesi è perciò dimostrata.

Per calcolare le stime dei coefficienti di autocorrelazione relativi alle serie temporali di portata, si è perciò utilizzato il metodo dei minimi quadrati. Se $(x_{t_1}, \dots, x_{t_n})$ è una realizzazione finita del processo stocastico, posto \mathbf{B} e \mathbf{C} essere rispettivamente le stime campionarie di $\mathbf{E}[\mathbf{x}_t \mathbf{x}_{t-1}^T]$ e $\mathbf{E}[\mathbf{x}_t \mathbf{x}_t^T]$, allora la stima della matrice di transizione è data da $\hat{\mathbf{F}} = \mathbf{B}\mathbf{C}^{-1}$, mentre la stima della matrice di varianza-covarianza dell'errore u_t da $\hat{\mathbf{Q}} = \mathbf{C} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T$

1.4 Il modello idrologico TOPKAPI

Il modello TOPKAPI è strutturato in base a quattro moduli principali che rappresentano rispettivamente la componente del suolo, la componente del deflusso superficiale, la componente del deflusso lungo la rete drenante e la componente di accumulo e scioglimento della neve (figura 1.1).

La componente del suolo riveste nel modello un ruolo fondamentale, sia come contributo diretto al deflusso nella rete drenante, sia come fattore che regola il bilancio idrico del suolo, in particolare per quanto riguarda l'attivazione della produzione di deflusso superficiale. Infatti in base alla saturazione raggiunta nel suolo si innesca il meccanismo del deflusso superficiale. Entrambe le componenti del suolo e del deflusso superficiale alimentano poi la componente della rete drenante.

Il modello è stato sviluppato sia in una versione distribuita, in cui le equazioni fondamentali del modello vengono risolte per ciascuna cella ad ogni passo

TOPKAPI model for each cell

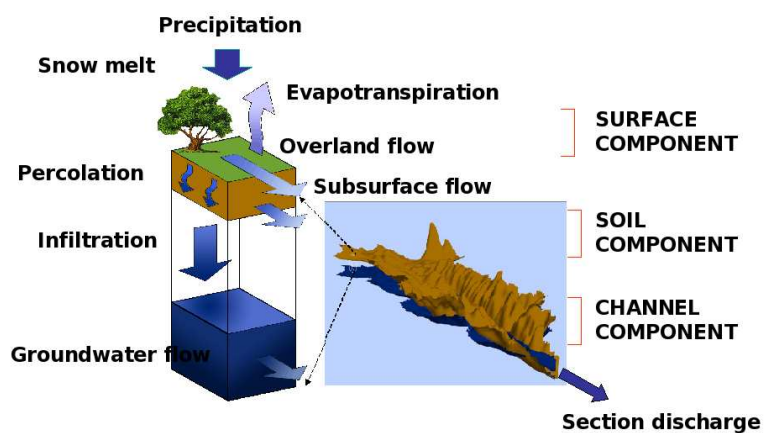


Figura 1.1: Schema del modello idrologico TOPKAPI.

di tempo, che in una versione aggregata, in cui le medesime equazioni vengono risolte per ciascun passo temporale sull'intero bacino o sottobacino. Il modello distribuito consente il controllo locale delle variabili di stato, e quindi il controllo locale della saturazione del suolo e dell'attivazione del deflusso superficiale. Si presta maggiormente alla calibrazione, ma ha dei tempi di calcolo che ne rendono difficile l'utilizzazione operativa. Il modello aggregato invece effettua il controllo della saturazione del suolo in forma globale, in base ad una curva area saturazione-volume immagazzinato che viene tarata sperimentalmente calibrando il modello distribuito; una volta che sia disponibile la calibrazione dei parametri fisici e dei parametri della curva, in fase operativa l'uso del modello aggregato comporta tempi di esecuzione del tutto compatibili con le esigenze operative in tempo reale.

1.4.1 La componente del suolo.

La componente del suolo è interessata da un flusso d'acqua in direzione orizzontale in condizioni di non saturazione, definito drenaggio, che ha luogo in uno strato superficiale, di spessore limitato, interessato da una conduttività idraulica elevata, dovuta alla macroporosità. Per la formazione dei deflussi superficiali si adotta un meccanismo di saturazione dal basso, altrimenti detto eccesso di saturazione o meccanismo di Dunne, trascurando invece l'eventuale attivazione del meccanismo hortoniano dovuto all'eccesso di infiltrazione. Infatti il meccanismo di eccesso di infiltrazione è caratteristico di una scala di modellazione locale, mentre il meccanismo di eccesso di saturazione, essendo legato ad un fenomeno di tipo cumulativo e condizionato da un movimento di redistribuzione laterale dell'acqua nel suolo, diviene dominante all'aumentare della scala di modellazione. Nel modello distribuito tale ipotesi si traduce nel fatto che in ciascuna cella la produzione di deflusso superficiale si attiva quando la cella stessa è completamente satura. Nel modello aggregato la produzione di deflusso superficiale si attiva su quella porzione di bacino che ha raggiunto la saturazione, determinata in funzione del contenuto d'acqua totale nel suolo, secondo una legge caratteristica del bacino.

Per la componente del suolo, come pure per quella del deflusso ipodermico e del deflusso in rete drenante, si adotta uno schema di propagazione ad onda cinematica. La descrizione del flusso orizzontale in condizioni insature dipende dalla conoscenza del profilo verticale del contenuto d'acqua nel suolo. Tuttavia lo spessore dello strato di suolo ad elevata conduttività idraulica, effettivamente interessato dal drenaggio, è pressoché trascurabile rispetto alle dimensioni orizzontali in gioco. Per tale ragione si è supposto che il flusso orizzontale possa essere valu-

tato assumendo la conduttività idraulica a saturazione costante su tutto lo strato ed utilizzando il contenuto d'acqua totale integrato sul profilo verticale. In accordo con tale ipotesi la trasmissività in condizioni di non saturazione viene calcolata con la seguente espressione approssimata:

$$T(\tilde{\Theta}) = k_S L \tilde{\Theta} \alpha \quad (1.4)$$

in cui k_S è la conduttività idraulica a saturazione, costante su tutto lo strato, α il parametro dipendente dalle caratteristiche del suolo, L lo spessore dello strato interessato dal flusso orizzontale, in metri, $\tilde{\Theta}$ è il valore medio lungo il profilo verticale del contenuto d'acqua ridotto:

$$\tilde{\Theta} = \frac{1}{L} \int_{L_0} \tilde{\theta}(z) dz \quad (1.5)$$

$$\tilde{\theta} = \frac{\theta - \theta_r}{\theta_s - \theta_r} \quad (1.6)$$

θ_r è il contenuto d'acqua residuo, ossia la quantità di acqua che non può essere rimossa per capillarità o per gravità, θ_s è il contenuto d'acqua a saturazione e θ il contenuto d'acqua nel suolo e z la direzione verticale. È importante sottolineare che la conduttività idraulica a saturazione è una permeabilità orizzontale, il cui valore è maggiore di circa un ordine di grandezza rispetto alla corrispondente permeabilità verticale, in genere misurata e fornita in letteratura.

L'abbinamento dell'equazione con un approccio di tipo cinematico porta ad un'equazione di serbatoio non lineare zero-dimensionale per ciascuna cella i nel modello distribuito:

$$\partial_t V_i = xp - \left(\frac{\alpha + 1}{\alpha x} \right)^\alpha \frac{1}{\left[k^{\frac{\alpha+1}{\alpha}} - (k-1)^{\frac{\alpha+1}{\alpha}} \right]^\alpha} C_i V_i^\alpha \quad (1.7)$$

e per l'intero bacino, o sottobacino, nel modello aggregato:

$$\partial_t V_{tot} = Nxp - \bar{C} \left[\frac{\alpha + 1}{\alpha Nx} \right]^\alpha V_{tot}^\alpha \quad (1.8)$$

dove

$$\bar{C} = \left[\sum_{i=1}^N \frac{\left(\frac{k}{N} \right)^{\frac{\alpha+1}{\alpha Nx}}}{\left(\frac{k-1}{N} \right)^{\frac{\alpha+1}{\alpha Nx}}} C_i^{\frac{1}{\alpha}} \right]^\alpha \quad (1.9)$$

è la conduttanza media equivalente. Nell'equazione i è l'indice della singola cella, k è la somma di tutte le celle drenate dalla cella i e rappresenta quindi la topologia dei collegamenti tra le celle, N è il numero totale di celle.

L'immagazzinamento dell'acqua nel suolo rappresenta la variabile di stato, ed in base al valore assunto è possibile calcolare ad ogni istante temporale, per ogni cella e per l'intero bacino, la portata uscente, ripartita tra suolo a valle e rete drenante.

1.4.2 La componente del deflusso superficiale.

Il raggiungimento delle condizioni di saturazione in una particolare zona determina l'attivazione del deflusso superficiale, ottenuto come l'eccesso di precipitazione rispetto alla quantità necessaria a rendere saturo lo strato di suolo. Si adotta anche in questo caso l'approssimazione dell'onda cinematica, e si utilizza come legge di resistenza la formula di Manning, nell'ipotesi di sezione rettangolare infinitamente larga:

$$Q = \frac{1}{n} i_f^{\frac{1}{2}} x y^{\frac{5}{3}} \quad (1.10)$$

in cui n è il coefficiente di Manning per i versanti, i_f pendenza della cella, x la larghezza della sezione, pari alla risoluzione della cella, ed y l'altezza del tirante idrico. Anche in questo caso si perviene ad un'equazione di serbatoio non lineare zero-dimensionale per ciascuna cella nel modello distribuito, e per l'intero bacino, o sottobacino, nel modello aggregato, formalmente analoghe a quelle per la componente del suolo.

1.4.3 La componente del deflusso in rete drenante.

Un certo numero di celle del modello digitale del terreno (DEM) sono interessate dalla presenza della rete drenante. In tali celle viene calcolato l'afflusso alla rete drenante per via ipodermica e superficiale, come frazione della portata totale uscente dalla cella che contiene un tronco di rete. Una volta che sia nota la portata che alimenta il deflusso lungo la rete drenante, questo viene descritto mediante un modello cinematico del tutto analogo a quello illustrato per la componente del deflusso superficiale.

Il TOPKAPI è un modello afflussi-deflussi, finalizzato alla rappresentazione degli idrogrammi di piena a partire dall'input meteorologico e dalle caratteristiche fisiche e morfologiche del bacino idrografico, più che alla descrizione accurata della propagazione dell'onda di piena lungo l'alveo; per tale ragione esso non va visto come un modello di tipo idraulico nel senso rigoroso del termine. Ne consegue che non è necessario descrivere in maniera accurata la geometria della sezione dell'alveo, ma è sufficiente darne una rappresentazione schematica. La re-

te drenante è perciò rappresentata mediante tronchi a sezione rettangolare, aventi la lunghezza della cella e larghezza costante lungo la cella, ma variabile da una cella all'altra. La larghezza variabile è l'unico aspetto che differenzia lo schema di aggregazione per la rete drenante, rispetto alle altre componenti. Si adotta anche in questo caso la formula di Manning, e si mantiene anche l'ipotesi di sezione rettangolare infinitamente larga:

$$Q = \frac{1}{n} i_f^{\frac{1}{2}} D h^{\frac{5}{3}} \quad (1.11)$$

in cui n è il coefficiente di Manning per i canali, i_f pendenza del fondo, assunta uguale alla pendenza della cella, D la larghezza della sezione ed h l'altezza del tirante idrico. Di nuovo si perviene ad equazioni di serbatoio non lineare zero-dimensionale analoghe alle precedenti.

1.4.4 La componente dell'accumulo e scioglimento della neve.

Per ogni singolo sottobacino viene usato un algoritmo che si basa sul bilancio energetico in termini di calorie a livello del manto nevoso e sul bilancio di massa del manto nevoso. Le equazioni che esprimono il bilancio di energia sono le seguenti:

per $T \leq T_{sn}$

$$E_f = E_i + C_{er} C_{sn} E T_0 + C_{sg} T_0 P \quad (1.12)$$

per $T > T_{sn}$

$$E_f = E_i + C_{er} C_{sn} E T_0 + [C_{sg} T_0 + C_{lf} + C_{sa} \Delta T] P \quad (1.13)$$

dove E_f è l'energia finale (Kcal), E_i l'energia iniziale (Kcal), P la precipitazione caduta nell'intervallo Δt (mm/Dt), C_{er} fattore di conversione del calore latente in radiazione, ovvero il calore latente di evaporazione ($\sim 606.5 - 0.695\Delta T$ Kcal Kg⁻¹), ΔT la variazione rispetto T_0 della temperatura media ragguagliata (°K), T_0 il punto di fusione della neve (273 °K), C_{sn} un fattore di riproduzione per tener conto dell'albedo, C_{sg} il calore specifico del ghiaccio (0.5 Kcal°K⁻¹Kg⁻¹), C_{lf} il calore latente di fusione dell'acqua (79.6 KcalKg⁻¹), C_{sa} il calore specifico dell'acqua (1 Kcal°K⁻¹Kg⁻¹), ΔT_{sn} la variazione rispetto a T_0 della temperatura di soglia della neve (°K) e ET_0 l'evapotraspirato.

L'equazione che esprime il bilancio di massa è la seguente:

$$E_f = C_{sg}W_{0sn}T_0 + C_{lf}R_{sn} \quad (1.14)$$

da cui si ricava lo scioglimento del manto nevoso nell'intervallo Δt :

$$R_{sn} = \frac{E_f - C_{sg}W_{0sn}T_0}{C_{lf}} \quad (1.15)$$

dove W_{0sn} è lo spessore del manto nevoso finale in millimetri.

1.4.5 La curva aree sature–volumi immagazzinati.

Nel modello aggregato la parte di bacino in grado di produrre deflusso superficiale viene determinata ad ogni istante temporale in base ad una opportuna curva, caratteristica del bacino. Tale curva viene determinata sperimentalmente, una volta calibrato il modello distribuito, rappresentando in un grafico in ascissa il numero delle celle sature ad ogni passo di tempo, ordinate in senso crescente, ed in ordinata i volumi immagazzinati in corrispondenza di ciascun valore di area

satura. Il grafico che si ottiene è fatto di punti sparsi che vengono interpolati con una funzione Beta.

1.4.6 L'uso del modello digitale del terreno.

Il modello TOPKAPI si basa sull'utilizzo di un modello digitale del terreno per descrivere la distribuzione spaziale delle caratteristiche fisiche e topografiche del bacino di applicazione. La funzione del DEM assegna ad ogni cella un valore per quanto riguarda le caratteristiche dei suoli, individua i percorsi di drenaggio, ossia i percorsi seguiti dall'acqua, per via ipodermica, superficiale e canalizzata, per raggiungere la sezione di chiusura del bacino, assegna ad ogni cella un valore per la pendenza, individua in maniera automatica la posizione e la dimensione del reticolo idrografico.

In relazione al primo punto, l'individuazione di zone omogenee per tipo ed uso del suolo permette di assegnare ad ognuna delle celle appartenenti a tale zona un medesimo valore per ciascuno dei parametri utilizzati dal modello. Il secondo punto consiste nell'individuare i collegamenti tra le celle secondo un criterio, detto relazione di vicinanza (*neighbourhood function*). Il terzo ed il quarto punto dipendono direttamente dal secondo, in quanto sfruttano i collegamenti individuati tra le celle. Inoltre è necessario definire un criterio appropriato, sia per assegnare un valore di pendenza, unico per ogni cella, sia per stabilire quali sono le caratteristiche che una cella deve avere per ospitare al suo interno la rete drenante. A tale proposito si deve tenere conto, tra l'altro, delle ipotesi alla base del meccanismo di aggregazione topologica delle celle: ciascuna cella può ricevere un contributo di deflusso da monte da un massimo di tre celle, mentre ciascuna cella fornisce un

contributo di deflusso verso valle ad una sola cella. Poiché lo scambio tra una cella e l'altra è distribuito su di una larghezza finita, il collegamento tra le celle può avvenire solo attraverso i lati delle celle medesime, e non attraverso gli angoli.

Ciò implica che nessuna cella può soltanto ricevere contributi dalle celle circostanti, senza che essa stessa riversi il proprio contributo ad una di esse; tale situazione si presenta quando tutte le celle circostanti si trovano a quota superiore alla cella esaminata, ossia quando si è in presenza di fosse o false uscite. Allo scopo di correggere situazioni di questo tipo, nonché allo scopo di assolvere le funzioni indicate nei primi tre dei quattro punti precedentemente elencati, sono stati sviluppati opportuni programmi di utility. L'individuazione automatica del reticolo idrografico e la dimensione dei tronchi di canale viene effettuata all'interno del codice di calcolo del modello TOPKAPI. Il metodo seguito si basa sull'individuazione dell'area drenata sottesa da ogni cella del DEM e dalla definizione di un valore di soglia per l'area drenata da una generica cella, in base a cui stabilire se quella determinata cella contiene al suo interno un tronco di rete drenante. L'area di soglia appropriata per l'estrazione della rete drenante viene determinata per tentativi, confrontando visivamente, tramite il GIS, il reticolo di volta in volta ottenuto con quello riportato nella cartografia tradizionale relativa alla medesima area geografica di interesse. Una volta stabilito che una determinata cella contiene un tronco di rete drenante, esso è genericamente considerato all'interno della cella, senza che sia possibile, né necessario, individuarne con esattezza la posizione. La rete drenante è rappresentata mediante tronchi di canale a sezione rettangolare; la sezione ha larghezza costante per ciascun tronco ma in generale può variare da un tronco all'altro. La larghezza del canale è anch'essa determinata in base all'area drenata, e precisamente essa aumenta in maniera proporzionale alla radice

quadrata dell'area drenata:

$$w = w_{max} + \left[\frac{w_{max} - w_{min}}{\sqrt{A_{tot}} - \sqrt{A_{th}}} \right] \left(\sqrt{A_{dr}} - \sqrt{A_{tot}} \right) \quad (1.16)$$

dove w larghezza del canale in una cella generica, w_{max} larghezza del canale nella cella di chiusura del bacino, w_{min} larghezza del canale in corrispondenza dell'area di soglia, A_{th} area di soglia, A_{tot} area totale ed A_{dr} area drenata dalla generica cella.

1.4.7 I parametri del modello.

I parametri di un modello sono tutte quelle grandezze che servono a descrivere le caratteristiche della realtà fisica a cui il modello viene applicato. Nel modello TOPKAPI i parametri possono essere suddivisi in tre gruppi: parametri che possono essere stabiliti a priori con una certa sicurezza, e quindi non necessitano di nessuna procedura di calibrazione, parametri che dovrebbero essere stabiliti a priori, e tuttavia sono soggetti ad un certo grado di incertezza e quindi possono essere soggetti a lievi aggiustamenti, ma non ad una vera e propria procedura di calibrazione, e parametri soggetti ad un grado più elevato di incertezza, tale da necessitare di una procedura di calibrazione più approfondita.

I parametri appartenenti al primo gruppo sono quelli connessi alle informazioni fornite dal DEM: pendenze e area di soglia per la definizione della rete drenante.

Le pendenze e le frazioni di divisione del deflusso vanno semplicemente calcolati una volta effettuato il trattamento del DEM; l'area di soglia viene definita per tentativi, ma ciò non va inteso come una forma di calibrazione, sia perché la solu-

zione di tale problema si raggiunge dopo pochissimi tentativi, sia perché comunque il valore appropriato viene definito in base al confronto con le informazioni di una mappa, e non in base alle portate simulate dal modello.

I parametri appartenenti al secondo gruppo sono costituiti essenzialmente dalle larghezze minime e massime dei canali; la larghezza massima corrisponde alla sezione di chiusura del bacino, ossia ad una sezione fluviale di qualche interesse, generalmente attrezzata con uno strumento di misura e per la quale si conosce una scala di deflusso. Tale dato è solitamente noto con una certa affidabilità. Tuttavia, poiché si adotta una legge di crescita della larghezza dei canali in base all'area drenata, e poiché tale legge dipende dalla larghezza massima e minima, è possibile che l'adozione del valore esatto della larghezza massima conduca ad una valutazione errata, in genere per eccesso, della larghezza dei tratti a monte; per tale ragione è possibile che sia necessario ridurre in una certa misura la larghezza massima. La larghezza minima è soggetta ad un margine di incertezza sicuramente maggiore e va definita in base alle conoscenze disponibili e a considerazioni di ordine generale; per tale ragione essa può subire alcuni aggiustamenti dettati dalla forma e dal ritardo dell'idrogramma di piena, ma tali aggiustamenti variano in un intervallo modesto.

I parametri appartenenti al terzo gruppo sono quelli che vanno effettivamente calibrati. A tale proposito va ribadito che in teoria, essendo il modello fisicamente basato, la calibrazione non dovrebbe essere necessaria, ma lo diventa a causa del grado di incertezza delle informazioni di cui si dispone sulle proprietà dei suoli. Per ogni cella, i parametri sono la permeabilità a saturazione, il contenuto d'acqua a saturazione, il contenuto d'acqua residuo, lo spessore dello strato di suolo interessato dal drenaggio, il coefficiente di scabrezza di Manning per il deflusso

superficiale, il coefficiente di scabrezza di Manning per il deflusso nella rete drenante. L'esponente α della legge di trasmissività è invece assunto uguale per tutte le celle.

I primi tre parametri esprimono le caratteristiche del suolo e dipendono quindi dal tipo di suolo. Per la permeabilità a saturazione il parametro considerato è una permeabilità orizzontale, generalmente di un ordine di grandezza maggiore rispetto a quella verticale.

L'esponente α può andare da un valore minimo di 2.5 fino ad un massimo di 9–10. I valori più alti sono caratteristici di suoli ad alta componente argillosa; tuttavia poiché questo parametro è l'unico che viene assegnato uguale per tutte le celle, in quanto ciò è richiesto dalle ipotesi necessarie per il meccanismo di aggregazione topologica delle celle, è consigliabile attestarsi su valori aventi una validità più generale, in genere tra 2.5 e 3.5, rappresentativi di un comportamento medio. Lo spessore del suolo è anch'esso un parametro importante per il drenaggio, poiché condiziona la capacità di immagazzinamento. Qualora non si disponga di misure affidabili di tale parametro, esso può venire valutato sulla base del tipo di terreno. Non ci sono a priori valori limite di tale parametro, tuttavia si tenga conto che si tratta di uno strato di suolo superficiale, il cui spessore può andare da alcune decine di centimetri a qualche metro; si può dire, senza assegnare a tale affermazione un valore troppo rigido, che 10 metri può essere visto come un limite superiore per tale parametro; tuttavia l'assegnazione di valori così alti va effettuata laddove ci siano conoscenze abbastanza affidabili, mentre se si effettua una calibrazione per tentativi è meglio attenersi a valori intermedi. I coefficienti di Manning per il deflusso superficiale vengono valutati in base alla copertura della vegetazione e all'uso del suolo, mentre quelli per il deflusso cana-

lizzato vengono in genere differenziati in base alle quote, attribuendo valori più elevati in zone montane. I valori del coefficiente di Manning per la rete drenante variano con continuità in un intervallo piuttosto ampio, in dipendenza dal tipo di sezione, di sedimenti che costituiscono il letto del fiume, di materiali naturali o riportati che costituiscono gli argini e infine dall'entità e dalla tipologia della vegetazione presente sulle scarpate.

Per quanto riguarda i coefficienti di Manning per i versanti si deve tener presente che l'applicazione della legge di Manning ai versanti va vista come un'estensione di tale legge, che a rigore si applica agli alvei, quindi anche i parametri di scabrezza vanno considerati in questa ottica. Certamente i versanti presenteranno una scabrezza molto più grande dei canali, poiché gli ostacoli al deflusso possono essere numerosi e di varia natura (vegetazione, manufatti, ecc.). Di conseguenza il coefficiente di Manning per il deflusso superficiale può andare oltre i valori di 0.08–0.09, fino ad arrivare ad un valore di 0.12, secondo il testo di Ven Te Chow (1959), per le scarpate.

1.4.8 I dati idro–meteorologici in ingresso.

I dati in ingresso al modello sono la precipitazione e la temperatura, o l'evapotraspirazione potenziale. I dati di precipitazione vengono forniti in continuo, con il passo temporale con cui sono disponibili, in corrispondenza dei pluviometri appartenenti alla rete di misura che copre l'area in esame. I dati puntuali vengono distribuiti sulle celle del DEM mediante le rispettive aree di influenza. È necessario costruire preventivamente tali poligoni per ciascun pluviometro ed individuare a quale poligono appartiene ciascuna cella del DEM. Ciò può essere

ottenuto agevolmente tramite un GIS, sovrapponendo al DEM una rappresentazione in formato vettoriale dei poligoni e facendo quindi un'operazione di mappatura dei poligoni sul DEM.

Anche i dati di temperatura vengono forniti al modello con il medesimo passo temporale, e vengono distribuiti sulle celle con la medesima tecnica, avendo però cura di correggere, per ogni cella, il dato di temperatura fornito dal termometro in base alla differenza di quota tra la cella stessa e il termometro, tenendo conto di un gradiente termico al variare dell'altitudine di $0,06^{\circ}\text{C}$ ogni metro. Il dato di temperatura così ottenuto viene utilizzato per il calcolo dell'evapotraspirazione potenziale. In alternativa è possibile fornire direttamente al modello dati di evapotraspirazione potenziale.

La risoluzione temporale dei dati di ingresso risulta generalmente inadeguata per l'integrazione delle equazioni di serbatoio non lineare sulle quali poggia il modello. Il codice di calcolo utilizza una procedura che raffinisce automaticamente il passo di integrazione sulla base di una soglia di precisione richiesta; tale raffinemento avviene per quelle celle e per quegli intervalli di tempo in cui le caratteristiche della cella, ad esempio la pendenza, o l'intensità della precipitazione inducono cambiamenti molto rapidi delle variabili di stato.

1.5 L'accoppiamento dei modelli idro-meteorologici

1.5.1 Modellistica dei processi atmosferici alla mesoscala

L'atmosfera terrestre è un fluido, il cui comportamento è regolato da leggi termodinamiche. La determinazione dello stato dell'atmosfera ha conosciuto un

forte miglioramento negli ultimi decenni, grazie all'iperbolico aumento delle osservazioni disponibili ed al notevole sviluppo dei modelli numerici, che, sulla base delle condizioni iniziali fornite dalle osservazioni, sono in grado di rappresentare, sempre con maggior dettaglio, i moti alle diverse scale spazio-temporali che concorrono nella circolazione atmosferica.

I sistemi a grande scala, o scala planetaria, hanno dimensioni spaziali confrontabili col raggio terrestre (~ 6000 km) ed una vita media pari a circa una settimana; tra essi sono comprese le onde di Rossby. I sistemi a scala sinottica sono, invece, caratterizzati da lunghezze d'onda di circa 1000–4000 km e da una scala temporale dell'ordine di qualche giorno; sono le strutture che si individuano su una carta sinottica, come i cicloni e gli anticicloni, e che determinano il tempo alle medie latitudini. I sistemi che non sono abbastanza grandi per essere visti in modo chiaro su una carta sinottica, pur se individualmente possono influenzare, anche in modo intenso, la situazione ad una o due stazioni vicine, sono chiamati sistemi a scala sub-sinottica. La scala sub-sinottica è molto ampia, e generalmente viene suddivisa in mesoscala, piccola scala e microscala. La mesoscala comprende strutture, come i fronti e la ciclogenesi orografica, con dimensioni comprese tra le decine e le centinaia di chilometri e vita media dell'ordine di qualche ora e fino a 1–2 giorni. I fenomeni convettivi, come le nubi a grande sviluppo verticale (cumulonembi), rientrano nella piccola scala (o scala locale) ed hanno dimensioni spaziali attorno al chilometro e temporali pari a circa un'ora. Infine ci sono i fenomeni di microscala, moti turbolenti non organizzati, caratterizzati da lunghezze d'onda inferiori al metro e che vivono per qualche minuto o qualche secondo.

Per potere descrivere nella maniera più accurata possibile i fenomeni atmosferici di interesse idrologico come le precipitazioni, i temporali associati a super-

celle, i sistemi convettivi alla mesoscala, le linee temporalesche prefrontali, il Consortium for small scale modelling (COSMO) ha creato il modello numerico ad area limitata Lokal Modell. Lo sviluppo di Lokal Modell è cominciato nel 1994 come progetto interno del Deutscher WetterDienst (DWD), il servizio meteorologico nazionale tedesco, il quale ritenne necessario sviluppare un nuovo modello numerico per risolvere le scale spaziali dove gli effetti non-idrostatici giocano un ruolo essenziale: meso- γ (5–50 km), meso- β (from 500 m to 5 km) e meso- α (50–500 m). Con un dominio abbastanza ampio ed una risoluzione spaziale della griglia orizzontale del modello tra 1 e 3 km, è possibile risolvere esplicitamente la convezione umida profonda ed i meccanismi di *feedback* sulle scale maggiori associati. In linea di principio Lokal Modell potrebbe avere la potenzialità di superare le limitazioni legate alla parametrizzazione della convezione nei modelli idrostatici con una risoluzione più bassa della griglia.

In questa tesi viene effettuato un accoppiamento tra il modello meteorologico, non-idrostatico e ad area limitata Lokal Modell, nella fattispecie la versione COSMO-LAMI, utilizzata operativamente presso il Servizio Idrometeorologico di ARPA con una risoluzione orizzontale di 7 km e con 35 livelli verticali (figura 1.2), ed il modello idrologico ad afflussi-deflussi TOPKAPI, allo scopo di prevedere nel breve termine le piene di portata del fiume Reno. La versatilità del modello TOPKAPI permette infatti di inserirlo in un sistema di preannuncio delle piene in tempo reale, nel quale i dati misurati di precipitazione e di temperatura vengono sostituiti con quelli previsti dal modello meteorologico.

1.5.2 Le previsioni di “ensemble”

Le previsioni idrologiche di ensemble ottenute con il modello afflussi–deflussi TOPKAPI si basano su di un sistema di previsioni meteorologiche di ensemble del modello non idrostatico ad area limitata COSMO–LAMI, integrato utilizzando come condizioni iniziali e condizioni al contorno le previsioni del modello di circolazione globale IFS, della corsa giornaliera delle 12 UTC, operativa presso ECMWF. La risoluzione orizzontale delle previsioni meteorologiche è pari a circa 10 km, mentre la discretizzazione verticale è data da una suddivisione in 32 strati. L’orizzonte previsionale è di 120 ore. Il dominio di integrazione è maggiore di quello di COSMO–LAMI e comprende l’intera regione europea (fi gura 1.2).

La metodologia Limited–Area Ensemble Prediction System (LEPS), descritta nei suoi principi essenziali in [17, 18, 19, 20], è costruita con l’intento di combinare i vantaggi del Ensemble Prediction System dell’ECMWF (EPS) con la capacità tipica dei modelli ad area limitata di rappresentare in dettaglio i fenomeni atmosferici su scale locali, in particolare per quelle regioni dove gli effetti legati ad una orografia a complessa rivestono un’influenza rilevante [29].

Per gli elevati costi computazionali non tutti gli elementi dell’EPS vengono utilizzati per produrre gli elementi di COSMO–LEPS. È perciò necessaria una strategia di riduzione del numero degli elementi, essenzialmente basata su un algoritmo di cluster analysis di tipo gerarchico. Viene stabilito a priori il numero dei centroidi che possano essere rappresentativi della dispersione dei 51 elementi dell’ensemble dell’ECMWF. In seguito, per ciascun raggruppamento l’elemento più prossimo al centroide secondo una data norma viene scelto come elemento rappresentativo ed utilizzato come condizioni iniziali ed al contorno per l’inte-

grazione del modello ad area limitata COSMO–LAMI. Fino al 2003 gli elementi dell’EPS erano suddivisi in 5 gruppi, mentre dal 2004 i gruppi sono diventati 10. Inoltre, la precipitazione è diventata una variabile prognostica e, per aumentare la variabilità dell’ensemble di previsioni ad area limitata, la convezione umida viene calcolata differentemente, scegliendo a caso per ogni elemento rappresentativo tra il metodo Tiedke e quello Keine–Fritsche.

Nello studio presente tutte le previsioni sono state considerate egualmente probabili. Così come per le previsioni deterministiche, anche per ciascuno degli elementi rappresentativi non è stata applicata alcuna procedura di “downscaling” per introdurre i valori di pioggia previsti nel modello idrologico.

1.6 Il bacino del fiume Reno

Il bacino imbrifero del fiume Reno, il cui corso si sviluppa complessivamente lungo 210 km, è il più esteso della Regione Emilia–Romagna, con una superficie di 4930 km². In Emilia–Romagna interessa per la maggior parte la Provincia di Bologna con una superficie di 3377 km² (68,5%), quella di Ravenna con 871 km² (11.7%), quella di Modena con 62 km² (1.3%) e per una piccolissima parte quella di Ferrara con 47 km² (0.9%) per una superficie complessiva di 4357 km² corrispondente all’88.4% dell’intero bacino. La parte ricadente nella Regione Toscana coinvolge la Provincia di Firenze con 378 km² (7.7%), di Pistoia con 155 km² (3.1%) e di Prato con 40 km², per un totale di 573 km², corrispondenti all’11.6% dell’intero bacino interregionale. La parte montana copre 1051 km² fino alla sezione della Chiusa di Casalecchio di Reno, dove il corso del fiume raggiunge una lunghezza di 84 km a partire dall’origine nel Comune di San Marcello

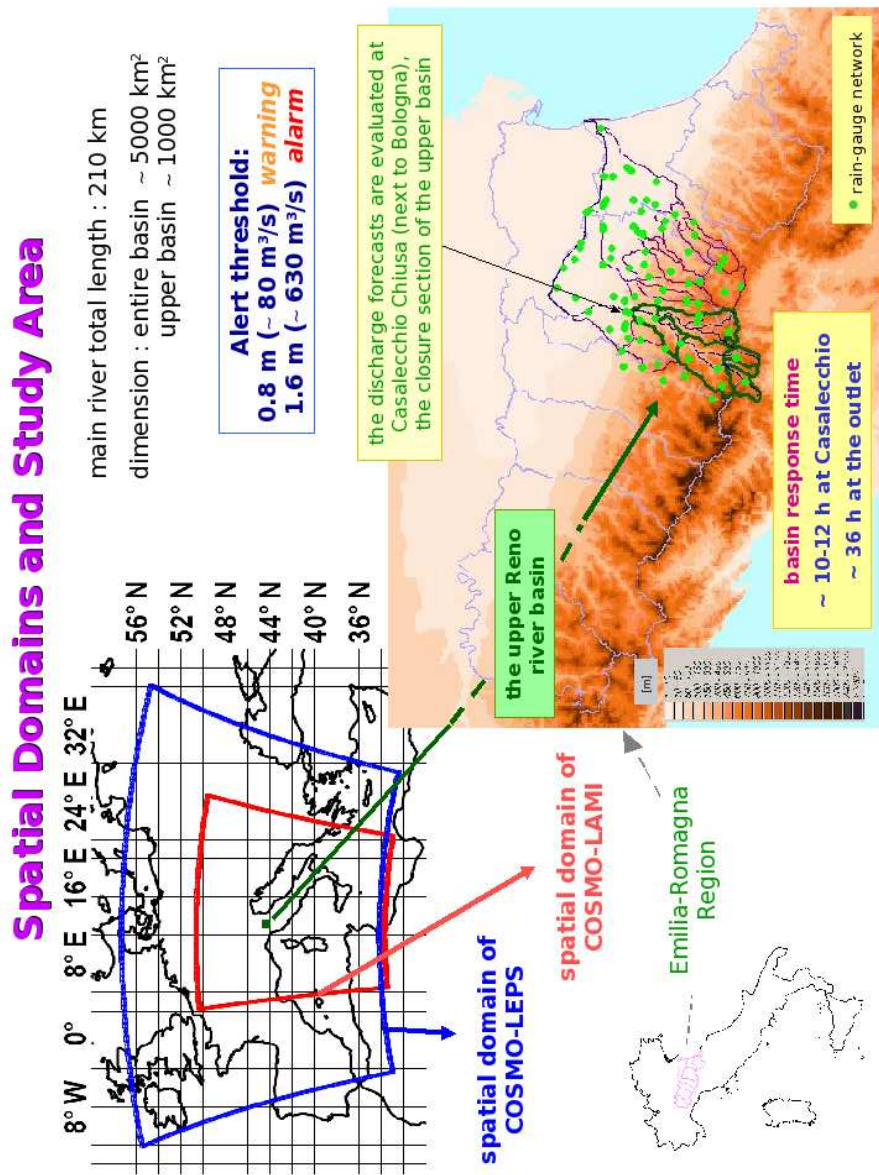


Figura 1.2: Dominio spaziale di COSMO-LAMI e di COSMO-LEPS, rete pluviometrica del bacino del fiume Reno.

Pistoiese. Dopo un tratto pedecollinare, lungo circa 6 km, il tratto vallivo conduce le acque, racchiuse da alte arginature, nello sfocio naturale del mare Adriatico sviluppandosi per altri 120 km nel territorio di pianura.

Nel suo tratto montano il fiume Reno riceve una serie di affluenti, per i maggiori di questi si può individuare una zona di pertinenza, detta sottobacino, a sua volta solcata da torrenti e rii secondari. I principali affluenti da sinistra sono, da monte verso valle, il Maresca e l'Orsigna in territorio toscano, il Randaragna, il Rio Maggiore, il Silla, il Marano, il Vergatello, il Croara, il Venola. Gli affluenti di destra, sempre da monte verso valle, sono il Limentra di Sambuca, il Limentra di Treppio, il Camperolo, il Setta con i subaffluenti Brasimone in sinistra ed il Gambellato, il Voglio ed il Sambro in destra idraulica. Qui sono presenti quattro invasi artificiali ad uso idroelettrico e potabile, i serbatoi di Molino del Palone sul Reno (474 m s.l.m.), di Pavana sul Limentra di Sambuca (470 m s.l.m.), di Suviana sul Limentra di Treppio (470 m s.l.m.) e di Brasimone sul Brasimone (842 m s.l.m.). I primi tre sono connessi in cascata, quello di Suviana ha la maggiore capacità di invaso (34 milioni di metri cubi), raccoglie le acque di Molino del Pallone e Pavana, collegati tra loro in cascata, ed è connesso, tramite condotte forzate, al bacino del Brasimone.

Il Reno nasce alla confluenza di due rami, Reno di Prunetta e Reno di Campolungo, e per i primi 10 km si presenta, di fatto, come un piccolo torrente montano in perfetto equilibrio per l'assoluta stabilità dei terreni attraversati e per la relativamente modesta pendenza che lo caratterizza (media del 3.7%). In questo tratto la valle è alquanto ampia e presenta la configurazione di un altopiano.

Segue un tratto di 15 km, con pendenza media dell'1.8%, lungo il quale la vallata si restringe progressivamente presentandosi profondamente incassata negli

ultimi 3 km prima del ponte della Venturina. In questa parte la vallata del Reno si sviluppa essenzialmente in sinistra e si presenta in soddisfacenti condizioni di stabilità generale di regime idraulico.

Nel tronco successivo a monte di Vergato, per circa 27 km, la pendenza media scende allo 0.8%; l'alveo attraversa dapprima formazioni costituite da un'argilla eminentemente galestrina, poi da argilla quasi plastica, per passare alle marne arenacee e rientrare infine nelle argille scagliose di tipo plastico, caratterizzate da pessime condizioni di stabilità.

Nei successivi 17 km fino alla confluenza del torrente Setta, con pendenza media dello 0.4%, la valle abbandona le argille scagliose per entrare nella formazione miocenica, ove dominano le marne sabbiose e, solo localmente, marne argillose e dure; l'alveo è incassato piuttosto profondamente entro questa formazione passando poi, oltre Pian di Venola, alla zona di deposito nella quale esso si presenta quasi ovunque inciso nelle alluvioni, antiche e recenti.

Negli ultimi 10 km con pendenza media dello 0.3% fino alla Chiusa di Casalecchio, limite del comprensorio montano, il fiume attraversa golene alluvionali recenti e depositi terrazzati antichi, più ampi in sinistra che in destra idraulica. Le formazioni geologiche presenti nelle pendici collinari sono arenarie del pliocene verso monte e marne dure e mioceniche verso valle. Questo tratto presenta marcati segni dell'intervento antropico che, a seguito delle consistenti estrazioni di materiale litoide nell'alveo e nel perialveo, ha dato luogo, tra l'altro, alla formazione di numerosi invasi alimentati dalle acque di subalveo.

Il tratto vallivo del fiume Reno, cioè a valle della Chiusa di Casalecchio, si suddivide in due aste, una lunga 5.5 km da Casalecchio al ponte della Via Emilia e l'altra, che si sviluppa per 124 km, dal ponte della linea ferroviaria Milano-

Bologna allo sfocio in mare. La prima rappresenta il tratto pedecollinare del corso del Reno e come tale riveste particolare importanza idraulica dovendo assolvere alla delicata funzione di raccordo fra il regime torrentizio del bacino montano e regime fluviale del corso arginato di valle. La sezione trasversale del fiume Reno ha una larghezza dell'ordine di 150–180 m. A definire l'importanza del tratto considerato contribuisce anche il fatto che esso si sviluppa praticamente all'interno della zona urbana di Bologna.

La seconda asta si snoda attraverso i territori di pianura delle province di Bologna, Ferrara e Ravenna; vi si immettono il torrente Samoggia, il Canale Navile, i torrenti Savena Abbandonato, Idice, Sillaro, Santerno e Senio. Le caratteristiche morfologiche dell'asta in argomento sono estremamente variabili in quanto risentono delle diverse vicende idrauliche che hanno, nel tempo, determinato l'attuale assetto del fiume Reno. È noto infatti che, alle origini, il bacino naturale del Reno si chiudeva alla confluenza con il torrente Samoggia divenendo poi, a valle, affluente di destra del fiume Po. A seguito di grandi lavori di riassetto idraulico, tesi al recupero ed alla bonifica dei territori vallivi della bassa bolognese, ferrarese e ravennate, il Reno venne inalveato attraverso il Cavo Benedettino ed il tratto terminale del Po di Primaro giungendo, con successive opere di sistemazione, ad assumere l'attuale configurazione, suddivisibile in quattro tratti. Il primo tratto fino Ponte Bagno (circa 19 km) è caratterizzato da un andamento tortuoso con ampie estensioni golenali, aventi una funzione modulatrice delle portate di piena, alternate a localizzate strettoie arginali. Il secondo tratto sino allo scolmatore del Cavo Napoleonico (circa 18 km) ha un andamento abbastanza regolare, con la geometria del cavo fluviale significativamente ristretta. Il terzo tratto sino alla Bastia (circa 47 km), con alveo canalizzato e argini ravvicinati e molto alti rispetto al

piano di campagna, racchiude all'interno uno sfioratore libero in corrispondenza di Gallo di Poggio Renatico, in corrispondenza del quale negli anni '50 si verificarono le rotte dell'argine sinistro. Questo sfioratore garantisce la decapitazione naturale delle piene con recapito delle acque di esubero nell'antistante canale di bonifica Cembalina. Infine, il quarto tratto corre sino al mare (circa 40 km), ed è caratterizzato da un alveo arginato relativamente ampio.

Capitolo 2

Il modello bayesiano di previsione

2.1 La teoria bayesiana della stima

La conoscenza di un determinato fenomeno è sempre incompleta, cosicchè diventa necessario descrivere l'incertezza del suo accadimento in termini probabilistici. Si possono tentare due diversi approcci, basati su altrettante definizioni di probabilità. L'approccio frequentista, per il quale l'accadimento di un determinato fenomeno è visto come la successione temporale di valori che può assumere una variabile casuale, e l'approccio bayesiano. Se nel primo caso la probabilità viene determinata come il valore limite a cui tende la frequenza relativa nel lungo periodo, nell'approccio bayesiano la probabilità è una misura reale della possibilità di accadimento di un determinato fenomeno, date le conoscenze di cui disponiamo. La distribuzione di probabilità diventa, cioè, una codifica della nostra incertezza attorno all'accadimento del fenomeno.

In termini più formali, si può affermare che se H_i è la proposizione che asserisce la verità di una determinata ipotesi, I la nostra conoscenza a priori, D la

proposizione rappresentativa dei dati, allora $p(D|H_i, I)$, detta funzione di verosimiglianza, è la probabilità di ottenere D se H_i ed I sono vere. Se la funzione $p(H_i|I)$ è la probabilità a priori, la probabilità a posteriori, $p(H_i|D, I)$, è data dal teorema di Bayes:

$$p(H_i|D, I) = \frac{p(H_i|I)p(D|h_i, I)}{\sum_i p(D|H_i, I)p(H_i|I)} \quad (2.1)$$

Il teorema di Bayes fornisce un modello di inferenza induttiva. Può essere considerato cioè un vero e proprio processo di apprendimento e la funzione di densità di probabilità a posteriori diviene una misura del nostro stato di conoscenza dell'accadimento di un determinato fenomeno e non una sua misura. Tra i principali vantaggi dell'approccio bayesiano si possono annoverare la considerazione della conoscenza a priori e l'eliminazione di eventuali componenti di errore mediante l'operazione di marginalizzazione [10].

Nel *Bayesian Forecasting System* l'incertezza totale è suddivisa in due categorie principali. La prima è associata a valori in ingresso al modello, mentre la seconda, aggettivata come idrologica, comprende tutti gli errori relativi alle stime dei parametri del modello, agli errori di misura, ed all'approssimazione con cui vengono descritti i processi idrologici [14]. Come evidenziato nella figura 2.1, si potrebbe pensare il *Bayesian Forecasting System* come due processori statistici, integrati dal modello idrologico, il primo dei quali trasforma le incertezze delle variabili in ingresso nell'incertezza dello stato in uscita, senza tenere conto dell'incertezza idrologica, mentre il secondo processore agisce in maniera assolutamente complementare.

Il *Bayesian Forecasting System* dà un fondamento metodologico per una pre-

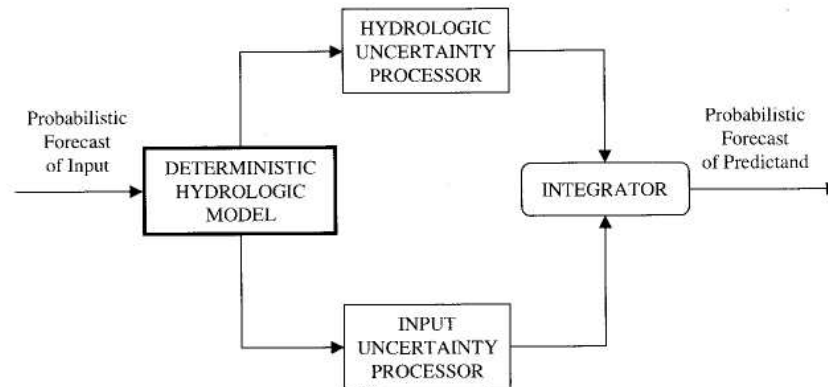


Figure 1. Structure of the Bayesian forecasting system.

Figura 2.1: Schema del Bayesian Forecasting System, tratta da [14].

visione probabilistica mediante l'utilizzo di un modello idrologico deterministico.

Questo sistema ha cinque proprietà generali:

- (i) decompone l'incertezza totale in due componenti, quantificate indipendentemente; la componente dovuta all'incertezza dei valori in ingresso e quella determinata dalla approssimazione del modello;
- (ii) la previsione probabilistica è nella forma di densità predittiva, che quantifica l'incertezza totale;
- (iii) la densità predittiva è una revisione della densità a priori; la densità a priori è un modello stocastico della previsione di piena e quantifica la sua incertezza se non si disponesse del modello idrologico; questa incertezza è una parte della variabilità del livello di portata;
- (iv) la previsione probabilistica è ben calibrata se la densità dell'errore in in-

gresso lo è; in altri termini in una simulazione di lungo periodo le proprietà statistiche delle simulazioni della portata di piena ricalcano le proprietà dei valori osservati e specificate nella densità a priori;

- (v) il sistema garantisce la proprietà di coerenza che è essenziale per chi deve prendere delle decisioni: se la densità in ingresso è non informativa, oppure il modello non ha nessuna capacità predittiva, allora la densità predittiva converge automaticamente alla densità a priori.

L'approccio bayesiano è quello che appare più naturale per combinare modelli fisici e statistici, e questo lavoro è rivolto principalmente allo studio di come minimizzare l'incertezza delle stime di piena.

In tutti i problemi di inferenza bayesiana si procede stabilendo quello che è il nostro stato di conoscenza a priori, ovvero la funzione di densità di probabilità a priori $p(H_i|I)$. In questo caso, la serie temporale delle osservazioni di portata può essere pensata come la parte finita di un'unica realizzazione di un processo stocastico $\{x_t\}$, dove t rappresenta l'indice della successione temporale, la cui memoria è determinata dal grado di correlazione tra le variabili casuali che lo compongono. Per realizzare una stima a priori dei livelli di portata è stato considerato un modello lineare, autoregressivo, con il quale determinare la funzione di distribuzione condizionata $g(x_t|x_{t-1})$ che dà la probabilità di avere x_t dato x_{t-1} .

Le veci delle informazioni di carattere sperimentale, che permettono di definire la funzione di verosimiglianza, sono fatte invece dalle stime di portata effettuate con il modello distribuito ad afflussi–deflussi TOPKAPI. Il modello è strutturato in tre moduli che rappresentano la componente di deflusso del suolo, di deflusso superficiale e di deflusso lungo la rete drenante. Per ciascun modulo si adot-

ta uno schema di propagazione ad onda cinematica. La componente di deflusso del suolo riveste il ruolo principale, in quanto è il bilancio idrico del suolo che regola la dinamica della produzione di deflusso superficiale e, quindi, della rete drenante. Poichè si tratta di un modello distribuito, basato su principi fisici, può essere facilmente applicato alla scala di versante, di bacino, oppure essere inserito all'interno di un modello di circolazione generale. La simulazione di portata è ottenuta alimentando il modello con i dati di precipitazione e di temperatura: $y_t = \mathcal{T}(x_0, p_0, p_1, \dots, p_t, T_0, T_1, \dots, T_t)$. Questi dati sono forniti in ingresso al modello con il passo temporale secondo cui sono resi disponibili dalla rete di misura e vengono distribuiti sulle celle mediante la determinazione delle aree di influenza.

In seguito, si arriva a determinare la funzione di densità di probabilità a posteriori. Posto che la stima della portata sia perfetta in senso statistico, e, condizionatamente all'osservazione che $X_t = x_t$, l'incertezza idrologica dello stato del modello y_t è caratterizzata dalla funzione di verosimiglianza $f(y_t|x_t)$, ottenuta mettendo in relazione le osservazioni storiche, x_t , e le corrispondenti simulazioni del modello, y_t :

$$y_t = \mathbf{H}x_t + v_t \quad (2.2)$$

Fissati x_{t-1} , p e T , la legge della probabilità totale stabilisce che la densità di probabilità dello stato del modello y_t è data da:

$$\kappa(y_t|x_{t-1}) = \int f(y_t|x_t)g(x_t|x_{t-1})dx_t \quad (2.3)$$

Da ciò segue che la funzione di densità a posteriori dello stato x_t è

$$\varphi(x_t|y_t, x_{t-1}) = \frac{f(y_t|x_t)g(x_t|x_{t-1})}{\kappa(y_t|x_{t-1})} \quad (2.4)$$

che, come richiesto, è in grado di minimizzare gli errori compiuti dai modelli.

La funzione di distribuzione dell'incertezza associata alla previsione di piena, in cascata alla densità di probabilità dell'errore della precipitazione in ingresso al modello idrologico, è indicata con $\pi(y_t|p, T)$. Ne segue per l'integrazione dei due processori che:

$$\begin{aligned} \psi(x_t|x_{t-1}, p, T) &= \int \varphi(x_t|x_{t-1})\pi(y_t|p, T)dy_t \\ &= \gamma(x_t|x_{t-1}, p, T)g(x_t|x_{t-1}) \end{aligned} \quad (2.5)$$

dove

$$\gamma(x_t|x_{t-1}, p, T) = \int \frac{f(y_t|x_t)\pi(y_t|p, T)}{\kappa(y_t|x_{t-1})} dy_t \quad (2.6)$$

La teoria bayesiana necessita che il sistema produca una previsione della piena ben calibrata, ossia che la densità di probabilità condizionata della piena osservata $X_{t-1} = x_{t-1}$ attesa debba essere uguale alla densità a priori:

$$\mathbf{E}[\psi(x_t|x_{t-1}, p, T)|X_{t-1} = x_{t-1}] = g(x_t|x_{t-1}) \quad (2.7)$$

2.2 Filtro di Kalman (H^2)

La strada seguita per minimizzare l'errore nelle stime di portata è stata quella di applicare un modello bayesiano, in cui la stima a posteriori viene data dalla combinazione tra la stima a priori data dal modello autoregressivo e dalla sti-

ma della funzione di verosimiglianza data dal modello TOPKAPI. Si consideri il sistema di equazioni nello spazio degli stati invariante nel tempo:

$$\begin{cases} \mathbf{x}_{t+1} = \mathbf{F}\mathbf{x}_t + \mathbf{\Gamma}\mathbf{u}_t \\ \mathbf{y}_t = \mathbf{H}\mathbf{x}_t + \mathbf{v}_t \end{cases} \quad (2.8)$$

dove $\mathbf{F} \in \mathcal{R}^{n \times n}$, $\mathbf{\Gamma} \in \mathcal{R}^{n \times m}$ e $\mathbf{H} \in \mathcal{R}^{p \times n}$ sono matrici note, $\mathbf{x}_t \in \mathcal{R}^n$ è il vettore di stato, $\mathbf{y}_t \in \mathcal{R}^p$ il vettore di osservazioni, o di “pseudo-osservazioni” nel caso si utilizzino le previsioni del modello, ed infine $\mathbf{u}_t \in \mathcal{R}^m$ e $\mathbf{v}_t \in \mathcal{R}^p$ sono vettori di variabili casuali a media nulla e varianze e covarianze note, rappresentativa degli errori.

Stabilito che \mathbf{Q} ed \mathbf{R}_t sono le matrici di covarianza, rispettivamente degli errori del processo markoviano e del modello idrologico, allora il problema da risolvere diventa trovare la stima lineare $\hat{\mathbf{x}}_t = \mathcal{F}_f(\mathbf{y}_0, \dots, \mathbf{y}_t)$ tale che la varianza dell'errore sia minima:

$$\min_{\mathcal{F}_f} \mathbf{E} \left[\sum_{t=0}^{\tau} (\hat{\mathbf{x}}_t - \mathbf{x}_t)^T (\hat{\mathbf{x}}_t - \mathbf{x}_t) \right] \quad (2.9)$$

Teorema 2.1 *Filtro di Kalman*

(a) *La soluzione del problema è data dallo stimatore dei minimi quadrati*

$$\hat{\mathbf{x}}_{t+1} = \mathbf{F}\hat{\mathbf{x}}_t + \mathbf{K}_{f,t+1}(\mathbf{y}_{t+1} - \mathbf{H}\mathbf{F}\hat{\mathbf{x}}_t), \text{ dove}$$

$$\mathbf{K}_{f,t+1} = \mathbf{Q}_{t+1}\mathbf{H}^T\mathbf{R}_{e,t+1}^{-1} \text{ è il guadagno del filtro, e}$$

$$\mathbf{Q}_{t+1} = \mathbf{F}\mathbf{P}_t\mathbf{F} + \mathbf{\Gamma}\mathbf{Q}\mathbf{\Gamma}^T$$

$$\mathbf{R}_{e,t+1} = \mathbf{H}\mathbf{Q}_{t+1}\mathbf{H}^T + \mathbf{R}_{t+1}.$$

(b) *La matrice di covarianza dell'errore è*

$$\mathbf{P}_{t+1} = (\mathbf{I} - \mathbf{K}_{f,t+1}\mathbf{H})\mathbf{Q}_{t+1}(\mathbf{I} - \mathbf{K}_{f,t+1}\mathbf{H})^T + \mathbf{K}_{f,t+1}\mathbf{R}_{t+1}\mathbf{K}_{f,t+1}^T.$$

Se $\mathbf{K}_{f,t+1} = \mathbf{Q}_{t+1} \mathbf{H} \mathbf{R}_{e,t+1}^{-1}$ si riduce alla equazione algebrica di Riccati discreta (DARE):

$$\begin{aligned} \mathbf{P}_{t+1} &= (\mathbf{I} - \mathbf{K}_{f,t+1} \mathbf{H}) \mathbf{Q}_{t+1} \\ &= \mathbf{F} \mathbf{P}_t \mathbf{F}^T + \Gamma \mathbf{Q} \Gamma^T - \mathbf{Q}_{t+1} \mathbf{H}^T \mathbf{R}_{e,t+1}^{-1} \mathbf{H} \mathbf{Q}_{t+1} \\ &= \mathbf{F} \mathbf{P}_t \mathbf{F}^T + \Gamma \mathbf{Q} \Gamma^T - \mathbf{K}_{f,t+1} \mathbf{R}_{e,t+1} \mathbf{K}_{f,t+1}^T \end{aligned}$$

(c) Il filtro di Kalman è equivalente al problema variazionale $\hat{\mathbf{x}}_{t+1} = \arg \min(J)$ dove

$$\begin{aligned} J(\mathbf{x}) &= (\mathbf{x} - \mathbf{F} \hat{\mathbf{x}}_t) \mathbf{Q}_{t+1}^{-1} (\mathbf{x} - \mathbf{F} \hat{\mathbf{x}}_t)^T + \\ &\quad (\mathbf{y}_{t+1} - \mathbf{H} \mathbf{x}) \mathbf{R}_{t+1}^{-1} (\mathbf{y}_{t+1} - \mathbf{H} \mathbf{x})^T \end{aligned}$$

Osservazioni:

- (i) Se gli errori seguono una distribuzione normale, allora $\hat{\mathbf{x}}_t$ è anche lo stimatore di massima verosimiglianza per \mathbf{x}_t ;
- (ii) \mathbf{K}_t è ben definita perché \mathbf{R}_t e \mathbf{P}_t e \mathbf{Q} sono definite positive o semidefinite positive;
- (iii) il problema di minimizzazione è ben posto perché J è una funzione convessa.

Dimostrazione. L'equivalenza tra (a) e (c) stabilisce che se $\hat{\mathbf{x}}_t$ è il valore ottimo, allora

$$\nabla(J(\hat{\mathbf{x}}_t)) = 2\mathbf{Q}_t^{-1}(\hat{\mathbf{x}}_t - \mathbf{F}\hat{\mathbf{x}}_{t-1}) - 2\mathbf{H}^T\mathbf{R}_t^{-1}(\mathbf{y}_t - \mathbf{H}\hat{\mathbf{x}}_t) = 0$$

$$\mathbf{Q}_t^{-1}(\hat{\mathbf{x}}_t - \mathbf{F}\hat{\mathbf{x}}_{t-1}) - \mathbf{H}^T\mathbf{R}_t^{-1}(\mathbf{y}_t - \mathbf{F}\hat{\mathbf{x}}_{t-1}) + \mathbf{H}^T\mathbf{R}_t^{-1}\mathbf{H}(\hat{\mathbf{x}}_t - \mathbf{F}\hat{\mathbf{x}}_{t-1}) = 0.$$

Ne consegue che

$$\hat{\mathbf{x}}_t = \mathbf{F}\hat{\mathbf{x}}_{t-1} + (\mathbf{Q}_t^{-1} + \mathbf{H}^T\mathbf{R}_t^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}_t^{-1}(\mathbf{y}_t - \mathbf{F}\hat{\mathbf{x}}_{t-1}).$$

È diretto da verificare che

$$\mathbf{H}^T\mathbf{R}_t^{-1}(\mathbf{H}\mathbf{Q}_t\mathbf{H}^T + \mathbf{R}_t) = (\mathbf{Q}_t^{-1} + \mathbf{H}^T\mathbf{R}_t^{-1}\mathbf{H})\mathbf{Q}_t\mathbf{H}^T,$$

e quindi che

$$(\mathbf{Q}_t^{-1} + \mathbf{H}^T\mathbf{R}_t^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}_t^{-1} = \mathbf{Q}_t\mathbf{H}^T(\mathbf{H}\mathbf{Q}_t\mathbf{H}^T + \mathbf{R}_t)^{-1}.$$

Si osservi che

$$\hat{\mathbf{x}}_t - \mathbf{x}_t = (\mathbf{I} - \mathbf{K}_{f,t}\mathbf{H})(\mathbf{F}\hat{\mathbf{x}}_{t-1} - \mathbf{x}_t) + \mathbf{K}_{f,t}(\mathbf{y}_t - \mathbf{x}_t),$$

allora la Traccia della matrice di covarianza dell'errore della stima a posteriori può essere scritta come

$$Tr[\mathbf{P}_t] = Tr[(\mathbf{I} - \mathbf{K}_{f,t}\mathbf{H})\mathbf{Q}_t(\mathbf{I} - \mathbf{K}_{f,t}\mathbf{H})^T] + Tr[\mathbf{K}_{f,t}\mathbf{R}_t\mathbf{K}_{f,t}^T]$$

Poiché la Traccia è una funzione scalare, continua e differenziabile dei coefficienti della matrice $\mathbf{K}_{f,t}$, si può definire il suo differenziale $d_{\mathbf{K}}$ considerando i termini del primo ordine in $\mathbf{K}_{f,t}$ della differenza $Tr[\mathbf{P}_t](\mathbf{K}_t + \mathbf{L}) - Tr[\mathbf{P}_t](\mathbf{K}_t)$.

$$\begin{aligned} d_{\mathbf{K}} &= 2Tr[\mathbf{K}_t \mathbf{H} \mathbf{Q}_t \mathbf{H}^T \mathbf{L}^T] - 2Tr[\mathbf{Q}_t \mathbf{H}^T \mathbf{L}^T] + 2Tr[\mathbf{K}_t \mathbf{R}_t \mathbf{L}^T] \\ &= 2Tr[\mathbf{K}_t \mathbf{H} \mathbf{Q}_t \mathbf{H}^T \mathbf{L}^T - \mathbf{Q}_t \mathbf{H}^T \mathbf{L}^T + \mathbf{K}_t \mathbf{R}_t \mathbf{L}^T] \\ &= 2Tr[(\mathbf{K}_t (\mathbf{H} \mathbf{Q}_t \mathbf{H}^T + \mathbf{R}_t) - \mathbf{Q}_t \mathbf{H}^T) \mathbf{L}^T] \end{aligned}$$

Affinché $d_{\mathbf{K}_t}$ sia nullo per \mathbf{L} qualsiasi deve essere

$$\mathbf{K}_t = \mathbf{Q}_t \mathbf{H}^T (\mathbf{H} \mathbf{Q}_t \mathbf{H}^T + \mathbf{R}_t)^{-1}$$

Nel filtro di Kalman viene assunto che le perturbazioni incognite siano variabili stocastiche a media nulla e con una statistica del secondo ordine nota. Posto che le componenti casuali siano *additive white Gaussian noise* (AWGN), allora $\hat{\mathbf{x}}_t$ è anche lo stimatore di massima verosimiglianza per \mathbf{x}_t .

Siano

$$g(\mathbf{x}_{t+1} | \hat{\mathbf{x}}_t) \sim \mathcal{N}(\mathbf{F} \hat{\mathbf{x}}_t, \mathbf{F} \mathbf{P}_t \mathbf{F}^T + \mathbf{\Gamma} \mathbf{Q} \mathbf{\Gamma}^T)$$

$$f(\mathbf{y}_t | \mathbf{x}_t) \sim \mathcal{N}(\mathbf{H} \mathbf{x}_t, \mathbf{R}_t)$$

rispettivamente la funzione di distribuzione a priori del modello autoregressivo e la funzione di verosimiglianza. Dal teorema di Bayes segue che

$$\varphi(\mathbf{x}_{t+1} | \hat{\mathbf{x}}_t) \sim \mathcal{N}(\mathbf{F} \hat{\mathbf{x}}_t + \mathbf{K}_t (\mathbf{y}_{t+1} - \mathbf{H} \mathbf{F} \hat{\mathbf{x}}_t), \mathbf{P}_{t+1})$$

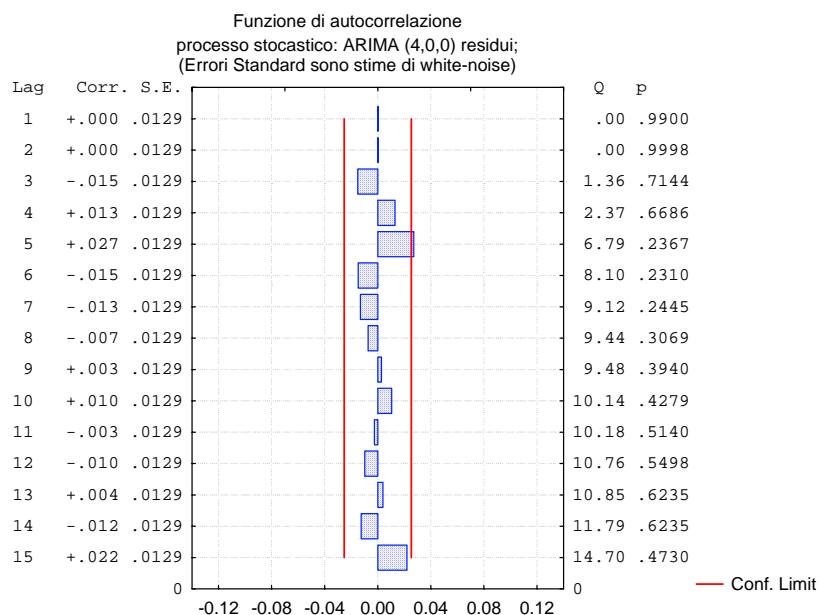


Figura 2.2: Funzione di autocorrelazione dell’errore nella simulazione di un processo stocastico, autoregressivo con rumore bianco, gaussiano ed additivo.

A titolo di esempio è stata generata una serie storica mediante una simulazione Monte Carlo con 1000 numeri pseudo-casuali, relativa ad un processo stocastico markoviano del secondo ordine. I coefficienti di autocorrelazione imposti erano rispettivamente $\alpha_1 = 1.60$ e $\alpha_2 = 0.65$. Utilizzando il programma Statistica le stime dei coefficienti ottenute ed i corrispondenti errori standard, calcolati sulla base della numerosità del campione, sono stati rispettivamente $a_1 = 1.6067 \pm 0.013$ e $a_2 = .6611 \pm .0241$. La stima della varianza dell’errore è stata $Q = .03549$. Le stime ottenute con il metodo dei minimi quadrati sono $a_1 = 1.5953$ e $a_2 = .6458$, coerenti con l’intervallo di confidenza. Anche la stima della varianza, $Q = .03487$, non si discosta molto da quella calcolata. Come si può notare dalla

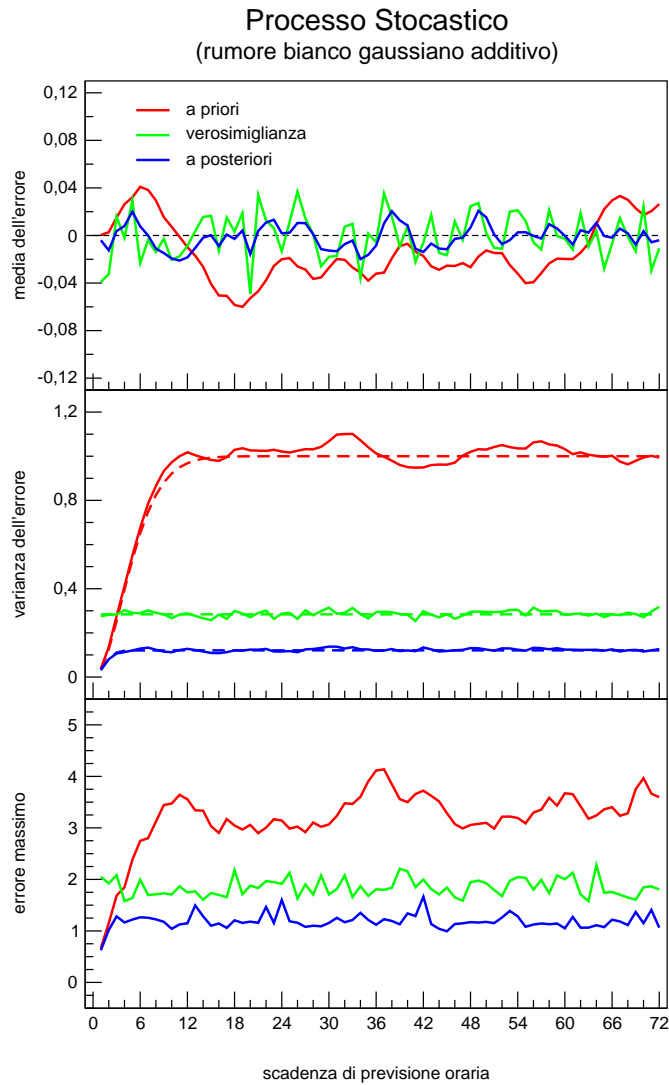


Figura 2.3: Medie (in alto), varianze (al centro) e valori massimi (in basso) dell'errore della stima a priori (in rosso, linea continua valore osservato, tratteggiata valore teorico), dell'errore della funzione di verosimiglianza (in verde) e dell'errore della stima a posteriori (in blu) relativi alla simulazione di un processo stocastico, autoregressivo con rumore bianco, gaussiano ed additivo.

fi gura 2.2, i valori della funzione di autocorrelazione degli errori rientrano tutti nell'intervallo di confi denza.

Nella figura 2.3 in alto sono stati riportati i valori medi per ogni scadenza temporale di simulazione, che dal test t-student, con livello di significatività $\alpha = 0.05$, risultano non essere significativamente diversi da zero. La figura centrale riporta invece le varianze degli errori ed in particolare è possibile osservare come le varianze campionarie si accordino molto bene con i corrispondenti valori teorici, sia per il modello autoregressivo, sia per la combinazione bayesiana calcolata con il filtro di Kalman. Come evidenziato nella figura in basso, la stima a posteriori del filtro riduce, inoltre, il valore massimo dell'errore.

2.3 Filtro di Kalman adattativo

Una prima generalizzazione è stata quella di utilizzare un modello nello spazio degli stati non stazionario, nel quale la matrice di transizione \mathbf{F} non è più invariante col tempo:

$$\begin{cases} \mathbf{x}_{t+1} = \mathbf{F}_t \mathbf{x}_t + \mathbf{\Gamma} \mathbf{u}_t \\ \mathbf{y}_t = \mathbf{H} \mathbf{x}_t + \mathbf{v}_t \end{cases} \quad (2.10)$$

Il vettore dei coefficienti di autoregressione, $\theta_t = (a_{1,t}, \dots, a_{p,t})^T$, che compaiono nella matrice \mathbf{F}_t vengono stimati ad ogni istante temporale mediante uno *smoother* che tiene conto della sequenza $(x_{t-p}, x_{t-p+1}, \dots, x_{t-1}, x_t)$ degli ultimi p valori osservati. La realizzazione di questo filtro adattativo è analoga al metodo *Mutually Interactive StateParameter Estimation* (MISP) [15, 30]. Nella stima dei coefficienti si assume, in primo luogo, che il vettore sia quasi-stazionario, ovvero che la stima a priori è uguale al valore assunto all'istante precedente $\hat{\theta}_t = \hat{\theta}_{t-1}$.

Quindi si calcola la stima $\hat{\mathbf{x}}_t$ combinando le p equazioni:

$$\mathbf{L}\hat{\mathbf{x}}_t = \mathbf{L}\mathbf{F}_{t-1}^p \mathbf{x}_{t-p} + \sum_{k=0}^{p-1} \mathbf{L}\mathbf{F}_{t-1}^k \mathbf{K}_{t-k} \nu_{t-k}$$

dove $\mathbf{L} = \begin{pmatrix} 0 & 0 & \dots & 1 \end{pmatrix}$. Esprimendo in forma scalare l'equazione precedente, è possibile dedurre il termine di innovazione all'istante t -esimo

$$\nu_t^{(\theta)} = \hat{x}_{t-p+1|t} - \mathbf{x}_{t-p}^T \hat{\theta}_{t-1} = \sum_{k=0}^{p-1} \mathbf{K}_{t-k}^{(p-k)} \nu_{t-k}$$

dove $\mathbf{K}_{t-k}^{(p-k)}$ esprime la componente $(p-k)$ -esima del guadagno dello *smoother*. La stima del vettore dei parametri all'istante t è data perciò da

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \mathbf{K}_t^{(\theta)} \nu_t^{(\theta)} \quad (2.11)$$

con la matrice del guadagno definita come

$$\mathbf{K}_t^{(\theta)} = (\mathbf{P}_{t-1}^{(\theta)} + \mathbf{Q}^{(\theta)}) \mathbf{x}_{t-p} \left(\mathbf{x}_{t-p}^T (\mathbf{P}_{t-1}^{(\theta)} + \mathbf{Q}^{(\theta)}) \mathbf{x}_{t-p} + \mathbf{R}_t^{(\theta)} \right)^{-1}$$

Nell'equazione precedente $\mathbf{P}_{t-1}^{(\theta)}$ è la matrice di covarianza dell'errore della stima a posteriori, $\mathbf{Q}^{(\theta)}$ è la matrice di covarianza della componente stocastica, mentre $\mathbf{R}_t^{(\theta)}$ è la matrice di covarianza dell'errore della stima data dalla funzione di verosimiglianza.

Partendo dalla definizione del termine di innovazione è possibile definire $\mathbf{R}_t^{(\theta)}$:

$$\mathbf{R}_t^{(\theta)} = \mathbf{E} \left[\nu_t^{(\theta)2} \right]$$

$$\begin{aligned}
&= \mathbf{E} \left[\left(\sum_{k=0}^{p-1} \mathbf{K}_{t-k}^{(p-k)} \nu_{t-k} \right) \left(\sum_{j=0}^{p-1} \mathbf{K}_{t-j}^{(p-j)} \nu_{t-j} \right) \right] \\
&= \mathbf{E} \left[\sum_{k=0}^{p-1} \sum_{j=0}^{p-1} \mathbf{K}_{t-k}^{(p-k)} \mathbf{K}_{t-j}^{(p-j)} \nu_{t-k} \nu_{t-j} \right] \\
&= \sum_{k=0}^{p-1} \sum_{j=0}^{p-1} \mathbf{K}_{t-k}^{(p-k)} \mathbf{K}_{t-j}^{(p-j)} \mathbf{E} [\nu_{t-k} \nu_{t-j}] \\
&= \sum_{k=0}^p \left(\mathbf{K}_{t-k}^{(p-k)} \right)^2 \mathbf{E} [(\mathbf{y}_t - \mathbf{H}\mathbf{F}_{t-1} \hat{\mathbf{x}}_{t-1})^2] \\
&= \sum_{k=0}^p \left(\mathbf{K}_{t-k}^{(p-k)} \right)^2 \mathbf{E} [(\mathbf{y}_t - \mathbf{H}\mathbf{x}_t + \mathbf{H}\mathbf{x}_t - \mathbf{F}_{t-1} \hat{\mathbf{x}}_{t-1})^2] \\
&= \sum_{k=0}^p \left(\mathbf{K}_{t-k}^{(p-k)} \right)^2 \left(\mathbf{E} [(\mathbf{y}_t - \mathbf{H}\mathbf{x}_t)(\mathbf{y}_t - \mathbf{H}\mathbf{x}_t)^T] + \right. \\
&\quad \left. \mathbf{F}_{t-1} \mathbf{E} [(\hat{\mathbf{x}}_{t-1} - \mathbf{x}_{t-1})(\hat{\mathbf{x}}_{t-1} - \mathbf{x}_{t-1})^T] \mathbf{F}_{t-1}^T + \mathbf{\Gamma} \mathbf{E} [\mathbf{u}_t \mathbf{u}_t^T] \mathbf{\Gamma}^T \right) \\
&= \sum_{k=0}^p \left(\mathbf{K}_{t-k}^{(p-k)} \right)^2 (\mathbf{R}_t + \mathbf{F}_{t-1} \mathbf{P}_{t-1} \mathbf{F}_{t-1}^T + \mathbf{\Gamma} \mathbf{Q} \mathbf{\Gamma}^T) \\
&= \sum_{k=0}^p \left(\mathbf{K}_{t-k}^{(p-k)} \right)^2 (\mathbf{R}_t + \mathbf{Q}_t)
\end{aligned}$$

L'equazione algebrica di Riccati per la stima della varianza dell'errore a posteriori resta la medesima.

2.4 Modellazione dell'errore

Il problema della modellazione della portata di un fiume è dovuto essenzialmente alla complessità del fenomeno. Questo ha delle forti implicazioni sulla struttura dell'errore della stima, poichè la portata del fiume è l'effetto osservabile di una numerosa quantità di processi fisici correlati tra loro, che in parte sono stati elencati nell'introduzione, e che soltanto in maniera approssimativa si è in grado

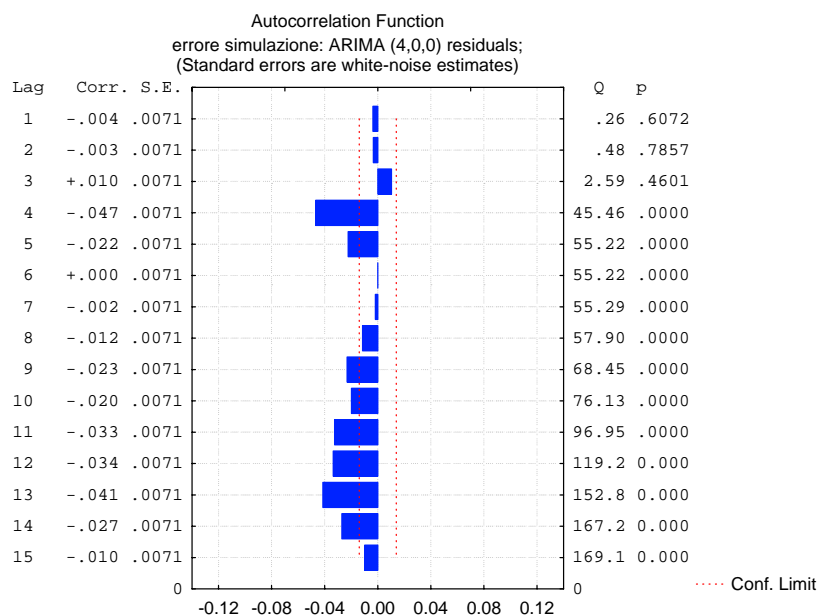


Figura 2.4: Funzione di autocorrelazione dell'errore nella simulazione del processo di piena con il modello TOPKAPI. Le piogge fornite in ingresso al modello idrologico sono quelle osservate dalla rete pluviometrica.

di modellare. Lo scostamento tra lo stato del modello e lo stato del sistema ad un determinato istante, ovvero l'errore compiuto dal modello nel simulare, o prevedere, la portata del fiume, dipende da come è evoluto lo stato del sistema e da quanto gradualmente il modello si è andato a discostare dalla realtà del bacino.

Sono state effettuate delle simulazioni con il modello idrologico TOPKAPI, nelle quali si è introdotto come variabile di ingresso la precipitazione osservata dalla rete pluviometrica del bacino, in modo tale da minimizzare l'incertezza dovuta a questa grandezza. L'errore viene così a dipendere soltanto dalle parametrizzazioni e dalle approssimazioni del modello. La dimensione del campione relativo alle simulazioni effettuate sono 273 records giornalieri. Ogni record è

composto da 72 valori orari di portata. L'esperienza dimostra che le stime delle piene di portata, ottenute con il modello TOPKAPI, hanno un errore autocorrelato, che si è cercato di rappresentare come un processo markoviano, mediante un modello autoregressivo AR(4).

$$v_t = \sum_{k=1}^q \beta_k v_{t-k} + w_t \quad (2.12)$$

I primi due coefficienti di autocorrelazione sono risultati significativamente diversi da zero, con un livello di confidenza di $\alpha = 0.05$. Nella figura 2.4 è riportata la funzione di autocorrelazione degli errori, dalla quale si evince che il modello autoregressivo può essere soltanto una prima approssimazione, poiché per distanze temporali superiori al tempo di correlazione del fiume tende a divergere oltre l'intervallo di confidenza. Ciò potrebbe essere indicativo di una correlazione di più lungo periodo, anche non lineare, ma non è stato possibile esplorare ulteriormente questo punto. Introducendo il modello autoregressivo per l'errore, le equazioni del filtro devono essere riscritte nella seguente maniera:

$$\begin{cases} \mathbf{x}_{t+1} = \mathbf{F}_t \mathbf{x}_t + \mathbf{d}_t \\ \mathbf{y}_t = \mathbf{H} \mathbf{x}_t \end{cases} \quad (2.13)$$

dove il vettore di stato \mathbf{x}_t è definito come

$$\mathbf{x}_t = \begin{pmatrix} x_t & x_{t-1} & \dots & x_{t-p+1} & v_t & v_{t-1} & \dots & v_{t-q+1} \end{pmatrix}^T$$

$$\mathbf{d}_t = \begin{pmatrix} u_t & 0 & \dots & 0 & w_t & 0 & \dots & 0 \end{pmatrix}^T$$

La matrice di transizione diventa la matrice a blocchi

$$\mathbf{F}_t = \begin{pmatrix} \mathbf{F}_{t,1} & 0 \\ 0 & \mathbf{F}_{t,2} \end{pmatrix}$$

Analogamente a quanto fatto nel paragrafo precedente, è stata generata una serie storica mediante una simulazione Monte Carlo con 1000 numeri pseudo-casuali, relativa ad un processo stocastico markoviano del secondo ordine per gli errori. I coefficienti di autocorrelazione imposti erano rispettivamente $\beta_1 = 1.20$ e $\beta_2 = 0.55$. Come si evidenzia nella figura 2.5, i valori della funzione di autocorrelazione degli errori per questo caso non sono contenuti nell'ambito dell'intervallo di confidenza. Utilizzando il programma Statistica le stime dei coefficienti ottenute ed i corrispondenti errori standard sono stati rispettivamente $b_1 = 1.2069 \pm 0.013$ e $b_2 = .5585 \pm .0202$. Con il metodo dei minimi quadrati le stime ottenute sono state $b_1 = 1.2184$ e $b_2 = .5634$, ancora coerenti con quelle individuate con il programma Statistica.

Le medie degli errori non risultano essere significativamente diversi da zero, ad un livello di significatività di $\alpha = 0.05$, per tutto il periodo temporale considerato (figura 2.6). Tuttavia, come evidenziato nella figura al centro, esiste una forte discrepanza tra la funzione della varianza campionaria della stima del filtro, calcolata sulla base delle simulazioni, e la corrispondente funzione teorica. Soltanto quando si applica questa versione del filtro, nel quale il vettore di stato è composto sia dai valori che dagli errori, si ottiene di nuovo un ottimo accordo tra i valori campionari ed i valori teorici (figura 2.7).

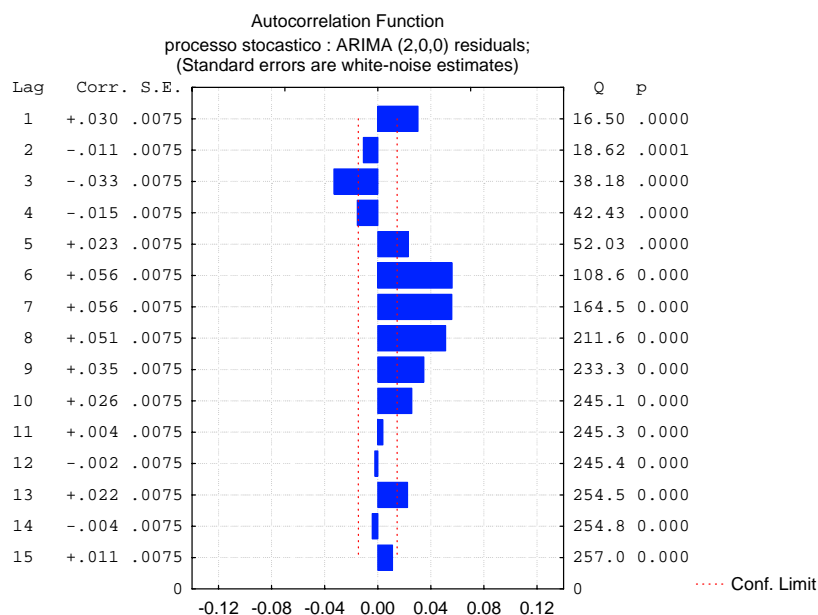


Figura 2.5: Funzione di autocorrelazione dell'errore nella simulazione di processo stocastico, autoregressivo con rumore colorato, gaussiano, additivo.

2.5 Filtro minimax (H^∞)

Un aspetto importante, ma non ancora considerato, è quello relativo alla distribuzione statistica degli errori. Le principali limitazioni all'utilizzo del filtro di Kalman sono due. La prima concerne la necessità di conoscere i primi due momenti, la media e la varianza, della distribuzione statistica degli errori. Si può ottenere una stima campionaria dei momenti della distribuzione da una serie temporale identificabile come un segmento di un processo stocastico stazionario, ma quando si trattano sistemi con un comportamento fortemente eteroschedastico, ossia con una varianza dell'errore non stazionaria, come nel caso delle serie di portata, legate alle fluttuazioni meteorologiche e climatiche, è piuttosto problematico

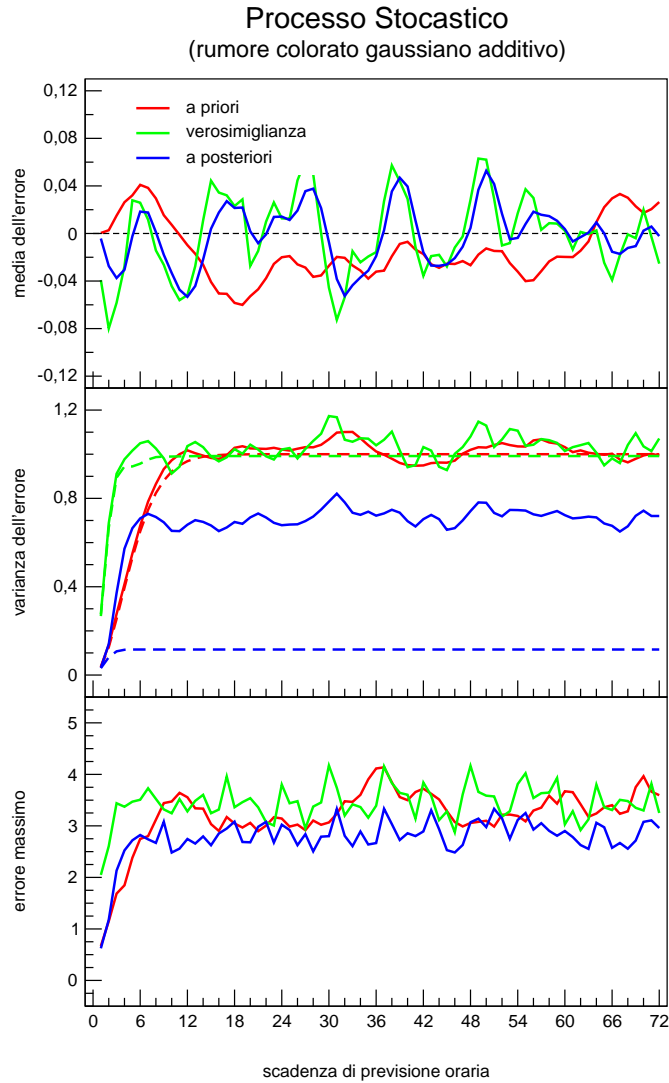


Figura 2.6: Idem della figura 2.3 per errori relativi alla simulazione di un processo stocastico, autoregressivo con rumore colorato, gaussiano ed additivo.

arrivare ad un buon grado di conoscenza dei momenti della distribuzione.

La seconda limitazione è invece dovuta alla costruzione del filtro, il quale minimizza la varianza media dell'errore, ma potrebbe essere poco significativa nel caso di una distribuzione con code più pesanti della distribuzione normale.

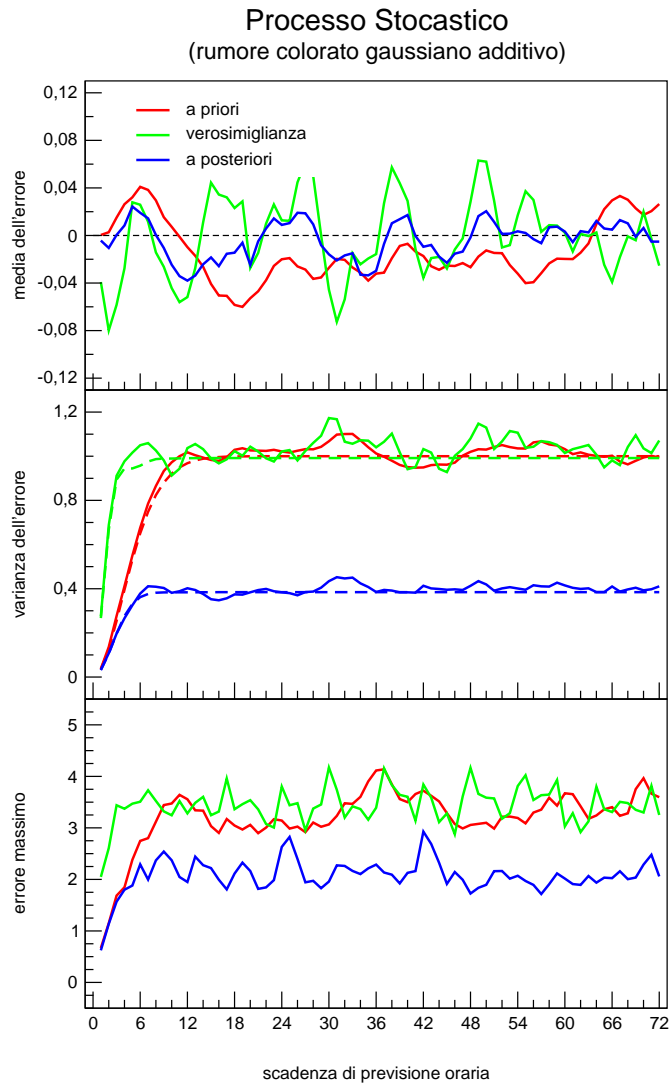


Figura 2.7: Idem della figura 2.6. Nell'equazione di stato si è introdotto il modello autoregressivo per l'errore.

Riprendendo le stime effettuate con il modello autoregressivo e le simulazioni del modello TOPKAPI, già descritte nel paragrafo precedente, si è utilizzato di nuovo Statistica per ottenere le distribuzioni campionarie degli errori e valutare quanto fossero simili ad una distribuzione normale. Si può osservare dalle figure

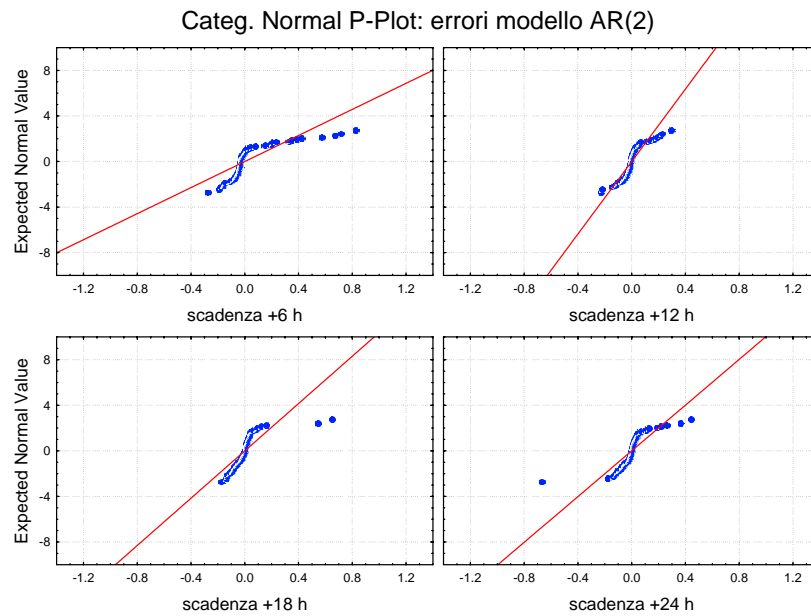


Figura 2.8: Diagramma P-P degli errori del modello autoregressivo AR(2). La linea continua rossa indica i valori relativi alla distribuzione normale, i cerchi blu i valori della distribuzione campionaria.

2.8 e 2.9 che gli errori del modello autoregressivo e quelli del modello idrologico, sebbene in maniera più contenuta, tendono a discostarsi dalla distribuzione gaussiana, soprattutto per i valori estremi.

In maniera analoga a quanto fatto per le simulazioni, si è compiuta un'analisi delle proprietà statistiche degli errori compiuti dal modello TOPKAPI, fornendo in ingresso non più le precipitazioni osservate dalla rete pluviometrica bensì i valori previsti dal modello meteorologico ad area limitata COSMO-LAMI. In questo modo le sorgenti di errore dell'errore sono due: quella dovuta al modello idrologico e quella legata alla previsione meteorologica. La dimensione del campione è di 218 records giornalieri, inferiore alla durata del periodo per la mancanza

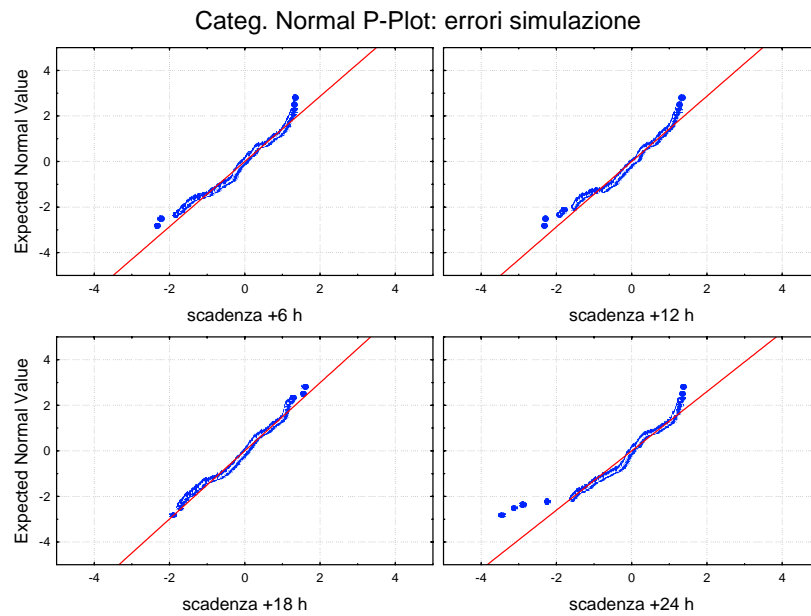


Figura 2.9: Diagramma P-P degli errori delle simulazioni di portata effettuate con il modello idrologico TOPKAPI, utilizzando le piogge osservate dalla rete pluviometrica.

delle previsioni del modello COSMO–LAMI per alcuni giorni. Anche per gli errori delle previsioni i valori estremi della distribuzione tendono a discostarsi dai valori normali (fi gura 2.10).

Come si evince dall’analisi statistica, per il comportamento eteroschedastico delle portate le due condizioni necessarie al fi ltro di Kalman non possono essere soddisfatte. Cosa è dunque necessario fare? Il fi ltro di Kalman naturalmente può essere utilizzato comunque, e può ancora continuare a fornire la stima migliore, tuttavia è sembrato opportuno nel proseguo del lavoro ricorrere ad un altro tipo di fi ltro, introdotto per la prima volta nel 1987 da Mike Grimlbe, dell’Università di Strathclyde, e denominato fi ltro H^∞ . Da allora il fi ltro è stato formulato in diversi

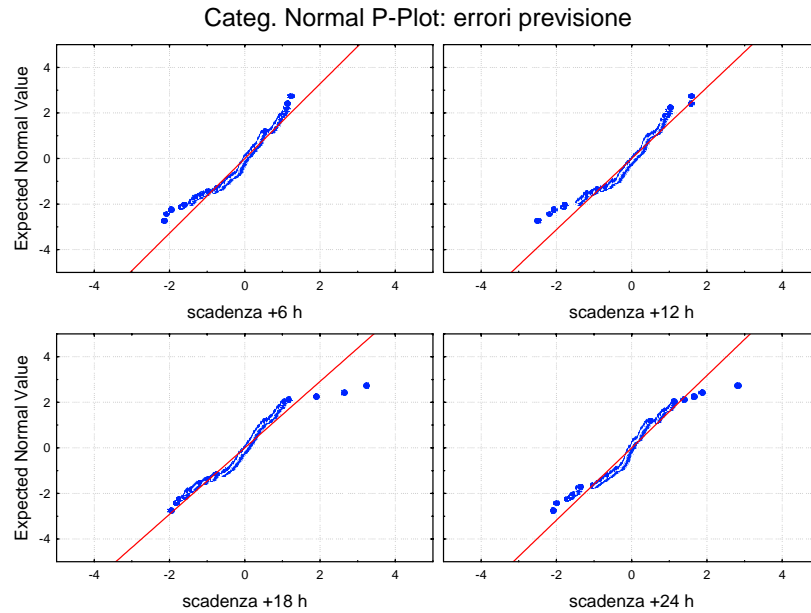


Figura 2.10: Idem della figura 2.9 per gli errori delle previsioni di portata effettuate con il modello idrologico TOPKAPI, utilizzando le piogge previste dal modello meteorologico COSMO-LAMI.

modi, ma qui, nel cercare di applicarlo al problema della previsione di portata per il fiume Reno, è stata utilizzata la versione introdotta da [11, 12, 15], in cui viene data una formale equivalenza con il filtro di Kalman, di seguito indicato con H^2 . Questo filtro ha la particolarità di non richiedere nessuna assunzione sulle distribuzioni.

Si supponga di utilizzare un modello non stazionario nello spazio degli stati:

$$\begin{cases} \mathbf{x}_{t+1} = \mathbf{F}_t \mathbf{x}_t + \mathbf{\Gamma} \mathbf{u}_t \\ \mathbf{y}_t = \mathbf{H} \mathbf{x}_t + \mathbf{v}_t \end{cases} \quad (2.14)$$

La norma che si vuole minimizzare nel filtro H_∞ è definita come

$$J_\infty = \min \left\{ \sup_{\mathbf{u}_t, \mathbf{v}_t} \frac{\sum_{t=0}^{\tau} \|\mathbf{L}\hat{\mathbf{x}}_t - \mathbf{L}\mathbf{x}_t\|^2}{\sum_{t=0}^{\tau} \|\mathbf{u}_t\|^2 + \|\mathbf{v}_t\|^2} \right\} \quad (2.15)$$

dove \mathbf{L} è un operatore lineare che lega l'osservabile alle componenti del vettore di stato \mathbf{x}_t e τ denota la lunghezza della serie temporale disponibile. Non sempre esiste una soluzione analitica al problema di minimizzazione, per cui è necessario ricorrere al criterio sub-ottimale

$$\sup_{\mathbf{u}_t, \mathbf{v}_t} \frac{\sum_{t=0}^{\tau} \|\mathbf{L}\hat{\mathbf{x}}_t - \mathbf{L}\mathbf{x}_t\|^2}{\sum_{t=0}^{\tau} \|\mathbf{u}_t\|^2 + \|\mathbf{v}_t\|^2} < \gamma_f^2 \quad (2.16)$$

dove γ_f^2 è da considerarsi come un livello di attenuazione del disturbo. È possibile perciò riscrivere l'equazione del problema sub-ottimale nella forma

$$J_{f,\tau} = \sum_{t=0}^{\tau} \|\mathbf{u}_t\|^2 + \|\mathbf{v}_t\|^2 - \gamma_f^{-2} \sum_{t=0}^{\tau} \|\mathbf{L}\hat{\mathbf{x}}_t - \mathbf{L}\mathbf{x}_t\|^2$$

Posto $\hat{\mathbf{s}}_t = \mathbf{L}\hat{\mathbf{x}}_t$ essere lo stimatore di $\mathbf{s}_t = \mathbf{L}\mathbf{x}_t$, $J_{f,\tau}$ diventa

$$\begin{aligned} J_{f,\tau} &= \sum_{t=0}^{\tau} \|\mathbf{u}_t\|^2 + \|\mathbf{v}_t\|^2 - \gamma_f^{-2} \sum_{t=0}^{\tau} \|\hat{\mathbf{s}}_t - \mathbf{s}_t\|^2 \\ &= \sum_{t=0}^{\tau} \mathbf{u}_t^T \mathbf{Q}^{-1} \mathbf{u}_t + \sum_{t=0}^{\tau} \mathbf{v}_t^T \mathbf{R}_t^{-1} \mathbf{v}_t - \gamma_f^{-2} \sum_{t=0}^{\tau} \|\hat{\mathbf{s}}_t - \mathbf{s}_t\|^2 \\ &= \sum_{t=0}^{\tau} \mathbf{u}_t^T \mathbf{Q}^{-1} \mathbf{u}_t + \sum_{t=0}^{\tau} (\mathbf{y}_t - \mathbf{H}\mathbf{F}_{t-1}\mathbf{x}_{t-1})^T \mathbf{R}_t^{-1} (\mathbf{y}_t - \mathbf{H}\mathbf{F}_{t-1}\mathbf{x}_{t-1}) \\ &\quad - \gamma_f^{-2} \sum_{t=0}^{\tau} \|\hat{\mathbf{s}}_t - \mathbf{s}_t\|^2 \\ &= \sum_{t=0}^{\tau} \mathbf{u}_t^T \mathbf{Q}^{-1} \mathbf{u}_t + \left[\begin{pmatrix} \mathbf{s}_t \\ \hat{\mathbf{y}}_t \end{pmatrix} - \begin{pmatrix} \mathbf{L} \\ \mathbf{H} \end{pmatrix} \mathbf{F}_{t-1} \mathbf{x}_{t-1} \right]^T \end{aligned}$$

$$\begin{pmatrix} -\gamma_f^{-2}\mathbf{I} & 0 \\ 0 & \mathbf{R}_t^{-1} \end{pmatrix} \left[\begin{pmatrix} \mathbf{s}_t \\ \hat{\mathbf{y}}_t \end{pmatrix} - \begin{pmatrix} \mathbf{L} \\ \mathbf{H} \end{pmatrix} \mathbf{F}_{t-1}\mathbf{x}_{t-1} \right]$$

dove \mathbf{Q} e \mathbf{R}_t sono scalari, o matrici, che diversamente dal filtro di Kalman non rappresentano la varianza degli errori, ma devono essere definite attraverso un processo di aggiustamento.

Il problema di minimizzazione diviene equivalente al problema di trovare uno stimatore minimax e può essere riscritto nella forma $\min\{\max_{\mathbf{u}_t, \mathbf{v}_t} J_{f, \tau}\}$. Lo stimatore minimax è uno stimatore il cui rischio massimo è minore o uguale al rischio massimo di ogni altro stimatore. È uno stimatore conservativo perché protegge contro la condizione peggiore, ovvero cerca di minimizzare il rischio massimo. In termini generali la funzione di rischio è una misura dell'errore associato allo stimatore. Lo stimatore minimax può essere inteso come uno stimatore bayesiano con rischio costante.

Il filtro H^∞ può essere trattato in modo unificata con il filtro H [11, 12, 15].

Teorema 2.2 *Filtro H^∞*

Per un dato livello $\gamma_f > 0$, lo stimatore $\hat{\mathbf{x}}_t$, tale per cui

$$\sup_{\mathbf{u}_t, \mathbf{v}_t} \frac{\sum_{t=0}^{\tau} \|\mathbf{L}\hat{\mathbf{x}}_t - \mathbf{L}\mathbf{x}_t\|^2}{\sum_{t=0}^{\tau} \|\mathbf{u}_t\|^2 + \|\mathbf{v}_t\|^2} < \gamma_f^2$$

esiste se e solo se è soddisfatta la condizione

$$\mathbf{P}_t^{-1} + \mathbf{H}^T\mathbf{H} - \gamma_f^{-2}\mathbf{L}^T\mathbf{L} > 0 \tag{2.17}$$

Posto

$$\mathbf{Q}_t = \mathbf{F}_{t-1} \mathbf{P}_{t-1} \mathbf{F}_{t-1}^T + \Gamma \mathbf{Q} \Gamma^T$$

e

$$\mathbf{R}_{e,t} = \begin{pmatrix} -\gamma_f^{-2} \mathbf{I} & 0 \\ 0 & \mathbf{R}_t \end{pmatrix} + \begin{pmatrix} \mathbf{L} \\ \mathbf{H} \end{pmatrix} \mathbf{Q}_t \begin{pmatrix} \mathbf{L}^T & \mathbf{H}^T \end{pmatrix}$$

la matrice \mathbf{P}_t è definita mediante l'equazione algebrica di Riccati ricorsiva

$$\mathbf{P}_{t+1} = \mathbf{F}_t \mathbf{P}_t \mathbf{F}_t^T + \Gamma \mathbf{Q} \Gamma^T - \mathbf{Q}_{t+1} \begin{pmatrix} \mathbf{L}^T & \mathbf{H}^T \end{pmatrix} \mathbf{R}_{e,t+1}^{-1} \begin{pmatrix} \mathbf{L} \\ \mathbf{H} \end{pmatrix} \mathbf{Q}_{t+1}$$

Osservazione:

È importante sottolineare che per $\gamma_f \rightarrow \infty$ lo stimatore H^∞ tende a coincidere con lo stimatore dei minimi quadrati definito attraverso il filtro \hat{H} . Infatti, la matrice $\mathbf{R}_{e,t}$, definita

$$\begin{aligned} \mathbf{R}_{e,t} &= \begin{pmatrix} -\gamma_f^{-2} \mathbf{I} & 0 \\ 0 & \mathbf{R}_t \end{pmatrix} + \begin{pmatrix} \mathbf{L} \\ \mathbf{H} \end{pmatrix} \mathbf{Q}_t \begin{pmatrix} \mathbf{L}^T & \mathbf{H}^T \end{pmatrix} \\ &= \begin{pmatrix} -\gamma_f^{-2} \mathbf{I} & 0 \\ 0 & \mathbf{R}_t \end{pmatrix} + \begin{pmatrix} \mathbf{L} \mathbf{Q}_t \mathbf{L}^T & \mathbf{L} \mathbf{Q}_t \mathbf{H}^T \\ \mathbf{L} \mathbf{Q}_t \mathbf{H}^T & \mathbf{H} \mathbf{Q}_t \mathbf{H}^T \end{pmatrix} \\ &= \begin{pmatrix} -\gamma_f^{-2} \mathbf{I} + \mathbf{L} \mathbf{Q}_t \mathbf{L}^T & \mathbf{L} \mathbf{Q}_t \mathbf{H}^T \\ \mathbf{L} \mathbf{Q}_t \mathbf{H}^T & \mathbf{R}_t + \mathbf{H} \mathbf{Q}_t \mathbf{H}^T \end{pmatrix} \end{aligned}$$

può essere considerata come una matrice a blocchi ed invertita utilizzando la

forma analitica

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & -(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \end{pmatrix}$$

Ponendo

$$\begin{aligned} \mathbf{M}_t &= (-\gamma_f^2 \mathbf{I} + \mathbf{L}\mathbf{Q}_t\mathbf{L}^T - \mathbf{L}\mathbf{Q}_t\mathbf{H}^T(\mathbf{R}_t + \mathbf{H}\mathbf{Q}_t\mathbf{H}^T)^{-1}\mathbf{H}\mathbf{Q}_t\mathbf{L}^T) \\ &= (-\gamma_f^2 \mathbf{I} + \mathbf{L}(\mathbf{I} - \mathbf{K}_{f,t}\mathbf{H})\mathbf{Q}_t\mathbf{L}^T)^{-1} \end{aligned}$$

dove $\mathbf{K}_{f,t} = \mathbf{Q}_t\mathbf{H}^T(\mathbf{R}_t + \mathbf{H}\mathbf{Q}_t\mathbf{H}^T)^{-1}$, si ottiene che l'inversa della matrice $\mathbf{R}_{e,t}$ è data da

$$\mathbf{R}_{e,t}^{-1} = \begin{pmatrix} \mathbf{M}_t & \mathbf{M}_t\mathbf{L}\mathbf{K}_{f,t} \\ \mathbf{K}_{f,t}^T\mathbf{L}^T\mathbf{M}_t & (\mathbf{R}_t + \mathbf{H}\mathbf{Q}_t\mathbf{H}^T)^{-1} + \mathbf{K}_{f,t}^T\mathbf{L}^T\mathbf{M}_t\mathbf{L}\mathbf{K}_{f,t} \end{pmatrix}$$

Per $\gamma_f \rightarrow \infty$ si evince che $\mathbf{M}_t \rightarrow \mathbf{0}$ e

$$\mathbf{R}_{e,t}^{-1} \rightarrow \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (\mathbf{R}_t + \mathbf{H}\mathbf{Q}_t\mathbf{H}^T)^{-1} \end{pmatrix}$$

Il filtro H^∞ tende al filtro H^2 e la equazione algebrica di Riccati diventa

$$\begin{aligned} \mathbf{P}_{t+1} &= \mathbf{F}_t\mathbf{P}_t\mathbf{F}_t^T + \mathbf{\Gamma}\mathbf{Q}\mathbf{\Gamma}^T \\ &\quad - \mathbf{Q}_{t+1} \begin{pmatrix} \mathbf{L}^T & \mathbf{H}^T \end{pmatrix} \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (\mathbf{R}_t + \mathbf{H}\mathbf{Q}_t\mathbf{H}^T)^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{L} \\ \mathbf{H} \end{pmatrix} \mathbf{Q}_{t+1} \end{aligned}$$

$$\begin{aligned}
&= \mathbf{F}_t \mathbf{P}_t \mathbf{F}_t^T + \mathbf{\Gamma} \mathbf{Q} \mathbf{\Gamma}^T - \mathbf{Q}_{t+1} \begin{pmatrix} \mathbf{L}^T & \mathbf{H}^T \end{pmatrix} \begin{pmatrix} 0 \\ (\mathbf{R}_t + \mathbf{H} \mathbf{Q}_t \mathbf{H}^T)^{-1} \mathbf{H} \end{pmatrix} \mathbf{Q}_{t+1} \\
&= \mathbf{F}_t \mathbf{P}_t \mathbf{F}_t^T + \mathbf{\Gamma} \mathbf{Q} \mathbf{\Gamma}^T - \mathbf{Q}_{t+1} \mathbf{H}^T (\mathbf{R}_t + \mathbf{H} \mathbf{Q}_t \mathbf{H}^T)^{-1} \mathbf{H} \mathbf{Q}_{t+1}
\end{aligned}$$

In seguito il coefficiente γ_f è stato posto uguale ad un valore superiore al massimo degli autovalori di $(\mathbf{Q}_t^{-1} + \mathbf{H}^T \mathbf{H})^{-1}$ e tale per cui vale la condizione aggiuntiva 2.17 [15].

Le differenze principali tra i filtri H^∞ e H^2 sono due. La prima è che deve essere soddisfatta la condizione aggiuntiva 2.17 perché possa esistere lo stimatore. Questa condizione non è presente nel filtro H^2 , in quanto la matrice di covarianza \mathbf{P}_t è sempre definita semi-positiva.

La seconda è data dalla presenza della matrice di covarianza $\begin{pmatrix} -\gamma_f^{-2} \mathbf{I} & 0 \\ 0 & \mathbf{R}_t^{-1} \end{pmatrix}$, indefinita e della matrice \mathbf{L} nell'equazione algebrica di Riccati. In particolare, la presenza di \mathbf{L} significa, ad esempio, che la stima della prima componente del vettore di stato con il filtro H^∞ , diversamente da quanto avviene per il filtro H^2 , non coincide con la prima componente della stima dell'intero vettore di stato. Infatti nel primo caso si ha $\mathbf{L} = \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix}^T$, mentre per il secondo $\mathbf{L} = \begin{pmatrix} 1 & 1 & \dots & 1 \end{pmatrix}^T$. Nel filtro H^2 accade, invece, che la stima di una qualsiasi combinazione lineare del vettore di stato è data semplicemente dalla medesima combinazione lineare della stima del vettore.

Ma ciò che appare sorprendente è invece il fatto che la soluzione del problema H^∞ è formalmente simile alla soluzione del problema del filtro di Kalman, fatto salva la condizione aggiuntiva per la sua esistenza. Nel problema H^2 la proiezione nello spazio di Hilbert dei vettori delle variabili casuali minimizzano la forma

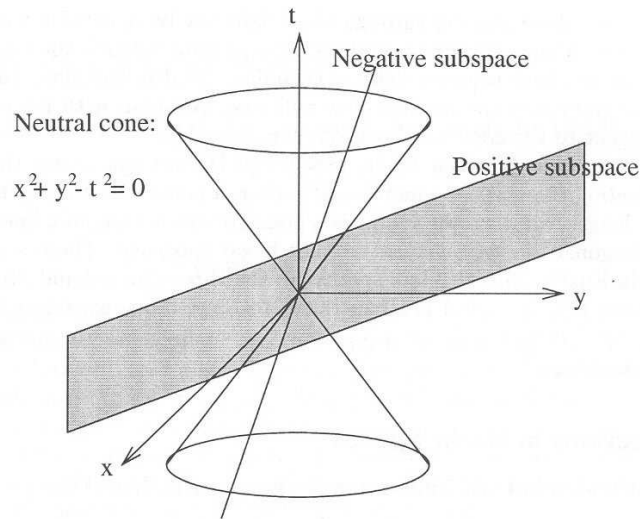


Figura 2.11: Decomposizione dei vettori positivi nello spazio di Krein, tratta da [12].

quadratica 2.9. Per il problema H^∞ può essere definito un tipo speciale di spazio metrico, indefinito, chiamato spazio di Krein, che differisce dallo spazio di Hilbert per alcune caratteristiche peculiari. Ci sono dei vettori neutri di lunghezza nulla che hanno norma non nulla. Esistono perciò dei sottospazi, denominati degeneri, che contengono dei vettori non nulli ortogonali a tutti gli altri vettori. Una conseguenza di questo fatto è che negli spazi di Krein non esiste sempre la proiezione di un vettore in un sottospazio, e quando esiste può non necessariamente essere in grado di minimizzare la forma quadratica. Qualora la proiezione esista il punto stazionario può essere effettivamente un minimo soltanto se la condizione aggiuntiva è soddisfatta (figura 2.11).

Capitolo 3

Risultati

3.1 Riepilogo meteorologico

Il modello bayesiano proposto, nelle sue diverse varianti, è stato quindi oggetto di una valutazione statistica per le stagioni autunnali dal 2003 al 2005. Le precipitazioni sono state per tutti e tre gli anni superiori al valore climatologico medio. Infatti, se prendiamo come riferimento le tabelle pubblicate sul Riepilogo Meteorologico di Arpa Rivista, è possibile osservare per la stazione di Bologna Borgo Panigale (codice synop 16140) una precipitazione media autunnale di 319,8 mm per il 2003, di 304,2 per il 2004 e di 289,2 mm per il 2005, a fronte di una media climatica stagionale di 214,7 mm.

Nel settembre 2003 il campo di geopotenziale presenta una fascia di anomalie positive, che si estende dal Bassopiano Sarmatico alla Penisola del Labrador, ed un dipolo di segno opposto con i due centri localizzati sull'Oceano Atlantico (figura 3.1). Da questa configurazione si può desumere che la Penisola Italiana sia stata prevalentemente interessata da una circolazione di masse d'aria umida, di origine

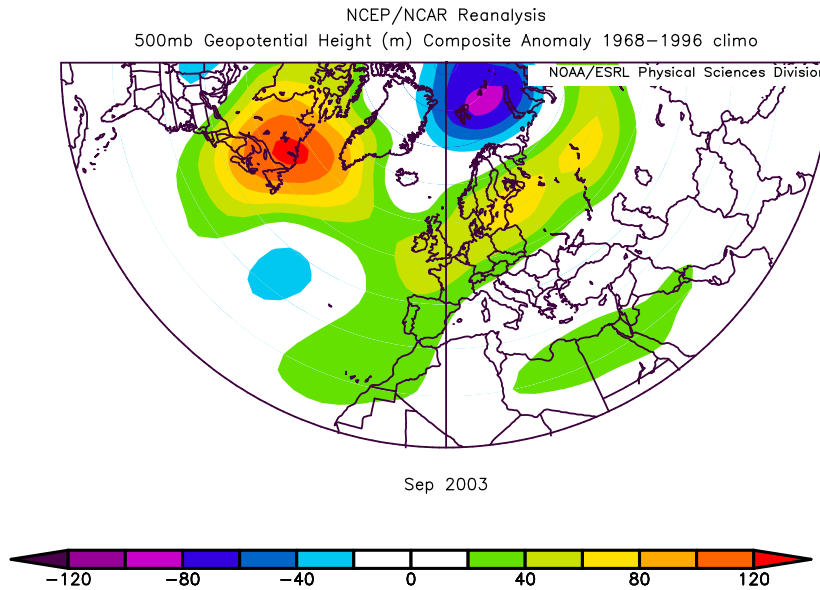


Figura 3.1: Anomalia media dell'altezza di geopotenziale del mese di settembre 2003.

atlantica, che hanno dato a luogo a precipitazioni di debole intensità, al di sotto della media mensile su tutto il territorio regionale. Soltanto nella Provincia di Bologna si sono verificati alcuni eventi temporaleschi.

Ad ottobre la configurazione del campo di anomalia del geopotenziale presenta un quadrupolo, con i centri allineati lungo l'asse (figura 3.2). Il mese è stato piuttosto piovoso, con valori oltre la media del periodo. Le piogge più intense si sono registrate al termine del mese, tra il 29 ed il 31, quando correnti umide ed instabili provenienti da Sud-Ovest hanno interessato il territorio regionale. In particolare, l'occlusione del sistema frontale ha dato luogo a eventi temporaleschi che hanno superato localmente anche i 100 mm in 24 ore lungo il crinale appenninico.

Durante il mese di novembre, caratterizzato da un dipolo per le anomalie di geopotenziale, l'afflusso di correnti umide ed instabili provenienti dal Mediterra-

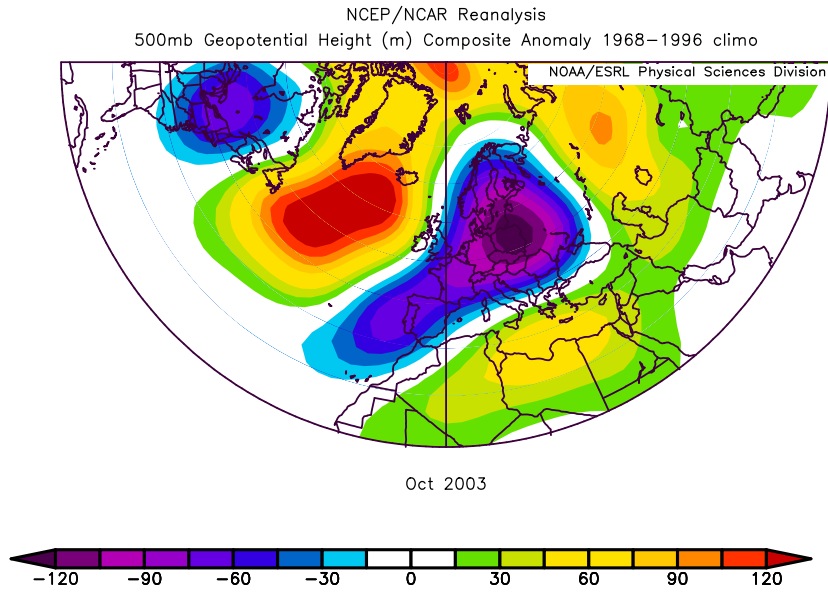


Figura 3.2: Anomalia media dell'altezza di geopotenziale del mese di ottobre 2003.

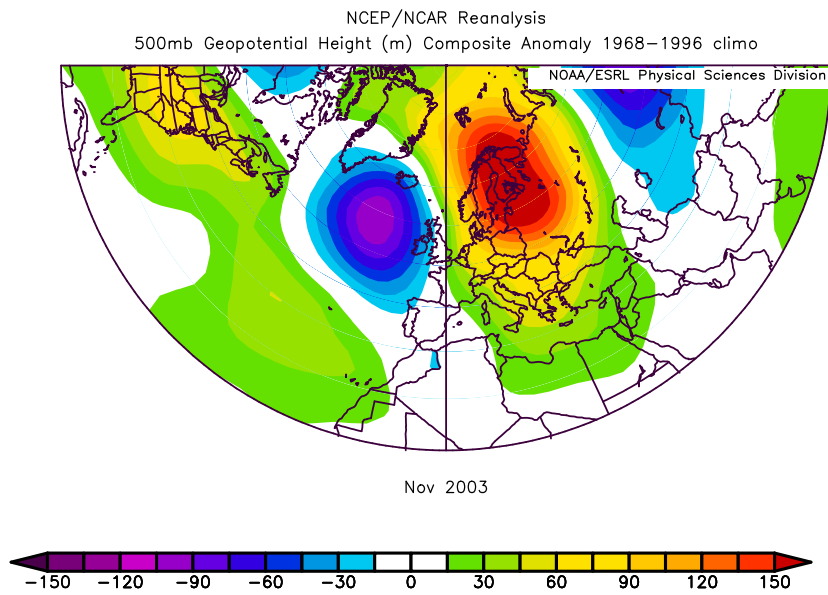


Figura 3.3: Anomalia media dell'altezza di geopotenziale del mese di novembre 2003.

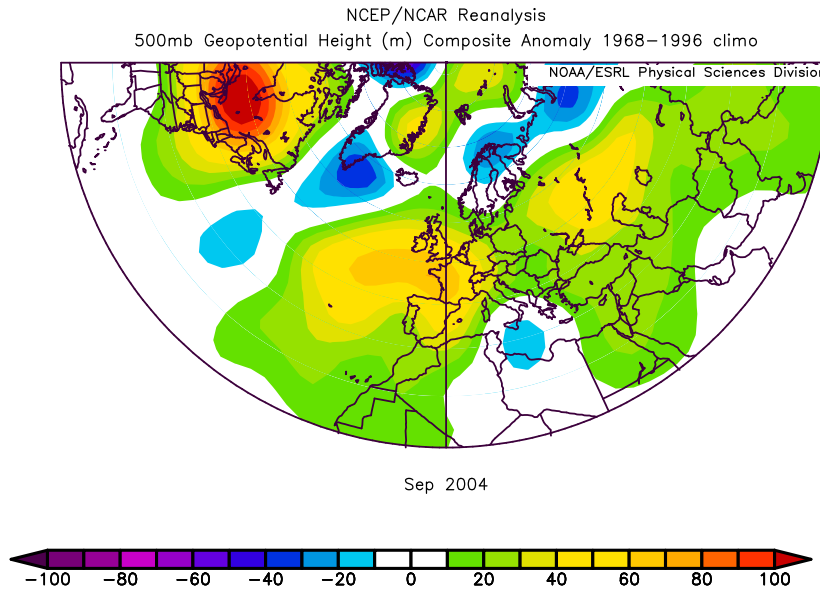


Figura 3.4: Anomalia media dell'altezza di geopotenziale del mese di settembre 2004.

neo, impattando sul versante meridionale della catena appenninica, hanno determinato frequenti precipitazioni, spesso di forte intensità (figura 3.3). In particolare sulla stazione di Montecauto delle Alpi, sull'alto Appennino bolognese si sono registrati 180 mm in 48 ore. A Sestola si sono accumulati nel mese 340 mm contro i 164 mm della media. Nel periodo autunnale i casi di piena, ovvero il superamento del livello arbitrario di soglia di $100 \text{ m}^3/\text{s}$, sono stati tre, e due di questi, presi poi in esame nel paragrafo successivo, hanno raggiunto e superato i $600 \text{ m}^3/\text{s}$.

Il mese di settembre del 2004 è stato contraddistinto da una anomalia positiva centrata a Sud-Ovest delle Isole Britanniche (figura 3.4). Questo ha permesso di avere condizioni di tempo stabile sul territorio regionale ad inizio del periodo e soltanto dopo la seconda decade del mese, quando il blocco originatosi sulla Penisola Iberica si è andato gradualmente a dissolversi, si sono verificati cate deboli

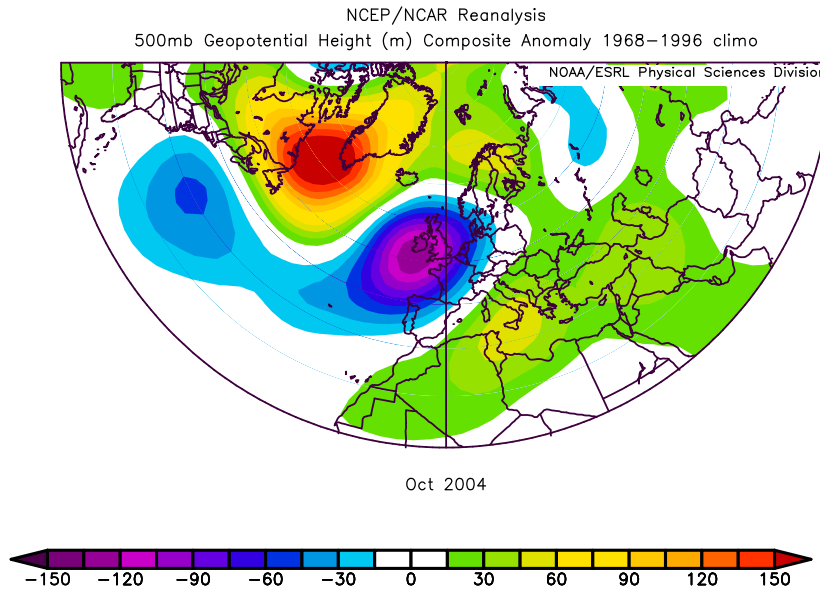


Figura 3.5: Anomalia media dell'altezza di geopotenziale del mese di ottobre 2004.

precipitazioni associate a correnti umide, relativamente instabili, provenienti dal Mediterraneo. Alla fine del mese, l'irruzione di masse d'aria fredda ha dato luogo ad eventi temporaleschi sulle regioni adriatiche della Penisola Italiana.

Ad ottobre la configurazione del campo dell'anomalia di geopotenziale presenta un marcato dipolo, il cui centro positivo è centrato tra la Groenlandia ed il Labrador, mentre quello negativo sulle Isole Britanniche (figura 3.5). Questo polo negativo è associato alla sequenza di onde depressionarie provenienti dall'Atlantico che hanno dato luogo sul territorio regionale a precipitazioni diffuse, spesso di carattere temporalesco. Al termine del mese di ottobre si è verificato il terzo evento di piena analizzato. Il mese di novembre è stato caratterizzato principalmente da una vasta area anticiclonica sull'Atlantico, come si può evincere dalla forte anomalia positiva centrata ad Ovest delle Isole Britanniche (figura 3.6). Le preci-

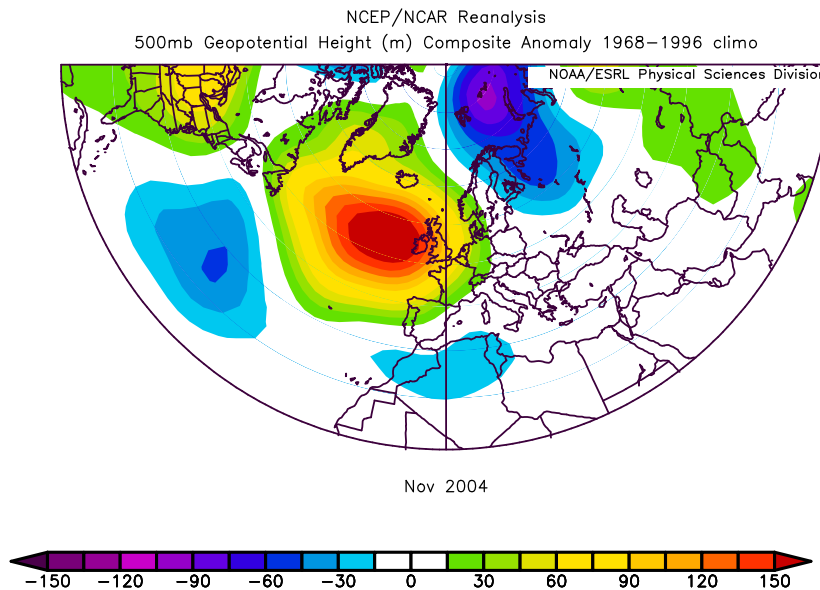


Figura 3.6: Anomalia media dell'altezza di geopotenziale del mese di novembre 2004.

pitazioni sono state relativamente scarse e concentrate prevalentemente al termine del mese. Anche per l'autunno 2004 i casi di piena sono stati tre.

L'autunno 2005 è stato quello meno piovoso del periodo preso in esame. Il campo del geopotenziale presenta una forte anomalia positiva, centrata sull'Europa centro-orientale (figura 3.7). Nel mese di settembre la scena meteorologica è dominata da condizioni di tempo stabile, fatto salvo due episodi, nei quali si sono registrate precipitazioni, anche a carattere di rovescio o temporalesco, soprattutto sulla parte occidentale della Regione all'inizio del mese e sulla Romagna nelle giornate del 17 e del 18.

La prima decade del mese di ottobre è caratterizzata da condizioni di forte instabilità, con piogge diffuse e persistenti, anche a carattere temporalesco. Le precipitazioni mensili sono state superiori alla media climatica e questo è coerente

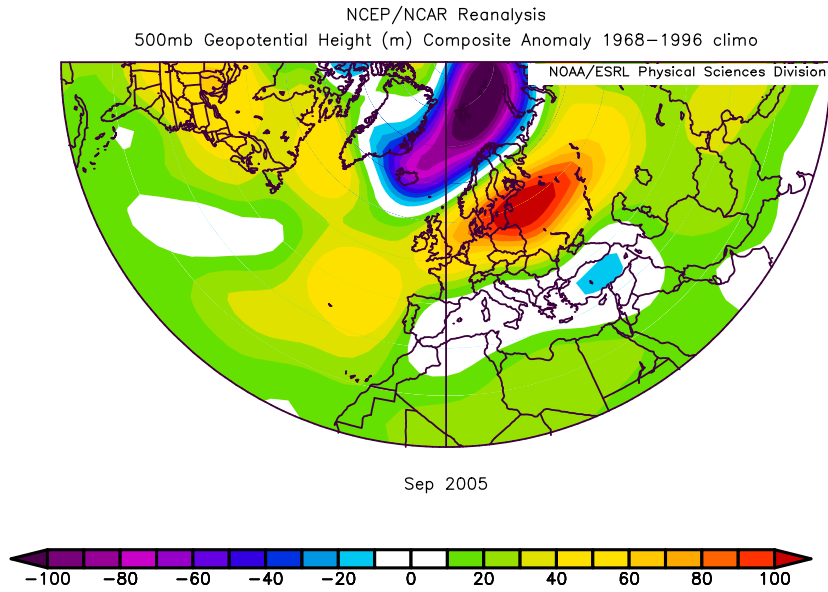


Figura 3.7: Anomalia media dell'altezza di geopotenziale del mese di settembre 2005.

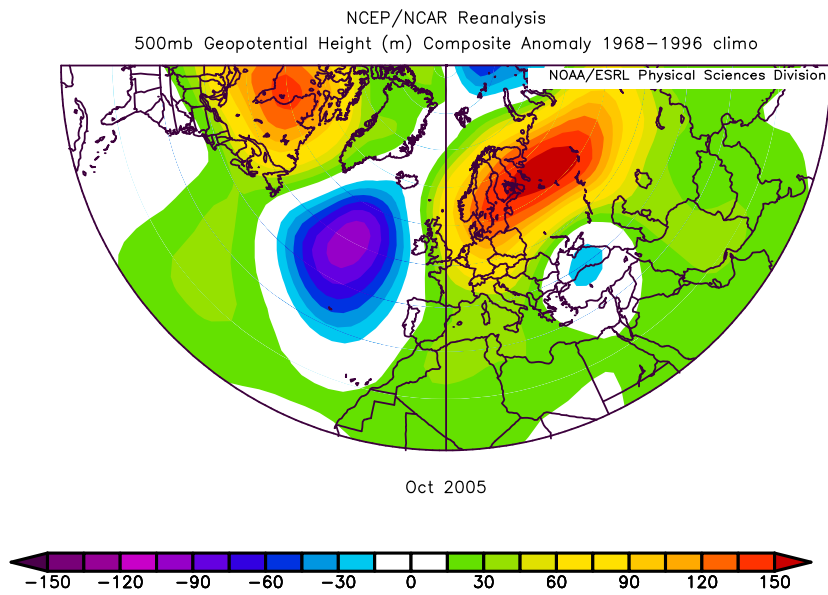


Figura 3.8: Anomalia media dell'altezza di geopotenziale del mese di ottobre 2005.

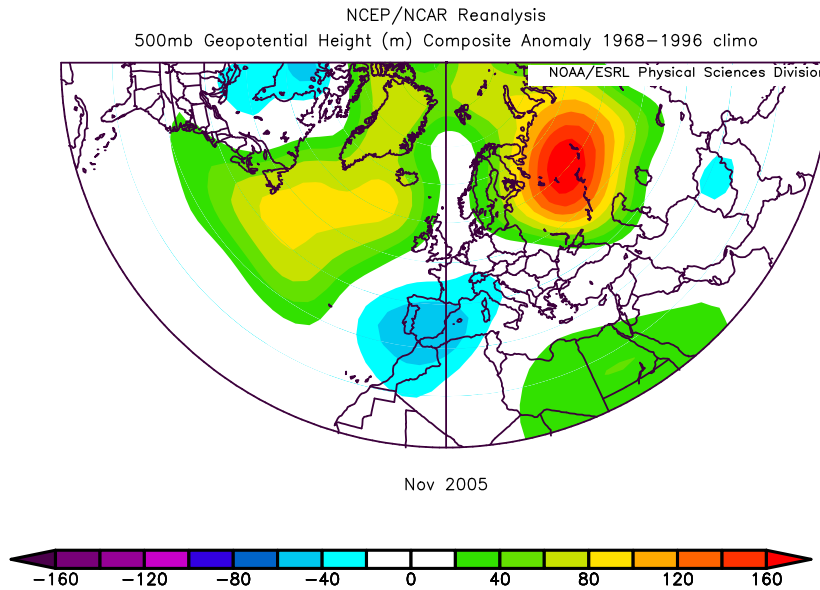


Figura 3.9: Anomalia media dell'altezza di geopotenziale del mese di novembre 2005.

con quanto si può osservare dalla mappa di anomalia del geopotenziale, nella quale è presente una struttura dipolare, con il polo positivo, più marcato, centrato sul Bassopiano Sarmatico, ed il polo negativo sul medio Atlantico (figura 3.8).

Nel mese di novembre si assiste alla persistenza dell'anomalia positiva sul Bassopiano Sarmatico, mentre un centro negativo si instaura sulla Penisola iberica (figura 3.9). Questa configurazione è associata ad una scena meteorologica caratterizzata da precipitazioni frequenti, particolarmente sul settore centrale e su quello orientale della Regione. Il numero delle giornate piovose a Bologna è stato di 10 contro le 7 di norma per il periodo. Abbondanti precipitazioni sono state osservate anche nell'ultima decade del mese, anche se hanno interessato più la parte orientale. In particolare, la pioggia caduta, associata allo scioglimento del manto nevoso sui rilievi dell'Appennino Tosco-Romagnolo, ha determinato un sensibile

innalzamento dei livelli idrometrici dei fiumi romagnoli, con particolare rilevanza per il fiume Senio. Per tutto il periodo autunnale 2005 gli eventi di piena osservati alla chiusa di Casalecchio sono stati 4, ma nessuno di essi ha superato i $300 \text{ m}^3/\text{s}$.

3.2 Analisi statistica

Il periodo preso in esame è stato più piovoso della norma del periodo. Tuttavia, il numero dei casi di piena registrati alla chiusa di Casalecchio per il sottobacino montano del fiume Reno sono stati relativamente pochi. Questa considerazione è importante per sottolineare che la dinamica delle piene, in un fiume a prevalente carattere torrentizio, come appunto il fiume Reno nella sua parte montana, sia un fenomeno raro, estremamente complesso da descrivere e da prevedere, poiché nella realizzazione del quale concorrono una quantità notevole di aspetti di carattere meteorologico ed idrologico. Non sempre una correlazione significativa tra le precipitazioni nel bacino e la portata misurata può essere sufficiente per prevedere l'evento di piena con un buon grado di accuratezza. Infatti, si è riscontrato che ad eventi meteorologici analoghi, non sempre è corrisposto un analogo evento di piena (dati non mostrati).

La comunità scientifica ha compiuto uno sforzo notevole per sviluppare famiglie di modelli idrologici deterministici. In questa tesi si è cercato di combinare un modello autoregressivo, basato soltanto sulle correlazioni statistiche delle serie temporali, ed il modello distribuito, ad afflussi–deflussi TOPKAPI, operativo presso il Servizio Idrometeorologico di ARPA Emilia–Romagna. La combinazione è stata effettuata in un'ottica bayesiana, secondo la quale la stima del modello autoregressivo viene considerata come stima a priori, mentre quella del modello

TOPKAPI come funzione di verosimiglianza. Il teorema di Bayes viene utilizzato per ottenere la stima a posteriori, che dovrebbe essere caratterizzata da una varianza dell'errore minore. In linea di principio, come visto nell'esempio del capitolo precedente, se la struttura dell'errore fosse ben descritta dalle equazioni del modello, allora la varianza campionaria non dovrebbe essere significativamente diversa dalla varianza teorica.

In prima istanza sono state effettuate delle simulazioni con il modello idrologico TOPKAPI, nelle quali si è introdotto come variabile di ingresso la precipitazione osservata dalla rete pluviometrica del bacino. Data la limitazione del campione, composto da 273 record giornalieri, si è pensato di utilizzare la tecnica statistica di ricampionamento, denominata bootstrap. Nell'adottare questa tecnica, che potrebbe essere descritta in maniera sommaria come un metodo non parametrico di campionamento con ripetizione, ogni volta sono stati realizzati due insiemi di dati, composti ciascuno rispettivamente da 100 records giornalieri scelti a caso. Il primo insieme viene utilizzato per calcolare i coefficienti di autocorrelazione della serie, mentre dal secondo si calcolano le proprietà statistiche degli errori: la media, la varianza ed il valore massimo. Il campionamento con ripetizione ha permesso di non dividere il campione in due parti, che difficilmente potevano avere le medesime proprietà statistiche, trattandosi di tre serie storiche non omogenee. La procedura di ricampionamento viene poi ripetuta per altrettante 100 volte.

Nella figura 3.10 la stima a posteriori è stata ottenuta tramite il filtro H^∞ stazionario, in cui gli errori sono stati trattati come se fossero non autocorrelati. Si può osservare in alto che le stime non sono significativamente deviate ($\alpha = 0.05$). Nella figura al centro la varianza campionaria dell'errore della stima a priori presenta una oscillazione periodica ogni 24 ore, ma si accorda in media con

la varianza teorica del modello autoregressivo. Questa periodicità giornaliera è dovuta al fatto che l'errore del modello si concentra soprattutto nei valori estremi. L'errore della simulazione con il modello TOPKAPI è costante in media. La varianza campionaria dell'errore della stima a posteriori sovrastima la varianza teorica calcolata con il filtro. L'errore massimo della stima a posteriori è minore per quasi tutte le scadenze temporali (in basso).

L'utilizzo di un filtro non stazionario, o adattativo, pur riducendo la varianza teorica data dalle equazioni del filtro, introduce, tuttavia, un aumento della varianza campionaria dell'errore dopo 6 ore (figura 3.11 al centro). Anche l'errore massimo è maggiore del corrispondente errore massimo del filtro di riferimento (in basso). Gli errori medi associati alle stime sono non devianti (in alto).

Il filtro con il modello autoregressivo per l'errore permette di ridurre la varianza della errore della stima a posteriori, come risulta dalla figura 3.12 al centro. Con questa variante del modello anche l'errore massimo risulta più contenuto (in basso). La media campionaria degli errori, inoltre, non è significativamente diversa da zero anche per un livello di significatività di $\alpha = 0.01$ (in alto). In maniera analoga a quanto ottenuto per il filtro senza modellazione dell'errore, l'utilizzo del filtro adattativo introduce un peggioramento della varianza campionaria della stima oltre le 6 ore di previsione (figura 3.13 al centro). L'errore massimo è superiore in misura ancora maggiore (in basso).

Per compiere l'analisi statistica degli errori di previsione si è di nuovo utilizzata la tecnica statistica del bootstrap. A differenza di quanto osservato per la figura 3.10, la varianza dell'errore della stima effettuata con il modello idrologico TOPKAPI non è più costante in media, ma tende a crescere linearmente. Allo stesso modo, anche la varianza dell'errore della stima a posteriori non tende più

Casalecchio chiusa
(simulazione valori autunnali 2003-2005)

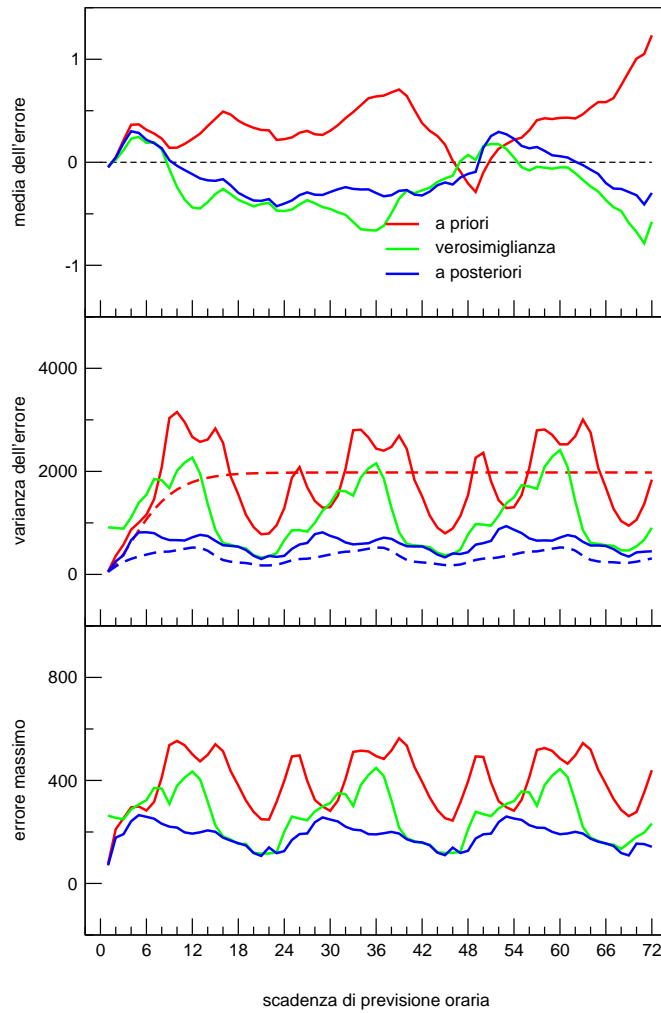


Figura 3.10: Medie (in alto), varianze (al centro) e valori massimi (in basso) dell'errore nella stima del processo di piena con un modello autoregressivo (in rosso, linea continua valore osservato, linea tratteggiata valore teorico), dell'errore nella simulazione del modello idrologico TOPKAPI (in verde), dell'errore della stima a posteriori (in blu).

Casalecchio chiusa
(simulazione valori autunnali 2003-2005)

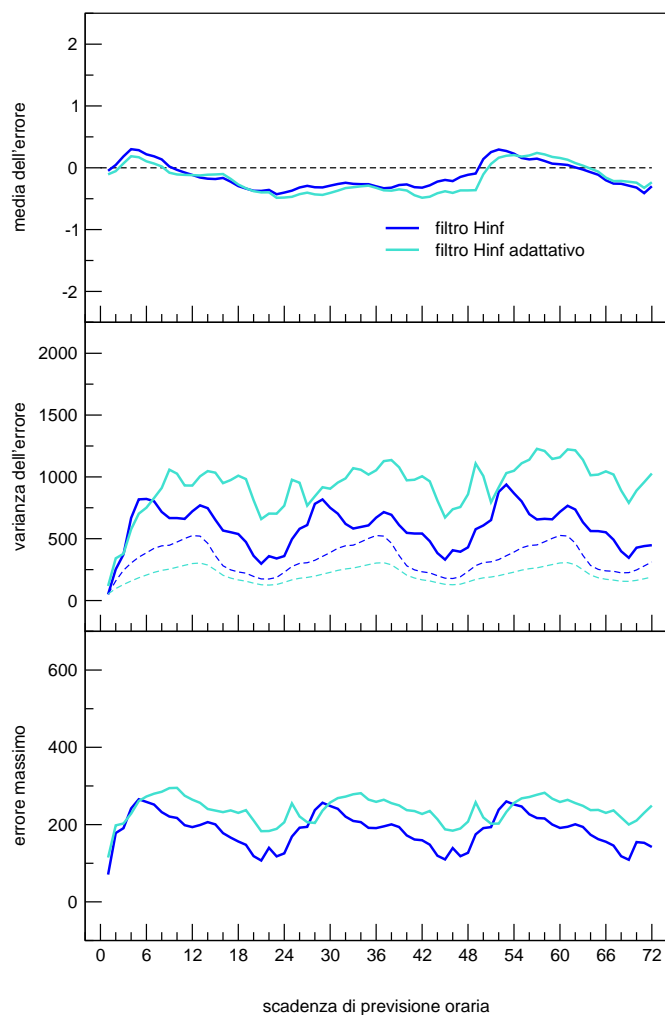


Figura 3.11: Medie (in alto), varianze (al centro) e valori massimi (in basso) dell'errore della stima a posteriori ottenuta con il filtro H^∞ stazionario (in blu) e dell'errore della stima a posteriori ottenuta con il filtro H^∞ adattativo (linea turchese). Le linee tratteggiate si riferiscono ai valori di varianza teorici.

Casalecchio chiusa
(simulazione valori autunnali 2003-2005)

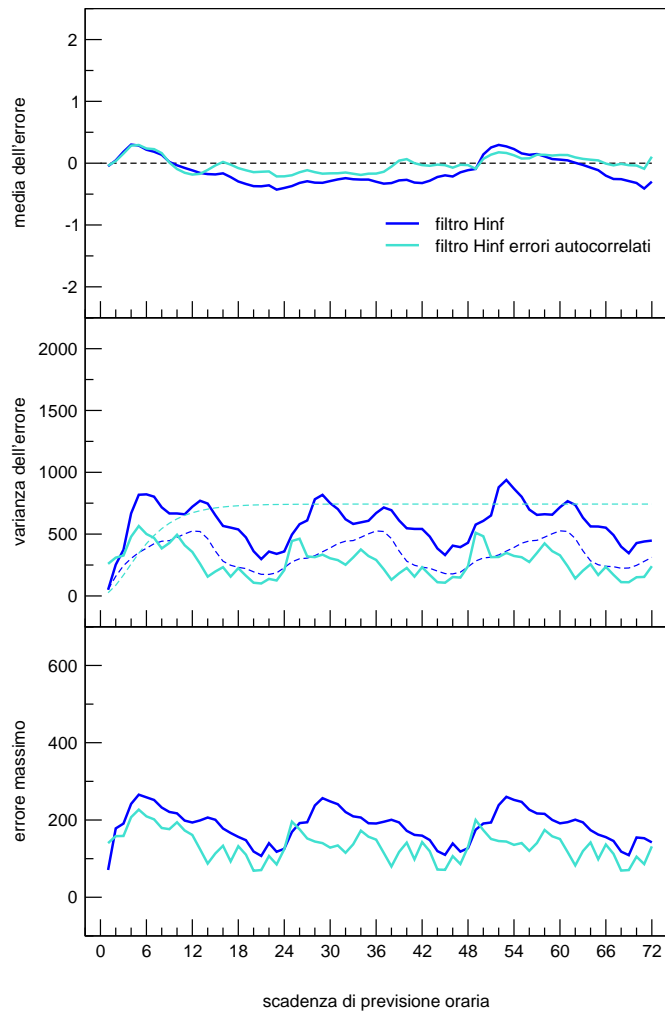


Figura 3.12: Idem della figura 3.11 per gli errori del filtro stazionario (in blu) e gli errori del filtro stazionario con gli errori autocorrelati (in turchese).

Casalecchio chiusa
(simulazione valori autunnali 2003-2005)

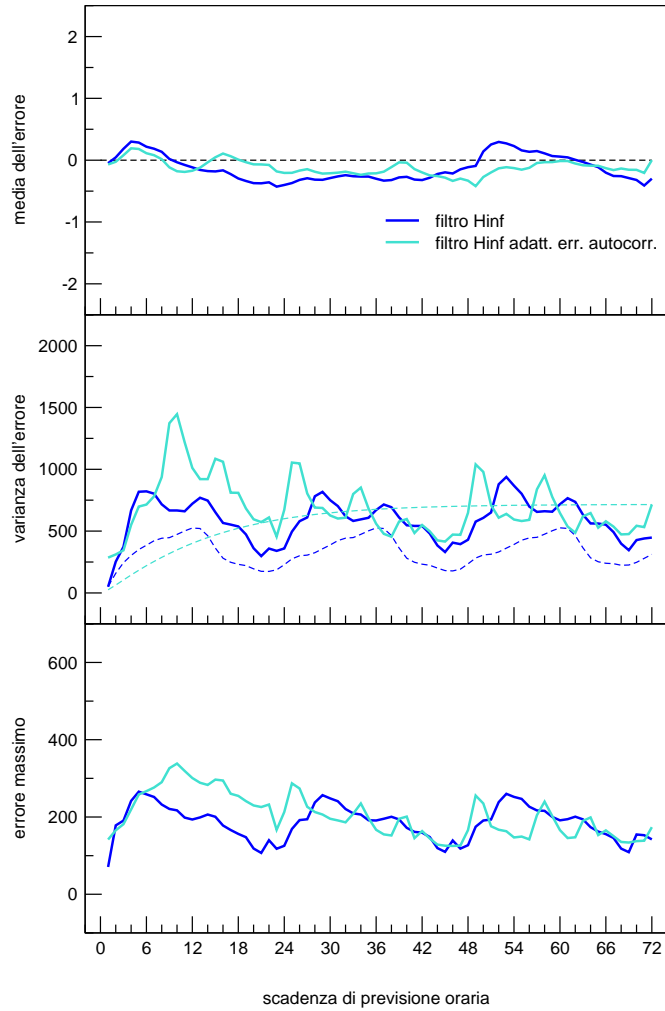


Figura 3.13: Idem della figura 3.11 per gli errori del filtro stazionario (in blu) e gli errori del filtro adattativo con gli errori autocorrelati (in turchese).

ad un asintoto ma segue la crescita della varianza del modello. Anche in questo caso si osserva uno scostamento significativo tra la varianza campionaria e la varianza teorica dell'errore della stima a posteriori (figura 3.14). Gli errori massimi riportati nella figura in basso non appaiono essere significativamente diversi.

In questo caso la stima a posteriori del filtro adattativo ha una varianza dell'errore che non differisce dalla varianza della corrispondente stima ottenuta con il filtro stazionario, sebbene venga ridotto il suo valore massimo (figura 3.15). Il modello con gli errori autocorrelati non introduce un miglioramento della varianza dell'errore della stima a posteriori come per le simulazioni. Infatti, come risulta evidente dalla figura 3.16 al centro, la varianza è simile a quella ottenuta dal modello precedente. Nemmeno il filtro adattativo, nel caso in cui si utilizzino i dati previsti di precipitazione, migliora la stima a posteriori. Infatti, come si può osservare dalla figura 3.17 al centro, nelle prime 36 ore di integrazione la varianza campionaria è sempre superiore alla corrispondente varianza campionaria ottenuta con il filtro stazionario.

3.3 Analisi degli eventi di piena

3.3.1 7–8 novembre 2003

L'evento si svolge dal 5 all'8 novembre 2003. Questo periodo è caratterizzato da precipitazioni diffuse, anche a carattere convettivo, che hanno interessato il sottobacino montano del fiume Reno.

Per descrivere, in maniera sintetica, le cause meteorologiche della precipitazione caduta occorre prendere in esame le mappe di geopotenziale e di tempera-

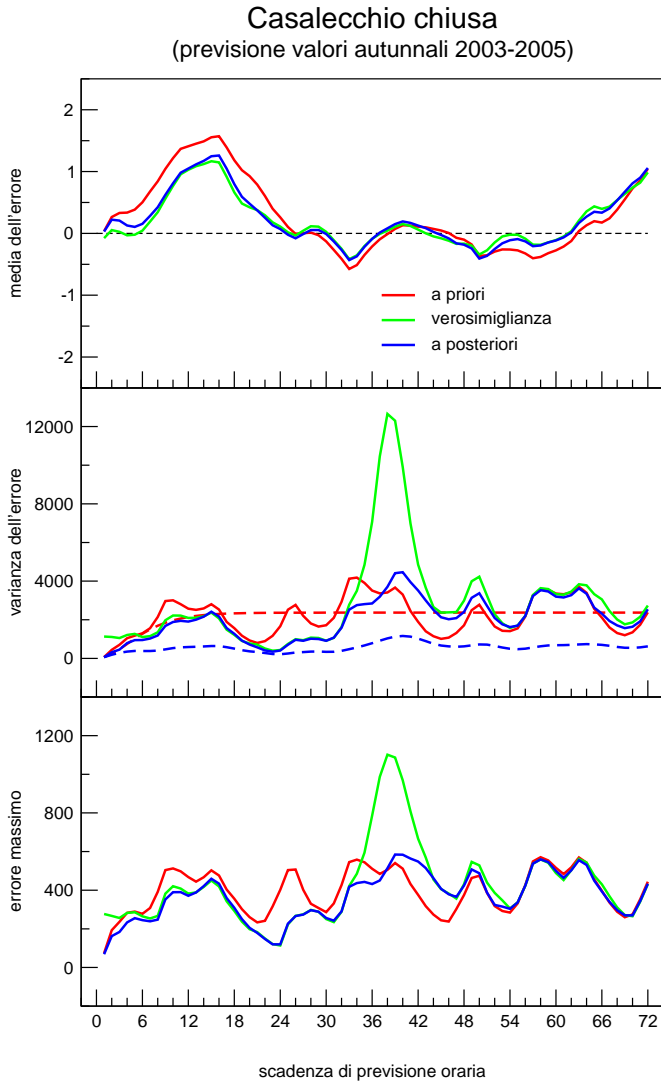


Figura 3.14: Idem della figura 3.10, per gli errori delle previsioni di portata.

Casalecchio chiusa
(previsione valori autunnali 2003-2005)

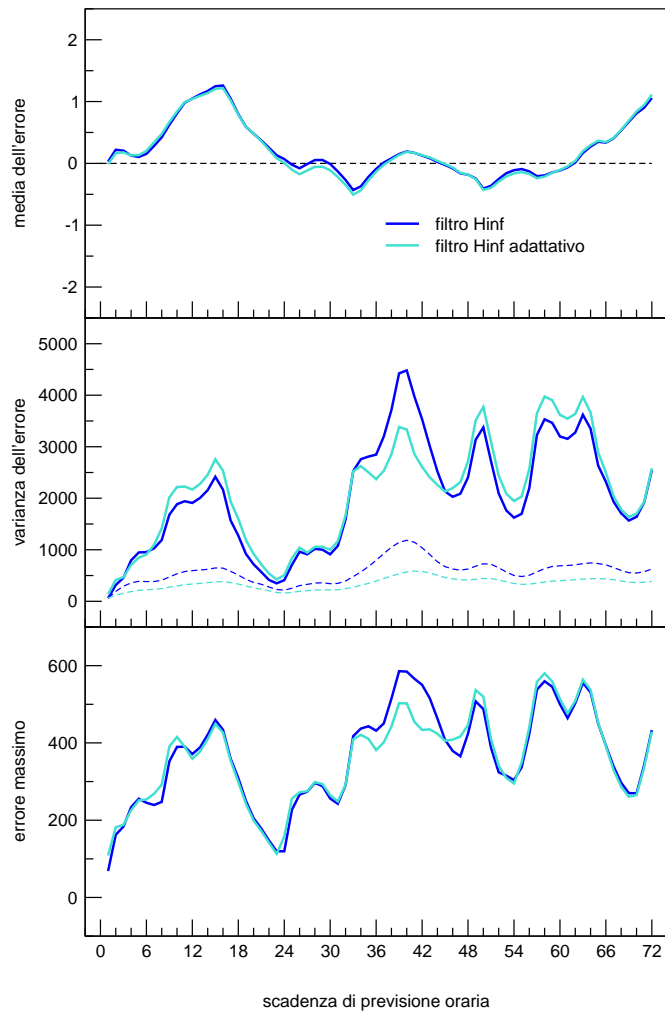


Figura 3.15: Idem della figura 3.11 per gli errori nelle previsioni di portata.

Casalecchio chiusa
(previsione valori autunnali 2003-2005)

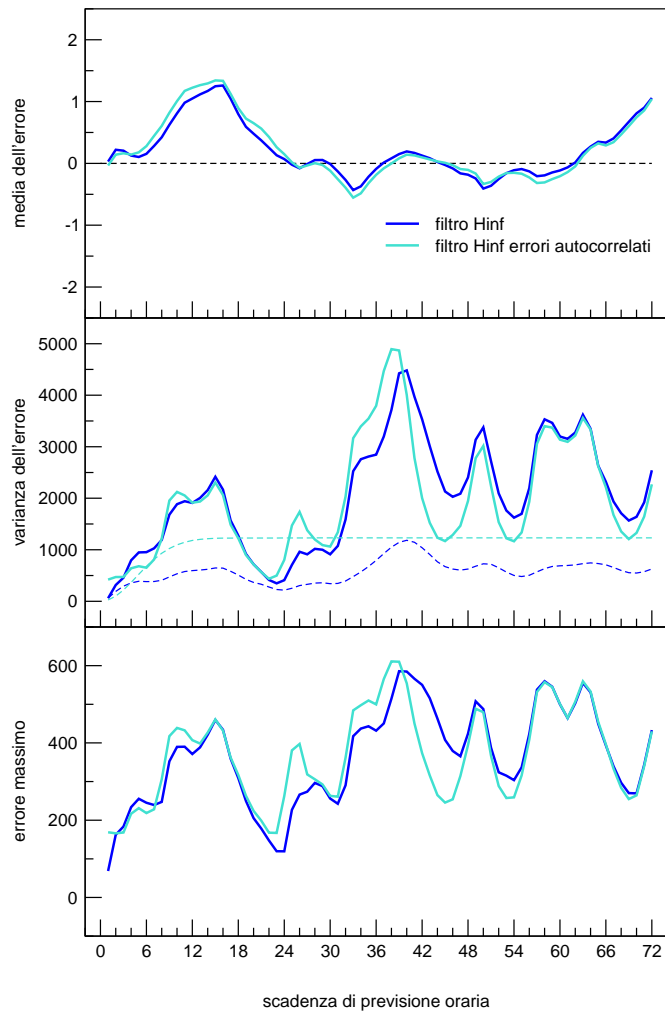


Figura 3.16: Idem della figura 3.12 per gli errori nelle previsioni di portata.

Casalecchio chiusa
(previsione valori autunnali 2003-2005)

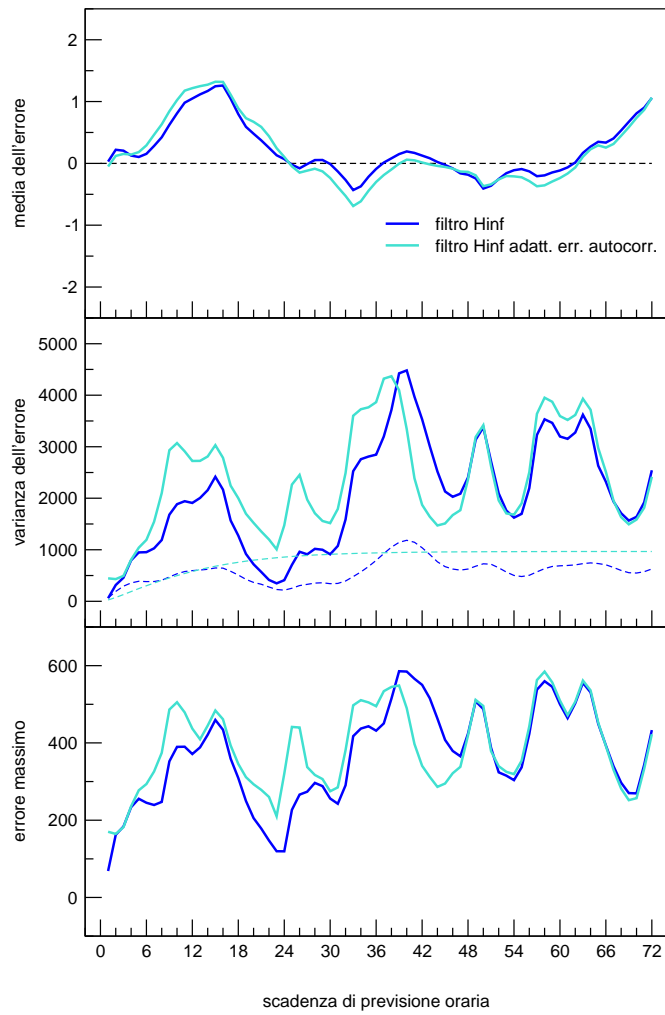


Figura 3.17: Idem della figura 3.13 per gli errori nelle previsioni di portata.

tura a 500 hPa. Nella figura 3.18, relativa alla situazione sinottica del 5 novembre 2003 sull'area Europea, definita da 28.5°N a 55.5°N e da 11.5°W a 31°E, si evidenziano due zone distinte. La prima, sulla sinistra della figura, caratterizzata da un promontorio, che dal Sud della Penisola iberica si protende oltre le Isole Britanniche, mentre la seconda si contraddistingue per la presenza di un nucleo di aria fredda, che in maniera obliqua si localizza dai Carpazi orientali fino al Bassopiano Sarmatico. La separazione delle due masse d'aria è bene evidenziata in figura dalla scala di colori con cui è riportata la temperatura. Associate al promontorio si possono infatti individuare masse d'aria di origine marittima, più calde ed umide, mentre sul bacino centrale del Mediterraneo, si distingue la presenza di masse d'aria fredda di origine continentale, con un nucleo di aria ancora più fredda, con forti caratteristiche di barotropicità, bene evidenziato in figura dal colore blu.

Nella giornata successiva la progressiva estensione del promontorio determina una rotazione ed un movimento retrogrado verso Sud-Ovest del nucleo di aria fredda, che comincia ad interessare la Penisola balcanica (figura 3.19).

Nella giornata del 7 novembre 2003 il coricamento del campo di anticlone sulla Penisola scandinava, purtroppo non visibile dalla figura, ha indotto nel nucleo di aria fredda un movimento retrogrado verso Ovest ed un'ulteriore rotazione, disponendosi longitudinalmente nella fascia che va dalla regione del Massiccio Centrale fino a Penisola Balcanica, attraversando tutta la catena montuosa alpina (figura 3.20). È da questo momento che il bacino del fiume Reno viene ad essere interessato da precipitazioni diffuse, a seguito dell'instabilità nei bassi strati dovuta all'intrusione di aria fredda.

Nella giornata dell'8 novembre, il nucleo freddo, con il suo movimento re-

ECMWF Analysis VT:Wednesday 5 November 2003 12UTC 500hPa temperature/ geopotential height

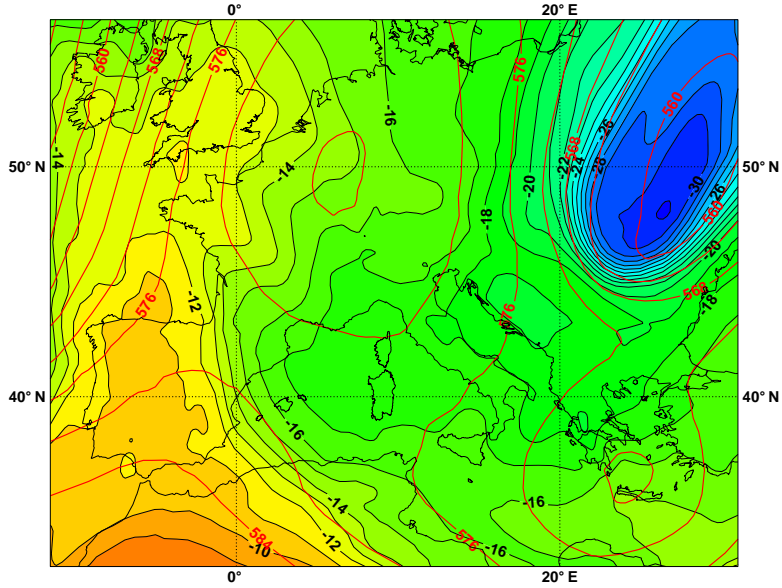


Figura 3.18: campo di geopotenziale e di temperatura a 500 hPa del 5 novembre 2003 12 UTC.

ECMWF Analysis VT:Thursday 6 November 2003 12UTC 500hPa temperature/ geopotential height

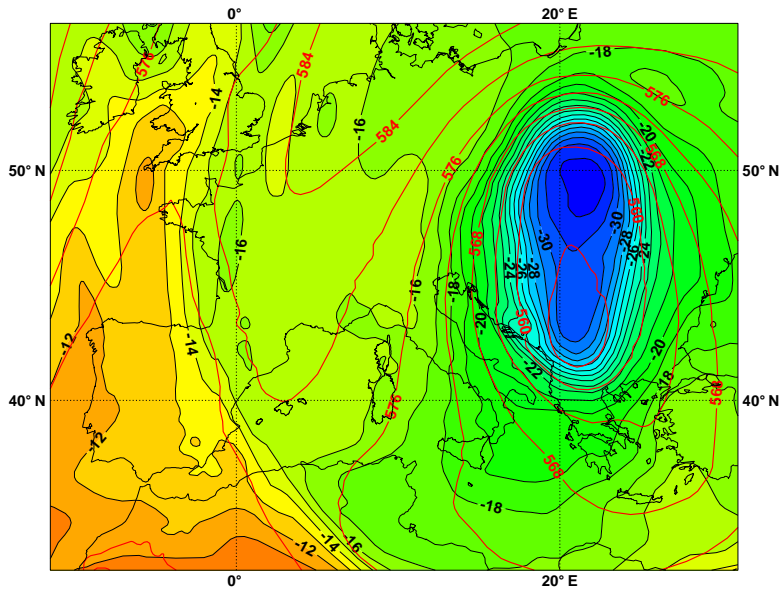


Figura 3.19: campo di geopotenziale e di temperatura a 500 hPa del 6 novembre 2003 12 UTC.

ECMWF Analysis VT:Friday 7 November 2003 12UTC 500hPa temperature/ geopotential height

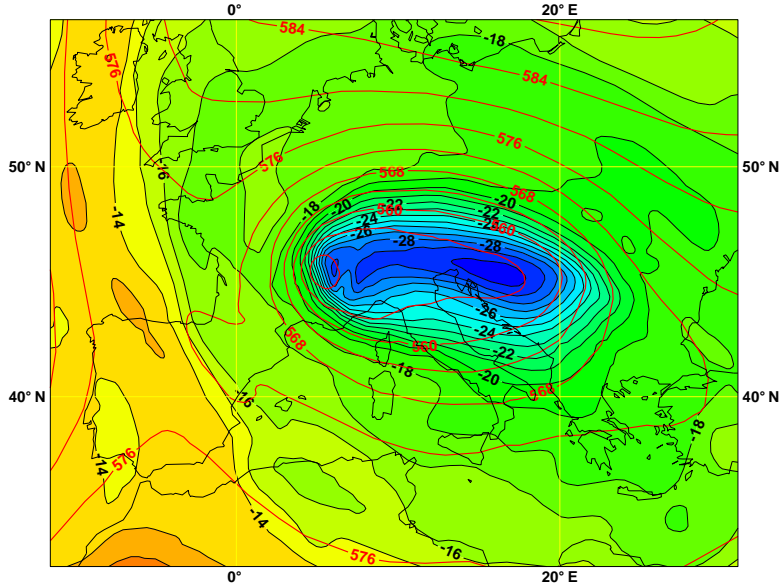


Figura 3.20: campo di geopotenziale e di temperatura a 500 hPa del 7 novembre 2003 12 UTC.

ECMWF Analysis VT:Saturday 8 November 2003 12UTC 500hPa temperature/ geopotential height

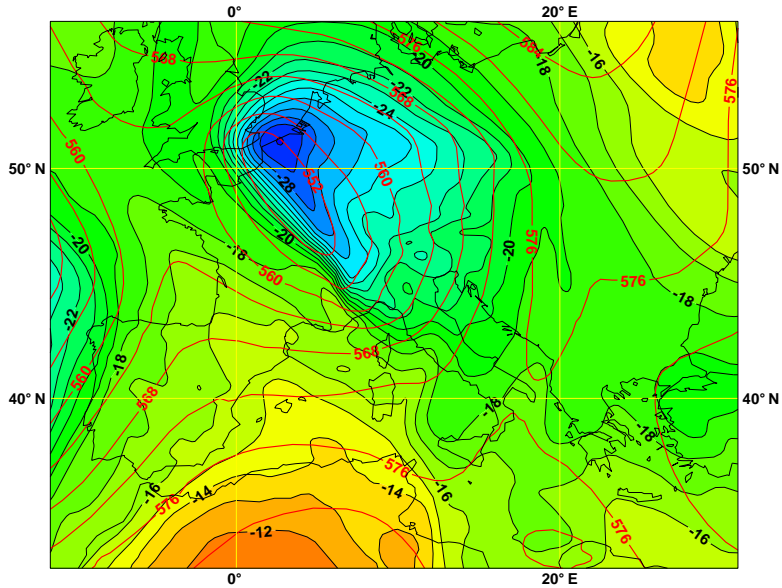


Figura 3.21: campo di geopotenziale e di temperatura a 500 hPa dell'8 novembre 2003 12 UTC.

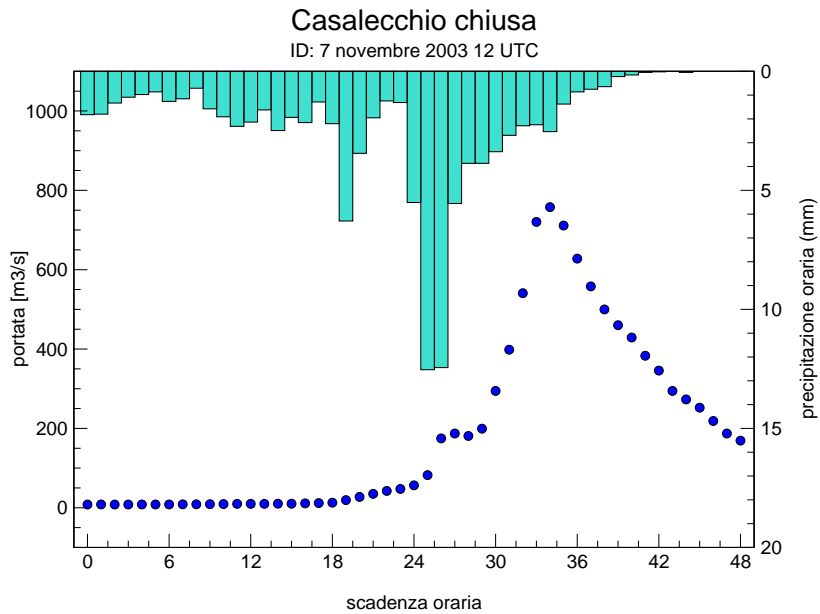


Figura 3.22: Valori di portata osservati a partire dal 7 novembre 2003 12 UTC (cerchi blu) e valore medio della precipitazione oraria caduta nel sottobacino Reno monte.

trogado e la sua ulteriore rotazione, oltre ad interrompere il promontorio che si estendeva dalla Penisola Iberica alla Penisola scandinava, comincia ad assumere anche caratteristiche barocline soprattutto sul suo lato occidentale, determinando così un aumento dell'instabilità sul Nord della Penisola Italiana (figura 3.21). Le precipitazioni diventano più intense ed assumono un maggiore carattere convettivo, come si può osservare nella figura 3.22, nella notte tra il 7 e l'8 novembre, in corrispondenza a +24 ore nella scala temporale del grafico, quando si raggiungono i valori massimi di intensità oraria, con una media areale di 15 mm. La figura riporta inoltre anche il grafico idrometrico della portata rilevata alla chiusa di Casalecchio, dove l'onda di piena si verifica tra +30 e +36 ore, con un valore massimo di circa $800 \text{ m}^3/\text{s}$. Dal diagramma si deduce anche il tempo di corrivazione del

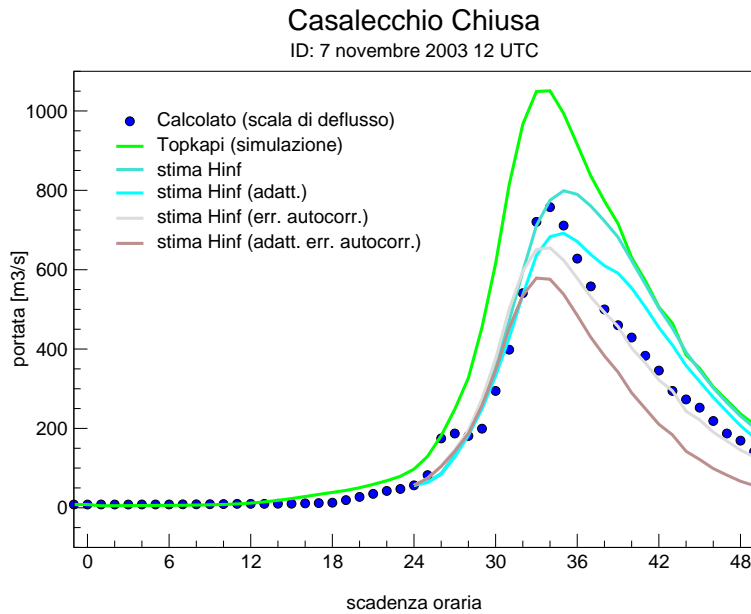


Figura 3.23: Valori di portata osservati a partire dal 7 novembre 2003 12 UTC (cerchi blu), valori simulati dal modello idrologico TOPKAPI utilizzando le precipitazioni osservate dalla rete pluviometrica (linea continua verde) e stime a posteriori con il filtro H^∞ .

sottobacino montano dell'ordine di 6–8 ore. In questo caso il terreno si trovava in condizioni sature, a seguito delle precipitazioni che avevano già interessato il bacino nelle giornate precedenti.

Nella figura 3.23 sono riportate le curve relative alla correzione con il modello bayesiano, nelle sue diverse varianti, alle stime della portata ottenute con il modello autoregressivo e con il modello TOPKAPI fornendo in ingresso le precipitazioni osservate. La simulazione del TOPKAPI (linea verde) sovrastima la portata osservata. Tutte le stime ottenute con il filtro H^∞ riescono a descrivere bene la curva di risalita della portata. Tuttavia, la portata stimata dal filtro H^∞ stazionario (linea turchese), se da un lato riesce a correggere bene il valore mas-

Casalecchio Chiusa

ID: 8 novembre 2003 12 UTC

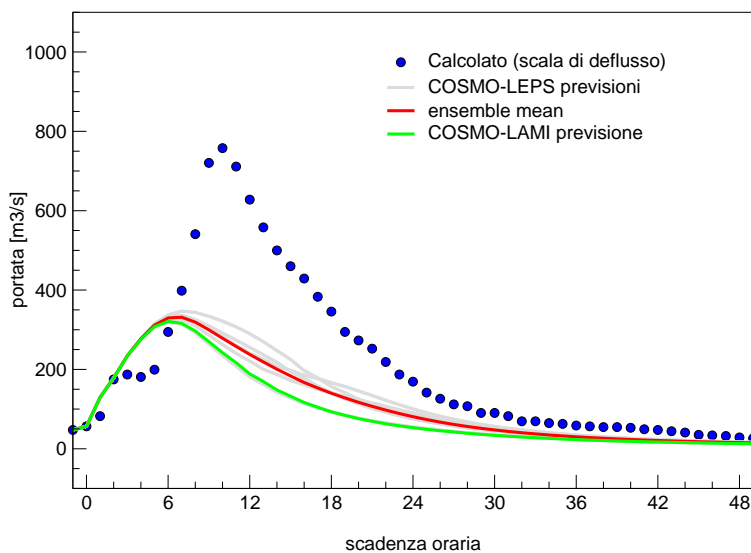


Figura 3.24: Previsioni di portata ottenute con il modello idrologico TOPKA-PI utilizzando le piogge previste dal modello COSMO-LAMI (linea continua verde) e dagli elementi di COSMO-LEPS (linee continue grigie). La media dell'ensemble idrologico è indicata dalla linea continua rossa.

simo della portata, dall'altro lo sposta leggermente avanti nel tempo e tende ad allungare il tempo di deflusso della piena. Le altre varianti sottostimano tutte il livello massimo di portata, ma il fitro con l'errore autocorrelato (linea grigia) riesce a descrivere molto bene anche la fase di deflusso.

Ben diverse sono invece le curve idrometriche di portata calcolate con le precipitazioni previste dal modello COSMO-LAMI e dagli elementi dell'ensemble COSMO-LEPS (figura 3.24). C'è una generale sottostima dell'evento di piena determinato da una sottovalutazione dell'intensità dell'evento meteorologico. Come si può osservare dalla figura viene prevista una portata massima di circa $400 \text{ m}^3/\text{s}$ a +6 ore dall'istante iniziale di riferimento, con un anticipo dell'evento di circa 6

ore e con una sottostima della portata di circa $400 \text{ m}^3/\text{s}$.

Nelle figure successive viene riportato come può essere utile il modello bayesiano in un'ottica operativa ai fini di una previsione a breve termine, rispettivamente a 4, 2 e zero ore dal colmo di piena (figure 3.26, 3.27 e 3.28). Risulta evidente come vengano aggiornate le previsioni dei modelli man mano che ci si approssima all'evento, utilizzando il filtro adattativo con gli errori autocorrelati. In particolare si può osservare come a due ore dall'evento e partendo da una previsione sbagliata, si riesca a prevedere un colmo di piena di circa $700 \text{ m}^3/\text{s}$ con un elemento di COSMO-LEPS. In corrispondenza del colmo di piena il modello correttamente non prevede un ulteriore aumento di portata. Il tempo di deflusso è abbastanza corretto sebbene la forma d'onda sia diversa.

3.3.2 27–28 novembre 2003

La dinamica meteorologica relativa a questo secondo caso di studio è completamente diversa. L'analisi si concentra principalmente sulla saccatura che dalla giornata del 26 novembre comincia ad estendersi dalle Isole Britanniche con asse in direzione Sud-Est verso il Continente Europeo (figura 3.29).

Il giorno successivo, a seguito del protendersi del promontorio sull'Oceano Atlantico, la saccatura trasla verso Est e si approfondisce fino a raggiungere la Penisola Iberica (figura 3.30). La Penisola Italiana comincia ad essere interessata da un flusso sud-occidentale di masse d'aria relativamente calde ed umide, che danno luogo a deboli precipitazioni, soprattutto sulle Regioni tirreniche.

L'onda depressionaria tende a separarsi in due parti come arriva in prossimità della catena alpina. La parte meridionale, ben visibile nella figura 3.31 per il

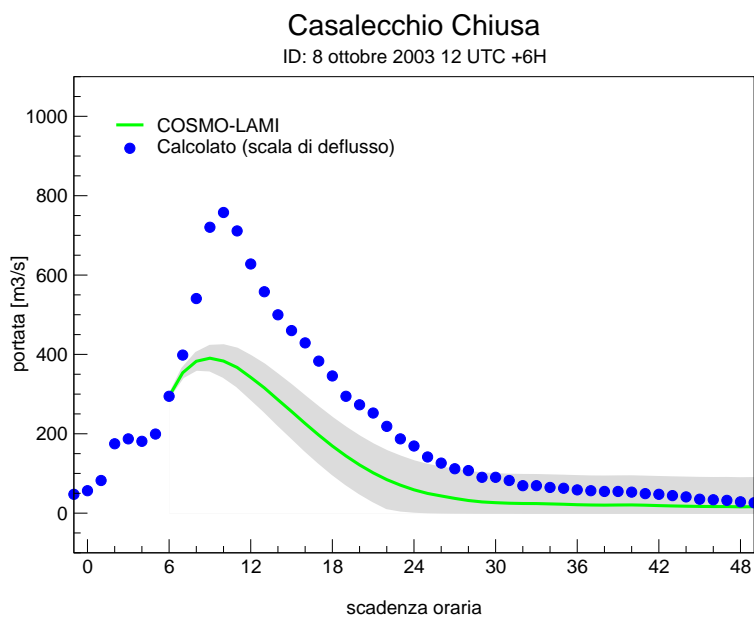


Figura 3.25: Previsioni di portata ottenute con il filtro H^∞ ad orizzonte infinito dopo 6 ore: stima a posteriori ottenuta con il modello idrologico TOPKAPI e le piogge previste dal modello COSMO-LAMI (linea continua verde). L'area colorata di grigio si riferisce all'intervallo di confidenza del 95%.

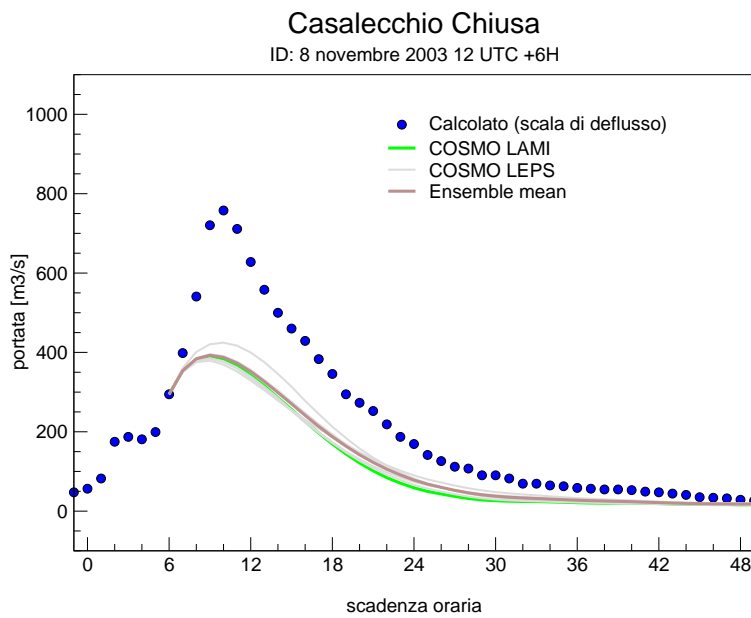


Figura 3.26: Previsioni di portata ottenute con il filtro H^∞ ad orizzonte infinito dopo 6 ore: stima a posteriori ottenuta con il modello idrologico TOPKAPI e le piogge previste dal modello COSMO-LAMI (linea continua verde) e dagli elementi di COSMO-LEPS (linee continue grigie) come funzione di verosimiglianza. La media dell'ensemble è indicata dalla linea continua marrone.

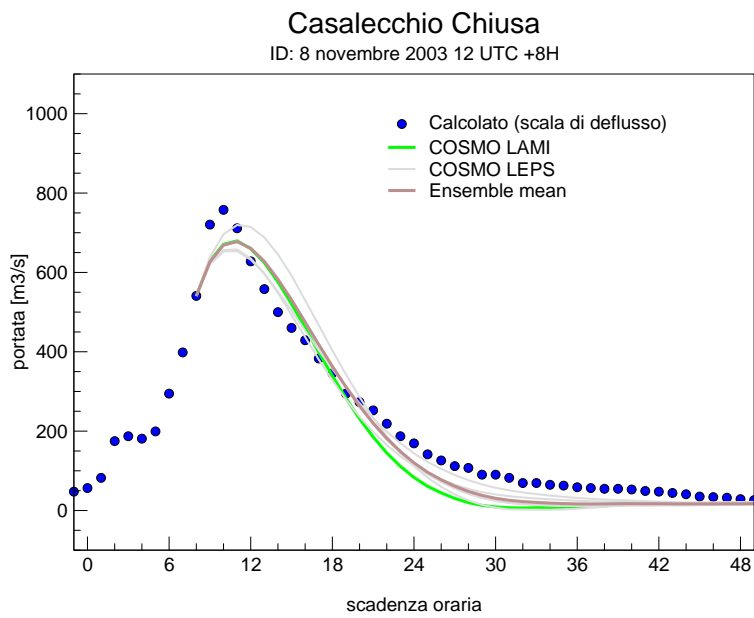
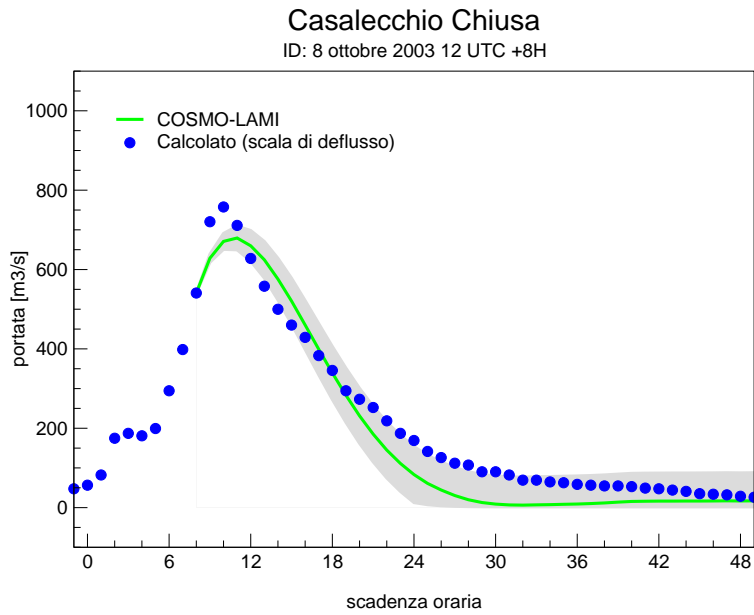


Figura 3.27: Idem per la figura 3.26, dopo 8 ore dall'inizio della previsione.

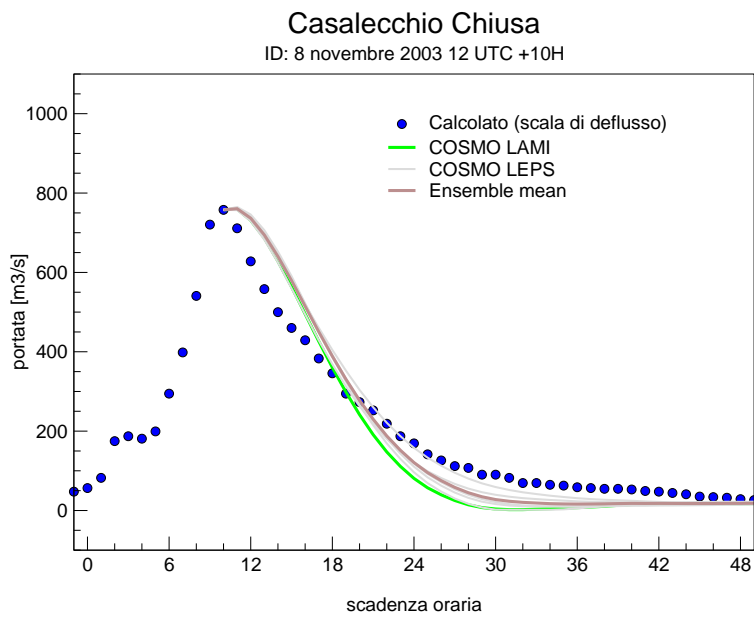
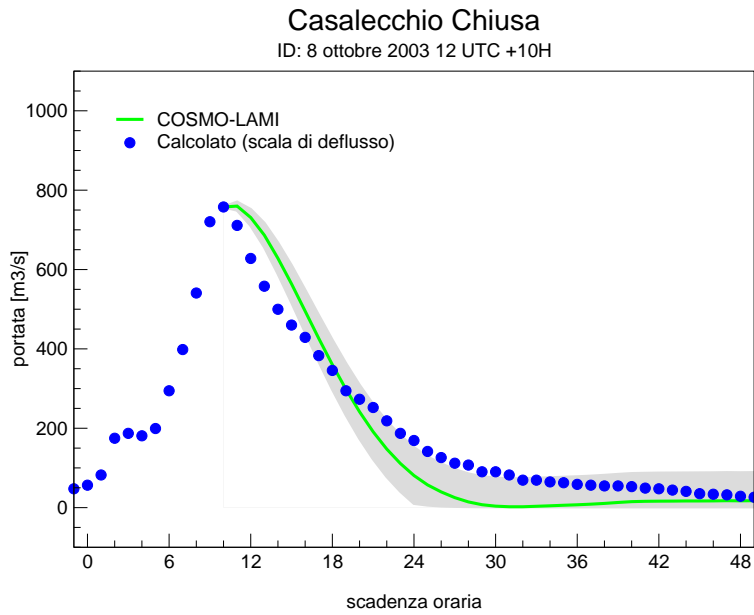


Figura 3.28: Idem per la figura 3.26, dopo 10 ore dall'inizio della previsione.

ECMWF Analysis VT:Wednesday 26 November 2003 12UTC 500hPa temperature/ geopotential height

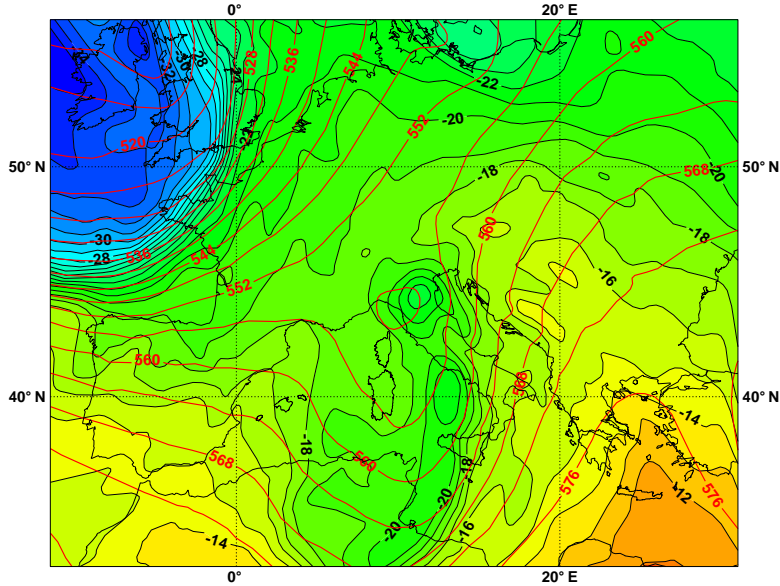


Figura 3.29: campo di geopotenziale e di temperatura a 500 hPa del 26 novembre 2003 12 UTC.

ECMWF Analysis VT:Thursday 27 November 2003 12UTC 500hPa temperature/ geopotential height

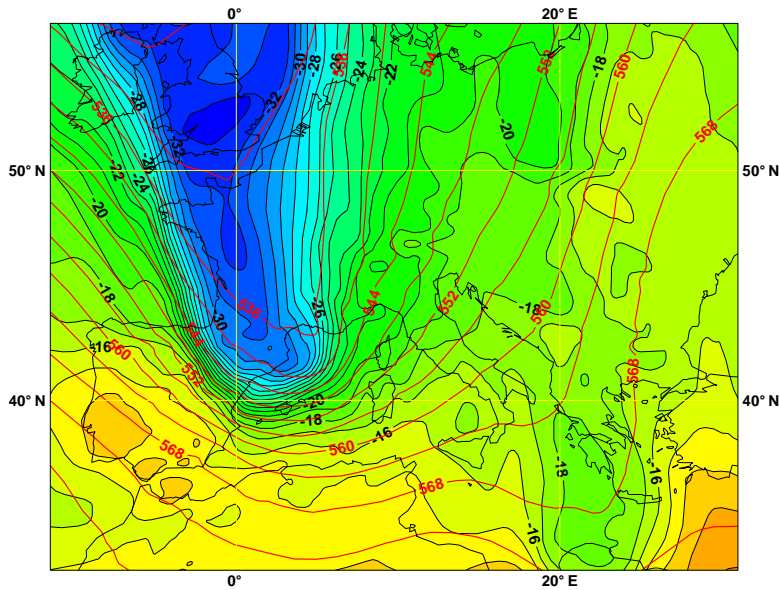


Figura 3.30: campo di geopotenziale e di temperatura a 500 hPa del 27 novembre 2003 12 UTC.

ECMWF Analysis VT:Friday 28 November 2003 12UTC 500hPa temperature/ geopotential height

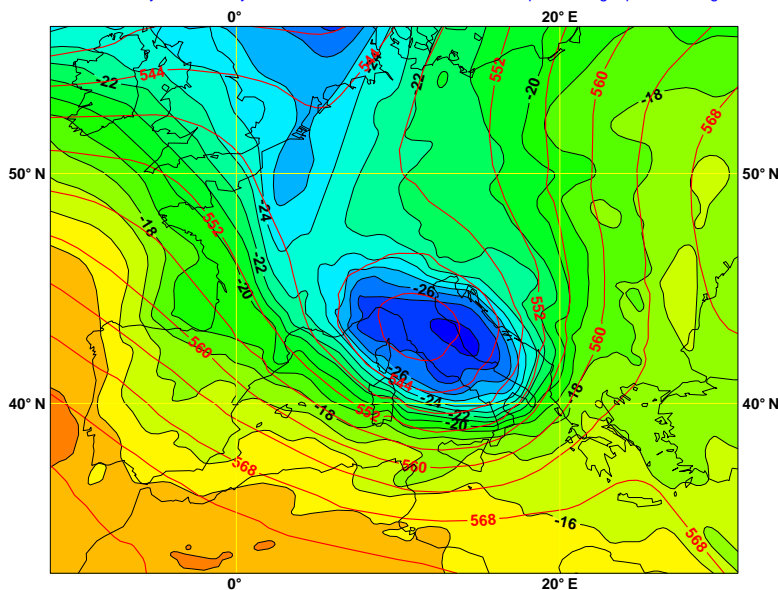


Figura 3.31: campo di geopotenziale e di temperatura a 500 hPa del 28 novembre 2003 12 UTC.

nucleo freddo, con il centro localizzato sul Nord della Penisola Italiana, alimenta l'instabilità degli strati inferiori dell'atmosfera determinando delle precipitazioni diffuse su tutta l'area coperta. Il sistema evolve verso una situazione di occlusione, come si evince dal carattere prevalentemente barotropico del nucleo nella figura 3.32 relativo al giorno 29 novembre, che ormai isolato, gradualmente va ad esaurirsi e, con la rimonta del promontorio di alta pressione, a spostarsi verso Est, sulla Penisola balcanica.

A differenza del caso precedente, le precipitazioni sono prevalentemente concentrate nelle prime 12 ore del periodo e la pioggia media areale ha un'intensità oraria inferiore, che oscilla tra 5 e 10 mm (figura 3.33). Il livello massimo di portata raggiunge gli $800 \text{ m}^3/\text{s}$ dopo circa 12 ore dall'inizio del periodo preso in esame. Anche per questo caso le stime a posteriori riescono a correggere la sovrastima

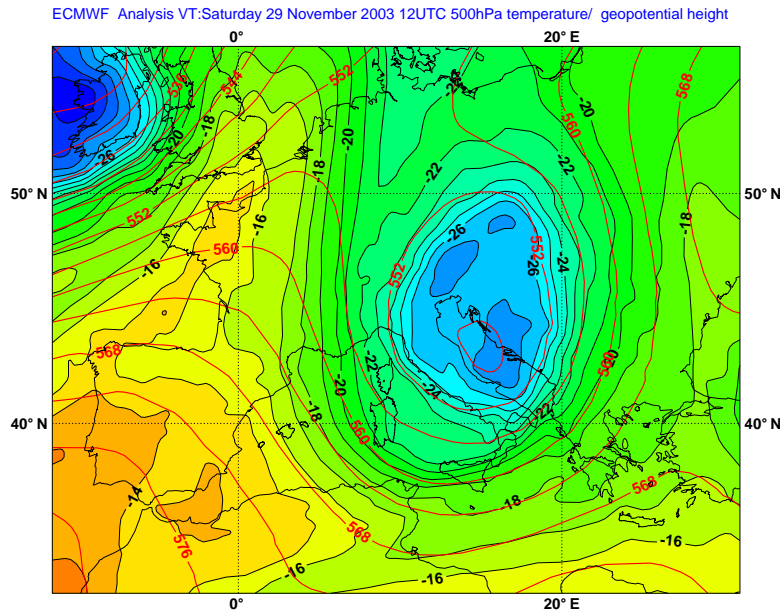


Figura 3.32: campo di geopotenziale e di temperatura a 500 hPa del 29 novembre 2003 12 UTC.

della simulazione del TOPKAPI (figura 3.34). In generale le curve riescono a seguire bene la curva della portata calcolata sulla base della scala di deflusso, salvo quella ottenuta con il filtro stazionario e gli errori autocorrelati che anticipa e sottostima il colmo di piena.

Il modello COSMO-LAMI e tutti gli elementi dell'ensemble non riescono a prevedere l'evento meteorologico anche per questo caso. Di conseguenza non viene visto nessun innalzamento del livello idrometrico dal modello TOPKAPI (figure 3.35). Utilizzando ancora il filtro adattativo con gli errori autocorrelati si riesce di nuovo a prevedere con qualche ora di anticipo la possibilità di un evento di piena di oltre $400 \text{ m}^3/\text{s}$ a 5 ore dal colmo di piena (figure 3.36). Approssimandosi al colmo di piena la curva idrometrica stimata tende ad adattarsi sempre più alla curva osservata (figure 3.37 e 3.38).

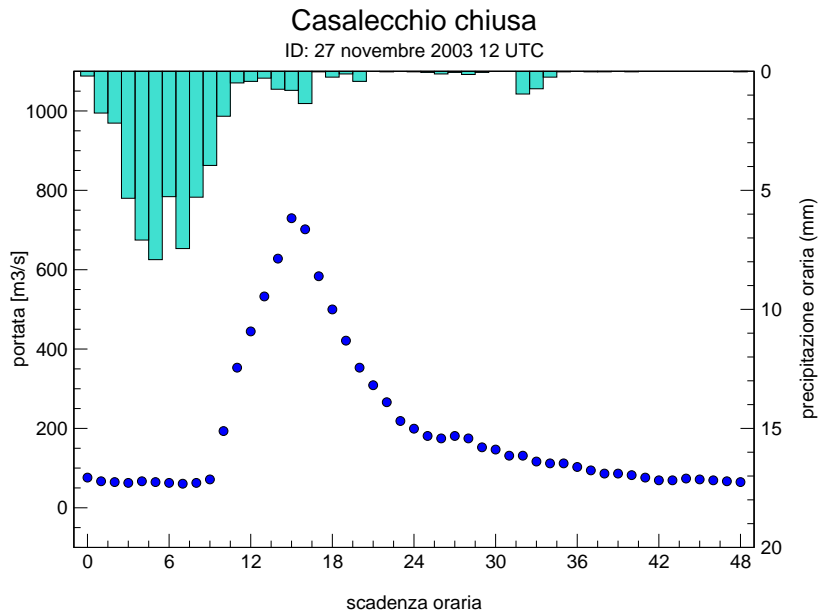


Figura 3.33: Idem per la figura 3.22, relativamente all'evento di piena del 28 novembre 2003.

3.3.3 31 ottobre – 1 novembre 2004

Il 29 novembre la situazione meteorologica è caratterizzata da un'ampia saccatura atlantica con il centro localizzato sulla Penisola della Cornovaglia (figura 3.39). Nella giornata successiva il centro della saccatura rimane pressoché stazionario, muovendosi lentamente dalla Penisola di Cornovaglia al Golfo di Bisaglia. L'asse della saccatura tende invece a ruotare verso Nord-Est, andando ad interessare anche la Penisola scandinava. Flussi Sud-occidentali, relativamente caldi ed umidi affluiscono sulla Penisola Italiana cominciando a dar luogo a deboli precipitazioni a partire dal versante tirrenico (figura 3.40).

L'avvezione di temperatura determina un relativo approfondimento della saccatura sulla Penisola Iberica, ma il corrispondente innalzamento del promontorio

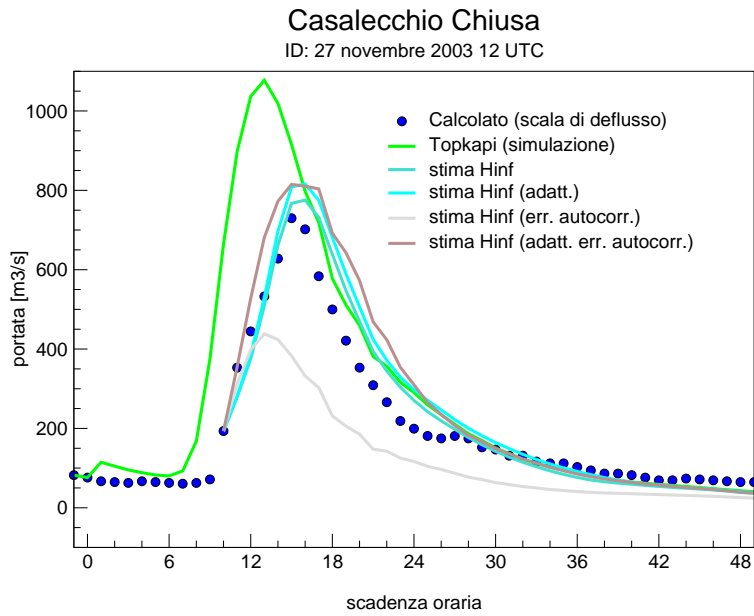


Figura 3.34: Idem per la figura 3.23 relativamente all'evento di piena del 28 novembre 2003.

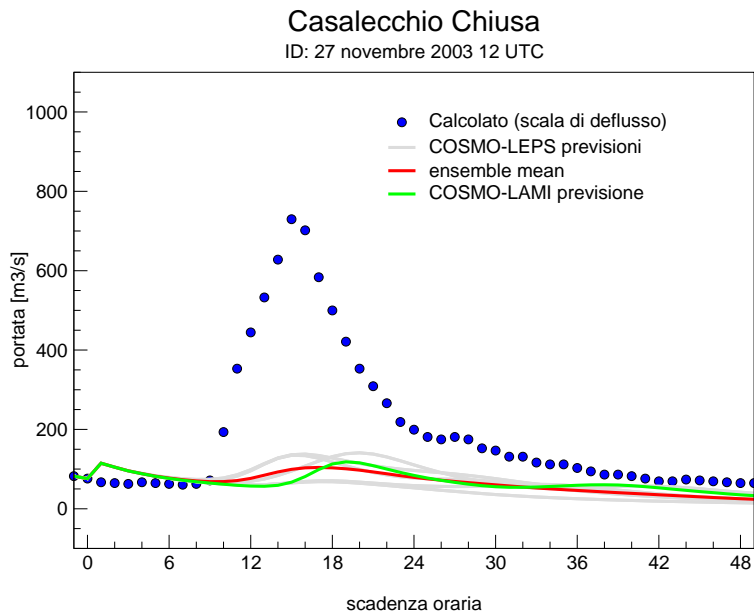


Figura 3.35: Idem per la figura 3.24, relativamente all'evento di piena del 28 novembre 2003.

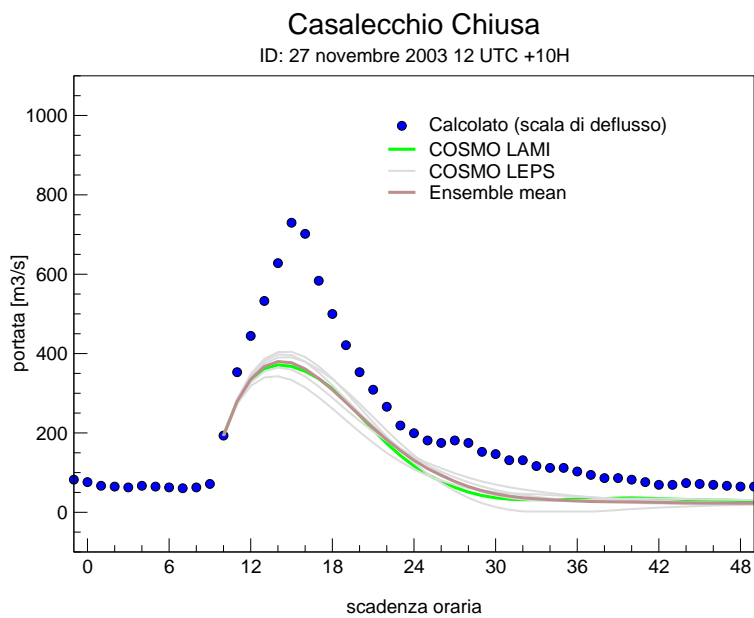
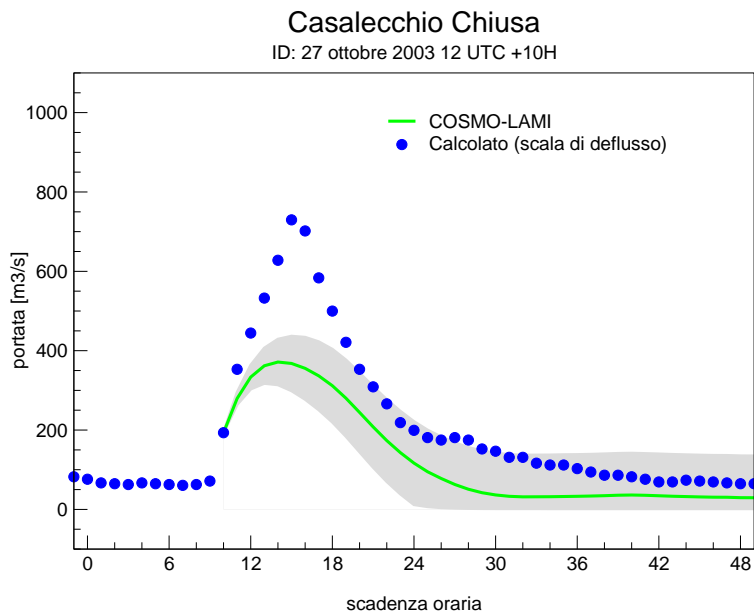


Figura 3.36: Idem per la figura 3.26, relativamente all'evento di piena del 28 novembre 2003 dopo 10 ore dall'inizio della previsione.

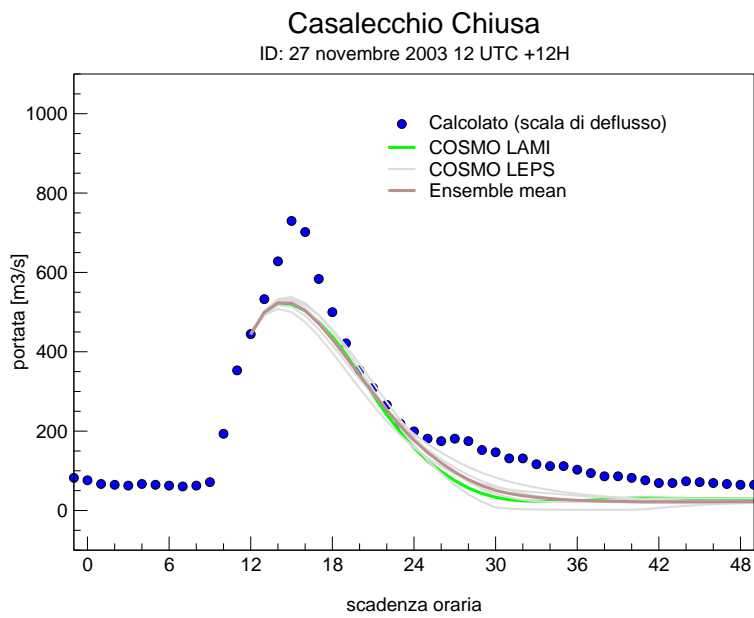
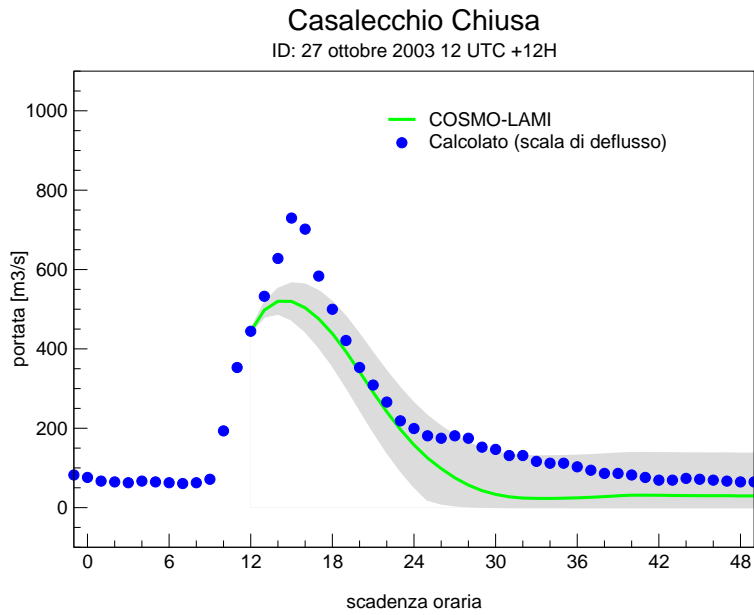


Figura 3.37: Idem per la figura 3.36, dopo 12 ore dall'inizio della previsione.

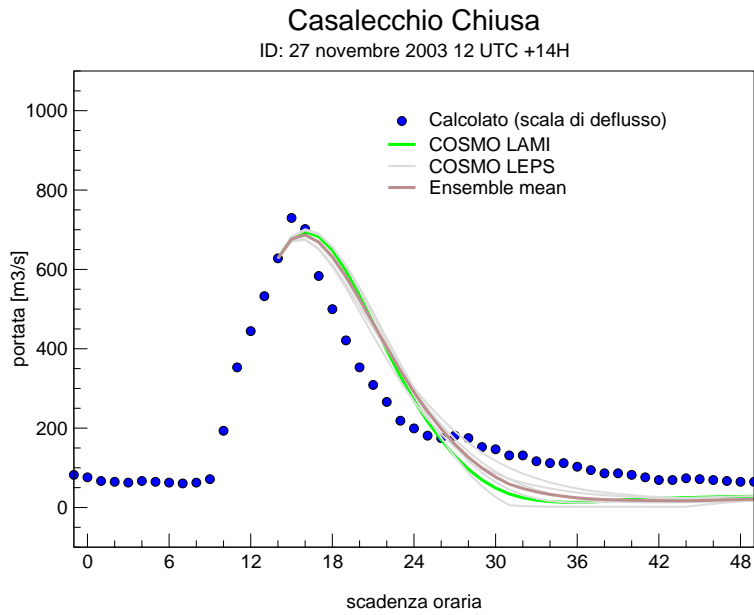
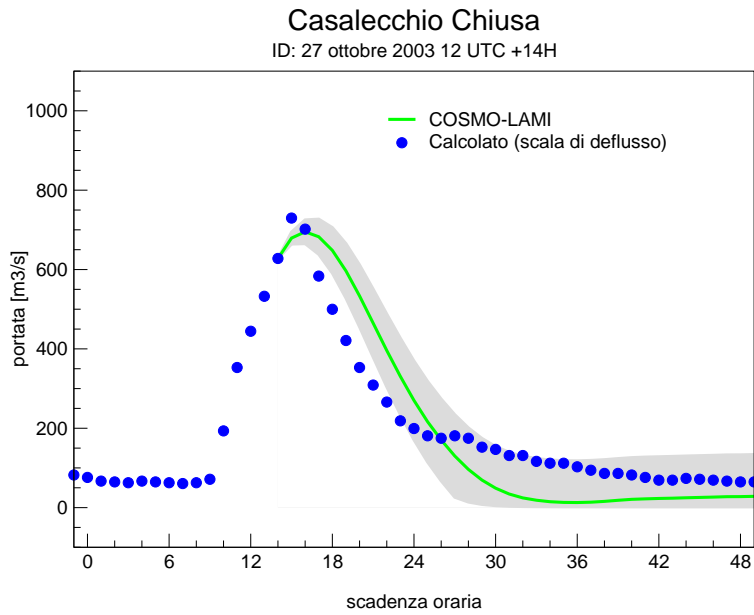


Figura 3.38: Idem per la figura 3.36, dopo 14 ore dall'inizio della previsione.

ECMWF Analysis VT:Friday 29 October 2004 12UTC 500hPa temperature/ geopotential height

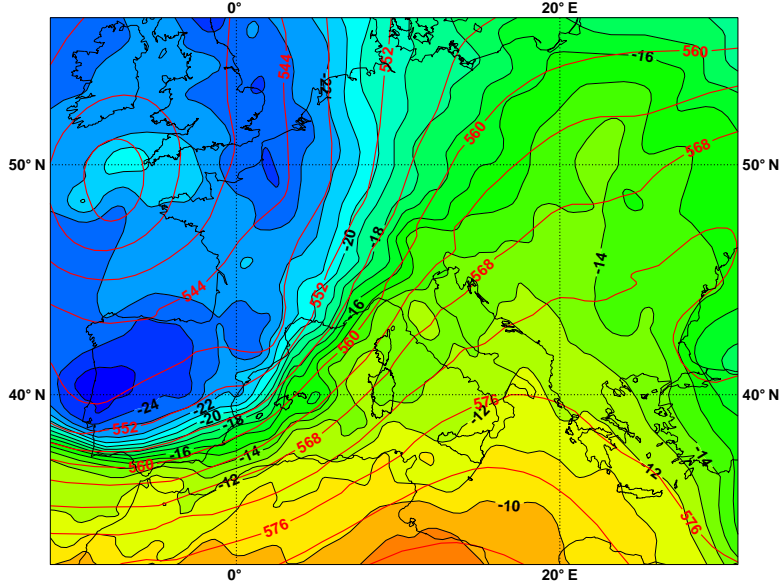


Figura 3.39: campo di geopotenziale e di temperatura a 500 hPa del 29 ottobre 2004 12 UTC.

ECMWF Analysis VT:Saturday 30 October 2004 12UTC 500hPa temperature/ geopotential height

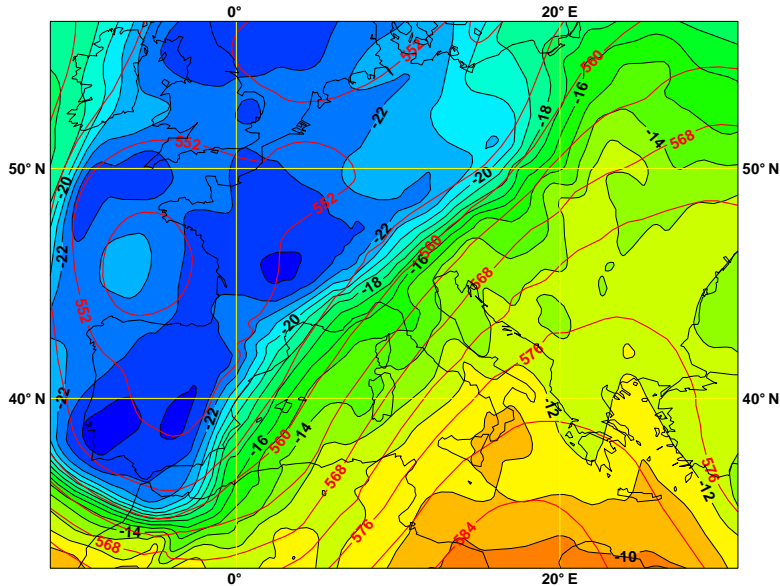


Figura 3.40: campo di geopotenziale e di temperatura a 500 hPa del 30 ottobre 2004 12 UTC.

ECMWF Analysis VT: Sunday 31 October 2004 12UTC 500hPa temperature/ geopotential height

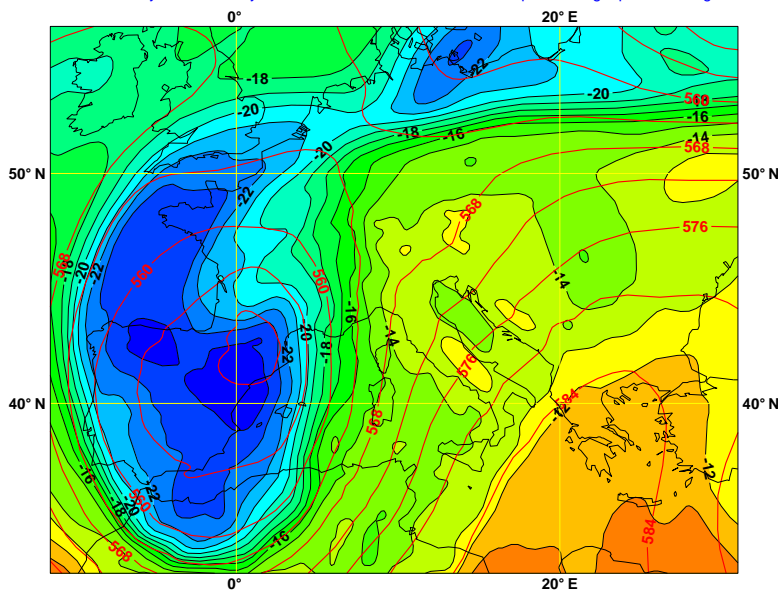


Figura 3.41: campo di geopotenziale e di temperatura a 500 hPa del 31 ottobre 2004 12 UTC.

sulla Penisola Balcanica determina la stazionarietà della situazione meteorologica che persiste anche nei due giorni successivi, come si può evincere dalla figura 3.41.

La situazione va in esaurimento nella giornata dell'1 novembre, quando si arriva all'occlusione e la circolazione ciclonica, ormai completamente separata, non viene più alimentata da masse d'aria fredda di origine polare. La circolazione in quota sull'Italia permane meridionale, ma non determina più quelle precipitazioni che hanno caratterizzato i giorni precedenti (figura 3.42). In particolare, dalla figura 3.43 si può osservare come la precipitazione abbia interessato il bacino del fiume Reno, ed in particolare la sua parte montana, nella giornata del 31 ottobre, cioè tra +12 e +24 ore nella scala temporale del grafico. Le precipitazioni medie areali sono state piuttosto intense, con punte di oltre 15 mm di intensità oraria.

ECMWF Analysis VT:Monday 1 November 2004 12UTC 500hPa temperature/ geopotential height

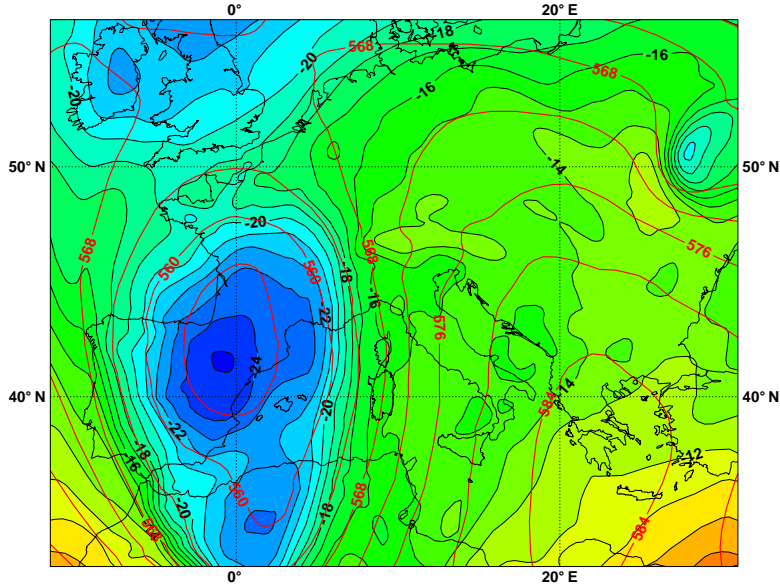


Figura 3.42: campo di geopotenziale e di temperatura a 500 hPa dell'1 novembre 2004 12 UTC.

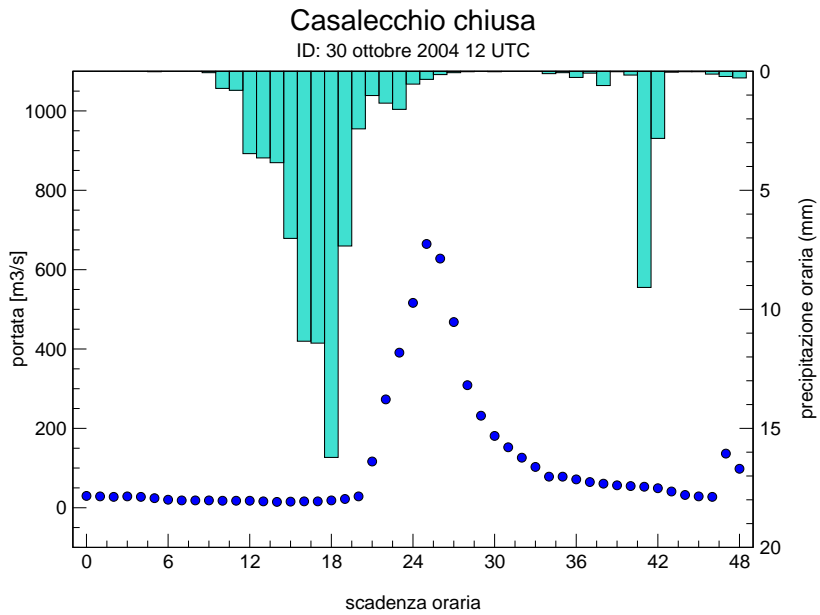


Figura 3.43: Idem per la figura 3.22, relativamente all'evento di piena del 31 ottobre 2004.

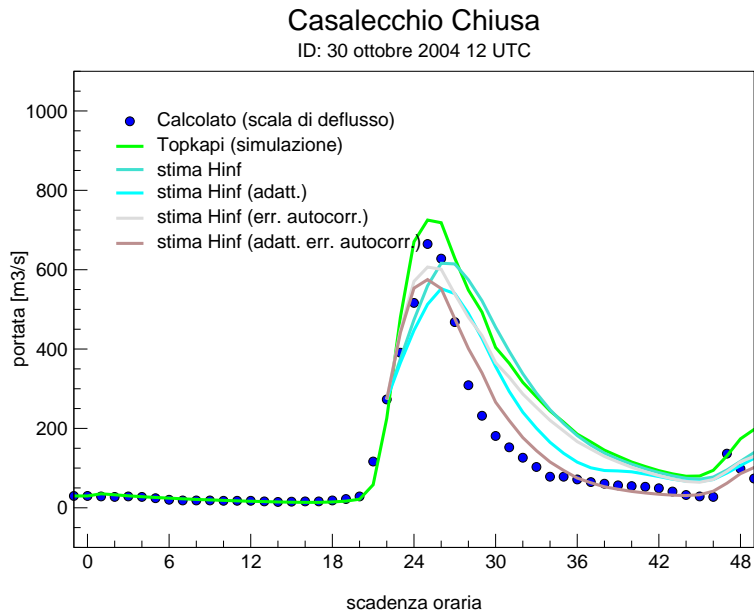


Figura 3.44: Idem per la figura 3.23 relativamente all'evento di piena del 31 ottobre 2004.

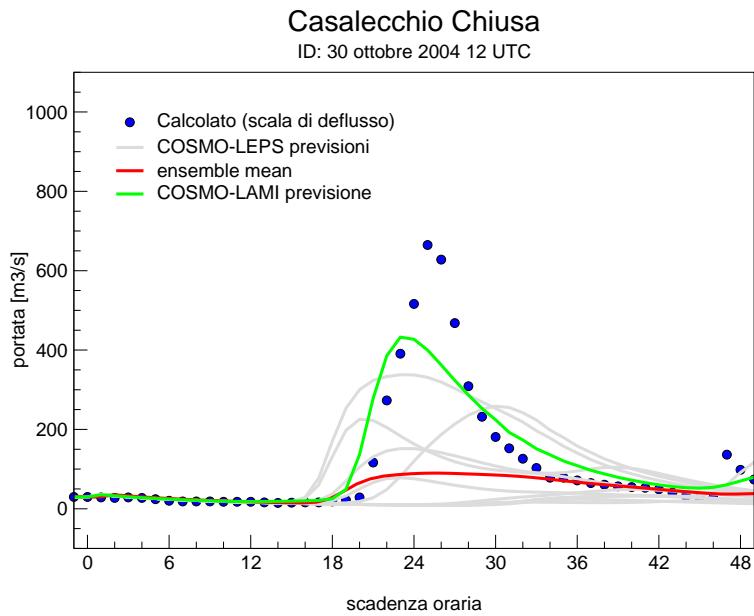


Figura 3.45: Idem per la figura 3.24, relativamente all'evento di piena del 31 ottobre 2004.

Analogamente ai casi precedenti, dalla figura 3.44 si evince come viene corretta la curva idrometrica simulata dal modello TOPKAPI con le precipitazioni osservate. In particolare si può osservare come il filtro senza il modello autoregressivo per gli errori, sia nella variante stazionaria che in quella adattativa, tende a sottostimare e ritardare il colmo di piena. La tempistica viene corretta utilizzando il filtro con gli errori autocorrelati, ma resta sempre una sottostima dell'evento, più accentuata per la variante adattativa.

Per questo caso il modello deterministico COSMO-LAMI riesce a prevedere in maniera abbastanza accurata la precipitazione sul bacino del fiume Reno, come si può osservare nella figura 3.45, in cui la portata calcolata raggiunge un colmo maggiore di $400 \text{ m}^3/\text{s}$. L'ensemble di COSMO-LEPS ha una maggiore variabilità, ma nessuno degli elementi riesce a produrre una previsione più accurata del modello deterministico. Come per i casi precedenti, si riportano le curve idrometriche stimate mediante il filtro adattativo, sulla base delle previsioni di portata a 3 ore ed 1 ora dal colmo di piena (figure 3.46 e 3.47).

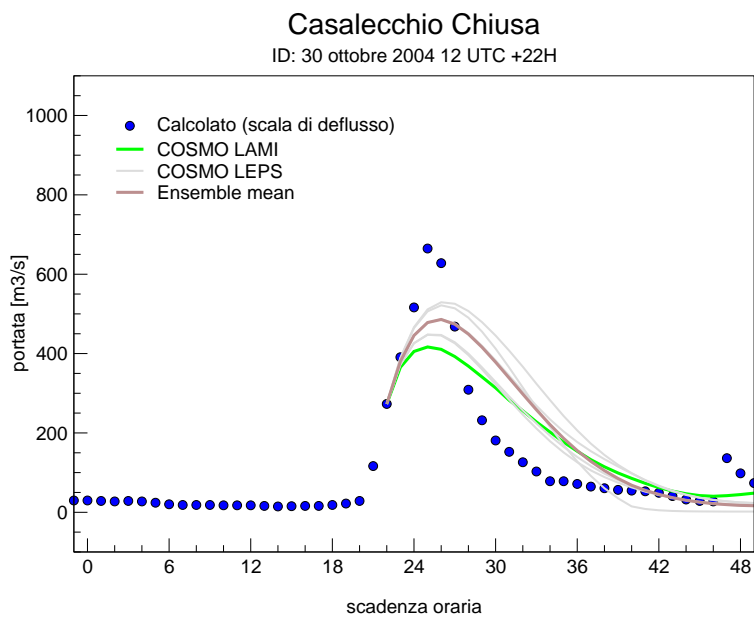
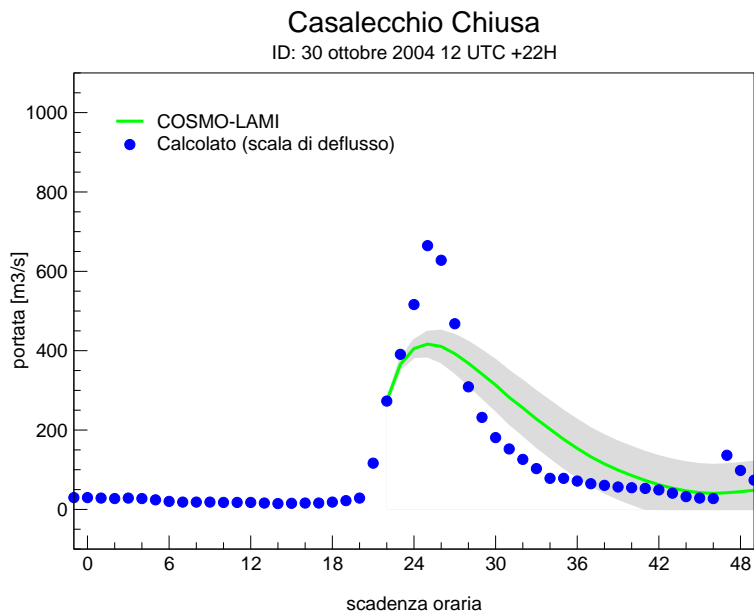


Figura 3.46: Idem per la figura 3.46, relativamente all'evento di piena del 31 ottobre 2004 dopo 22 ore dall'inizio della previsione.

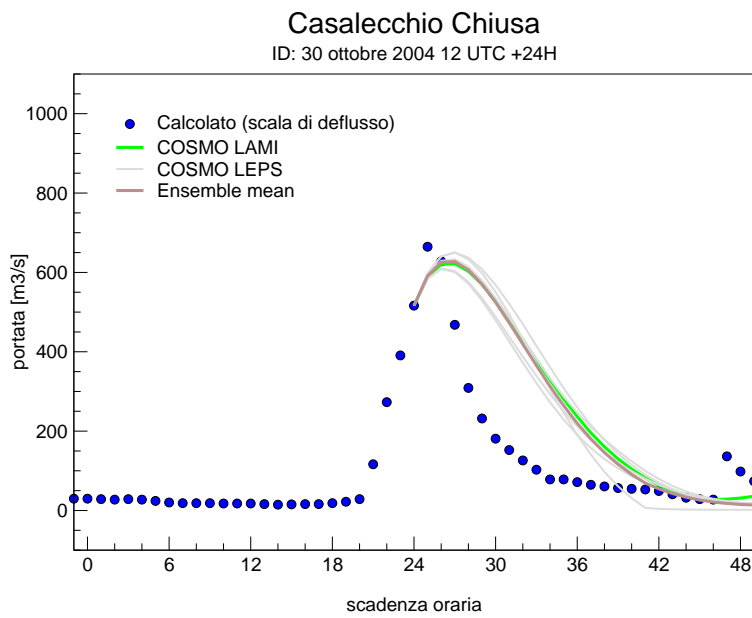
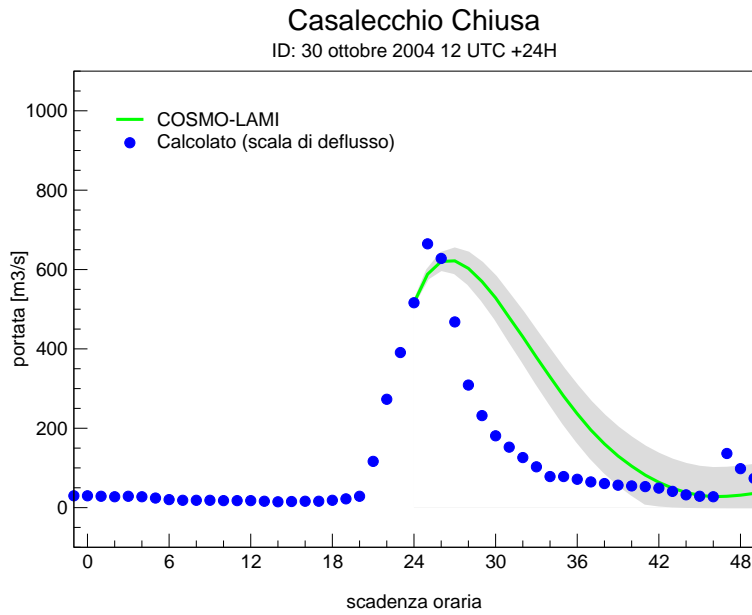


Figura 3.47: Idem per la figura 3.46, dopo 24 ore dall'inizio della previsione.

Capitolo 4

Conclusioni e prospettive future

Il lavoro di ricerca svolto ha avuto come obiettivo la formulazione di un modello bayesiano per la stima dell'incertezza nella previsione idrologica di portata alla chiusa di Casalecchio, al termine del sottobacino montano del fiume Reno. Il bacino è di media grandezza, dell'ordine di 10^3 km², localizzato nel versante nord dell' Appennino Tosco-Emiliano, e la previsione delle piene di portata è relativamente complessa, poichè legata al carattere prevalentemente torrentizio del fiume. I tempi di corrivazione alla chiusa di Casalecchio sono dell'ordine di 6–8 ore.

Il punto di partenza del lavoro è stato l'articolo di Krzysztofowicz: *Bayesian theory of probabilistic forecasting via deterministic hydrologic model*, pubblicato su *Water Resources Research* nel 1999, nel quale l'autore determina previsioni idrologiche di tipo probabilistico sulla base dei principi dell'inferenza statistica bayesiana. L'evoluzione nel tempo della portata di un fiume è vista come lo svolgimento di un processo markoviano, nel quale lo stato della grandezza idrologica ad un determinato istante di tempo è dato da un modello autoregressivo che combina il valore assunto dalla medesima grandezza all'istante precedente ed il valore

di una grandezza casuale che esprime una misura dell'errore.

La serie temporale delle osservazioni di portata è stata perciò considerata come la parte finita di un'unica realizzazione di un processo stocastico, la cui memoria è determinata dal grado di correlazione tra le variabili casuali che lo compongono. Inizialmente si è preso in esame un modello lineare, autoregressivo e stazionario, i cui coefficienti sono stati calcolati mediante il metodo dei minimi quadrati. Tuttavia, il modello stocastico stazionario assume che la dinamica delle variazioni dei livelli di portata segua un processo diffusivo, ossia gaussiano, mentre dall'analisi della distribuzione statistica degli errori sono state osservate larghe deviazioni dalla distribuzione normale, soprattutto in corrispondenza dei valori estremi. È stata introdotta una prima generalizzazione, per la quale i coefficienti di autoregressione non sono più costanti ma vengono stimati tramite uno smoother, che tiene conto, di volta in volta, dei valori di portata osservati nelle ore precedenti all'istante di riferimento (filtro adattativo). L'esperienza ha dimostrato inoltre che nelle stime di portata l'errore è fortemente autocorrelato ed è stato perciò necessario rappresentarlo con un secondo processo markoviano.

Il modello bayesiano sviluppato combina la stima di portata compiuta con lo schema autoregressivo (a priori) con quella data dal modello distribuito, ad afflussi–deflussi TOPKAPI, al fine di giungere ad un'unica stima a posteriori, cui associare il valore minimo di una determinata funzione di rischio.

Se i parametri dei modelli fossero perfettamente conosciuti e le componenti casuali AWGN (*additive white Gaussian noise*), allora lo stimatore ottimo in senso bayesiano sarebbe dato dal filtro di Kalman, in quanto minimizza la varianza dell'errore della stima. Ma come avviene nella maggior parte dei casi reali, anche in idrologia per la previsione delle piene, le assunzioni necessarie per il filtro di

Kalman sono troppo restrittive. In generale le statistiche e le distribuzioni relative agli errori nelle stime processi idrologici non sono conosciute, per la notevole variabilità dei fenomeni e inoltre la serie temporale a disposizione può essere troppo breve per ottenere una stima dei parametri robusta. Per questa ragione si è cercato di utilizzare il filtro H^∞ , in alternativa al filtro di Kalman, poiché permette di non fare nessuna assunzione relativa alle componenti di errore. È conservativo, si propone di minimizzare il rischio massimo, e fornisce una stima più robusta rispetto al filtro di Kalman. Il filtro H^∞ corrisponde ad uno stimatore bayesiano con rischio costante. Molti metodi sono stati proposti, ma qui è stato utilizzato quello basato sull'equazione algebrica di Riccati, che permette di implementare il filtro in maniera ricorsiva e consente un raffronto diretto con il filtro di Kalman.

Il periodo di studio preso in esame comprende le stagioni autunnali dal 2003 al 2005. I risultati ottenuti introducendo nel modello TOPKAPI le precipitazioni osservate hanno mostrato che le stime calcolate con il filtro H^∞ , nel quale l'errore non è di tipo AWGN ma rappresentato mediante un processo markoviano, posseggono una varianza campionaria dell'errore minore e nei casi di piena hanno in generale una maggiore capacità di adattamento alle curve di deflusso, sia nella fase di risalita che in quella di discesa. Il filtro, sia nella forma stazionaria che nella forma adattativa, permette inoltre di correggere anche eventuali problemi legati al ritardo, o all'anticipo del colmo presente nella stima del modello idrologico.

L'aspetto più rilevante è comunque lo scostamento tra la varianza campionaria dell'errore e la varianza teorica per tutte le varianti del filtro. Questo può essere indicativo che, come già osservato dalla funzione di autocorrelazione dell'errore, la struttura dell'errore può essere compresa in un processo autoregressivo soltanto in prima approssimazione. La natura fortemente eteroschedastica del processo

necessita di transitare da un modello di questo tipo ad un modello non lineare, cosiddetto “autoregressive conditional heteroschedasticity (ARCH)”. Questa considerazione parte dal presupposto che le variazioni della varianza dell’errore potrebbero essere generate, oltre che da cause esogene, anche da autocorrelazioni non lineari.

Il filtro adattativo, adottato per ovviare al problema della sottostima, si è dimostrato scarsamente efficace, in quanto se per le stime oltre il tempo di convergenza del filtro la varianza dell’errore generalmente è superiore a quella del filtro stazionario, nelle prime 6 ore il miglioramento non appare significativo.

La versatilità del modello TOPKAPI permette di inserirlo in un sistema di preannuncio delle piene in tempo reale, nel quale i dati misurati di precipitazione e di temperatura sono sostituiti da quelli previsti dal modello meteorologico COSMO-LAMI. Nel periodo considerato soltanto per l’evento del 31 ottobre 2004 il modello COSMO-LAMI riesce a cogliere la corretta dinamica meteorologica, riproducendo in maniera sufficientemente accurata il campo di precipitazione (dati non mostrati), sebbene poi la previsione idrologica di TOPKAPI sottostimi il colmo di piena.

Krzysztofowicz decompone le molteplici fonti di errore di una previsione in due componenti principali: la prima legata all’incertezza delle precipitazioni sull’area del bacino; la seconda, invece, che raccoglie le incertezze dovute sia ad errori nelle misure e nelle stime dei parametri del modello idrologico, sia ad errori del modello nel descrivere i processi fisici in esame. Dai risultati ottenuti si evince che alla scala dei bacini di media dimensione, la previsione della quantità di precipitazione che può accadere è la prima fonte di incertezza. Analogamente a quanto avviene in meteorologia, si è cercato di utilizzare un ensemble di previsioni

idrologiche, ottenute accoppiando il modello TOPKAPI alle uscite di COSMO-LEPS, con l'intento di associare il grado di incertezza della stima alla dispersione degli elementi dell'ensemble. Tuttavia, nei tre eventi presi in esame l'ensemble di previsioni di portata, oltre a presentare una generale sottostima degli eventi di piena, ha già nella distribuzione non condizionata una variabilità limitata e non calibrata che rende conseguentemente la distribuzione a posteriori non calibrata.

La ragione della limitata variabilità potrebbe essere nella costruzione dell'ensemble meteorologico. Infatti, il periodo scelto per massimizzare gli errori dei "singular vectors" è di 48 ore. Le traiettorie nello spazio delle fasi degli elementi dell'ensemble tendono a divergere soltanto dopo 120 ore. Sono tempi che vanno oltre la scala di tempo di interesse idrologico, che difficilmente, per i bacini di medie dimensioni, possono superare le 24 ore.

A questo proposito, uno delle attività prioritarie del consorzio COSMO per il periodo 2006-2008 è il progetto di ricerca COSMO-SREPS, che si prefigge lo scopo di sviluppare un sistema di previsioni di ensemble per il breve termine (Short-Range Ensemble Prediction System) basato sul modello non-idrostatico ad area limitata COSMO-LAMI. La previsione di ensemble per scadenze temporali inferiori a 3 giorni può infatti consentire di quantificare l'incertezza associata alla previsione meteorologica dei modelli deterministici, ad alta risoluzione, in grado di descrivere i processi alla mesoscala. La metodologia scelta per il sistema COSMO-SREPS permette di comprendere nella costruzione dell'ensemble la descrizione di sorgenti di errore di natura diversa e soprattutto agenti su una vasta gamma di scale spazio-temporali.

Infine, l'utilizzo del filtro H^∞ ha permesso di ovviare alle limitazioni richieste dal filtro di Kalman, tuttavia, poichè la metrica del filtro è indefinita, i risultati di-

pendono da un lavoro di aggiustamento dei parametri che entrano nelle equazioni del filtro fatto ad hoc.

L'esperienza fatta, seppur producendo un risultato parzialmente negativo, è tuttavia stata utile per evidenziare alcuni aspetti legati alla struttura dell'errore, che solo in prima approssimazione può essere trattato come un processo markoviano, all'utilizzo dell'ensemble di previsioni, ma soprattutto per individuare nuove strade da percorrere e nuovi strumenti da utilizzare ai fini della definizione di un sistema di allerta accurato per scopi di Protezione civile.

Bibliografia

- [1] Abbott et al. 1986. *An introduction to the European Hydrological System-Système Hydrologique Européen, SHE - History and philosophy of a Physically Based Distributed Modelling Systems*, Journal of Hydrology, 87, 45–59.
- [2] Bartholmes J, Todini E, 2005. *Coupling meteorological and hydrological models for flood forecasting*, Hydrol. Earth Syst. Sci. 9(4), 333–46.
- [3] Beven KJ, Binley AM, 1992. *The future of distributed models: model calibration and uncertainty prediction*, Hydrological Processes 6, 279–98.
- [4] Beven KJ, Kirby MJ, 1979. *A physically based, variable contributing area model of basin hydrology*. Hydrolog. Sci. Bull., 14, 1–3.
- [5] Blackie JR, Eeles CW, 1985. *Lumped Catchment Models. hydrological forecasting*, eds. Anderson and Burt, John and Wiley & Sons Ltd.,
- [6] Burnasch RJC, Ferral RL, Mc Guire RA, 1973. *A general streamflow simulation system. Conceptual modelling for digital computers*, Report by the Joint Federal State River Forecast Center, Sacramento.

- [7] Crawford NH, Linsley RK, 1966. *Digital simulation in Hydrology, Stanford Watershed model IV*, Technical Report 39, Dept. Civil Eng. Stanford University.
- [8] De Roo APJ, Wesseling CG, Van Deursen WPA, 1998. *Physically based river modelling within a GIS. The LISFLOOD model*, Proc. 3rd Int. Conf. On Geo-Computation.
- [9] Diomede T, 2004. *Usa degli analoghi meteorologici nella previsione operativa delle piene*, Tesi di Dottorato, XV ciclo, Dipartimento di Scienze della Terra, Università di Bologna
- [10] Gregory PC, 2005. *Bayesian Logical Data Analysis for the Physical Sciences*, Cambridge University Press.
- [11] Hassibi B, Sayed AH, Kailath T., 1996. *H^∞ Optimality of the LMS Algorithm*. IEEE Trans. on Signal Processing 44, 267–80.
- [12] Hassibi B, Sayed AH, Kailath T., 1999. *Indefinite-Quadratic Estimation and Control: A unified Approach to H^2 and H^∞ Theories*, Eds. Society for Industrial and Applied Mathematics.
- [13] Kowen N, 2000. *WATFLOOD/SPL: Hydrological model and flood forecasting system*, Dept. of Civil Eng., University of Waterloo, Waterloo, Ont.
- [14] Krzystofowicz R, 1999. *Bayesian theory of probabilistic forecasting via deterministic hydrological model*, Water Resour. Res. 35, 2739–50.

- [15] Labarre D, Grivel E, Najim M, Christov N. *Dual H^∞ Algorithms for Signal Processing, Application to Speech Enhancement*. IEEE Trans. on Signal Processing, in press.
- [16] Mantovan P, Todini E, 2006. *Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology*, Journal of hydrology 330, 368–81.
- [17] Marsigli C, Montani A, Nerozzi F, Paccagnella T, Tibaldi S, Molteni F, Buizza R. 2001. *A strategy for high-resolution ensemble prediction. Part II: Limited-area experiments in four Alpine flood events*. Q.J.R. Meteorol. Soc. in press.
- [18] Molteni F, Buizza R, Palmer TN, Petroliagis T. 1996. *The ECMWF Ensemble Prediction System: Methodology and validation*. Q.J.R. Meteorol. Soc. 122, 73–119.
- [19] Molteni F, Buizza R. 1999. *Validation of ECMWF Ensemble Prediction System using empirical orthogonal function*. Q.J.R. Meteorol. Soc. in press.
- [20] Molteni F, Buizza R, Marsigli C, Montani A, Nerozzi F, Paccagnella T. 2001. *A strategy for high-resolution ensemble prediction. Part I: Definition of Representative Members and Global Model Experiments*. Q.J.R. Meteorol. Soc. in press.
- [21] Mood AM, Graybill FA, Boes DC, 1988. *Introduzione alla Statistica*, Eds. McGraw Hill.

- [22] Montanari A, Brath A, 2004. *A stochastic approach for assessing the uncertainty of rainfall–runoff simulations*, Water Resour. Res. 40, 1–11.
- [23] Montanari A, 2006. *What do we mean by “uncertainty”? The need for a consistent wording about uncertainty assessment in hydrology.*, Invited commentary to HPToday.
- [24] Moore RJ, 1985. *The probability-distributed principle and runoff production at point and basin scales*, Hydrology Science 20, 273–297.
- [25] Natale L, Todini E, 1977. *A constrained parameter estimation technique for linear models in hydrology in Mathematical Models for Surface Water Hydrology*, John Wiley & Sons, 115–147.
- [26] Rossa A, 2004. *Propagation of uncertainty in advanced meteo–hydrological forecast systems*, Proposal for a new COST Action, COST 731
- [27] Rockwood DM, Nelson ML, 1966. *Computer application to Streamflow Synthesis and Reservoir regulation*, Proc. IV Int. Conference on Irrigation and Drainage.
- [28] Sugawara M, 1995. *Tank model in Computer models of watershed hydrology*, Chapter 6, edited by Singh VP *et al.*, Water Resources Publications, Littleton, Colorado.
- [29] Tibaldi S, Paccagnella T, Marsigli C, Montani A, Nerozzi F, 2006. *Limited–Area Ensemble Forecasting: the COSMO–LEPS system*, Predictability of weather and climate, Cambridge University Press, 489–513.

- [30] Todini E, 1978. *Mutually interactive state/parameter (MISP) estimation in hydrological applications*, Modelling, Identification and Control in Environmental Systems, Varstenskiste eds., 699–713.
- [31] Todini E, 1989. *Flood Forecasting Models*, Excerpta, 4, 117–162.
- [32] Todini E, Ciarrapica L, 2001. *The TOPKAPI model in Mathematical Models of Large Watershed Hydrology* Chapter 12, edited by Singh VP et al., Water Resources Publications, Littleton, Colorado.
- [33] Todini E, 2002. *The CLS model in Mathematical Models of Large Watershed Hydrology*, Chapter 20, edited by Singh, V.P. et al., Water Resources Publications, Littleton, Colorado.
- [34] Verdin J, Klaver R, 2002. *Grid-cell based crop water accounting for the Famine Early Warning System*, Hydrological Processes, 16(8), 1617–1630.
- [35] Wigmosta MS, Vail LW, Lettenmaier DP, 1994. *A distributed hydrology-vegetation model for complex terrain*, Water Resour. Res., 30(6), 1665–1679.
- [36] Wikle CK, Mark Berliner L, Cressie N, 1998, *Hierarchical Bayesian space–time models*, Environmental and Ecological Statistics 5, 117–54.
- [37] Wilks SS, 1963. *Mathematical Statistics*, Eds. John Wiley & Sons.
- [38] Zhao RJ, 1992. *The Xinanjiang model applied in China*, Journal of Hydrology, 135, 371–281.

Ringraziamenti

Un sentito ringraziamento va al Prof. Ezio Todini, che è stato il “motore primo” di questo lavoro. L’intuizione è sua, il resto del lavoro è stato un mio inseguimento nel cercare di realizzare tecnicamente ciò che ho potuto cogliere.

Un sentito ringraziamento alla Prof.ssa Daniela Cocchi, del Dipartimento di Statistica dell’Università di Bologna, per i suoi suggerimenti e per avermi dato la possibilità di partecipare al suo gruppo di lavoro.

Un altrettanto sentito ringraziamento al Dott. Carlo Cacciamani che, nel suo ruolo di responsabile Area Previsioni Meteo e Climatologia del Servizio Idrometeorologico di ARPA Emilia–Romagna, ha favorito in tutti i modi possibili la realizzazione di questo lavoro di tesi. Inoltre, il suo interessamento, le domande sempre acute che mi ha posto, le discussioni che ne sono seguite, mi hanno permesso di fare luce sul lavoro che stavo facendo.

Ringrazio l’Ing. Tommaso Diomede, “the last but not the least”, per avere effettuato le elaborazioni del TOPKAPI, poi utilizzate in questo lavoro. Senza la sua ampia pazienza e disponibilità non avrei mai potuto dirimere tutte le questioni di carattere idrologico che mi si sono presentate.