

UNIVERSITÀ DEGLI STUDI DI PADOVA
FACOLTÀ DI INGEGNERIA

—
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

—
TESI DI LAUREA MAGISTRALE IN INGEGNERIA
DELL'AUTOMAZIONE

METODI DI IDENTIFICAZIONE
NON-PARAMETRICA E LA LORO
APPLICAZIONE SU DATI CLIMATICI
DI EDIFICI

RELATORE: CH.MO PROF. ALESSANDRO CHIUSO

TUTORI IN AZIENDA: ING. IRENE PARADISI

DIPL.-ING. LORENZ VORHAUSER

LAUREANDO: MARKUS AUSSERER

DATA DI LAUREA: 17.04.2012

ANNO ACCADEMICO 2011-2012

*Il lavoro di tesi è stato svolto presso l'istituto di ricerca Fraunhofer Italia a
Bolzano.*

Indice

Sommario	IX
Introduzione	XI
1 Modellazione termica di edifici: considerazioni teoriche	1
1.1 Software di simulazione energetica EnergyPlus TM	6
2 Identificazione di sistemi: considerazioni teoriche	9
2.1 Considerazioni teoriche relative al PEM	11
2.1.1 Descrizione del principio di funzionamento di PEM	11
2.1.2 Determinazione della struttura del modello	13
2.2 Considerazioni teoriche relative a SSEH	14
2.2.1 Descrizione del problema	15
2.2.2 Kernels per l'identificazione di sistemi	19
2.2.3 Un modello Bayesiano per l'identificazione di sistemi lineari sparsi	22
2.2.4 Stima degli iperparametri e delle risposte impulsive del predittore	24
2.2.5 Generalizzazione a sistemi MIMO	27
3 Setup sperimentale per la validazione di SSEH	33
3.1 Setup degli studi Monte Carlo	34
3.2 Il monitoraggio dell'edificio	38
3.3 Metodi di identificazione considerati	44
3.3.1 Diverse versioni di SSEH e PEM	44
3.3.2 Procedure di selezione delle variabili per SSEH	48
4 Risultati degli studi Monte Carlo	57
4.1 Influenze delle lunghezze del set di dati e delle risposte impulsive .	58
4.1.1 Ingressi non correlati e modelli con livello di sparsità ridotto	58

4.1.2	Ingressi correlati e modelli con livello di sparsità ridotto . .	64
4.1.3	Ingressi non correlati e modelli con livello di sparsità aumentato	68
4.1.4	Ingressi correlati e modelli con livello di sparsità aumentato	73
4.2	Diverse procedure di selezione di SSEH	77
4.2.1	Procedure di selezione di variabili	78
4.2.2	Procedure di selezione di singole risposte impulsive	85
4.3	Confronto tra diverse versioni di SSEH e PEM	87
5	Risultati dell'identificazione di modelli termodinamici	95
5.1	Confronto tra Stable Spline with Exponential Hyperprior (SSEH) e Prediction Error Minimization (PEM)	96
5.1.1	Configurazione di riferimento: $N = 500$ e $T = 6$ minuti . .	96
5.1.2	Influenza della lunghezza del set di dati e del tempo di campionamento	101
5.1.3	Selezione manuale di sensori	104
5.2	Confronto tra diverse procedure di selezione	106
6	Discussione e confronto dei risultati	113
6.1	Considerazioni relative all'efficienza computazionale	113
6.2	Analisi dei risultati degli studi Monte Carlo	116
6.3	Analisi dei risultati delle identificazioni su dati reali	123
	Conclusioni	129
	Bibliografia	131

Elenco delle abbreviazioni

AIC	Akaike Information Criterion
AICc	corrected Akaike Information Criterion
ARMAX	Autoregressivo a media mobile con ingressi esogeni
ARPAV	Agenzia Regionale per la Prevenzione e Protezione Ambientale del Veneto
ARX	Autoregressivo con ingressi esogeni
BEMS	Building Energy Management System
BIBO	Bounded Input Bounded Output
BIC	Bayesian Information Criterion
DTG	Dipartimento di Tecnica e Gestione dei Sistemi Industriale dell'Università degli Studi di Padova
FIR	Finit Impulse Response
GLAR	Group Least Angle Regression
GLasso	Group Lasso
HVAC	Heating Ventilation Air-Conditioning
IIR	Infinet Impulse Response
LAR	Least Angle Regression
Lasso	Least Absolute Shrinkage and Selection Operator
MAP	Maximum A Posteriori
MIMO	Multiple Input Multiple Output
MISO	Multiple Input Single Output
MKL	Multiple Kernel Learning
NREL	National Renewable Energy Laboratory
PEM	Prediction Error Minimization
PID	Proporzionale, Integrale e Derivativo
RKHS	Reproducing Kernel Hilbert Space
SIMEA	Sistema Integrato/distribuito di Monitoraggio Energetico ed Ambientale
SNR	Signal to Noise Ratio

SSEH Stable Spline with Exponential Hyperprior
VARX Vector AutoRegressive with eXogenous inputs

Sommario

La modellazione trova molte applicazioni nell'edilizia, come nella diagnosi energetica, nel management di edifici, nel controllo ottimo e nella caratterizzazione di componenti in loco. Invece di eseguire una modellazione basata su una precisa descrizione fisica dell'edificio è spesso conveniente impiegare metodi di identificazione, che ricavano il modello da una serie di dati ottenuti da un monitoraggio dell'edificio.

Il monitoraggio può essere eseguito facilmente, installando una rete con decine o centinaia di sensori wireless, che forniscono una risoluzione spaziale e temporale dei dati molto precisa. Si ottiene così una grande quantità di dati, che può rendere il problema di identificazione del modello difficilmente risolvibile con metodi PEM, che costituiscono lo stato dell'arte nel settore dell'identificazione. Inoltre questi metodi hanno bisogno di intervalli di osservazioni di lunga durata per fornire modelli accettabili.

Per effettuare il monitoraggio e il controllo di un edificio conviene usare un numero minore di sensori, che possono essere un sottoinsieme dei sensori utilizzati in fase di modellazione; eseguire la selezione dei sensori può essere però un problema critico. Per questa ragione nella presente tesi sono stati impiegati metodi di identificazione non-parametrica che favoriscono modelli sparsi per eseguire la selezione di sensori e sono adattati a situazioni in cui la durata delle osservazioni è ridotta.

Nella prima parte delle simulazioni i metodi di identificazione sono stati valutati impiegando diversi studi Monte Carlo, dove per ogni esperimento sono stati generati andamenti di ingressi, di rumori e un modello di riferimento sparso e sono stati calcolati i relativi andamenti delle uscite. Dagli andamenti di ingresso e di uscita sono stati identificati modelli impiegando metodi non-parametrici e metodi PEM, confrontandone le prestazioni. Per set di dati di dimensioni ridotte con i metodi non-parametrici si sono ottenuti risultati nettamente migliori di quelli forniti dai metodi PEM. All'aumentare della dimensione, i risultati dei metodi PEM si sono avvicinati a quelli dei metodi non-parametrici.

Nella seconda parte delle simulazioni i diversi metodi sono stati applicati a dati ottenuti da un monitoraggio di un edificio per il terziario esistente, confrontando ancora le prestazioni dei modelli ottenuti impiegando i diversi metodi di identificazione. Anche in questo caso i risultati ottenuti con i metodi non-parametrici sono migliori di quelli raggiunti dai diversi metodi PEM impiegati.

Introduzione

Motivazione

L'aumento del prezzo delle risorse energetiche e l'accresciuta sensibilità ambientale hanno attratto l'attenzione sulla riduzione dei consumi energetici. A questo proposito la Commissione Europea si è posta gli obiettivi di ridurre il consumo energetico del 20%, di ridurre la produzione di CO_2 del 20% e di fornire il 20% dell'energia totale da fonti energetiche rinnovabili fino al 2020. In Europa gli edifici sono i maggiori responsabili del consumo finale di energia, seguiti dai trasporti, l'industria e l'agricoltura, come rappresentato in Figura 1 (sinistra). In particolare, nel settore edile circa 40% dell'energia primaria è usata per la climatizzazione degli ambienti [1], più del 10% per la produzione di acqua calda sanitaria (in Figura 1 (destra) è riportata la ripartizione del consumo per edifici residenziali): in questi ausili si trovano le maggiori potenzialità di risparmio.

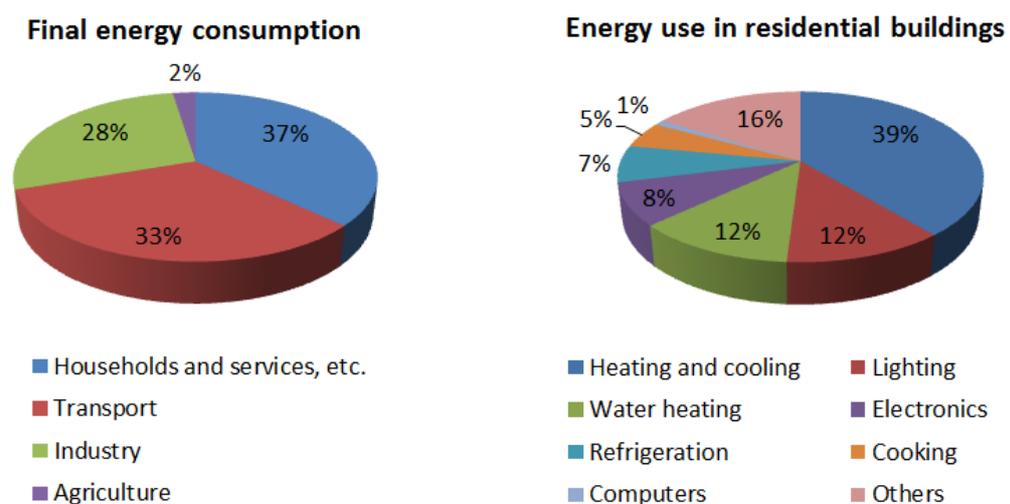


Figura 1 – Ripartizione del consumo energetico nei diversi settori (sinistra) [2] e ripartizione del consumo energetico per applicazioni nel settore edile residenziale (destra) [1].

Curando l'efficienza energetica nelle fasi di progettazione e di costruzione di edifici nuovi, ma soprattutto la ristrutturazione energetica di edifici esistenti, si può ridurre sensibilmente il consumo energetico [3, 4]. In entrambi i casi l'ottimizzazione del controllo del sistema edificio-impianto riveste un ruolo determinante per raggiungere specifici obiettivi prestazionali. È dimostrato infatti che la combinazione di metodi di controllo evoluti, monitoraggio di andamenti fisici e tecnologie di comunicazione possono migliorare sia il consumo energetico, sia il comfort [5, 6]. Oltre alla progettazione di moderni Building Energy Management System (BEMS), tale combinazione può servire per la progettazione di strumenti per la valutazione delle prestazioni energetiche di un edificio esistente. In questo contesto la modellazione delle caratteristiche di edifici assume un ruolo cruciale.

Stato dell'arte

In generale, ed in particolare anche nel settore edile, il problema della modellazione matematica può essere affrontato seguendo due approcci differenti, come spiegato in dettaglio nel capitolo 1: la modellazione in avanti e la modellazione basata su dati.

La modellazione in avanti si basa su una descrizione dettagliata dell'edificio come per esempio delle sue caratteristiche fisiche, degli impianti installati e dei profili di occupazione, e utilizza metodi di calcolo stazionari e dinamici. I metodi stazionari impiegano misure mediate nel tempo (tipicamente sul mese) e servono a calcolare il fabbisogno di energia da soddisfare con specifiche applicazioni quali riscaldamento, raffrescamento e produzione di acqua calda sanitaria. Questi metodi sono di norma impiegati per eseguire la certificazione energetica [7], per la diagnosi delle prestazioni energetiche e per la progettazione di impianti. Esistono norme che specificano come devono essere applicati tali metodi [8], che costituiscono la base di programmi di calcolo che supportano la progettazione energetica e impiantistica, come ad esempio *MC4Suite*[©] e *EdilClima*[®].

Con i metodi dinamici di modellazione in avanti si possono calcolare invece andamenti in funzione del tempo delle diverse grandezze di interesse: anche questi sono di solito usati per simulare il consumo energetico in edifici. Per riuscire a determinare gli andamenti temporali, questi metodi hanno bisogno di dati di input più dettagliati rispetto ai metodi stazionari: per esempio servono dati meteo con una precisa risoluzione temporale, deve essere descritto in maniera dettagliata il comportamento dell'impianto e devono essere specificate le masse termiche dei

vari componenti di un edificio. I programmi di calcolo più noti che impiegano questi metodi sono *EnergyPlus*TM e *Trnsys*[©]. Tutti questi programmi sono molto complessi da usare; sono utilizzabili solo da persone esperte e richiedono molto tempo per essere accurati. In molti casi gli andamenti simulati sono comunque diversi da quelli reali per fenomeni non considerati e per l'incertezza associata all'uso reale dell'edificio.

Un buon catalogo di programmi di calcolo che impiegano metodi stazionari o dinamici di modellazione in avanti sono presentati sul sito del Dipartimento dell'energia degli Stati Uniti d'America, sezione Efficienza Energetica e Energia Rinnovabile [9].

Per le difficoltà associate ai metodi di modellazione in avanti può essere conveniente impiegare metodi basati su dati monitorati per ricavare modelli utilizzabili per la valutazione delle prestazioni energetiche di edifici esistenti [10, 11]. Anche per l'implementazione di BEMS, l'utilizzo di modelli basati su dati assume un aspetto cruciale. Infatti la capacità della predizione di andamenti termici e igrometrici di tali modelli può essere utilizzata per esempio per progettare controllori predittivi, che hanno mostrato avere prestazioni superiori a quelle di controllori del tipo Proporzionale, Integrale e Derivativo (PID) [5, 6]. Un ulteriore ambito (nel settore edilizio) in cui si utilizzano i metodi basati su dati è la diagnosi di determinate componenti [12, 13]. Bisogna sottolineare il fatto che in tutti questi casi i metodi basati su dati possono essere applicati ovviamente solo su edifici già esistenti, dato che l'utilizzo di tali metodi ha senso solo se i dati sono ricavati da un monitoraggio reale.

Nel contesto della modellazione basata su dati nel settore edile, due diverse strutture tecnologiche per il monitoraggio possono essere applicate [14]: sistemi wireless e sistemi cablati. I sistemi cablati prevedono l'installazione permanente di pochi sensori in posizioni favorevoli scelte con cura e servono a fornire i dati di monitoraggio al BEMS. Il sistema wireless invece è di solito costituito da un gran numero di sensori, posizionati spesso in modo quasi-casuale e sono spesso installati temporaneamente. Rispetto ai sistemi cablati l'installazione di questi sistemi è meno costosa e meno dispendiosa poiché non intrusiva, proprio per la mancanza del cablaggio. Tali sistemi wireless sono usati spesso per valutare le prestazioni energetiche, inoltre spesso supportano la progettazione di BEMS nella fase di modellazione e la determinazione delle posizioni strategiche per l'installazione di sensori cablati e fissi. Siccome per sistemi wireless si impiegano di solito un gran numero di sensori, anche per monitoraggi di breve durata (giorni/settimane) si

ottiene una grande quantità di dati, che i metodi di identificazione devono essere in grado di gestire. In particolare alcuni metodi effettuano una selezione del sottoinsieme di sensori che fornisce le maggiori informazioni, fornendo quindi un'indicazione sulla localizzazione più favorevoli dei sensori.

Gli strumenti più usati nella modellazione basata su dati nel settore edilizio si basano su modelli Autoregressivo con ingressi esogeni (ARX), Autoregressivo a media mobile con ingressi esogeni (ARMAX) (ottenuti di solito impiegando metodi PEM) e reti neurali [5, 11, 15, 16]. Questi tipi di modelli sono però difficili da gestire nel caso si abbia un gran numero di ingressi e uscite in confronto al numero di dati a disposizione, situazione comune nel caso appena descritto, in cui il monitoraggio è eseguito per qualche giorno/settimana con molti sensori wireless. Recentemente sono state proposte numerose tecniche di identificazione che, favorendo modelli sparsi, effettuano la selezione di un sottoinsieme di sensori, adattandosi bene a situazioni in cui il numero di ingressi e uscite (sensori) è grande se paragonato al numero di dati disponibili. Risultati molto promettenti sono stati raggiunti da Least Absolute Shrinkage and Selection Operator (Lasso) [17], Least Angle Regression (LAR) [18] e le loro estensioni a gruppi [19]. In [20] è stato sviluppato il metodo SSEH per sistemi Multiple Input Single Output (MISO). Questo metodo è stato confrontato in [20–22] con diverse versioni PEM e con altri metodi di identificazione come Group Lasso (GLasso), Group Least Angle Regression (GLAR) e Multiple Kernel Learning (MKL). Rispetto ad essi SSEH si è mostrato più adatto alla condizione di aver un gran numero di ingressi e uscite in confronto al numero di dati.

A differenza dei metodi Lasso, LAR e le loro estensioni a gruppi, SSEH adotta un punto di vista Bayesiano per eseguire una stima non-parametrica delle risposte impulsive in uno spazio infinito-dimensionale. Questo metodo si basa su un paradigma sviluppato in [23]: al fine di garantire la condizione di stabilità Bounded Input Bounded Output (BIBO), le risposte impulsive sono modellate come processi Gaussiani che adattano come autocovarianza i cosiddetti *stable spline kernels* introdotti in [23] con fattori di scala incogniti e differenti. A questi fattori di scala si assegnano descrizioni “a priori” di tipo esponenziale (chiamati in letteratura “iperpriori”) che equivale a includere un termine di penalità di tipo l_1 che è ben noto favorire la sparsità. Una volta che tutti gli iperparametri del kernel sono noti, le risposte impulsive si ottengono applicando il “Representation Theorem” [24].

Obiettivi

Come già accennato, l'applicazione della modellazione basata su dati a edifici esistenti ha diversi vantaggi rispetto alla modellazione in avanti e può essere impiegata per esempio per la certificazione energetica, per la progettazione di BEMS e per effettuare diagnosi di componenti di edifici. Per motivi di costo e di facile applicazione conviene impiegare sistemi wireless per eseguire il monitoraggio, posizionando un grande numero di sensori, ma effettuandolo solo per qualche giorno. Ci si trova così spesso in situazioni in cui si ha un gran numero di ingressi e uscite, e quindi un grande numero di parametri da stimare, in confronto alla durata dei dati a disposizione. Occorre quindi individuare un sottoinsieme di sensori capace di contenere tutte le informazioni utili per un'applicazione successiva. In [20–22] è stato sviluppato il metodo SSEH per sistemi MISO ed è stato testato in condizioni simili a quelle richieste, mostrandosi migliore di altri metodi di identificazione che attualmente costituiscono lo stato dell'arte.

L'obiettivo della presente tesi è di generalizzare inizialmente il metodo SSEH, sviluppato in [20] per sistemi MISO, a sistemi Multiple Input Multiple Output (MIMO) e di testarlo e confrontarlo con il metodo PEM impiegando diversi studi Monte Carlo. Successivamente si vuole applicare il metodo SSEH a dati reali, monitorati in un edificio esistente, analizzando in particolare la selezione di sensori effettuata da diverse versioni di SSEH e confrontando ancora i risultati con quelli ottenuti impiegando il metodo PEM.

Struttura della tesi

La presente tesi si divide fondamentalmente in due parti: una parte teorica comprendente i capitoli 1 e 2, e una parte relativa alle simulazioni comprendente i capitoli da 3 a 6.

In capitolo 1 si riporta una panoramica dei diversi approcci di modellazione che sono impiegati nel settore edile. In particolare, come esempio dell'approccio di modellazione in avanti, non considerato più in seguito, è presentato brevemente il programma di simulazione energetica EnergyPlus. Nel capitolo 2 è presentata la teoria dell'identificazione dei sistemi, descrivendo inizialmente il metodo PEM e considerando in particolare il problema della selezione della struttura del modello. Di seguito si introduce il metodo non-parametrico SSEH, limitandosi prima a sistemi MISO e generalizzando poi la trattazione ai sistemi MIMO.

Segue la seconda parte più applicativa della tesi. Nel capitolo 3 sono descritte le impostazioni adoperati nelle diverse simulazioni, spiegando prima il setup degli studi Monte Carlo, poi il monitoraggio effettuato nell'edificio e infine le diverse versioni di SSEH con le differenti procedure di selezione e di PEM utilizzate nelle simulazioni. I risultati degli studi Monte Carlo sono riportati in capitolo 4. Inizialmente si evidenzia la dipendenza delle prestazioni dei metodi SSEH e PEM dalla lunghezza del set di dati e del metodo SSEH dalle lunghezze delle risposte impulsive. Si mostrano poi le prestazioni di diverse procedure di selezione delle variabili e selezione di risposte impulsive del metodo SSEH. Nella parte finale del capitolo si confrontano diverse versioni dei metodi di identificazione SSEH e PEM. In capitolo 5 sono riportati i risultati delle simulazioni eseguite impiegando i dati monitorati nell'edificio. Si confrontano ancora i metodi di identificazione SSEH e PEM e si mostrano poi le selezioni effettuate da procedure di selezione delle risposte impulsive. In capitolo 6 si affronta inizialmente l'onere computazionale dei metodi SSEH e PEM e poi si discutono i risultati delle simulazioni riportate nei capitoli 4 e 5.

Infine si traggono le conclusioni su quanto visto nella presente tesi e si delineano prospettive su possibili lavori futuri da svolgere.

Notazioni

Di seguito sono spiegate le notazioni utilizzate nella trattazione della presente tesi.

Il simbolo $\mathbf{p}(\cdot)$ denota la funzione di densità di probabilità di una qualche variabile. Con $\mathbb{E}[\cdot]$ si denota l'aspettazione, mentre $\hat{\mathbb{E}}[\cdot|\cdot]$ denota il miglior stimatore lineare, che nel caso Gaussiano coincide con l'aspettazione condizionata. Con $\hat{\cdot}$ in generale si denota la stima di una qualche variabile.

Con I si intende, se non diversamente specificato, la matrice identità di dimensione opportuna. Per una qualche matrice $A \in \mathbb{R}^{n \times m}$, $A^{[i,j]}$ denota l'elemento di A nella i -esima riga e j -esima colonna, A' denota la trasposta della matrice A . Se A è un vettore (che si assume sempre essere un vettore colonna) si usa la notazione $A^{[i]}$ invece di $A^{[i,1]}$ o $A^{[1,j]}$. Con $A^{[-i]}$ si denota un vettore in cui l' i -esimo elemento è stato soppresso; il simbolo $\|\cdot\|_p$ denota la norma p di un qualche vettore; se $p = 2$ il pedice viene omissso. La notazione $l_1(\mathbb{Z}^+)$ è usata per indicare lo spazio delle sequenze reali a dimensione infinita $g := [g_1, g_2, \dots]'$ per le quali vale $\|g\|_1 := \sum_{i=1}^{\infty} |g_i| < \infty$.

Dato un processo stocastico $\{s_t\}_{t \in \mathbb{Z}^+}$ e un intero T si usano le seguenti notazioni per denotare rispettivamente i vettori del passato e del futuro (finito) all'istante t :

$$s_t^- := [s'_{t-1}, s'_{t-2}, \dots], \quad s_t^+ := [s'_t, \dots, s'_{t+T-1}].$$

Per lo specifico istante temporale $t = t_0$, che sarà usato successivamente, si usano le notazioni semplificate $s^- := s_{t_0}^-$ e $s^+ := s_{t_0}^+$.

Capitolo 1

Modellazione termica di edifici: considerazioni teoriche

Con il termine “modello” di un qualche sistema fisico, biologico, economico o altro si intende una descrizione matematica delle relazioni tra alcune variabili e proprietà del sistema. Un modello consiste di tre elementi fondamentali [10]:

1. **Variabili di ingresso:** Chiamate anche variabili esogene, regressori o variabili indipendenti, agiscono sul sistema. Le variabili di ingresso possono essere classificate in variabili controllabili (come per esempio la velocità di rotazione di un ventilatore) e variabili non controllabili dall'utente (come i dati climatici).
2. **Struttura e parametri/proprietà del sistema:** Fornisce una descrizione fisica del sistema (per esempio la massa termica o le proprietà meccaniche di un elemento).
3. **Variabili di uscita:** Chiamate anche variabili endogene o variabili dipendenti, descrivono la reazione del sistema. Variabili di uscita sono per esempio la temperatura all'interno di una stanza o il consumo di energia di un edificio.

La modellazione matematica applicata a sistemi fisici si occupa di determinare uno dei tre elementi quando gli altri due sono noti. Fondamentalmente esistono due approcci distinti per affrontare questo problema, la modellazione in avanti (o diretta) e la modellazione all'indietro (basata su dati), e un approccio ibrido che combina i due metodi [11]. I differenti metodi sono schematizzati in Figura 1.1. In letteratura, i modelli ottenuti seguendo l'approccio in avanti sono anche chiamati di tipo “white-box”, quelli ottenuti seguendo un approccio ibrido “gray-box” e

infine quelli ottenuti da una modellazione basata su dati di tipo “black-box”.

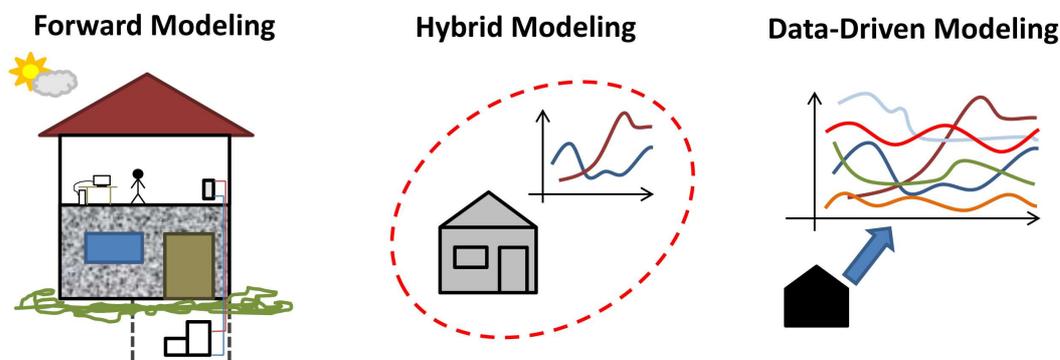


Figura 1.1 – *Diversi approcci di modellazione: La modellazione in avanti (forward modeling) necessita di una precisa descrizione delle proprietà fisiche, degli impianti installati, degli profili orari di occupazione e di funzionamento delle apparecchiature, delle condizioni climatiche e altro. Per la modellazione basata su dati (data-driven modeling) servono andamenti temporali che riescano a descrivere il modello. La modellazione ibrida (hybrid modeling) è una composizione degli altri due metodi: si usano semplici legami fisici derivando i parametri mancanti da qualche andamento monitorato (ne serve un numero ridotto rispetto al metodo basato su dati) [11].*

Modellazione in avanti L’approccio classico della modellazione in avanti ha come obiettivo la predizione di qualche variabile di uscita per variabili di ingresso specificate e richiede la conoscenza dettagliata della struttura e dei parametri del sistema. L’aumento della potenza computazionale ha consentito la costruzione di modelli sempre più complessi, assicurando un’accuratezza accettabile. Il grande vantaggio di questo approccio sta nel fatto che il sistema non deve esistere fisicamente per poter prevedere il suo comportamento. Per questo motivo è l’approccio ideale nella fase di progettazione e di analisi preliminare degli edifici per ottenere una stima delle prestazioni, in particolare per la modellazione del consumo energetico di un edificio. In questo caso la modellazione parte da una descrizione fisica molto dettagliata del sistema e delle componenti di interesse, come la localizzazione geografica e la geometria dell’edificio, le caratteristiche fisiche (stratigrafie dei componenti opachi e descrizione dei componenti vetriati), il tipo di impianto per la climatizzazione Heating Ventilation Air-Conditioning (HVAC), i profili orari di occupazione e di uso delle apparecchiature. Usando la descrizione fisica e i dati climatici relativi alla zona in cui è situato l’edificio come variabili di ingresso, si possono simulare o predire il carico di picco e il consumo istantaneo di energia

(variabili di uscita), impiegando il modello “white-box”. Esempi di software di simulazione che si basano su questo approccio sono *BLAST*, *DOE-2*, *TRNSYS*[®] e in particolare *EnergyPlus*[™], che sarà presentato nella sezione seguente. In Figura 1.2 è illustrato l’ordine in cui i programmi di simulazione energetica lavorano tipicamente.

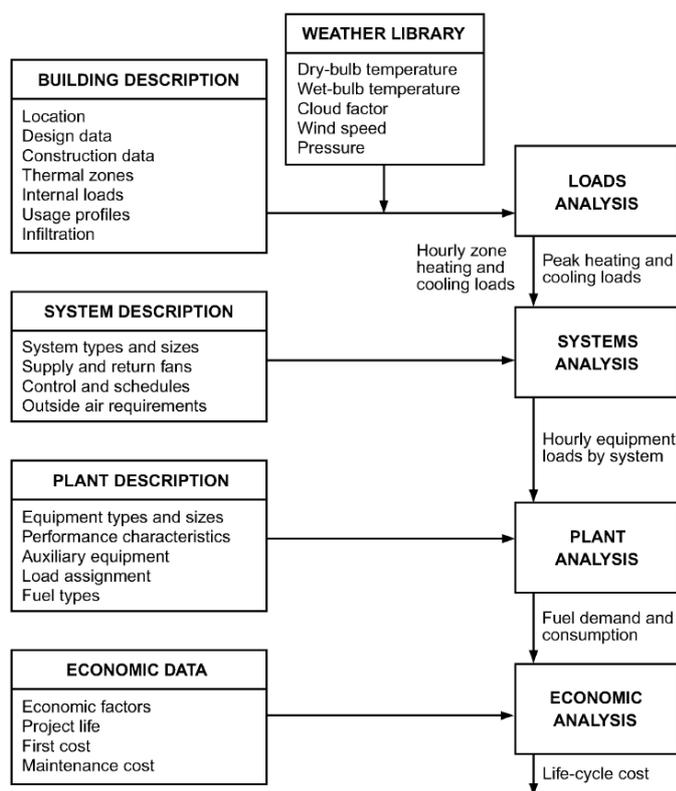


Figura 1.2 – Diagramma di flusso eseguito da programmi di simulazione energetica di edifici impiegando l’approccio in avanti [10].

Modellazione basata su dati In questo approccio, i dati di ingresso e di uscita, ricavati da misure e quindi noti, determinano una descrizione matematica cioè un modello del sistema da cui sono stati rilevati. Di questo modello non solo i parametri sono sconosciuti, ma spesso pure la struttura (sia la forma che l’ordine) deve essere determinata dai dati. Il processo di determinazione del modello (parametri e se sconosciuta anche la struttura) è chiamato identificazione di modelli del sistema. Al contrario del metodo in avanti, l’approccio all’indietro è rilevante solo se il sistema è già esistente e si possono ottenere dati di prestazione appropriati per l’identificazione di un modello. Con l’impiego di dati ottenuti da

esperimenti specifici, eseguiti con l'obiettivo di eccitare il sistema con una gamma di situazioni che durante il suo funzionamento normale non si presentano, si ottengono modelli più accurati. Spesso il sistema non permette tali esperimenti e i modelli devono essere identificati da dati ottenuti monitorando il funzionamento standard del sistema, come accade per edifici abitati.

Modelli ottenuti da un processo di identificazione del sistema sono più semplici e spesso anche più accurati nella predizione del comportamento del sistema di modelli ottenuti seguendo un approccio in avanti. L'identificazione è molto usata in diversi ambiti come la fisica, la biologia, l'ingegneria e l'economia ma non è ancora molto impiegata nell'ambito della progettazione degli edifici [10].

I potenziali settori di applicazione dell'identificazione di modelli per edifici sono i seguenti [11]:

- diagnosi energetica: per esempio per correggere i dati di consumo presunti in relazione alle condizioni climatiche correnti o per determinare certe caratteristiche di un edificio (come il coefficiente di perdita del calore $[W/K]$);
- management di edifici: i modelli possono condurre alla diagnosi dei componenti di un edificio;
- controllo ottimo: i modelli permettono la gestione intelligente e più efficace degli accumuli termici;
- caratterizzazione di componenti sul posto: i modelli possono determinare per esempio la trasmittanza termica di un muro esistente;
- sviluppo di modelli semplificati: l'identificazione può definire modelli computazionalmente efficienti. A differenza degli altri settori sopra elencati, spesso i modelli semplificati sono ricavati da modelli ottenuti dall'approccio in avanti e quindi non basati su edifici esistenti.

Nei seguenti capitoli si presentano metodi di identificazione studiati per la predizione e per il controllo del comportamento del sistema: si confrontano il cosiddetto PEM, che rappresenta lo stato dell'arte in questo ambito, con un nuovo metodo non parametrico, chiamato SSEH, che esegue una selezione dei dati ed è quindi adatto a situazioni in cui non sono disponibili molti dati in confronto al numero di parametri da stimare.

Modellazione ibrida

Questo approccio usa legami fisici semplici tra gli ingressi e le uscite del sistema, determinando i parametri sconosciuti del modello matematico con l'uso di metodi di identificazione. Quindi questo approccio, come l'approccio all'indietro, è basato su dati misurati di sistemi esistenti, ma ha il vantaggio di avere modelli con un significato fisico noto e quindi più facili da gestire. Per questo motivo metodi ibridi sono molto usati nell'ambito del controllo degli impianti di edifici [5,6]. Svantaggi rispetto ai metodi basati su dati sono che spesso non si conoscono legami facili tra ingressi e uscite e serve sempre un grande sforzo per ricavarli.

Validazione e test dei modelli Un aspetto importante riguarda la validazione dei modelli ottenuti tramite i diversi metodi e quindi la validazione dei diversi approcci di modellazione. Sono state sviluppate tre metodologie fondamentali per testare l'accuratezza dei programmi di simulazione energetica [25]:

- Validazione empirica: si confrontano i risultati ottenuti da un programma o un algoritmo con dati monitorati di un edificio reale, in un laboratorio o in apposite camere di prova (per esempio celle di test PASLINK [26]).
- Validazione analitica: si confrontano i risultati di una simulazione con i risultati di una soluzione analitica nota o con metodi di calcolo numerico.
- Test comparativi: Si confronta il programma o l'algoritmo con se stesso e con altri programmi che sono già stati validati.

Ciascuno di questi metodi presenta svantaggi specifici rispetto agli altri [25]. Il metodo analitico, per esempio, può essere applicato solo in casi molto semplici di cui si conosce la soluzione analitica. Il test comparativo non prende in considerazione soluzioni reali o misurate. La validazione empirica comporta un alto costo dovuto al setup sperimentale e, nonostante ciò, si possono manifestare errori sperimentali che introducono delle incertezze. Per limitarle, spesso si combinano le diverse metodologie, come nell'esempio illustrato in Figura 1.3. L'interconnessione tra i diversi metodi di validazione può essere diversa, per esempio anticipando il test comparativo alla verifica analitica, così da poter definire esperimenti più efficaci.

Esistono anche diversi metodi standardizzati per validare software di simulazione energetica come il *BESTest* delle National Renewable Energy Laborato-

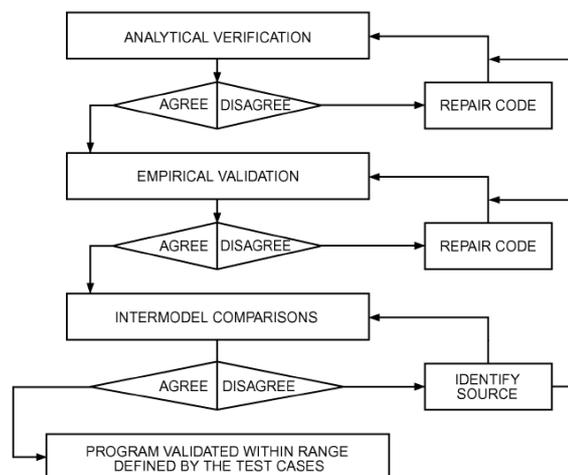


Figura 1.3 – Esempio di una procedura che integra i metodi fondamentali per ottenere una validazione di programmi di simulazione energetica più accurata [25].

ry (NREL) o lo ASHRAE *Standard 104*. Lavoro considerevole in questo ambito è stato fatto dalla rete *DYNASTEE - PASSYS*, supportato dalla Commissione Europea.

1.1 Software di simulazione energetica Energy-PlusTM

Uno dei programmi di calcolo più usato per eseguire simulazioni energetiche di edifici è EnergyPlusTM [27] ed è presentato qui come esempio di programmi di modellazione in avanti dinamici. EnergyPlusTM è disponibile gratuitamente sul sito del Dipartimento dell'energia degli Stati Uniti d'America, sezione Efficienza Energetica e Energia Rinnovabile [28].

EnergyPlusTM ha le sue radici nei due programmi di simulazione BLAST e DOE-2, sviluppati alla fine degli anni '70 e inizio degli anni '80. Entrambi si rivolgono a ingegneri e architetti che vogliono dimensionare impianti HVAC, sviluppare un'analisi dei costi di una ristrutturazione, ottimizzare il consumo energetico e altro.

EnergyPlusTM ha conservato gli obiettivi dei due programmi precedenti e la prima versione del programma di simulazione energetica dinamica è stata presentata nell'Aprile 2001. Partendo da una descrizione geometrica, fisica e impiantistica

dell'edificio, EnergyPlus™ calcola i carichi di riscaldamento e di raffreddamento determinati da un setpoint di temperatura, il consumo di energia primaria, l'illuminamento di zone specifiche e altre grandezze legate al funzionamento di un edificio. Per dare un'idea delle potenzialità di EnergyPlus si elencano qui alcune caratteristiche:

- Passo di tempo definibile dall'utente;
- Soluzioni basati sull'equilibrio termico;
- Modello di trasferimento di calore verso il suolo;
- Modello di trasferimento di massa e di calore;
- Modello solare anisotropo;
- Modelli di comfort termico;
- Controllo della luce diurna;
- Sistemi HVAC controllabili in catena chiusa;
- Calcolo delle emissioni inquinanti prodotte dall'edificio;
- Interfaccia con altri programmi di simulazione;

Per ulteriori dettagli si guardi [27] e la documentazione tecnica di EnergyPlus™ disponibile sul sito [28].

Si noti che EnergyPlus™ non ha una interfaccia utenti grafica; i file di input, del tempo e di output sono file di testo in formato ASCII. Il codice è però "open source" ed è formattato in modo da consentire a terzi lo sviluppo di interfacce grafiche che facilitino l'uso del software di simulazione da parte di persone non esperte di programmazione. Esempi di interfacce grafiche sono il Plug-in *Open-Studio* per Google SketchUp e *DesignBuilder*.

EnergyPlus™ è stato validato tramite diversi metodi standardizzati come il BESTest sviluppato dell'IEA o l'HVAC test sviluppato dall'ASHRAE. Per ulteriori informazioni si consulti la pagina web di EnergyPlus™ [28].

Capitolo 2

Identificazione di sistemi: considerazioni teoriche

Nel campo dell'identificazione dei sistemi, si cerca di proporre algoritmi capaci di costruire in maniera automatica un modello dinamico che descriva l'evoluzione temporale di qualche variabile di interesse, a partire da una serie di dati osservati. Come già accennato nel capitolo 1, i dati osservabili si classificano in ingressi o variabili esogene, spesso definiti dall'utente e denotati con $u(t)$, e uscite o variabili endogene, di norma misurabili in istanti discreti e indicate con $y(t)$.

Esistono motivazioni diverse alla base della definizione del modello di un sistema. In questa tesi l'interesse è incentrato su modelli utili per la predizione o per il controllo di un sistema esistente. In altre parole, l'obiettivo della tesi è sviluppare algoritmi per costruire modelli che predicano o controllino l'evoluzione futura delle variabili del sistema. Questo genere di richiesta fatta al modello impone delle restrizioni sul tipo di dati che possono essere usati per l'identificazione.

Ogni modello appartiene a una classe specifica ovvero a un insieme di modelli che rispondono a determinate caratteristiche, come per esempio la classe dei modelli lineari, tempo-invarianti e di dimensione finita. Il problema di identificazione sta quindi nello scegliere il miglior modello appartenente ad una certa classe (come quella appena accennata) avendo a disposizione un set di dati con certe proprietà statistiche minime (che riescono a spiegare la scelta di un modello).

Nell'ambito dell'identificazione dei sistemi, i metodi parametrici, basati sulla minimizzazione dell'errore di predizione (PEM), sono da lungo tempo lo stato dell'arte [29–31]. Il principio sul quale si basano è molto semplice ed è descritto nella sezione 2.1. Le proprietà statistiche di PEM sono note quando è fissata la struttura del modello, appartenente alla classe di modelli suddetta, evento che raramente si verifica in applicazioni reali. Deve quindi essere risolto un problema

di selezione da un'insieme di strutture potenziali del modello, problema che si riduce spesso alla selezione di un'ordine appropriato di una funzione di trasferimento o di un modello in spazio di stato, che di norma è affrontato impiegando criteri come Akaike Information Criterion (AIC) e Bayesian Information Criterion (BIC) [32, 33]. Se si applica PEM con selezione a dati sperimentali spesso si ottengono risultati molto scarsi anche se, come noto dalla teoria, l'algoritmo PEM senza selezione di modello è asintoticamente efficiente per processi con innovazione Gaussiana. Questi risultati scarsi si ottengono soprattutto nel caso di sistemi con un gran numero di parametri in confronto al numero di dati disponibili.

Numerose tecniche di stima nuove, adatte proprio a sistemi con un grande numero di parametri in confronto al numero di dati disponibili, sono state proposte di recente. Risultati molto promettenti sono stati raggiunti da Lasso [17], LAR [18] e le loro estensioni a gruppi [19]. In questa tesi è stato utilizzato il cosiddetto algoritmo SSEH, che a differenza dei precedenti, adotta il punto di vista Bayesiano per costruire un predittore lineare ottimo. Il punto di partenza è un nuovo paradigma di identificazione sviluppato in [23], che esegue una stima non-parametrica delle risposte impulsive. Invece di assumere a posteriori una struttura della funzione di trasferimento a dimensione finita (per esempio di tipo "Box-Jenkins", ARMAX o ARX) la risposta impulsiva del sistema è cercata in uno spazio a dimensione infinita. Siccome il problema in questi termini è mal posto, si utilizzano i metodi di regolarizzazione Bayesiana per correggerlo. In particolare, le risposte impulsive sono modellate come processi Gaussiani che adottano come autocovarianza i cosiddetti *stable spline kernel*, comprendenti la condizione di stabilità BIBO.

L'approccio SSEH, che è stato sviluppato in [20] per sistemi MISO, estende il paradigma non-parametrico alla costruzione di predittori lineari ottimi per sistemi sparsi. Per un sistema con m ingressi e p uscite, si devono stimare $(m + p)p$ risposte impulsive (dato che il predittore ha come ingressi gli m ingressi del sistema più le p uscite passate), che sono modellate come realizzazioni di processi Gaussiani. Come autocovarianze dei processi Gaussiani si impongono *stable spline kernels* con fattori di scala differenti ed incogniti. A questi fattori di scala si assegnano degli iperpriori esponenziali con un'ipervarianza comune. La sparsità è favorita da una penalità l_1 sugli iperparametri del kernel, al contrario di GLAR, dove la sparsità è favorita considerando la somma delle norme l_1 delle singole risposte impulsive. Una volta selezionati i kernel, le risposte impulsive si ottengono applicando il "Representer Theorem" [24]. L'algoritmo SSEH per sistemi MISO è stato confrontato in [21] con PEM (con selezione di variabili impiegando AIC e BIC) e con GLAR e si è mostrato migliore.

L'approccio SSEH è spiegato in dettaglio nella sezione 2.2, riferendosi prima al caso MISO e generalizzando poi il procedimento al caso MIMO.

2.1 Considerazioni teoriche relative al PEM

2.1.1 Descrizione del principio di funzionamento di PEM

In questa sezione è descritto il principio su cui si basano i metodi parametrici di minimizzazione dell'errore di predizione (PEM) [29]. Fissata la struttura del modello $M(\theta)$ appartenente a una classe parametrica $\mathcal{M} = \{M(\theta); \theta \in \Theta\}$ e date le sequenze di dati $\{y_t\}_{t=t_0, \dots, N}$ e $\{u_t\}_{t=t_0, \dots, N}$ si esegue il seguente procedimento:

1. Si costruisce un predittore per y_t all'istante $t-1$ per un valore di θ generico. Questo predittore può essere denotato col simbolo $\hat{M}(\theta)$ ed è una funzione deterministica dei dati fissando θ . Dal predittore è emessa la predizione di y_t in base ai dati misurati

$$\hat{M}(\theta) : (\{y_l\}_{l=t_0, \dots, t-1}, \{u_l\}_{l=t_0, \dots, t-1}) \rightarrow \hat{y}_{t|t-1}(\theta)$$

Poiché la predizione $\hat{y}_{t|t-1}(\theta)$ è una funzione dei dati, oltre che del parametro θ , prima di aver i dati di misura può essere pensata come una quantità aleatoria.

2. Al variare di θ si calcolano gli errori di predizione

$$e_t(\theta) := y_t - \hat{y}_{t|t-1}(\theta), \quad t = t_0, \dots, N$$

3. Si minimizza rispetto a θ una cifra di merito, che fornisce un'informazione su quanto bene il predittore (che è funzione di θ) predice y_t . Esempi di cifre di merito sono l'errore quadratico medio di predizione

$$V_N(\theta) := \frac{1}{N} \sum_{t=t_0}^N e_t(\theta)^2 \quad (2.1)$$

oppure la media pesata degli errori di predizione quadratici,

$$V_N(\theta) := \frac{1}{N} \sum_{t=t_0}^N \beta(N, t) e_t(\theta)^2, \quad \beta(N, t) \geq 0$$

dove il peso $\beta(N, t)$ è una qualche funzione non negativa, che, se N è piccolo, attribuisce pesi minori agli errori di predizione computi nella fase iniziale,

quando maggiore è l'influenza delle condizioni iniziali stimate in modo approssimate. Al crescere di N il peso diventa inutile e quindi $\beta(N, t) \rightarrow 1$. In ogni caso, minimizzando la cifra di merito si ottiene una stima del parametro θ

$$\hat{\theta}_N = \arg \min_{\theta} V_N(\theta) \quad (2.2)$$

che è la stima PEM del parametro del modello.

4. Come ultimo passo, si prende come stima della varianza dell'innovazione l'errore quadratico medio residuo dato da (2.1).

Nel caso in cui:

- si possa calcolare il predittore lineare a minima varianza d'errore;
- il modello reale che genera i dati appartenga alla stessa classe di modelli parametrici \mathcal{M} scelti per il PEM;
- i dati soddisfino determinate proprietà come la stazionarietà e la proprietà di persistente eccitazione,

si può dimostrare che per $N \rightarrow \infty$ lo stimatore PEM è consistente, cioè che il parametro stimato $\hat{\theta}_N$ converge al parametro vero. Sperimentazioni pratiche dimostrano che, se la struttura del modello vero è nota e il numero di parametri da stimare non è troppo elevato, per un set di dati di dimensione attorno a 10^2 , si ottengono già risultati molto soddisfacenti. Per la dimostrazione della consistenza e per altri risultati teorici interessanti dello stimatore PEM come la distribuzione asintotica delle stime, si rimanda a [29–31].

Il PEM è legato ad altri approcci impiegati nell'identificazione dei sistemi:

- nel caso di modelli della struttura ARX, l'identificazione PEM si riduce a un problema ai minimi quadrati, che può essere risolto esplicitamente. Per questo l'uso di modelli ARX è molto conveniente;
- se si assume che il disturbo ha una distribuzione Gaussiana, allora il PEM diventa identico al metodo “maximum likelihood” (stimatore di massima verosimiglianza) definito da

$$\hat{\theta} = \arg \max_{\theta} L(\theta),$$

dove $L(\theta)$ è una funzione di “likelihood”, cioè una densità di probabilità delle osservazioni condizionata al parametro θ .

2.1.2 Determinazione della struttura del modello

Per la determinazione di una appropriata struttura di un modello da dati osservati, in generale sembra che manchino dei risultati soddisfacenti. Un'eccezione è la determinazione dell'ordine per modelli lineari, per la quale esistono vari metodi [30]. Un approccio molto usato per determinare un ordine adeguato è testare un certo numero di modelli diversi e scegliere quello che fornisce il risultato più soddisfacente. In pratica, prima si fissa la struttura del modello, per esempio ARMAX, ARX o modelli in spazio di stato, poi si aumenta iterativamente l'ordine del modello e infine si confrontano i risultati dei singoli modelli. Aumentando l'ordine e quindi usando un numero maggiore di parametri, migliora il "FIT" (3.1), cioè la bontà con la quale il modello riesce a descrivere i dati osservati. Ad un certo punto l'aumento di ordine non è più di nessuna utilità, poiché al crescere del numero di parametri aumenta anche la varianza delle stime. Questo aumento della varianza rende il modello inservibile quando è usato su nuovi dati, non impiegati per l'identificazione.

Si fa di seguito riferimento al caso di identificazione PEM, dove la stima del parametro $\hat{\theta}_N$ è ottenuta minimizzando la funzione di merito $V_N(\theta)$ come descritto dall'equazione (2.2). Se si aumenta il numero di parametri del modello, il minimo della funzione decresce al variare di θ , perché si aggiungono nuovi gradi di libertà al problema di ottimizzazione. Un andamento tipico del minimo della funzione di merito $V_N(\hat{\theta})$ all'aumentare dell'ordine del modello è riportato in Figura 2.1.

Il confronto tra i modelli può essere visto come un test sulla significanza del decrescere di $V_N(\hat{\theta})$, eseguito per esempio mediante un "F-test" [11]. Un altro approccio, che sfrutta comunque lo stesso principio, è usare un criterio che dia informazioni sulla bontà dei diversi modelli. Un tale criterio penalizza in qualche maniera il decrescere della funzione $V_N(\hat{\theta})$ all'aumentare del numero di parametri del modello. Si seleziona il modello che dà il minimo valore per questo criterio. Il criterio può per esempio essere della forma

$$W_N = N \log V_N(\hat{\theta}) + \gamma(N, p),$$

dove p è il numero di parametri da stimare e $\gamma(N, p)$ è il termine aggiuntivo che penalizza un p grande. Varie scelte per la funzione $\gamma(N, p)$ sono state proposte come per esempio:

- AIC è determinato della scelta $\gamma(N, p) = 2p$ [32];

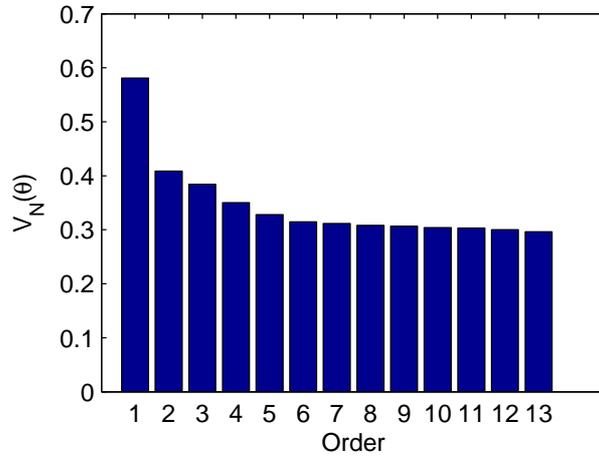


Figura 2.1 – Tipico andamento di una funzione di merito $V_N(\hat{\theta})$ all'aumentare dell'ordine del modello. Questo andamento è stato ottenuto usando un modello vero in spazio di stato del 5° ordine, mentre la struttura dei modelli d'identificazione era del tipo ARX.

- corrected Akaike Information Criterion (AICc) è dato dalla scelta $\gamma(N, p) = 2p \left(\frac{N}{N - p - 1} \right)$ [34,35]. Questa versione generalizzata di AIC è da preferire nel caso in cui il numero di parametri p è piccolo in confronto al numero di dati osservati N [36];
- BIC è dato da $\gamma(N, p) = p \cdot \log N$.

Oltre alla selezione dell'ordine appropriato, questi metodi sono adatti anche a determinare ritardi nel modello.

È importante notare che in condizioni reali spesso non esiste una struttura e un'ordine del modello “giusto”, ma la struttura e l'ordine scelti dipendono spesso dalle circostanze nelle quali sono impegnati.

2.2 Considerazioni teoriche relative a SSEH

Il metodo SSEH è stato introdotto all'inizio di questo capitolo e qui è spiegato in dettaglio. Inizialmente il problema è posto per sistemi MISO, cioè sistemi che hanno più ingressi e una sola uscita, seguendo [20, 21, 37]. Nella parte finale di questa sezione si descrive come generalizzare l'approccio a sistemi MIMO.

2.2.1 Descrizione del problema

Siano $\{u_t\}_{t \in \mathbb{Z}}$ e $\{y_t\}_{t \in \mathbb{Z}}$ due processi stocastici Gaussiani congiuntamente stazionari, che per il momento abbiano le dimensioni $u_t \in \mathbb{R}^m$ e $y_t \in \mathbb{R}$. Con un piccolo abuso di notazione con u_t si indica sia una variabile causale (del processo stocastico $\{u_t\}_{t \in \mathbb{Z}}$) che il suo valore campionato. Lo stesso vale per y_t . Questi due processi siano rispettivamente uscita e ingresso di un sistema dinamico tempo-invariante, del quale si cerca di identificare un modello dinamico tempo-invariante lineare della forma

$$y_t = \sum_{k=1}^{\infty} f_k u_{t-k} + \sum_{k=0}^{\infty} g_k e_{t-k} \quad (2.3)$$

da un set di dati misurati $\{u_t, y_t\}_{t=1, \dots, N}$. In (2.3) $f_k \in \mathbb{R}^{1 \times m}$ e $g_k \in \mathbb{R}$ sono le componenti sconosciute delle risposte impulsive e e_t è l'innovazione Gaussiana, cioè l'errore di predizione lineare a un passo

$$e_t := y_t - \hat{y}_{t|t-1} = y_t - \hat{\mathbb{E}}[y_t | y_t^-, u_t^-] \quad (2.4)$$

dove

$$\hat{\mathbb{E}}[y_t | y_t^-, u_t^-] := \sum_{j=1}^m \left[\sum_{k=1}^{\infty} h_k^{[j]} u_{t-k}^{[j]} \right] + \sum_{k=1}^{\infty} h_k^{[m+1]} y_{t-k}. \quad (2.5)$$

Si noti che e_t è un processo Gaussiano con media zero e si può definire $\sigma^2 := \mathbb{E}e_t^2$ la varianza di e_t . È utile definire la matrice di Toeplitz $A_{ti} \in \mathbb{R}^{T \times \infty}$, $i = 1, \dots, m+1$ dove

$$\begin{aligned} A_{ti}^{[j,k]} &:= u_{t+j-k-1}^{[i]}, & i = 1, \dots, m \\ A_{tm+1}^{[j,k]} &:= y_{t+j-k-1}, & j, k \in \mathbb{Z}^+ \end{aligned}$$

e per $t = t_0$ è usata la notazione semplificata $A_i := A_{t_0 i}$. Impiegando la matrice di Toeplitz il predittore può essere scritto nella forma

$$\hat{\mathbb{E}}[y_t | y_t^-, u_t^-] = \sum_{i=1}^m A_{ti} h^{[i]} + A_{t(m+1)} h^{[m+1]}. \quad (2.6)$$

Dalle equazioni (2.4) e (2.6) si può ricavare l'espressione per il futuro finito y^+ , cioè y_t nell'intervallo $t \in [t_0, t_0 + T - 1]$, che servirà in futuro:

$$y^+ = \sum_{i=1}^m +1 A_i h^{[i]} + e^+. \quad (2.7)$$

Le sequenze $h_k := [h_k^{[1]}, \dots, h_k^{[m]}, h_k^{[m+1]}] \in \mathbb{R}^{1 \times m}$, $k \in \mathbb{Z}^+$ sono i coefficienti delle risposte impulsive del predittore che sono tenuti a descrivere un sistema

BIBO stabile, cioè $h^{[i]} \in l_1(\mathbb{Z}^+)$, $i = 1, \dots, m+1$. Come descritto in sezione 2.1 l'identificazione nell'ambito PEM può essere vista come la stima di queste risposte impulsive h_k da un insieme finito di misure di ingressi e uscite. L'assunzione di non avere il feedback diretto (cioè $f_0 = 0$) in (2.3) semplifica il problema in quanto in condizioni deboli garantisce identificabilità. Si noti che il prodotto di matrici semi-infinite in (2.6) e (2.7) è da intendere come il limite di sequenze finite. Sotto l'assunzione $h^{[i]} \in l_1(\mathbb{Z}^+)$ comunque l'operazione di limite è ben posta e si può quindi usare formalmente l'espressione di limite.

Come accennato nell'introduzione di questo capitolo, si considerano situazioni in cui si ha un gran numero di ingressi m in relazione al numero di misure disponibili e solo alcuni di tali ingressi servono a predire y_t . Questo può accadere quando gli andamenti delle variabili d'uscita sono completamente indipendenti da quelli di uno specifico ingresso, cioè quando le due variabili non sono correlate, oppure in situazioni in cui tutte le informazioni utili per predire l'uscita, contenute nell'andamento di un ingresso, possono essere ottenute da combinazioni di altri ingressi più ricchi di informazioni. In termini matematici questo significa che, se l'ingresso $u^{[i]}$ non serve, si ha $\|h^{[i]}\|_1 = 0$. Graficamente queste situazioni possono essere illustrate in un grafo orientato, dove appare un arco diretto nel caso in cui $\|h^{[i]}\|_1 \neq 0$ per $i = 1, \dots, m$ e appare un cappio nel caso in cui $\|h^{[m+1]}\|_1 \neq 0$. In Figura 2.2 è rappresentato per esempio un sistema con $\|h^{[1]}\|_1 = \|h^{[5]}\|_1 = 0$, mentre $\|h^{[i]}\|_1 \neq 0$ per $i \in \{2, 3, 4, 6\}$. Questo significa che per predire y_t servono solo le misure passate di $u^{[2]}$, $u^{[3]}$, $u^{[4]}$ e di y stesso.

In pratica non è sempre evidente se un segnale misurato è significativo per la predizione di y_t . Metodi PEM standard, descritti per esempio in [29–31], non effettuano alcuna selezione di variabili e stimano l'intero modello usando tutti gli ingressi. Nelle situazioni di riferimento, in cui si ha un gran numero di ingressi in confronto alle misure a disposizione, questo può portare a risultati scarsi. Per questo motivo nel campo della selezione di variabili è stata fatta una ricerca molto vasta. I metodi classici sono descritti per esempio in [38,39], mentre una carrellata dei metodi più recenti è data in [40].

Il metodo di identificazione trattato in questo elaborato favorisce modelli sparsi. Dovrebbe quindi essere in grado di identificare la struttura di una rete, come per esempio quella in Figura 2.2, stimando tutte le risposte impulsive non nulle in (2.7).

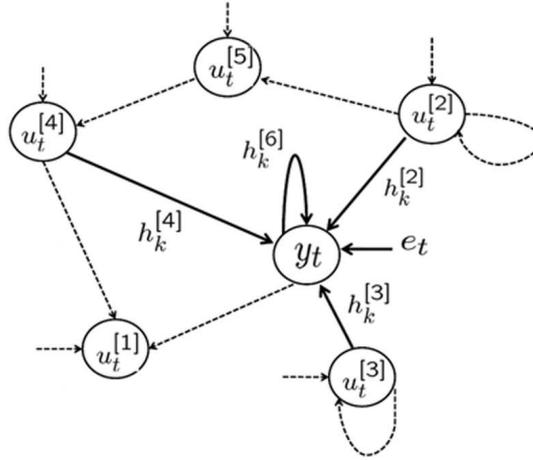


Figura 2.2 – Rete dinamica che rappresenta l'interazione (causalità di Granger) tra 6 variabili. Gli archi solidi rappresentano le connessioni di una rete dinamica relativi al nodo y_t . Con riferimento all'equazione (2.6) l'assenza degli archi da $u_t^{[i]}$, $i = 1, 5$ a y_t evidenzia che $h_k^{[1]} = h_k^{[5]} = 0$, $\forall k \in \mathbb{Z}^+$. Si noti la presenza di un arco entrante in y_t che mostra la dipendenza dall'errore a un passo di predizione e_t . Considerando invece una variabile $u_t^{[i]}$ gli archi tratteggiati indicano quali variabili servono a predire tale specifica variabile [21].

Rapporto di SSEH con altri metodi di identificazione sparsa

Per introdurre le idee principali alla base degli algoritmi di identificazione sparsa, si consideri il problema di ricostruire il parametro $\phi \in \mathbb{R}^m$ nel modello lineare

$$y_i = X_i' \phi + e_i, \quad i = 1, \dots, T \quad (2.8)$$

dove $X_i \in \mathbb{R}^m$ sono i T regressori, mentre $\{e_i\}$ è rumore Gaussiano a media nulla e varianza σ^2 . Nelle sezioni seguenti sarà spiegato in dettaglio come questi concetti sono usati nel campo dell'identificazione dei sistemi.

Se il numero m di regressori è grande in relazione al numero T di dati a disposizione, ottenere predittori accurati e stabili e modelli interpretabili diventa un problema impegnativo. In questo ambito negli ultimi anni è stata sviluppata una ricerca molto vasta, si guardi per esempio [17, 39, 41, 42].

Un lavoro pionieristico in questa direzione è il cosiddetto Lasso [17], dove la selezione di regressori è effettuata risolvendo un problema della seguente forma:

$$\hat{\phi} := \arg \min_{\phi} \sum_{i=1}^T (y_i - X_i' \phi)^2 + \gamma \|\phi\|_1, \quad (2.9)$$

dove γ è il cosiddetto parametro di regolarizzazione. Nell'ambito Bayesiano $\hat{\phi}$ può essere visto come lo stimatore Maximum A Posteriori (MAP)

$$\hat{\phi} := \arg \max_{\phi} \mathbf{p}(\{y_i\}|\phi)\mathbf{p}(\phi) \quad (2.10)$$

dato che ϕ è indipendente dal rumore di misura, imponendo che ϕ abbia una densità di probabilità esponenziale del tipo

$$\mathbf{p}(\phi) \propto e^{-\epsilon\|\phi\|_1}$$

e settando $\gamma = 2\epsilon\sigma^2$ in modo che (2.9) e (2.10) si equivalgono.

Recentemente il Lasso è stato esteso anche alla stima di modelli con rumore auto-regressivo e per modelli Vector AutoRegressive with eXogenous inputs (VARX). Nonostante le buone proprietà, il Lasso non è molto usato nella pratica per la sua scarsa efficienza computazionale. È stato sviluppato l'algoritmo LAR, che con una piccola modifica può essere visto come una implementazione efficiente del Lasso [18], e la sua versione a gruppi GLAR [19].

Seguendo un approccio Bayesiano la stessa formulazione può essere ottenuta modellando le componenti di ϕ come variabili Gaussiane indipendenti con densità di probabilità

$$\mathbf{p}(\phi^{[i]}|\lambda_i) = \frac{1}{\sqrt{2\pi\lambda_i^2}} e^{-(\phi^{[i]})^2/2\lambda_i^2}$$

Si assume poi che anche gli iperparametri λ_i sono variabili causali con una certa densità di probabilità $p(\lambda_i)$. Questi iperparametri $\hat{\lambda}_i$ possono essere stimati impiegando la marginale di λ_i (chiamata anche *marginal likelihood*)

$$\mathbf{p}(\lambda) = \int \prod_i p(\phi^{[i]}|\lambda_i)p(\lambda_i)d\phi^{[i]}. \quad (2.11)$$

La stima degli iperparametri può poi essere usata per stimare il modello ϕ seguendo il paradigma empirico di Bayes [43]. Questo approccio forma la base per l'algoritmo SSEH.

Il problema dell'identificazione sparsa Il problema di identificazione qui considerato consiste nello stimare le risposte impulsive $h^{[i]}$ in (2.7) soggette alla condizione di stabilità $h^{[i]} \in l_1(\mathbb{Z}^+)$ per $i = 1, \dots, m + 1$. Come accennato in precedenza, l'interesse è incentrato su stimatori adatti a predire y selezionando in automatica le variabili che portano informazione tra $u^{[1]}, \dots, u^{[m]}, y$. In altre parole ci si aspetta che alcune risposte impulsive $\hat{h}^{[i]}$ saranno nulle. La risoluzione di questo problema richiede la stima di variabili raggruppate [19], ma la

peculiarità qui è che i gruppi stanno in spazi di dimensione infinita. Questo è simile all'ottenimento di soluzioni sparse del problema lineare, però con qualche complicazione non triviale:

- Nell'identificazione parametrica, cioè nel caso in cui le risposte impulsive siano modellate in classi di modelli a dimensione finita [29–31], ogni gruppo di parametri descrive una risposta impulsiva. Restringendosi alla classe di modelli ARX/FIR si ottiene un algoritmo di selezione di variabili, che può essere chiamato ARX-GLAR. Nel caso generale di modelli IIR la parametrizzazione non è lineare e, per stimare l'ordine della classe parametrica che descrive le risposte impulsive, un'ulteriore problema di selezione di modello dovrebbe essere affrontato. Per questo qui si preferisce lavorare in ambito non parametrico come descritto in sezione 2.2.2 in modo che gli $h^{[i]}$ facciano parte di uno spazio a dimensione infinita.
- I parametri incogniti sono le risposte impulsive modellate come processi Gaussiani seguendo i principi spiegati in sezione 2.2.2. La formulazione del problema che seguirà sarà simile al *multiple kernel learning* [44].

2.2.2 Kernels per l'identificazione di sistemi

Nell'identificazione parametrica le risposte impulsive $h^{[i]}$ sono parametrizzate in un dominio di dimensione finita, per esempio usando modelli ARMAX. Seguendo questo approccio sorgono due problemi. Il primo problema è che bisogna selezionare un modello, cioè affrontare un problema di stima dell'ordine, che di solito è risolto tramite criteri come AIC o BIC (si guardi [32, 33]). Il secondo problema è che in generale la parametrizzazione di un predittore non è lineare e quindi spesso l'identificazione richiede la soluzione di un problema di minimizzazione non lineare e non convesso. Nell'ambito di regressione Gaussiana invece una risposta impulsiva $h^{[i]}$ è considerata una realizzazione di un processo stocastico \mathbf{h} . In particolare questo processo è modellato come processo Gaussiano indipendente a media nulla e con covarianza $\lambda_i^2 K$ (vedi [23, 45, 46]). Il kernel K definisce un Reproducing Kernel Hilbert Space (RKHS) \mathcal{H}_K , al quale appartengono gli stimatori Bayesiani delle risposte impulsive $h^{[i]}$ (si guardi [24] e Proposizione 4)

$$\hat{h}^{[i]} = \mathbb{E}[h^{[i]} | \{\hat{y}_t, u_t\}], \quad i = 1, \dots, m + 1. \quad (2.12)$$

Dato che la funzione di covarianza (2.17) delle risposte impulsive $h^{[i]}$ è proporzionale al kernel K , di seguito i termini kernel e covarianza saranno usati come

sinonimi. Nel RKHS è definita una norma $\|\cdot\|_{\mathcal{H}_K}$ che, come sarà mostrato dopo, controlla la complessità della funzione da ricostruire regolarizzando il processo di stima.

Definendo una struttura adeguata di un kernel in un contesto Bayesiano, è possibile incorporare delle informazioni a priori sul processo stocastico. Spesso queste informazioni a priori sono limitate al fatto che il processo stocastico \mathbf{h} , e possibilmente qualche sua derivata, sono a tempo continuo con energia limitata. In questo caso un approccio molto usato è modellare il processo stocastico come un integrale multiplo di dimensione j di un rumore bianco Gaussiano. Se il rumore bianco ha varianza unitaria, l'autocorrelazione di \mathbf{h} è data da [46]

$$W_j(s, t) = \int_0^1 G_j(s, u)G_j(t, u)du, \quad (2.13)$$

$$G_j(r, u) = \frac{(r - u)_+^{j-1}}{(j-1)!}, \quad (u_+) = \begin{cases} u & \text{per } u \geq 0 \\ 0 & \text{per } u < 0 \end{cases}$$

Nel caso particolare $j = 2$ si ottiene il cosiddetto “cubic spline kernel”. In Figura 2.3 (sinistra) sono riportati 100 realizzazioni di un processo Gaussiano che ha un'autocovarianza proporzionale a W_2 .

Nell'ambito dell'identificazione di sistemi, si ha l'informazione aggiuntiva sulla stabilità delle risposte impulsive, che non è fornita dal kernel definito in (2.13). In effetti, come si può vedere da Figura 2.3 (sinistra), la varianza delle realizzazioni cresce nel tempo. In particolare è facile vedere che, se l'autocovarianza di \mathbf{h} è proporzionale a W_j , la varianza di una sua realizzazione è nulla per $t = 0$ e tende all'infinito al crescere di t . Nell'ambito dell'identificazione dei sistemi invece, sapendo che il processo stocastico rappresenta una risposta impulsiva stabile, si vuole avere una varianza finita per $t = 0$ che tende a zero al crescere di t . Queste proprietà sono assicurate considerando la classe di kernels

$$K_j(t, \tau) = W_j(e^{-\beta t}, e^{-\beta \tau}), \quad t, \tau \in \mathbb{R}^+ \quad (2.14)$$

dove il carattere β è un scalare positivo che regola la velocità di decadimento della varianza [23]. In pratica β sarà un parametro sconosciuto, che deve essere stimato dagli dati di misura. Nel caso particolare $j = 2$ in (2.14) si ottiene il cosiddetto “stable spline kernel” introdotto in [23]:

$$K_2(t, \tau) = \frac{e^{-\beta(t+\tau)}e^{-\beta \max(t,\tau)}}{2} - \frac{e^{-3\beta \max(t,\tau)}}{6}. \quad (2.15)$$

La stabilità delle risposte impulsive è garantita dalla seguente proposizione [23]:

Proposizione 1. *Sia \mathbf{h} Gaussiano a media nulla e autocovarianza K_2 . Allora con probabilità uno le realizzazioni di \mathbf{h} sono risposte impulsive continue di sistemi dinamici BIBO stabili.*

In Figura 2.3 (centro) sono riportati 100 realizzazioni di un processo \mathbf{h} con autocovarianza proporzionale a una stable spline kernel (2.15) con $\beta = 0.4$.

Come discusso in [46] spesso i predittori hanno qualche “dinamica veloce”. Infatti per modelli ARMAX, gli zeri del modello di rumore (se non avvengono delle cancellazioni) diventano i poli del predittore che sono condivisi da tutte le funzioni di trasferimento del predittore stesso. Nel caso in cui il modello di rumore è di tipo “low pass”, esso ha gli zeri in alta frequenza che poi diventano i poli in alta frequenza del predittore. In queste condizioni la varianza a priori (2.15) potrebbe risultare troppo lenta, il che imporrebbe un bias allo stimatore. È quindi utile arricchire il kernel con una componente parametrica che modella le componenti ad alta frequenza. Questo può essere fatto modellando le risposte impulsive $h^{[i]}$ come la convoluzione di un processo Gaussiano indipendente a media nulla e con autocovarianza data dalla versione campionata di $\lambda_i^2 K_2$ con una risposta impulsiva parametrica r . Qui si fa la restrizione a componenti parametriche r , la cui zeta-trasformata $R(z)$ ha la seguente forma:

$$R(z) = \frac{z^2}{z^2 + \theta_1 z + \theta_2}, \quad \theta \in \Theta \subset \mathbb{R}^2 \quad (2.16)$$

dove la regione ammissibile Θ costringe i due poli di $R(z)$ ad appartenere al semicerchio unitario aperto sinistro del piano complesso. Realizzazioni di un processo con autocovarianza data dalla convoluzione di un stable spline kernel con una componente parametrica sono riportate in Figura 2.3 (destra), dalla quale si può osservare il ruolo della componente parametrica.

Si noti che potrebbe essere considerato anche un ordine più alto della parte parametrica. Questo richiederebbe la risoluzione di un problema di stima dell’ordine, ma come discusso in [46] questo porterebbe a una complicazione non necessaria, in quanto già usare due poli aggiunge abbastanza flessibilità allo stimatore. Concludendo, l’autocovarianza di ogni risposta impulsiva $h^{[i]}$ del predittore è definita da (2.15) e (2.16) come descritto in sezione 4.3 di [46] ed è denotata con

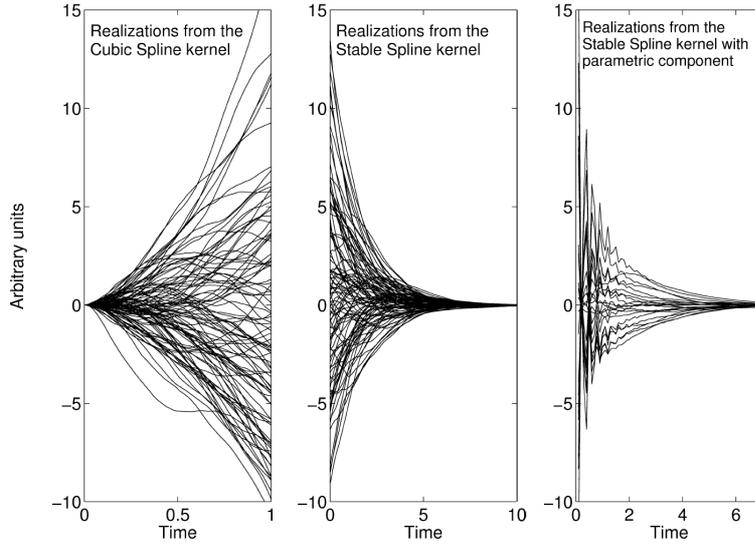


Figura 2.3 – Realizzazioni di un processo stocastico con autocovarianza proporzionale al “standard cubic spline kernel” (sinistra, si guardi [46]), al “stable spline kernel” (centro) e alla sua versione campionata arricchita con una componente parametrica con poli $-0.5 \pm 0.6j$ (destra) [20].

$K : N \times N \rightarrow \mathbb{R}$. Si ha quindi:

$$\mathbb{E}[h_i^{[i]} h_k^{[i]}] = \lambda_i^2 K(l, k, \theta, \beta), \quad i = 1, \dots, m+1, \quad l, k \in \mathbb{N}. \quad (2.17)$$

Si noti che il kernel in (2.17) dipende dagli iperparametri sconosciuti $\lambda_1, \dots, \lambda_{m+1}, \theta, \beta$. Per aumentare la flessibilità del modello, i fattori di scala λ_i possono essere differenti per le diverse risposte impulsive. I fattori relativi a K invece sono condivisi da tutte le $h^{[i]}$. Nel capitolo 2.2.3 sarà definito un iperparametro causale vettoriale ζ che raccoglie tutti questi parametri.

2.2.3 Un modello Bayesiano per l’identificazione di sistemi lineari sparsi

La varianza di rumore σ^2 sarà stimata in un passo preliminare usando un modello ARX come descritto in [47]. Nella descrizione Bayesiana σ^2 è quindi assunto come noto. Gli iperparametri β, θ e $\{\lambda_i\}$ invece sono descritti come vettori causali mutuamente indipendenti con le seguenti proprietà:

- β è data da una densità di probabilità non informativa in \mathbb{R}^+ ;
- θ ha una distribuzione uniforme nell’intervallo Θ che obbliga le due radici di $P_\theta(z)$ a stare nel semicerchio aperto sinistro del piano complessi (2.16);

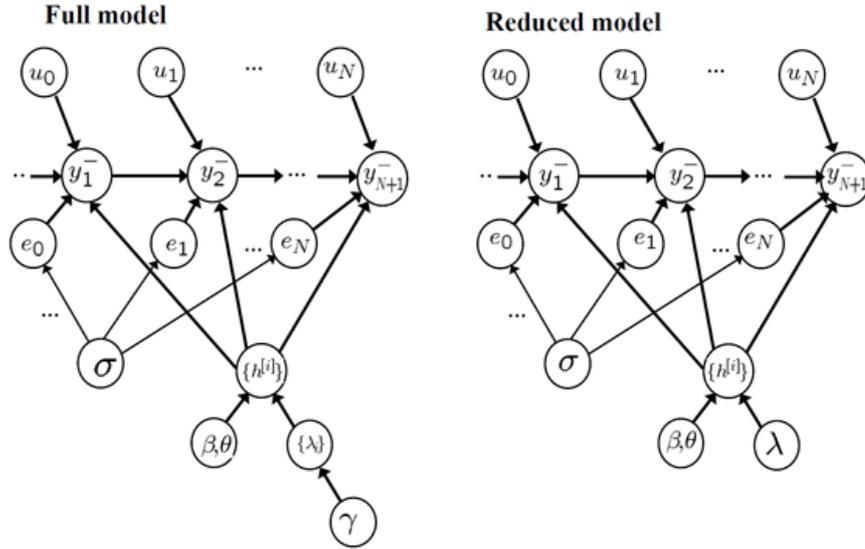


Figura 2.4 – Rete Bayesiana che descrive il modello non parametrico per l'identificazione di sistemi lineari sparsi, dove $y_t^- = [y_{t-1}, y_{t-2}, \dots]$ e, nel modello ridotto, $\lambda := \lambda_1 = \dots = \lambda_m + 1$ [21].

- ogni λ_i è una variabile causale indipendente con densità esponenziale e con l'inverso della media $\gamma \in \mathbb{R}^+$, cioè:

$$\mathbf{p}(\lambda_i) = \gamma e^{-\gamma \lambda_i} \chi(\lambda_i), \quad i = 1, \dots, m + 1 \quad (2.18)$$

con $\chi(x)$ la funzione indicatrice con $\chi(x) = 1$ se $x \geq 0$ e $\chi(x) = 0$ se $x < 0$. Anche il parametro γ è interpretato come una variabile causale con una densità di probabilità non informativa in \mathbb{R}^+ ed è stimato come descritto al “passo 2)” sotto.

È utile definire un iperparametro vettoriale causale ζ , che raccoglie i parametri da stimare appena citati:

$$\zeta := [\lambda_1, \dots, \lambda_{m+1}, \theta, \beta, \gamma]. \quad (2.19)$$

Il modello stocastico che descrive l'interazione delle diverse variabili è riportato in Figura 2.4 (sinistra).

Il problema della stima delle risposte impulsive $h^{[i]}$ da dati misurati $\{y_t, u_t\}$ può essere formulato come stimatore a minima varianza (2.12). Dato che in pratica y^- e u^- non sono mai completamente noti, bisogna introdurre delle approssimazioni. Per ottenere un'espressione computabile si introduce la seguente densità di

probabilità:

$$\begin{aligned} \mathbf{p}(y^+, \{h^{[i]}\}, y_t^-, u|\zeta) &\propto [\Pi_{t=t_0}^N \mathbf{p}(y_t|\{h^{[i]}\}, y_t^-, \zeta, u_t^-)] \mathbf{p}(y^-, \{h^{[i]}\}, u^-|\zeta) \\ &\approx [\Pi_{t=t_0}^N \mathbf{p}(y_t|\{h^{[i]}\}, y_t^-, \zeta, u_t^-)] \mathbf{p}(\{h^{[i]}\}|\zeta) \mathbf{p}(y_t^-, u_t^-). \end{aligned} \quad (2.20)$$

il primo \propto deriva dal fatto che il predittore di u_t dato il passato u_t^- e y_{t+1}^- non dipende dall'iperparametro ζ . Il secondo \approx deriva dall'assunzione che il passato (y^-, u^-) non contiene informazioni sulle risposte impulsive del predittore e sull'iperparametro ζ .

È ora richiamato un lemma introdotto in [46] che poi servirà in Proposizione 4 e Proposizione 3:

Lemma 2. *Siano i poli di $R(z)$ in (2.16) stabili. Allora, se $\{y_t\}$ e $\{u_t\}$ sono processi stocastici stazionari, a media nulla e varianza finita, ogni operatore $\{A_i\}$ è quasi sicuramente continuo in \mathcal{H}_K .*

2.2.4 Stima degli iperparametri e delle risposte impulsive del predittore

Stima degli iperparametri

Come descritto nella proposizione seguente, l'iperparametro vettoriale ζ è stimato ottimizzando la marginale a posteriori, che è la densità congiunta di y^+ , ζ e $h^{[i]}$ integrato rispetto agli $h^{[i]}$. La proposizione deriva da semplici cambiamenti delle densità di probabilità in gioco e il Lemma 2 garantisce che sia ben posto. Con un piccolo abuso di notazione ora K è un elemento di $\mathbb{R}^{\infty \times \infty}$, cioè il suo elemento i -esimo è la sequenza $K(\cdot|i)$, $i \in \mathbb{N}$. La dimostrazione della proposizione è simile a quella della Proposizione 4 in [46] con piccole modifiche valide in assenza di retroazione ed è qui omessa.

Proposizione 3. *Siano $\{y_t\}$ e $\{u_t\}$ due processi stocastici stazionari a media nulla e varianza finita. Allora, assumendo che valgano le condizioni di Lemma 2 e l'approssimazione (2.20), la stima “maximum a posteriori” di ζ dato y^+ vale*

$$\hat{\zeta} = \arg \min_{\zeta} J(y^+; \zeta), \quad \theta \in \Theta, \beta > 0, \lambda_i \geq 0, i = 1, \dots, m+1 \quad (2.21)$$

dove J è quasi certamente puntualmente ben posto ed è dato da (usando (2.20))

$$\begin{aligned}
J(y^+; \zeta) &:= -\log \mathbf{p}(y^+ | y^-, u, \zeta) \\
&= -\log \left[\int \mathbf{p}(y^+, \{h^{[i]}\}, y^-, u | \zeta) \mathbf{p}(\zeta) dh^{[1]}, \dots, dh^{[m+1]} \right] \\
&\approx \frac{1}{2} \log (\det[2\pi V[y^+]]) + \frac{1}{2} (y^+)' (V[y^+])^{-1} y^+ + \\
&\quad + \gamma \sum_{i=1}^{m+1} \lambda_i + \text{const} \tag{2.22}
\end{aligned}$$

con $V[y^+] = \sigma^2 I + \sum_{i=1}^{m+1} \lambda_i^2 A_i K A_i'$.

La funzione obiettivo (2.22) è una versione Bayesiana di quella legata al *multiple kernel learning* [48]. Un termine aggiuntivo è $1/2 \log (\det[2\pi V[y^+]])$, che penalizza la complessità del modello. Infatti questo termine è proporzionale a $\{\lambda_i\}$. Il secondo termine in (2.22) $1/2 (y^+)' (V[y^+])^{-1} y^+$ tiene conto dell'aderenza dei dati sperimentali e decresce quando $\{\lambda_i\}$ aumenta. Il termine $\gamma \sum_{i=1}^{m+1} \lambda_i$ ha l'effetto di includere un'aggiuntiva penalità l_1 su $\{\lambda_i\}$.

Una questione cruciale per l'uso pratico sta nel trovare un buon punto di partenza per il processo di ottimizzazione. Una soluzione molto usata è l'impiego di un modello Bayesiano ridotto (riportati in Figura 2.4 a destra), che porta a una soluzione sub-ottimale del problema di ottimizzazione. Questo modello consiste nel fissare $\lambda_i = \lambda$ per $i = 1, \dots, m+1$ comune a tutte le risposte impulsive $h^{[i]}$ e di rimuovere la penalità l_1 sugli λ_i , $i = 1, \dots, m+1$ settando γ a zero. Ottenuta la stima per λ si impone $\hat{\gamma} = 1/\hat{\lambda}$ che poi a sua volta è usato per fare la selezione delle variabili (impiegando una procedura Bayesiana di *forward selection*). Questo descritto può essere ottenuto seguendo la procedura seguente:

1. Si ottiene $\hat{\lambda}_i$ per $i = 1, \dots, m+1$, $\hat{\theta}$ e $\hat{\beta}$ risolvendo il seguente problema modificato di (2.22):

$$\arg \min_{\zeta} [J(y^+; \zeta) - (m+1)\gamma\lambda] \quad \text{con } \theta \in \Theta, \beta > 0, \lambda_1 = \dots = \lambda_{m+1} = \lambda \geq 0;$$

2. Si denota con I l'insieme degli indici delle variabili selezionate. Si impone $\hat{\gamma} = 1/\hat{\lambda}$ e si definisce $\bar{\zeta}_I := [\tilde{\lambda}_1, \dots, \tilde{\lambda}_{m+1}, \hat{\theta}, \hat{\beta}, \hat{\gamma}]$, dove $\tilde{\lambda}_i = \hat{\lambda}_i$ se $i \in I$ e $\tilde{\lambda}_i = 0$ altrimenti. Si inizializza $I = \emptyset$ e si ripete la seguente procedura:

- (a) **per** $j \in \{1, \dots, m+1\} \setminus I$ si definisce $I_j := I \cup j$ e si calcola $J(y^+; \bar{\zeta}_{I_j})$;

(b) si seleziona

$$\bar{j} := \arg \max_{j \in \{1, \dots, m+1\} \setminus I} J(y^+; \bar{\zeta}_{I_j}) - J(y^+; \bar{\zeta}_I); \quad (2.23)$$

(c) **se** $J(y^+; \bar{\zeta}_{I_j}) - J(y^+; \bar{\zeta}_I) > 0$ si pone $I := I_j$ e si torna al passo (a);
Altrimenti si esce dalla procedura.

Questo algoritmo cerca, ad ogni ciclo, tra gli indici che non sono ancora stati inclusi nel modello, quello che dà il maggior miglioramento della funzione obiettivo. Se la funzione obiettivo migliora effettivamente (cioè se $J(y^+; \bar{\zeta}_{I_j}) - J(y^+; \bar{\zeta}_I) > 0$), quell'indice \bar{j} è incluso e si inizia a cercare il successivo indice da includere. Altrimenti, cioè se $J(y^+; \bar{\zeta}_{I_j}) - J(y^+; \bar{\zeta}_I) \leq 0$, il set degli indici è completo e si esce dalla procedura. L'iperparametro $\bar{\zeta}_I$ trovato alla fine della procedura è usato come punto di partenza per il problema di ottimizzazione usando il modello Bayesiano intero. Si noti che a questo punto ci sono due possibili scelte. Una è di non considerare più le variabili che non sono state scelte dalla procedura appena vista. L'altra è di usare $\bar{\zeta}_I$ solo come punto di partenza per il problema di ottimizzazione, ottimizzando anche i λ_i che in $\bar{\zeta}_I$ erano nulle.

Oltre alla *forward selection* procedure standard di selezione di variabili, in statistica ci sono la *backward elimination*, che parte da un set di variabili pieno e toglie indici finché la funzione obiettivo migliora, e la *stepwise selection*, che è una combinazione degli altri due. Diverse versioni di questi algoritmi saranno spiegate e confrontate nella parte sperimentale di questo elaborato. Si noti comunque che esiste una letteratura molto vasta nell'ambito della selezione di variabili, noto anche sotto il nome di *feature selection*. È stato proposto un gran numero di algoritmi alternativi, che in determinati casi sono vantaggiosi rispetto alle procedure appena citate [49–51].

Stima delle risposte impulsive di un predittore per ζ fissato

Per un'iperparametro $\zeta = [\lambda_1, \dots, \lambda_{m+1}, \theta, \beta, \gamma]$ fissato, lo stimatore (2.12) può essere calcolato come descritto nella seguente proposizione [46]:

Proposizione 4. *Nelle stesse assunzioni di Lemma 2, approssimazione (2.20) e ipotizzando che e_t sia innovazione Gaussiana, quasi certamente si ha*

$$\{\hat{h}^{[i]}\}_{i=1}^{m+1} = \arg \min_{\{h^{[i]} \in \mathcal{H}_K\}_{i=1}^{m+1}} \|y^+ - \sum_{i=1}^{m+1} A_i h^{[i]}\|^2 + \sigma^2 \sum_{i=1}^{m+1} \frac{\|h^{[i]}\|_{\mathcal{H}_K}^2}{\lambda_i^2} \quad (2.24)$$

dove $\|\cdot\|$ indica la norma Euclidea. Inoltre, quasi certamente per $i = 1, \dots, m+1$ si ha

$$\hat{h}^{[i]} = \lambda_i^2 K A_i' c, \quad c = \left(\sigma^2 I + \sum_{i=1}^{m+1} \lambda_i^2 A_i K A_i' \right)^{-1} y^+. \quad (2.25)$$

Si noti che in (2.24) ogni σ^2/λ_i^2 rappresenta un parametro di regolarizzazione che impone un compromesso tra il fit di $y_t - \hat{y}(t|t-1)$ e la regolarità di $h^{[i]}$. Una volta stimate le risposte impulsive del predittore $h^{[i]}$, semplici equazioni possono essere usate per calcolare le risposte impulsive f e g in (2.3) e quindi il predittore di k -passi in avanti. Per dettagli si guardi [30].

Osservazione. Si noti che le risposte impulsive $h^{[i]}$ stanno nello spazio a dimensione infinita \mathcal{H}_K . Per questioni di implementazioni queste devono essere troncate a una certa lunghezza t_0 . Si sottolinea che questo troncamento non comporta nessun errore di bias e, se si hanno a disposizione sufficienti risorse computazionali, si può scegliere t_0 sufficientemente alto per considerare una lunghezza “adoperabile” delle risposte impulsive. La lunghezza “adoperabile” (si guardi [46]) dipende dall’iperparametro β in (2.15). Si noti infatti che il parametro di regolarizzazione $\|h^{[i]}\|_{\mathcal{H}_K}$ obbliga $\hat{h}^{[i]}$ a decadere in maniera esponenziale rendendo il troncamento irrilevante per usi pratici. Ovviamente se $h^{[i]}$ è troncato alla lunghezza t_0 , anche l’operatore A_i deve essere troncato. Per non appesantire le notazioni è stato però scelto di non esplicitare questo troncamento nelle formule.

2.2.5 Generalizzazione a sistemi MIMO

Quanto visto nelle sezioni precedenti per sistemi MISO è generalizzato in questa sezione a sistemi MIMO. Si considerino due processi stocastici Gaussiani congiuntamente stazionari $\{u_t\}_{t \in \mathbb{Z}}$ e $\{y_t\}_{t \in \mathbb{Z}}$ sia $u_t \in \mathbb{R}^m$ e $y_t \in \mathbb{R}^p$. Questi siano rispettivamente gli m ingressi e le p uscite (a differenza di quanto visto in precedenza, dove si aveva una sola uscita) di un sistema dinamico tempo-invariante che si cerca di descrivere mediante un modello lineare della forma

$$y_t = \sum_{k=1}^{\infty} f_k u_{t-k} + \sum_{k=0}^{\infty} g_k e_{t-k}. \quad (2.26)$$

$f_k \in \mathbb{R}^{p \times m}$ e $g_k \in \mathbb{R}^{p \times p}$ sono le componenti sconosciute della risposta impulsiva e $e_t \in \mathbb{R}^{p \times 1}$ è l’innovazione Gaussianiana definito da (2.4), cioè l’errore di predizione

lineare a un passo. Il predittore lineare a un passo può essere scritto nella forma

$$\begin{aligned}\hat{\mathbb{E}}[y_t^{[l]} | y_t^-, u_t^-] &:= \sum_{i=1}^m \left[\sum_{k=1}^{\infty} h_k^{[l,i]} u_{t-k}^{[i]} \right] + \sum_{i=1}^p \left[\sum_{k=1}^{\infty} h_k^{[l,i+m]} y_{t-k}^{[i]} \right] \\ &= \sum_{i=1}^m A_{ti} h^{[l,i]} + \sum_{i=1}^p A_{t(m+i)} h^{[l,m+i]}\end{aligned}\quad (2.27)$$

con $l = 1, \dots, p$. Le sequenze $h_k \in \mathbb{R}^{p \times m+p}$, $k \in \mathbb{Z}^+$ sono le risposte impulsive del predittore e $A_{ti} \in \mathbb{R}^{T \times \infty}$ sono le matrici di Toeplitz definite da:

$$\begin{cases} A_{ti}^{[j,k]} := u_{t+j-k-1}^{[i]}, & \text{per } i = 1, \dots, m \\ A_{ti}^{[j,k]} := y_{t+j-k-1}^{[i]}, & \text{per } i = m+1, \dots, m+p \end{cases} \quad \text{con } j, k \in \mathbb{Z}^+$$

Come nelle sezioni precedenti, oggetto di interesse sono ancora situazioni in cui il numero di variabili $m+p$ è grande rispetto al numero di dati a disposizione e alcune variabili sia di ingresso che di uscita non forniscono informazioni per la predizione di qualche variabile di uscita. In termini matematici questo vuol dire che, se una qualche variabile $i \in \{1, \dots, m, m+1, \dots, m+p\}$ non serve per predire una qualche variabile d'uscita $j \in \{1, \dots, p\}$, si ha $h^{[j,i]} = 0$. Similmente a quanto visto per sistemi MISO in Figura 2.2 questo può essere rappresentato graficamente da un grafo orientato, come quello in Figura 2.5. In tale figura per esempio per la predizione della uscita $y_t^{[1]}$ informazioni sono contenute in $u_t^{[1]}$, $u_t^{[1]}$, $y_t^{[2]}$ e $y_t^{[3]}$, mentre $u_t^{[3]}$, $u_t^{[4]}$ e il proprio passato $y_t^{[1]}$ non servono. Si noti che l'ingresso $u_t^{[3]}$ non contiene informazioni per la predizione di alcuna variabile di uscita.

Come visto nelle sezioni precedenti il metodo di identificazione qui considerato è in grado di identificare la struttura del sistema, simile a quello di Figura 2.5, favorendo modelli sparsi. Inoltre sono stimate le risposte impulsive non nulle del sistema

$$\hat{h}^{[i,j]} = \mathbb{E}[h^{[i,j]} | \hat{y}_t, u_t], \quad i = 1, \dots, p, \quad j = 1, \dots, m+p. \quad (2.28)$$

seguendo un approccio non parametrico: le risposte impulsive del predittore $h^{[i,j]}$ sono modellate come processi Gaussiani indipendenti a media nulla e con autocovarianza dato da

$$\mathbb{E}[h_l^{[i,j]} h_k^{[i,j]}] = \lambda_{ij}^2 K(l, k, \theta, \beta), \quad i = 1, \dots, p, \quad j = 1, \dots, m+p, \quad l, k \in \mathbb{N}. \quad (2.29)$$

La funzione $K : N \times N \rightarrow \mathbb{R}$ è il kernel visto in sezione 2.2.2, che definisce il RKHS, al quale appartengono le stime delle risposte impulsive $\hat{h}^{[i,j]}$ [24]. Assumendo

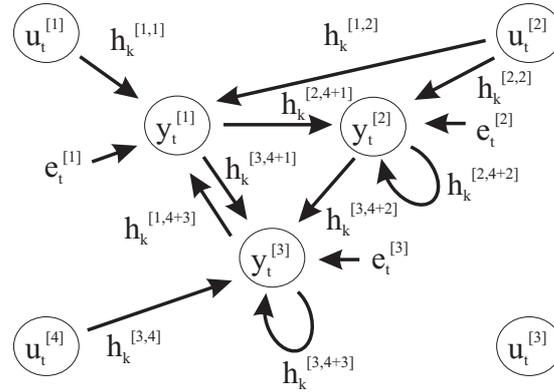


Figura 2.5 – Rete dinamica che rappresenta l'interazione (causalità di Granger) tra $p = 3$ variabili di uscita e $m = 4$ variabili di ingresso. Gli archi solidi rappresentano una relazione esistente nella predizione tra una variabile di ingresso j e una variabile di uscita i (in tale casi $h_k^{[i,j]} \neq 0$) oppure tra una variabili di uscita w e un'altra variabile di uscita v (in tale casi $h_k^{[v,m+w]} \neq 0$). Gli archi non riportati in figura indicano che la relativa risposta impulsiva è nulla. Si noti che possono esserci situazioni in cui una qualche variabile non influenza la predizione del sistema, come la variabile d'ingresso $u_t^{[3]}$ in figura. In grafico è rappresentato anche l'influenza dell'errore di predizione e_t .

ancora nota la varianza del rumore Σ^2 (dato che è stimata nell'inizializzazione) e assumendo che ogni λ_{ij} sia una variabile causale con distribuzione esponenziale dato da $\mathbf{p}(\lambda_{ij}) = \gamma e^{-\gamma \lambda_{ij}} \chi(\lambda_{ij})$ per $i = 1, \dots, p$ e $j = 1, \dots, m + p$, nel caso MIMO l'iperparametro vettoriale causale è dato da:

$$\zeta := [\lambda_{1,1}, \dots, \lambda_{1,m+p}, \dots, \lambda_{p,1}, \dots, \lambda_{p,m+p}, \theta, \beta, \gamma]. \quad (2.30)$$

Facendo gli ovvi adattamenti si può impiegare quindi il modello Bayesiano riportato in Figura 2.4 (sinistra) nel caso di sistemi MIMO: oltre a u_t anche y_t e e_t sono da intendere processi vettoriali, σ^2 diventa una covarianza matriciale Σ (matrice covarianza del processo di innovazione e_t) e gli $m + 1$ elementi di $h^{[i]}$ e di λ_i si estendono a $h^{[i,j]}$ e λ_{ij} , dove entrambi sono di numerosità $p(m + p)$. Le rimanenti variabili sono invece condivise per tutte le risposte impulsive delle diverse uscite. Il parametro β , che regola la velocità di decrescita del kernel, è scelto in maniera tale da riuscire a modellare la risposta impulsiva $h^{[i,j]}$ con la dinamica più alta. Così questo parametro è adeguato anche per la modellazione delle altre risposte impulsive. Il parametro γ individua la densità di tutte le λ_{ij} . Anche il parametro (vettoriale) θ è comune a tutte le risposte impulsive dato che serve a modellare le dinamiche veloci del predittore, che sono indotte dal modello

di rumore condiviso dalle risposte impulsive.

Il Lemma 2 e l'assunzione (2.20) rimangono validi nel caso MIMO e possono essere quindi applicati nella seguente proposizione, che è l'estensione di Proposizione 3 e descrive come si stima l'iperparametro vettoriale ζ nel caso di sistemi MIMO. Per dimostrare la seguente proposizione si assume inoltre che e_t sia un processo di innovazione Gaussiana a media nulla a variabili scorrelate, cioè un processo la cui matrice di covarianza sia una matrice diagonale:

$$\mathbb{E}[e_t e_t'] = \Sigma = \text{diag}\{\sigma_1^2, \dots, \sigma_p^2\}. \quad (2.31)$$

Proposizione 5. *Siano $\{y_t\}$ e $\{u_t\}$ due processi stocastici stazionari a media nulla e varianza finita. Allora, assumendo che valgono le condizioni di Lemma 2, (2.31) e (2.20), la stima "maximum a posteriori" di ζ dato y^+ vale*

$$\hat{\zeta} = \arg \min_{\zeta} J(y^+; \zeta) = \arg \min_{\zeta} \sum_{i=1}^p J(y^{[i]+}; \zeta), \quad (2.32)$$

$$\theta \in \Theta, \beta > 0, \lambda_{ij} \geq 0, i = 1, \dots, p, j = 1, \dots, m+p$$

dove J è quasi certamente puntualmente ben posto ed è dato da (usando (2.20))

$$\begin{aligned} J(y^{[i]+}; \zeta) &:= -\log \mathbf{p}(y^{[i]+} | y^{[i]-}, u, \zeta) \\ &= -\log \left[\int \mathbf{p}(y^{[i]+}, \{h^{[i,j]}\}, y^{[i]-}, u | \zeta) \mathbf{p}(\zeta) dh^{[i,1]}, \dots, dh^{[i,m+p]} \right] \\ &\approx \frac{1}{2} \log (\det [2\pi V[y^{[i]+}]]) + \frac{1}{2} (y^{[i]+})' (V[y^{[i]+}])^{-1} y^{[i]+} + \\ &\quad + \gamma \sum_{j=1}^{m+p} \lambda_{ij} + \text{const} \end{aligned} \quad (2.33)$$

con $V[y^{[i]+}] = \sigma_i^2 I + \sum_{j=1}^{m+p} \lambda_{ij}^2 A_{ij} K A_{ij}'$.

Dimostrazione. Dato che $J(y^+; \zeta) := -\log \mathbf{p}(y^+ | y^-, u, \zeta)$, se si riesce a dimostrare che $\mathbf{p}(y^+ | y^-, u, \zeta)$ può essere fattorizzato come

$$\mathbf{p}(y^+ | y^-, u, \zeta) = \prod_{i=1}^p \mathbf{p}(y^{[i]+} | y^{[i]-}, u, \zeta)$$

si ha che $J(y^+; \zeta) = \sum_{i=1}^p J(y^{[i]+}; \zeta)$. Questo equivale a considerare tutte le uscite del sistema separatamente, così da ricondurre il caso MIMO al caso MISO, per il quale vale la Proposizione 3.

Inizialmente è stato assunto che $\{u_t\}_{t \in \mathbb{Z}}$ e $\{y_t\}_{t \in \mathbb{Z}}$ siano due processi stocastici Gaussiani congiuntamente stazionari. Utilizzando anche l'assunzione (2.31) vale

$$\begin{aligned} \mathbf{p}(y^+ |, y^-, u, \zeta) &= \frac{1}{\sqrt{2\pi \det(\Sigma)}} \exp\left(-\frac{1}{2}(y_t - \hat{y}_{t|t-1})' \Sigma^{-1} (y_t - \hat{y}_{t|t-1})\right) \\ &= \frac{1}{\sqrt{2\pi \prod_{i=1}^p (\sigma_i^2)}} \exp\left(-\frac{1}{2} \sum_{i=1}^p \frac{(y_t^{[i]} - \hat{y}_{t|t-1}^{[i]})^2}{\sigma_i^2}\right) \\ &= \prod_{i=1}^p \mathbf{p}(y^{[i]+} |, y^{[i]-}, u, \zeta) \end{aligned}$$

e quindi la Proposizione 5 risulta dimostrata. \square

Come nel caso MISO, è importante avere un valido punto di partenza per il problema di ottimizzazione. Un buon metodo è usare un modello Bayesiano ridotto, simile a quello rappresentato in Figura 2.4 (destra) (con le modifiche equivalenti da fare per il modello Bayesiano pieno) ed eseguire una procedura equivalente a quella nel caso MISO:

1. Si ottengono $\hat{\lambda}$, $\hat{\theta}$ e $\hat{\beta}$ risolvendo il seguente problema modificato di (2.33):

$$\arg \min_{\zeta} [J(y^+; \zeta) - p(m+p)\gamma\lambda]$$

con $\theta \in \Theta$, $\beta > 0$ e $\lambda_{1,1} = \dots = \lambda_{p,m+p} = \lambda \geq 0$;

2. Si denota con I l'insieme delle coppie di indici (i, j) delle variabili selezionate. Si impone $\hat{\gamma} = 1/\hat{\lambda}$ e si definisce $\bar{\zeta}_I := [\lambda_{1,1}, \dots, \lambda_{p,m+p}, \hat{\theta}, \hat{\beta}, \hat{\gamma}]$, dove $\tilde{\lambda}_{ij} = \hat{\lambda}_{ij}$ se $(i, j) \in I$ e $\tilde{\lambda}_{ij} = 0$ altrimenti. Si inizializza $I = \emptyset$ e si ripete la seguente procedura:

(a) **per** $l \in \{1, \dots, m+p\} \setminus I$ si definisce $I_l := I \cup (i, l)$ con $i = 1, \dots, p$ e si calcola $J(y^+; \bar{\zeta}_{I_l})$;

(b) si seleziona

$$\bar{l} := \arg \max_{l \in \{1, \dots, m+p\} \setminus I} J(y^+; \bar{\zeta}_{I_l}) - J(y^+; \bar{\zeta}_I); \quad (2.34)$$

(c) **se** $J(y^+; \bar{\zeta}_{I_l}) - J(y^+; \bar{\zeta}_I) > 0$ si pone $I := I_l$ e si torna al passo (a); **Altrimenti** si esce dalla procedura.

Si noti che al passo (a) si aggiungono tutti i p indici relativi ad un ingresso al set delle variabili selezionate I , cioè qualora un ingresso sia considerato, è utilizzato per predire tutte le variabili di uscita. Nella fase di validazione dell'algoritmo

SSEH questa procedura è stata confrontata con altri metodi di identificazione. Nella seconda parte della validazione e nella fase di identificazione termica invece sono state testate anche altre procedure di selezione, che permettono di individuare l'insieme delle uscite, per le quali uno specifico ingresso può essere utile nella fase di predizione. Si riesce così a individuare la struttura di un sistema dinamico simile a quello riportato in Figura 2.5.

Una volta stimato l'iperparametro ζ , (facendo gli ovvi adattamenti per potere usare nel caso MIMO) Proposizione 4 può essere usata per ottenere le stime delle risposte impulsive $h^{[i,j]}$ con $i = 1, \dots, p$ e $j = 1, \dots, m + p$. Impiegando semplici formule dalle $h^{[i,j]}$ possono essere dedotte le risposte impulsive del sistema f e g e quindi anche il predittore a k passi [30].

Capitolo 3

Setup sperimentale per la validazione di SSEH

La validazione dell'approccio di identificazione SSEH si esegue in due fasi. La prima fase prevede studi Monte Carlo, durante i quali si eseguono un gran numero di esperimenti causali e si confrontano i risultati per poter trarre delle informazioni sulla bontà dei metodi impiegati: per ogni esperimento si generano in maniera causale ingresso, rumore e modello di riferimento (sparso) e si calcola l'uscita del corrispondente modello. Ai dati di ingresso e di uscita si applicano poi i diversi metodi di identificazione. Dopo aver eseguito tutti gli esperimenti di uno studio, si confrontano le prestazioni dei modelli ottenuti dai diversi metodi di identificazione nei vari esperimenti.

Si considerano studi Monte Carlo per valutare l'influenza del modo in cui sono generati i dati (correlati o no), la dimensione dei dati, la lunghezza delle risposte impulsive per l'approccio SSEH e diversi livelli di sparsità del modello di riferimento. Mediante lo studio Monte Carlo si confrontano inoltre diverse versioni sia dell'approccio SSEH che dell'approccio PEM e si confrontano diverse procedure di selezione delle variabili legate al metodo SSEH.

La seconda fase di validazione prevede di identificare modelli termodinamici da dati monitorati in un edificio esistente. Si confrontano ancora i modelli ottenuti dall'identificazione impiegando i diversi approcci in modo da analizzare le prestazioni in situazioni reali. In questo caso si presterà particolare attenzione alla selezione dei sensori eseguita dal metodo SSEH.

Il confronto tra i modelli ottenuti impiegando i diversi metodi di identificazione è condotta sulla base dell'indice FIT_k (a k passi di predizione (3.1)). Questo dà valori vicino ad uno se la predizione e l'uscita reale nel test set sono simili e

decrese al diminuire della somiglianza degli due andamenti (potendo assumere anche valori negativi). Il FIT_k è dato dalla formula

$$FIT_k := 1 - \frac{\|y_t^{test} - \hat{y}_{t|t-k}^{test}\|}{\|y_t^{test} - \bar{y}_t^{test}\|} \quad (3.1)$$

dove \bar{y}_t^{test} indica la media delle uscite del set di test $\{y_t^{test}\}_{t=1}^{N_{test}}$ e $\hat{y}_{t|t-k}^{test}$ è la predizione a k passi di tale uscite usando il modello identificato.

Per valutare le prestazioni delle procedure di selezione di SSEH negli studi Monte Carlo si calcolano le percentuali delle risposte impulsive, che sono state considerate in maniera corretta, separando i dati relativi agli ingressi (cioè relativi a $h^{[i,j]}$ per $i = 1, \dots, p$ e $j = 1, \dots, m$) e relative alle uscite (cioè relativi a $h^{[i,j]}$ per $i = 1, \dots, p$ e $j = m + 1, \dots, m + p$). Questa scelta è frutto dell'osservazione che, durante le simulazioni, la percentuale di risposte impulsive considerate in maniera corretta relative agli ingressi differisce da quella relativa alle uscite. Le percentuali si ottengono dalla somma del numero di risposte impulsive nulle sia nel modello di riferimento sia nel modello identificato, con il numero di risposte impulsive diverse da zero in entrambi i modelli, diviso il numero totale di risposte impulsive.

In certi casi invece, della percentuale di risposte impulsive, è stata calcolata la percentuale di ingressi o uscite considerate in maniera corretta, dato che determinate versioni della procedura di selezione (anche quella presentata in sezione 2.2.5) considerano le sole variabili e non le singole risposte impulsive. In questo caso con “considerato in maniera corretta” si intende che, se un ingresso o il passato di un'uscita dà contributo alla predizione del sistema, allora tale variabile deve essere selezionata dalla procedura di selezione.

Ovviamente non ha senso calcolare questi indici per i modelli ottenuti dalle identificazioni utilizzando il metodo PEM che non prevede la possibilità di eseguire la selezione delle variabili.

3.1 Setup degli studi Monte Carlo

Come accennato sono stati effettuati studi Monte Carlo per confrontare diverse versioni dell'approccio di identificazione non parametrico SSEH e dell'approccio parametrico PEM. Le diverse versioni utilizzate sono presentate in sezione 3.3.

In ogni studio Monte Carlo sono stati effettuati 200 esperimenti, per ciascuno dei quali sono stati generati:

- il modello sparso di riferimento;
- il set dei dati di training;
- il set dei dati di test.

In seguito si applicano i metodi di identificazione al set di dati di training, ottenendo un modello differente per ogni versione considerata. Successivamente, impiegando i dati del set di test, si calcolano le uscite di ciascun modello. Dalle diverse uscite dei modelli identificati e dall’uscita “vera” del modello di riferimento del set dei dati di test si possono quindi determinare gli indici *FIT*. Come accennato, i diversi *FIT* forniscono un’informazione sulla bontà di approssimazione delle uscite ottenute dai modelli identificati all’uscita del modello di riferimento. Confrontando i diversi *FIT* si possono classificare le versioni di identificazione impiegate. Questo procedimento è schematizzato in Figura 3.1.

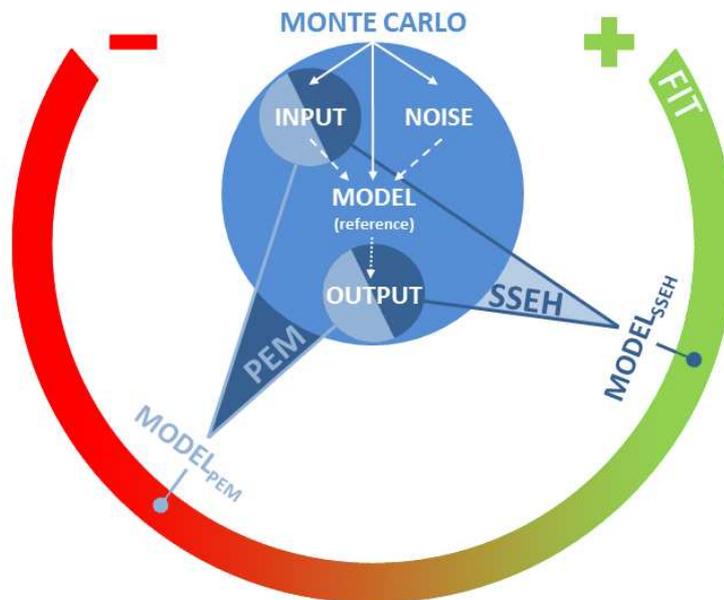


Figura 3.1 – Schema del procedimento seguito per ciascun esperimento dei diversi studi Monte Carlo.

Il modello sparso di riferimento Il modello lineare di riferimento nello spazio di stato di ordine $n = 30$ (ordine verosimile per sistemi reali) con $m = 10$ ingressi e $p = 3$ uscite ha la forma

$$\begin{cases} x_{t+1} = Ax_t + Bu_t + Ke_t \\ y_t = Cx_t + e_t \end{cases} \quad (3.2)$$

dove e_t è rumore bianco Gaussiano a media nulla e con varianza $\mathbb{E}(e_t e_t') = 0.04 \cdot I$ (scelta per avere una SNR accettabile per applicare i diversi processi di identificazione). Questo modello è generato impiegando la funzione `drmodel.m` di *MATLAB*[®] che genera in modo casuale un modello a tempo discreto, nello spazio di stato, della forma:

$$\begin{cases} x_{t+1} = Ax_t + \bar{B}\bar{u}_t \\ y_t = Cx_t + Du_t \end{cases} \quad (3.3)$$

poi trasformato per ottenere la forma (3.2). Per questo si annulla la matrice D e si usa il comando `drmodel.m` in modo da generare un modello della forma (3.3) con p uscite e $m + p$ ingressi. Si ottiene così una matrice \bar{B} di dimensione $n \times (m + p)$ che può essere pensata come $\bar{B} = [B \ K]$, con B e K delle dimensioni opportune. Si noti comunque che deve essere garantiti che il fattore $A - KC$ sia stabile. Altrimenti la matrice K deve essere sostituita con il guadagno di Kalman K_{kal} del sistema, che può essere calcolato impiegando la funzione `kalman.m` di *MATLAB*[®]. Considerando quindi le prime m variabili di \bar{u}_t come ingressi u_t e le rimanenti p variabili come rumori e_t il modello è della forma (3.2).

Un modello sparso, dove certi ingressi non contribuiscono all'evoluzione del sistema (equivale a dire che sono nulle certe $h^{[i,j]}$ per $i = 1, \dots, p$ e per le $j \in \{1, \dots, m\}$ relative agli ingressi) si ottiene annullando semplicemente le relative colonne della matrice B ed è chiamato in questa tesi *modello con livello di sparsità ridotto*. Negli studio Monte Carlo, dove è stato usato un tale modello, ad ogni ciclo sono stati annullate un numero causale di colonne di B compreso tra 1 e 9, così che almeno un ingresso non contasse e che almeno un ingresso contribuisse all'evoluzione del sistema. In questo caso la percentuale di ingressi e uscite e di risposte impulsive considerate in maniera corretta si equivale.

Negli studi Monte Carlo sono stati usati anche modelli chiamati in questa tesi *modelli con livello di sparsità aumentato*. Con questo nome si intendono modelli dove non solo qualche ingresso non contribuisce in alcun modo all'evoluzione dell'intero sistema, ma anche qualche altro ingresso contribuisce alla predizione di solo qualche uscita. Inoltre c'è la possibilità che il passato di qualche uscita non dia contributo alla predizione oppure ne dia un contributo parziale, contando solo per la predizione di qualche uscita (non più di tutte). In Figura 2.5 nel capitolo 2 è schematizzato l'esempio di un'interazione di un sistema in tale condizioni. Ottenere un modello con livello di sparsità aumentato da un modello generato mediante il comando `drmodel.m` non è più così immediato come ottenere un mo-

dello in cui solo qualche ingresso non contribuisce. Per questo nelle simulazioni si è proceduto nel modo seguente: ad ogni ciclo Monte Carlo è stato generato un modello di riferimento random usando `drmodel.m` (per evitare problemi computazionali, è stato controllato se qualche autovalore della matrice del sistema avesse un valore assoluto maggiore di 0.999 e in tal caso è stato generato un nuovo modello usando `drmodel.m`). È poi stato calcolato il predittore di Kalman in spazio di stato, che è stato trasformato in funzione di trasferimento ottenendo i legami tra predizioni e ingressi e tra predizioni e uscite visto in equazione (2.27):

$$\hat{\mathbb{E}}[y_t^{[l]} | y_t^-, u_t^-] = \sum_{i=1}^m A_{ti} h^{[l,i]} + \sum_{i=m+1}^{m+p} A_{ti} h^{[l,i]} \quad (3.4)$$

dove l'operatore A_{ti} per $i = 1, \dots, m$ raggruppa il passato degli ingressi e per $i = m + 1, \dots, m + p$ raggruppa il passato delle uscite. Questo predittore è poi stato sparsificato annullando certe risposte impulsive $h^{[i,j]}$ per $i \in \{1, \dots, p\}$ e per $j \in \{1, \dots, m + p\}$, dove $j \in \{1, \dots, m\}$ si riferiscono alle risposte impulsive tra ingressi e predittore e $j \in \{m + 1, \dots, m + p\}$ alle risposte impulsive tra uscite e predittore: per gli $j \in \{1, \dots, m\}$ un set causale di dimensione causale tra uno e sei è stato scelto e tutte le risposte impulsive $h^{[i,j]}$ relative a quel set sono state annullate (per $i = 1, \dots, p$ in modo che gli relativi ingressi non danno alcun contributo all'evoluzione del sistema). Delle rimanenti risposte impulsive non nulle tra ingresso e predizione un numero causale, diverso per ogni $i \in \{1, \dots, p\}$, è stato annullato in modo che la predizione di ogni uscita dipende dall'evoluzione di almeno un ingresso. Per le uscite invece ne sono state selezionate al massimo due in modo che il loro passato non contribuisce alla predizione di nessuna uscita, cioè è stato scelto un set di dimensione al massimo due tra $j \in \{m + 1, \dots, m + p\}$ per il quale tutte le $h^{[i,j]}$, $i = 1, \dots, p$ sono state annullate. Delle rimanenti risposte impulsive non nulle tra uscite e predizioni, è stato annullato un numero causale in modo che la predizione di ogni uscita dipende almeno da un'altra o di se stessa. Dopo aver sparsificato le risposte impulsive del predittore, da queste si può ricavare il modello di riferimento nello spazio di stato per generare i dati che però, per la sparsificazione del predittore, può non essere stabile. Per aggirare questo problema, ad ogni ciclo di Monte Carlo è stato ripetuto il processo di generazione di un modello, di sparsificazione del relativo predittore e di ritrasferimento in un modello fino a che questo non risultava stabile. Sono state necessarie, in media, cinque ripetizioni di questo processo per ottenere un modello di riferimento stabile.

Il set dei dati di training Dal set dei dati di training si identificano i modelli impiegando i diversi approcci. Si generano i dati di ingresso u_t e di rumore bianco e_t di dimensione variabile (per confrontare le prestazioni degli algoritmi di identificazione in funzione della dimensione dei set di dati) a partire dai quali sono calcolate le relative uscite y_t del modello di riferimento.

Nei diversi studi Monte Carlo che sono stati considerati, cambia il modo di generare gli ingressi u_t del set di training. Inizialmente si considerano ingressi di rumore bianco dove le singole componenti sono scorrelate. Per rendere il problema di selezione più impegnativo si considerano ingressi, che sono sempre bianchi, ma hanno componenti correlate e sono stati generati usando il modello seguente:

$$u_t^{[i+1]} = u_t^{[i]} + v_t^{[i]}, \quad i = 1, \dots, m-1$$

con $\{u_t^{[1]}\}$ un processo bianco che ha varianza unitaria e v_t un processo vettoriale bianco, indipendente da $\{u_t^{[1]}\}$ e con varianza $\mathbb{E}(v_t v_t') = 0.04 \cdot I$. In questa maniera per i coefficienti di correlazione lineare ρ_i , $i = 1, \dots, m-1$ vale

$$\rho_i := \text{var}\{u_t^{[i]}\} \left[\sqrt{\text{var}\{u_t^{[i]}\} \text{var}\{u_t^{[i+1]}\}} \right]^{-1} \in [0.980, 0.988],$$

valori leggermente superiori a quanto è stato controllato sulle misure reali di temperatura, umidità e irraggiamento di un edificio.

Il set dei dati di test I dati di test servono per validare i diversi modelli identificati. Si generano di nuovo l'ingresso u_t e il rumore bianco Gaussiano e_t , di dimensione $N_{test} = 1000$ dagli quali si calcolano le relative uscite del modello di riferimento y_t . In questo caso $u(t)$ è sempre generato come processo bianco con componenti indipendenti in modo che l'eventuale scelta di una variabile o di una risposta impulsiva sbagliata della procedura di selezione di SSEH si evidenzia in maniera più chiara in termini di FIT_k .

Dopo aver identificato i modelli impiegando i diversi metodi, si calcolano le uscite di tali modelli usando gli ingressi e i rumori del test set generati prima. L'uscita di questi modelli si confrontano poi con le uscite del modello di riferimento sulla base dell'indice di FIT_k (3.1) al fine di valutare i metodi di identificazioni utilizzati.

3.2 Il monitoraggio dell'edificio

Diversamente da quanto visto per gli studi Monte Carlo, in questo caso non si conosce il modello "vero" di riferimento. La procedura da eseguire per confronta-

re le diverse versioni dei metodi di identificazione è quindi leggermente diversa. I dati ottenuti dal monitoraggio si suddividono in dati di training e dati di test. Si eseguono le identificazioni impiegando le differenti versioni dei dati di training e per ciascuna di esse si determina un modello specifico. Per ciascuno di questi si calcolano le uscite sul set di training e gli indici *FIT*. Per classificare le diverse versioni dei metodi di identificazione si confrontano poi i diversi *FIT* ottenuti. Il procedimento seguito è schematizzato in Figura 3.2.

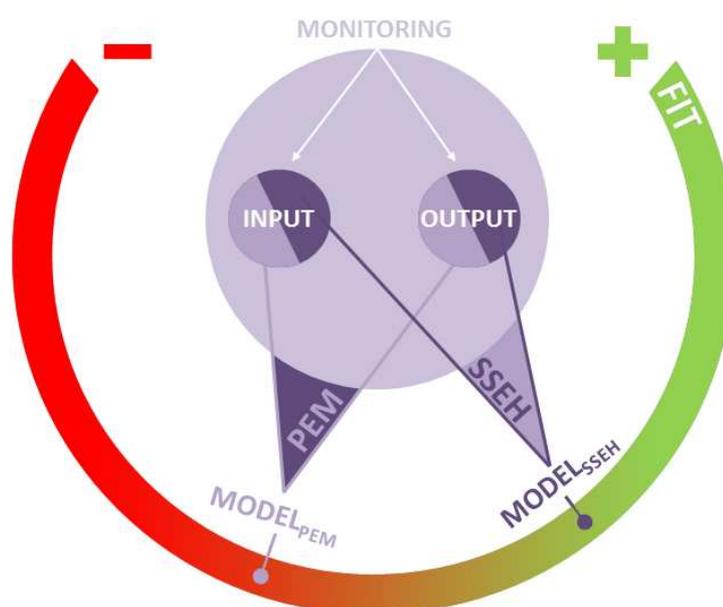


Figura 3.2 – Schema del procedimento seguito per identificare modelli da dati ottenuti dal monitoraggio dell'edificio.

Il monitoraggio è stato eseguito nell'ambito del progetto Sistema Integrato/distribuito di Monitoraggio Energetico ed Ambientale (SIMEA) [52], finanziato dalla Regione Veneto, il cui responsabile scientifico è prof. Angelo Cenedese (del Dipartimento di Tecnica e Gestione dei Sistemi Industriale dell'Università degli Studi di Padova (DTG)). L'obiettivo del progetto SIMEA è la realizzazione e il test di un sistema innovativo di sensori e relativi algoritmi di elaborazione dei dati, che permetta di effettuare una rilevazione e una valutazione precisa dei principali parametri ambientali ed energetici negli edifici. Il progetto mira a produrre un profilo energetico e operativo degli edifici, con un triplice scopo:

- ai fini della certificazione (energetica e sanitaria);
- suggerire e implementare delle modifiche strutturali agli ambienti e ai materiali per migliorare le condizioni operative e di comfort;

- incentivare la sensibilità sul consumo energetico e l'utilizzo ottimizzato delle risorse attraverso un controllo automatico intelligente e la possibilità di ricavare informazioni utili alle procedure di manutenzione.

Sono state monitorate due zone dell'edificio che ospita il laboratorio Ri.Cert. S.p.A. situato a Monte di Malo in provincia di Vicenza [53]. Si considerano qui solo i sensori relativi a una zona al primo piano dell'edificio di circa $150m^2$, composta da quattro uffici, una sala riunioni e un laboratorio leggero di analisi chimiche. In questa zona sono state distribuite 19 schede multisensore, la cui locazione è illustrata in Figura 3.3, che comunicano wireless usando lo standard ZigBee con un PC per la registrazione dei dati.

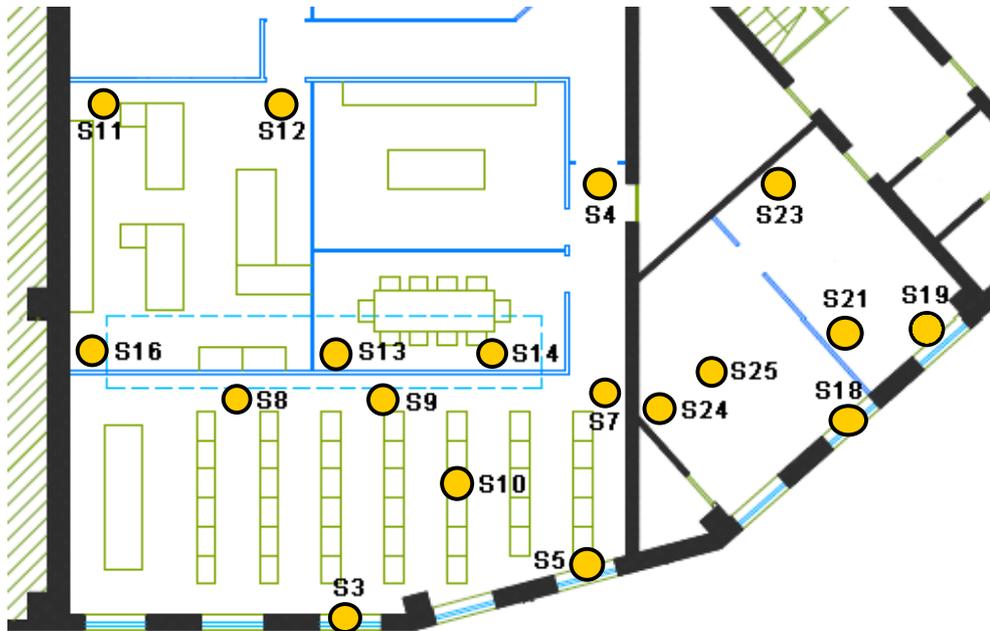


Figura 3.3 – Localizzazione delle schede multisensore utilizzate per monitorare i dati impiegati nel test degli algoritmi di identificazione. Questa zona si trova al primo piano dell'edificio del laboratorio Ri.Cert. S.p.A. situato a Monte di Malo in provincia di Vicenza.

Ogni scheda multisensore usata per il monitoraggio dispone di un microprocessore che pre-elabora i dati ed è equipaggiata con quattro sensori standard di temperatura, umidità, irraggiamento e pressione. Inoltre ogni scheda ha tre possibili estensioni per fare misure per esempio di CO, CO₂, fumi, etc. Nelle identificazioni però sono state utilizzate le sole misure di temperatura, umidità e irraggiamento. Per ulteriori dettagli sulla struttura del sistema di monitoraggio e

sulle apparecchiature impiegate si consulti [54].

I dati utilizzati per testare gli algoritmi di identificazione sono stati registrati nel periodo invernale tra il 03.01.2012 e il 20.02.2012. Siccome il sistema di monitoraggio è stato spento diverse volte per uno o più giorni, il monitoraggio non è continuo e i dati disponibili per le simulazioni sono relativi a registrazioni eseguite in 28 giorni distribuiti nei periodi di monitoraggio.

Il tempo di campionamento T , con cui sono prese le misure degli sensori della scheda multisensore, è di 10 secondi. Di questi dati il microprocessore calcola la media ogni 5 minuti e prova a trasmetterli al PC con un ritardo casuale tra 1 e 3 secondi per cercare di evitare collisioni, cioè per evitare che un'altra scheda trasmetta nello stesso istante. Se il PC non reagisce con una risposta di "trasmissione avvenuta con successo" un altro campione è aggiunto alla media e la scheda multisensore riprova a trasmettere i dati. Se la trasmissione fallisce ancora l'operazione si ripete altre due volte e se comunque non va a buon fine il dato si perde e si procede a calcolare la media dei 5 minuti successivi. Per dare un'idea dell'andamento temporale dei dati, in Figura 3.4 sono illustrati quattro giorni tipici delle registrazioni.

Dalla Figura 3.4 si vede che alcune registrazioni non presentano un comportamento realistico. C'è da aspettarsi che questi andamenti non daranno un grosso contributo negli modelli identificati; in particolare nei modelli identificati tramite il metodo SSEH si prevede che questi siano scartati dalla procedura di selezione.

Per la procedura, con la quale è effettuata la trasmissione, i dati registrati sul PC non si presentano con un tempo di campionamento T fisso, requisito necessario per poterli utilizzare in un algoritmo di identificazione. I dati necessitano quindi di una pre-elaborazione, che li riporti a un passo di campionamento T costante. Per trasformarli nel formato adeguato per l'utilizzo nell'identificazione si è eseguita la seguente operazione: è stato generato un nuovo vettore temporale con un tempo di campionamento di $T = 6$ minuti, in modo da avere 10 campioni all'ora. Per ogni elemento del vettore temporale sono state cercate le tre registrazioni più vicine (il che è possibile perché con ogni dato è stato registrato anche l'istante temporale in cui è stato ricevuto) ed è stato fatto un'interpolazione con un polinomio del secondo ordine di questi dati. Dall'interpolazione è poi stato ricavato il valore della funzione nell'istante specifico corrispondente all'elemento del vettore temporale. Si ottiene così l'andamento temporale delle misure con

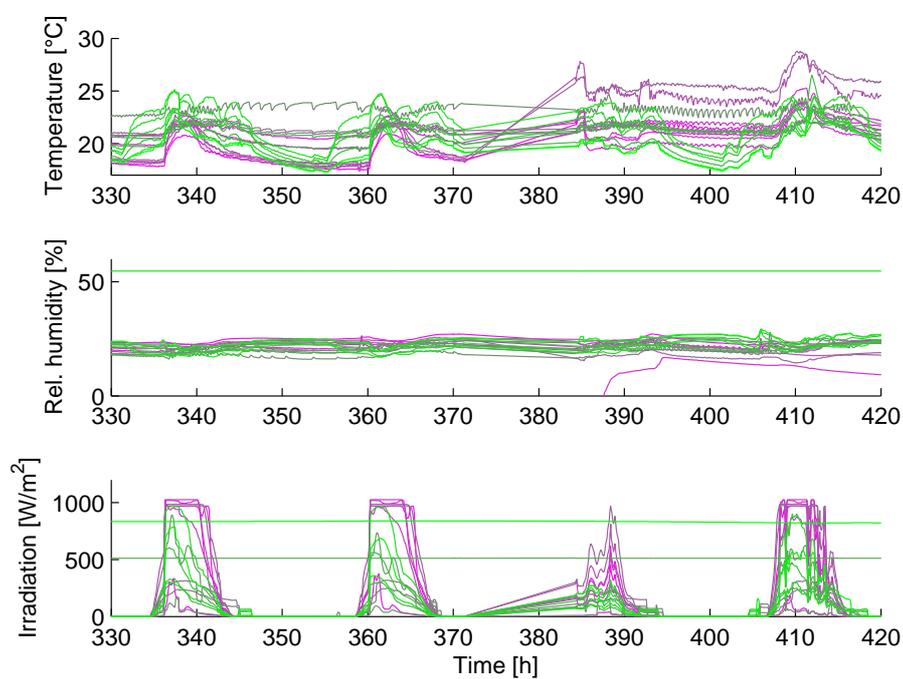


Figura 3.4 – Andamenti temporali di quattro giorni tipici delle registrazioni. Si noti che il sistema è stato spento per una notte e che due sensori di umidità e due sensori di irraggiamento non hanno registrato dati realistici.

un tempo di campionamento T fisso e con rumore sovrapposto, in maniera tale da renderli adeguati al test degli algoritmi di identificazione. Dato che il rumore sovrapposto è di piccola ampiezza non dà grossi impatti sull'esito delle identificazioni. In Figura 3.5 è riportato un confronto tra l'andamento di temperatura registrato e quello ottenuto dopo l'interpolazione.

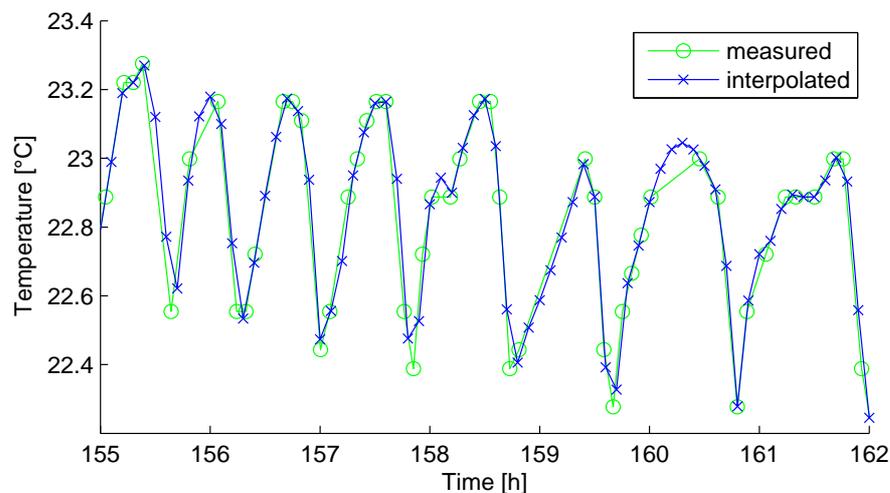


Figura 3.5 – Confronto tra l'andamento di una temperatura registrata e l'andamento ottenuto da un'interpolazione del secondo ordine per ottenere un tempo di campionamento T costante, pari a 6 minuti.

Oltre ai dati monitorati all'interno dell'edificio, nelle simulazioni sono stati impegnati anche dati meteo del periodo 03.01.2012 - 20.02.2012. Tali dati meteo sono stati misurati da una stazione agrometeorologica dell'Agenzia Regionale per la Prevenzione e Protezione Ambientale del Veneto (ARPAV) [55], situata a Malo. Per il progetto SIMEA questi dati sono stati messi a disposizione gratuitamente dall'ARPAV, contengono misure di vento (direzione e velocità), temperatura, pressione, umidità relativa, pioggia e irraggiamento, e sono disponibili per ogni ora. Per poter essere utilizzati nell'identificazione, i dati devono avere un tempo di campionamento T fisso, uguale per tutti gli andamenti impiegati, anche questi necessitano di una pre-elaborazione. Questa è stata eseguita come per i dati di misura dell'edificio, interpolando i dati con polinomi del secondo ordine e valutando tali polinomi negli istanti temporali fissi.

3.3 Metodi di identificazione considerati

Nelle simulazioni sono state considerate differenti versioni dell'approccio SSEH e differenti versioni dell'approccio PEM classico per modelli in spazio di stato.

La prima versione di SSEH impiega il modello Bayesiano ridotto, le altre due versioni invece utilizzano il modello Bayesiano intero. Mentre una di queste due ottimizza tutte le variabili dell'iperparametro, l'altra ottimizza solo quelle relative alle variabili che sono state selezionate nella fase di inizializzazione. Come procedura standard di selezione delle variabili si impiega la cosiddetta forward selection Bayesiana presentata in sezione 2.2.5, la cui versione MISO è stata usata negli articoli [20–22]. Oltre a questa si considerano anche la versione di forward selection sequenziale e le procedure backward elimination e stepwise selection, anche queste nelle versioni sequenziali e Bayesiane.

Dell'approccio PEM si considerano due versioni con oracolo (una che dispone dell'informazione su quali risposte impulsive sono nulle nel modello di riferimento e l'altra invece no), che hanno accesso al set di test per selezionare l'ordine del modello. Si considerano poi due versioni che utilizzano l'indice BIC e una che considera l'indice AICc per la selezione dell'ordine.

Le diverse versioni degli metodi di identificazione e poi anche delle procedure di selezione delle variabili per SSEH sono presentate in dettaglio qui di seguito.

3.3.1 Diverse versioni di SSEH e PEM

Si considerano le seguenti versioni dell'approccio SSEH:

1. **SSEH sub-ottimo:** considera il modello Bayesiano ridotto rappresentato in Figura 2.4 (destra). Dato che la dimensione del problema di ottimizzazione è molto ridotta rispetto a quella del modello Bayesiano intero, questa versione è computazionalmente più efficiente rispetto alle altre versioni. Lo svantaggio di usare il modello Bayesiano ridotto è che peggiora la flessibilità e quindi ci si attendono prestazioni del modello identificato, in termini di *FIT*, peggiori di quelle ottenibili usando il modello Bayesiano intero. Si vedrà però che le prestazioni di questa versione di SSEH sono comunque buone in confronto alle altre versioni SSEH e alle versioni PEM.
2. **SSEH ottimo:** considera il modello Bayesiano intero per avere una flessibilità maggiore rispetto al modello ridotto. Come punto di partenza del problema di ottimizzazione si usa l'iperparametro identificato impiegando

il metodo SSEH sub-ottimo, con $\lambda_{ij} = 0$ nel caso in cui una variabile non è considerata e un valore comune $\lambda_{ij} = \lambda$ altrimenti. Si ottimizzano tutti i valori dell'iperparametro di partenza e quindi anche le variabili che dal metodo sub-ottimo sono state scartate in questo caso possono influire nel modello identificato. Dopo aver eseguito la procedura di selezione, si annullano le λ_{ij} inferiori a un certo valore (nelle simulazioni sono state annullate se risultava $\lambda_{ij} < e^{-10}$), poiché ininfluenti sul sistema.

3. **SSEH ottimo con preselezione:** considera il modello Bayesiano intero e utilizza come punto di partenza l'iperparametro ottenuto dall'identificazione con il metodo SSEH sub-ottimo. A differenza dell'algoritmo SSEH ottimo, le variabili non selezionate nell'identificazione SSEH sub-ottima non sono più considerate dal problema di ottimizzazione. Per questo motivo questa versione non solo è computazionalmente più efficiente di SSEH ottimo, ma siccome devono essere ottimizzate meno variabili è atteso un risultato in generale migliore.

Lo svantaggio è che nella fase di inizializzazione può capitare che una certa variabile non sia considerata, anche se in realtà contribuisce all'evoluzione del sistema reale (succede soprattutto se le variabili sono fortemente correlate). Mentre la versione SSEH ottima ottimizza tutte le variabili, questa versione non considera più le variabili non selezionate nella fase di inizializzazione e compie quindi un errore.

Anche in questa versione si annullano le λ_{ij} che dopo la procedura di selezione risultano inferiori a un determinato valore (nelle simulazioni sono state annullate se risultava $\lambda_{ij} < e^{-10}$).

Come accennato in sezione 2.2 per scopi implementativi le risposte impulsive del predittore $h^{[i,j]}$ devono essere troncate ad una certa lunghezza. Per valutare l'effetto di questo troncamento negli studi Monte Carlo sono state considerate diverse lunghezze fisse delle risposte impulsive. In più è stata stimata una lunghezza adeguata tramite un'identificazione PEM su modelli di struttura ARX (che è computazionalmente più efficiente rispetto ad altre versioni PEM, perché si riduce a un problema dei minimi quadrati), impiegando l'indice AIC per la stima dell'ordine.

Dell'approccio PEM si considerano le seguenti versioni:

4. **PEM+Oracolo:** usa il metodo PEM classico impiegando modelli in spazio di stato (implementato nella funzione `pem.m` del *System Identification Tool-*

box di *MATLAB*[®] [56]), che ha accesso al test set: si seleziona il modello e quindi l'ordine del modello, che dà le migliori prestazioni in termini di FIT_k (può essere anche diverso per ogni passo di predizione k considerato) in predizione sul test set.

5. **PEM+Oracolo+set:** come la versione PEM+Oracolo usa il metodo classico di identificazione PEM in spazio di stato e ha accesso ai dati di test per ricavare l'ordine migliore. Inoltre si cerca di equipaggiare questa versione con l'informazione di quali risposte impulsive $h^{[i,j]}$ del modello di riferimento sparso sono nulle.

Nel caso in cui i dati sono generati da un modello di riferimento con livello di sparsità ridotto, cioè che ha solo qualche ingresso che non conta in alcun modo nell'evoluzione del sistema (se un certo $h^{[i,j]}$ è nullo per qualche $j \in \{1, \dots, m+p\}$, allora è nullo per ogni $i = 1, \dots, p$), per tener conto nell'identificazione PEM dell'informazione sulla sparsità basta eliminare tali ingressi superflui. Questa versione dà quindi il risultato migliore ottenibile utilizzando l'approccio PEM per un tale modello. Nel caso invece, in cui i dati sono stati generati da un modello con livello di sparsità aumentato, non è affatto semplice equipaggiare il metodo PEM con l'informazione sulla sparsità. Il motivo è che da PEM sono identificate le matrici di un modello in spazio di stato e il legame tra le variabili che non contano nella predizione di solo qualche altra variabile e queste matrici non sono per niente semplici. È stato scelto per questo di utilizzare la stessa versione usata prima, non considerando semplicemente gli ingressi che non contribuiscono in alcun modo all'evoluzione del sistema. Per tali modelli quindi questa versione PEM non è quella ottimale, in quanto viene trascurata l'informazione delle $h^{[i,j]}$ nulle non per tutti gli $i = 1, \dots, p$, ma ci si avvicina.

Per tenere conto anche delle $h^{[i,j]}$ nulle non per tutti gli $i = 1, \dots, p$, è stata provata anche una versione MISO, identificando il modello relativo solo ad una specifica uscita e agli ingressi che contano a predirla, non considerando quelli che non contribuiscono alla predizione. Nelle simulazioni si è visto comunque che tale versione ha delle prestazioni inferiori alla versione PEM+Oracolo (migliori comunque delle versioni PEM che impiegano un qualche criterio di selezione del modello) dato che si perde l'informazione sul contributo delle altre uscite del sistema a quella usata per l'identificazione. Siccome l'obiettivo dell'utilizzo delle versione PEM+Oracolo+set è stato valutare le prestazioni ottimali del metodo PEM e questa versione ha delle prestazioni inferiori a PEM+Oracle, nella parte della descrizione dei

risultati della presente tesi questa versione non è stata riportata.

6. **PEM+Oracolo+preselezione:** identico alla versione PEM+Oracolo+set, solo che al posto di usare il set di variabili che contano effettivamente per la predizione del sistema, si usa il set di variabili selezionati da una procedura di selezione di SSEH. Questa versione è stata utilizzata nelle simulazioni, nelle quali sono stati impiegati i dati reali monitorati, al posto della versione PEM+Oracolo+set, perché ovviamente non si conoscono le variabili superflui per la predizione del sistema.
7. **PEM+BIC:** usa il metodo classico PEM in spazio di stato impiegando l'indice BIC per la selezione dell'ordine del modello.
8. **PEM+BIC+preselezione:** come PEM+BIC usa il metodo PEM in spazio di stato equipaggiato con BIC per effettuare la stima dell'ordine del modello. In più usa le informazioni dalla selezione di variabili del metodo SSEH sub-ottimo per annullare le risposte impulsive $h^{[i,j]}$ nel modello identificato (in contrasto con la versione PEM+Oracolo+set che conosce il set delle $h^{[i,j]}$, che sono effettivamente nulle nel modello di riferimento).
9. **PEM+AICc:** usa il metodo PEM classico in spazio di stato equipaggiato con AICc per effettuare la stima dell'ordine del modello.

Si noti che le versioni Oracolo del metodo PEM non sono di solito applicabili in condizioni reali. Specialmente se il monitoraggio è di breve durata, condizione studiata in questa tesi, non si ha a disposizione un set di variabili di test. Se la durata del monitoraggio è invece abbastanza lunga, un modo spesso adoperato è di dividere i dati in due parti e di usare una porzione dei dati per effettuare la selezione dell'ordine. Però anche procedendo così, la versione PEM+Oracolo+set non è impiegabile, dato che non si conosce il set di variabili che danno effettivamente contributo alla predizione del sistema.

Per i modelli di riferimento con livello di sparsità ridotto, come ordine massimo dei modelli identificati mediante i diversi metodi PEM è stato considerato 12. Questo limite è stato scelto perché ordini più alti hanno creato problemi nel caso di set di dati N a dimensioni ridotte. Il motivo è che in tali casi il numero di parametri da stimare è troppo alto in confronto ai dati disponibili e quindi la procedura PEM non è ben definita. Il limite di 12, che può sembrare piccolo in confronto all'ordine 30 del modello di riferimento scelto, non ha comportato comunque delle limitazioni nelle simulazioni dove sono stati impiegati modelli con livello di sparsità ridotto. È stato osservato che gli ordini effettivamente scelti

erano inferiori a tale limite, anche per le dimensioni massime del set di dati considerate.

Per le simulazioni, dove sono stati impiegati invece i modelli di riferimento con livello di sparsità aumentato, è stato notato che già per $N = 300$ l'ordine 12 è stato scelto diverse volte e per $N = 800$ tale limite è stato raggiunto quasi sempre. Per questo motivo, dove nelle simulazioni sono stati impiegati i modelli di riferimento con livello di sparsità aumentato, per $N \geq 300$ si è scelto di aumentare l'ordine massimo dei modelli identificati mediante PEM a 30.

Come accennato in precedenza, per stimare una lunghezza adeguata delle risposte impulsive è stata usata un'altra versione di PEM, della quale però non sono stati riportati le prestazioni in termini di FIT_k :

10. **PEM/ARX+AIC**: usa il metodo PEM per modelli ARX e impiega l'indice AIC per selezionare l'ordine adeguato del modello. Per questa versione l'ordine massimo considerato è stato 50. Nel caso in cui il metodo PEM dava errore per un numero di parametri da stimare troppo alto, l'identificazione è stata arrestata e l'ordine adeguato è stato selezionato tra quelli dove l'identificazione è stata eseguita con successo.

3.3.2 Procedure di selezione delle variabili per SSEH

Nell'ambito della regressione lineare gli algoritmi più usati per fare la selezione di variabili (in letteratura anche nota come feature selection) sono i procedimenti euristici **forward selection**, **backward elimination** e **stepwise selection**. Un'altra procedura standard, chiamata **all-best selection**, seleziona il miglior set di variabili, considerando tutte gli insiemi possibili. È stato proposto un gran numero di algoritmi alternativi, come procedure di "branch and bound" ed elaborazioni della procedure di stepwise selection [49–51, 57], che in certi casi sono vantaggiose rispetto alle procedure standard appena citate.

La procedura di forward selection parte da un insieme di variabili vuoto e valuta ad ogni passo (impiegando un qualche indice di validazione) se aggiungere o meno una certa variabile all'insieme delle variabili selezionate fino a considerare tutte le variabili oppure fino a che non si verifica una qualche condizione di arresto. Nel caso SSEH questo significa partire con un iperparametro con $\lambda_{i,j} = 0$ per $i = 1, \dots, p$ e $j = 1, \dots, m + p$ e settare un certo $\lambda_{i,j}$ uguale al valore ottenuto dal problema di ottimizzazione se il valore dalla marginale $J(y^{[i]+}; \zeta)$ definita in (2.33) decresce.

La procedura di backward elimination invece parte da un insieme in cui tutte le variabili sono considerate e si procede togliendo delle variabili (nel caso SSEH ciò significa partire con tutti i $\lambda_{i,j}$ uguale al valore ottenuto dal problema di ottimizzazione e valutare se settare a zero un $\lambda_{i,j}$ alla volta).

La terza procedura, detta stepwise selection, è una combinazione delle altre due: si parte con un insieme vuoto e ad ogni passo si valuta se aggiungere una certa variabile all'insieme delle variabili considerate (come per la forward selection). Se una certa variabile è aggiunta si valuta se qualche altra variabile, già appartenente all'insieme delle variabili selezionate, può essere di nuovo tolta.

Per capire meglio come lavorano le diverse procedure in Figura 3.6 sono illustrati tutti i possibili stati di un problema con quattro variabili, ordinati per numero di variabili selezionate (a sinistra nessuna variabile è inclusa, a destra tutte le variabili sono state selezionate). Assumendo che uno di questi stati sia quello ottimale la procedura di forward selection, aggiungendo ad ogni passo una variabile, si avvicina da sinistra allo stato ottimo come la procedura di stepwise selection (avendo però anche la possibilità di tornare successivamente verso sinistra togliendo certe variabili). Quella di backward elimination invece procede da destra verso sinistra togliendo ad ogni passo una variabile dal set di variabili selezionate (chiamato così con un piccolo abuso di notazione ed è in realtà un set che include tutte le risposte impulsive $h^{[i,j]}$ relative alle variabili $j \in \{1, \dots, m + p\}$ selezionate). La procedura all-best selection infine, considera tutti i sottoinsiemi di variabili e seleziona quello che dà le prestazioni migliori.

Da Figura 3.6 sembra che, nel caso in cui esiste un determinato stato ottimo, impiegando una qualunque delle tre procedure si arriva comunque allo stato ottimo. La procedura forward selection è più efficiente rispetto alla procedura di backward elimination nel caso in cui poche delle variabili servono a predire l'evoluzione del sistema (situazione che si presenta spesso nel caso in cui c'è un gran numero di variabili e quindi l'efficienza della procedura di selezione ha un peso sull'efficienza totale dell'algorithmo di identificazione) e si utilizza un qualche criterio di arresto della procedura.

Lo svantaggio della procedura di forward selection si presenta però nel caso in cui le variabili sono tra di loro correlate e la selezione è effettuata in maniera sequenziale (si valuta la prima variabile, se dà un contributo positivo si include la variabile nel set delle variabili selezionate, altrimenti no; poi si valuta la seconda, poi la terza e così via fino all'ultima). Se una qualche variabile non è strettamente legata all'evoluzione del sistema, ma è correlata con un'altra variabile che è

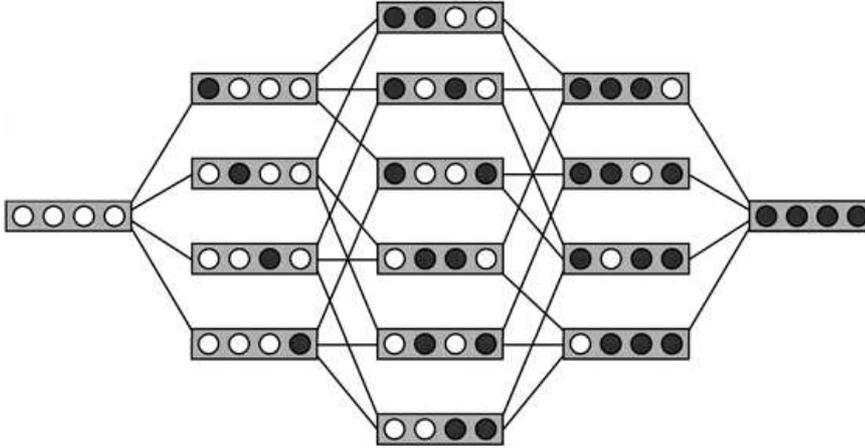


Figura 3.6 – Possibili stati di un problema di selezione delle variabili con quattro variabili ordinati rispetto al numero di elementi selezionati (cerchio bianco per variabile selezionata e nero per variabile non selezionata) [58]. La procedura di forward selection valuta i stati da destra a sinistra, la procedura di backward elimination da sinistra a destra. La procedura di stepwise selection procede inizialmente verso destra e riconsidera poi gli stati a sinistra.

legata all'evoluzione che segue nella sequenza con cui sono valutate le variabili, entrambe sono selezionate. Questo è però un'errore, dato che la seconda variabile porta tutta l'informazione della prima aggiungendone anche un'altra e quindi la prima non dovrebbe essere inclusa.

L'analogo del problema appena descritto per la procedura di forward selection non si presenta nella procedura di backward elimination, dato che una certa variabile non è mai tolta se è strettamente legata all'evoluzione di un sistema e quindi non si presenta l'eventuale problema di doverla aggiungere successivamente. Un altro approccio per aggirare il problema, ma che nel caso in cui poche variabili servono per la predizione del sistema può essere computazionalmente più efficiente della procedura di backward elimination, è dato dalla procedura di stepwise selection: ad ogni aggiunta di una qualche variabile si rivalutano tutte le altre variabili già selezionate. Se tale variabile appena aggiunta porta tutta l'informazione inclusa in una qualche altra variabile già aggiunta, si toglie di nuovo questa altra variabile.

Per m variabili di ingresso e p variabili di uscita in genere la procedura di stepwise selection è comunque computazionalmente più onerosa rispetto alle altre due. Infatti, mentre le procedure sequenziali di selezione di variabili forward selection e backward elimination hanno complessità $O(m + p)$ (si valuta ogni variabile una volta), la procedura stepwise selection ha complessità $O((m + p)^2)$, dato che nel

caso peggiore in cui si aggiungono tutte le variabili, ad ogni passo bisogna rivalutare tutte le variabili già aggiunte. La procedura all-best selection è la procedura meno efficiente di tutte quelle considerate, dato che deve considerare tutti i possibili set di variabili e ha quindi complessità $O((m+p)!)$. Da tale procedura sono però attese le prestazioni migliori.

Nelle diverse procedure di selezione può anche essere cambiata la sequenza con cui si valutano le variabili. Invece di valutarle sequenzialmente conviene utilizzare un qualche criterio, che determina quale tra le variabili dà il massimo contributo all'evoluzione del sistema (per le procedure di forward selection) oppure quale invece ne dà il minimo (per le procedure di backward elimination). Nel caso dell'approccio SSEH come criterio si può usare il delta di variazione della marginale definita in (2.33) aggiungendo o togliendo una qualche variabile. Una procedura che impiega un tale criterio, si indica con il nome "Bayesiana", dato che si utilizza l'informazione a priori della entità di decrescita della marginale per effettuare la selezione delle variabili.

Nella procedura di forward selection Bayesiana in confronto a quella sequenziale si parte sempre da un set di variabili selezionate vuoto. Invece di valutare le variabili sequenzialmente, ad ogni passo a tale set si aggiunge la variabile non ancora contenente, che provoca il maggior decremento della marginale. Quando si giunge alla situazione in cui, aggiungendo una qualunque delle variabili (non ancora selezionate) la marginale non decresce più, al set delle variabili non si aggiunge più nessuna variabile e la procedura si arresta.

La procedura backward elimination Bayesiana è molto simile a quella di forward selection Bayesiana, ma parte da un insieme di variabili selezionate contenente tutte le variabili e toglie ad ogni passo la variabile che dà il maggior decremento della marginale (2.33). Questa eliminazione è eseguito fino a che, togliendo una qualunque delle variabili contenute nel set, la marginale non decresce più e si arresta quindi la procedura avendo trovato il set finale. Anche la procedura di stepwise selection è simile: partendo da un set di variabili selezionate vuoto, si aggiunge ad ogni passo quella che dà il maggior decremento della marginale e si controlla se tale aggiunta rende superflua qualche variabile già contenuta nel set (controllando se togliendo qualche variabile la marginale decresce ulteriormente), che può quindi di nuovo essere tolta.

Si noti che la convergenza della procedura stepwise selection Bayesiana non è così ovvia. Potrebbe capitare un ciclo infinito in cui ciclicamente si aggiungono certe variabili che poi sono di nuovo tolte. Queste situazioni sono però impedito dal

fatto che le variabili si aggiungono oppure si tolgono solo se la marginale decresce strettamente. È quindi garantita la convergenza della procedura stepwise selection Bayesiana.

La procedura che in questa tesi è chiamata, con un piccolo abuso di notazione, all-best selection Bayesiana non è una vera procedura all-best, dato che si impiega una condizione di arresto. Gli stati che non si sono verificati prima del verificarsi della condizione d'arresto non sono considerati. La procedura all-best selection Bayesiana parte da un insieme vuoto di variabili selezionate (scelta fatta qui, ma ugualmente si poteva partire da un'insieme contenente tutte le variabili). Ad ogni passo, per ogni variabile non contenuta nel set attuale delle variabili selezionate, si valuta l'effetto dell'inserimento di una variabile e per ogni variabile contenuta nel set attuale, si valuta l'effetto dell'eliminazione di una variabile. L'operazione che dà la maggior decrescita della marginale è effettuata. Si continua a togliere e aggiungere variabili finché la marginale non decresce più. La convergenza di questa procedura è ancora garantita dalla stretta decrescita della marginale.

In sezione 2.2.5 è riportato il deflusso dettagliato della procedura forward selection Bayesiana. Non si riportano i dettagli delle altre procedure, dato che basta fare le ovvie modifiche.

Lo svantaggio accennata prima della procedura forward selection sequenziale, che si presenta soprattutto se le variabili sono tra di loro correlate, nel caso della procedura Bayesiana non si presenta più. Il motivo è che la procedura Bayesiana considera prima le variabili che contengono più informazioni. Una variabile che non contribuisce all'evoluzione del sistema ma è correlata ad una che ne contribuisce è considerata solo successivamente e dato che tutta l'informazione è già contenuta in quella che contribuisce all'evoluzione, l'altra non è aggiunta più. Per questo motivo e dato che la procedura di forward selection nel caso in cui poche variabili contano (cosa che in casi reali si determina soprattutto se si ha un gran numero di variabili) è computazionalmente più efficiente rispetto alla procedura di backward elimination, come procedura standard di selezione delle variabili per SSEH si considera la forward selection Bayesiana. La versione MISO di questa procedura è stata usata anche negli articoli [20–22] per confrontare l'approccio SSEH con altri metodi di identificazione (parametrici e non-parametrici).

Rispetto alle procedure sequenziali lo svantaggio delle procedure Bayesiane è ovviamente la complessità computazionale. Come accennato, per m variabili di ingresso e p variabili di uscita, le procedure di selezione delle variabili forward selection e backward elimination sequenziali considerano ogni variabile una sola volta e hanno quindi complessità $O(m + p)$, le versioni Bayesiane di queste procedure

hanno invece complessità $O((m+p)^2)$. Questo perché nel caso peggiore in cui si aggiungono o tolgono rispettivamente tutte le variabili, ad ogni passo devono essere calcolate le marginali rispettivamente di ogni variabile non ancora contenente, o già contenuta nel set delle variabili selezionate. Nello stesso modo per la procedura stepwise selection la complessità aumenta da $O((m+p)^2)$ nel caso sequenziale a $O((m+p)^3)$ nel caso Bayesiano, e per la procedura all-best selection da $O((m+p)!) nel caso sequenziale a $O((m+p) \cdot (m+p)!) nel caso Bayesiano.$$

Procedure di selezione considerate nelle simulazioni Come accennato diverse volte come procedura di selezione standard si usa la procedura forward selection Bayesiana, presentata in dettaglio in sezione 2.2.5. Sono stati eseguiti però studi Monte Carlo con lo scopo di confrontare la bontà di diverse procedure di selezione, alcune di queste procedure sono state utilizzate poi anche nell'identificazione di modelli da dati misurati in edifici.

Nelle simulazioni relative alle procedure di selezione, in una prima fase sono state considerate procedure che effettuano la selezione di variabili, cioè procedure che prevedono solo di annullare tutte (per ogni $i = 1, \dots, p$) le risposte impulsive del predittore $h^{[i,j]}$, per certi j . Sono state utilizzate le procedure forward selection, backward elimination e stepwise selection, che effettuano la selezione valutando le variabili sequenzialmente. Per evidenziare meglio i pregi e i difetti di queste procedure sequenziali, sono state considerate le versioni che cominciano a valutare le $h^{[i,j]}$ dagli ingressi (da $j = 1$) aumentando ad ogni passo la j , e le versioni che cominciano dalle uscite (da $j = m+p$) e riducono la j ad ogni passo. Le procedure si indicano per questo relativamente con “sequenziali in avanti” e con “sequenziali all'indietro”.

Si noti che nel caso sequenziale la procedura all-best selection non è stata utilizzata. Un motivo è che le procedure sequenziali sono state considerate in prima linea per mostrare lo svantaggio della procedura forward selection, che si evidenzia già nel considerare le altre procedure. Un altro motivo è che nel caso sequenziale la procedura all-best selection è computazionalmente molto più onerosa rispetto alle altre procedura considerate. Si considera però la procedura all-best selection Bayesiana, oltre alle procedure forward selection Bayesiana, backward elimination Bayesiana e stepwise selection Bayesiana.

Nella seconda fase delle simulazioni relative al confronto delle procedure di

selezione, le procedure considerate effettuano la selezione delle singole risposte impulsive, cioè sono in grado di valutare ogni singola risposta impulsiva $h^{[i,j]}$. Si ottiene così l'interconnessione (causalità di Granger) tra le variabili di un sistema, del tipo schematizzato in Figura 2.5. Di queste procedure non si fa un'analisi così accurata come per le procedure di selezione delle variabili dato che ovviamente i risultati di tale procedure si possono trasferire alle procedure di selezione delle singole risposte impulsive.

La complessità computazionale di queste procedure di selezione delle risposte impulsive è ovviamente superiore alle procedure di selezione delle variabili, a cui si è fatto riferimento in precedenza, analizzando la complessità delle diverse procedure. Mentre prima si è visto che la complessità è funzione di $m + p$ (numero totale di variabili), per le procedure di selezione delle risposte impulsive la complessità è funzione di $p(m + p)$ (numero totale di risposte impulsive): mentre per esempio la forward selection sequenziale per la selezione delle variabili ha complessità $O(m + p)$ e quella Bayesiana $O((m + p)^2)$, le versioni di selezione delle risposte impulsive hanno complessità $O(p(m + p))$ e $O((p(m + p))^2)$ rispettivamente.

Nella seconda fase sono state usate solo procedure Bayesiane. È stato considerato la procedura di selezione delle risposte impulsive all-best selection Bayesiana: partendo da un set di risposte impulsive $h^{[i,j]}$ vuoto, ad ogni passo si calcola per ogni risposta impulsiva il cambiamento della marginale inserendola o togliendola (se rispettivamente non è, oppure è già inclusa) da tale set. Si inserisce o toglie dal set di risposte impulsive scelte la $h^{[i,j]}$ che dà la maggior decrescita della marginale definita in (2.33), affinché questa non decresce più.

Le altre procedure di selezione delle risposte impulsive considerate sono in pratica delle estensioni della procedura standard di forward selection Bayesiana:

- **FORWARD-AllBest**: è simile alla procedura all-best selection, ma invece di partire dal set vuoto di risposte impulsive selezionate, come set iniziale si usano le variabili selezionati dalla procedura forward selection Bayesiana, eseguita in un passo preliminare.
- **FORWARD-AllBest-Input**: esegue la selezione delle variabili impiegando la procedura forward selection Bayesiana, il cui risultato è poi usato come set iniziale di diverse procedure all-best selection: si valutano tutte le singole risposte impulsive eseguendo $m + p$ procedura all-best selection. Partendo da $j = 1$ e aumentandolo ad ogni passo, ciascuna di queste procedure di selezione delle risposte impulsive considera le p risposte impulsive $h^{[i,j]}$ relative a un determinato j .

- **FORWARD-AllBest-Output:** esegue come la procedura FORWARD-AllBest-Input, partendo dal set iniziale di variabili selezionate dalla procedura forward selection Bayesiana, diverse procedure all-best selection. In questo caso si eseguono però p procedura all-best selection, partendo da $i = 1$ e aumentandolo ad ogni passo, considerando le $m + p$ risposte impulsive $h^{[i,j]}$ relative all'attuale i .
- **FORWARD-backward:** usa come set iniziale le variabili selezionati dalla procedura forward selection Bayesiana e valuta se conviene togliere singole risposte impulsive da tale set impiegando una procedura di selezione delle risposte impulsive del tipo backward elimination.

Queste estensioni della procedura di selezione standard delle variabili forward selection Bayesiana sono state considerate perché in problemi di ottimizzazione non-lineare un criterio cruciale per avere un buon esito è l'impiego di un'adeguata condizione iniziale. In confronto alla pura procedura all-best selection, dove si parte dal set iniziale vuoto, nelle estensioni della procedura forward selection Bayesiana il set iniziale è dato dalle variabili selezionate da tale procedura.

In teoria la procedura forward selection Bayesiana seleziona tutte le variabili che contribuiscono alla predizione del sistema. Nel caso in cui certe variabili non contribuiscono alla predizione di tutte le uscite, si possono annullare le relative risposte impulsive $h^{[i,j]}$ applicando la procedura backward elimination di singole risposte impulsiva sul set delle variabili scelte dalla forward selection Bayesiana. Dato che può capitare che una certa variabile dà un contributo molto ridotto e solo alla predizione di poche altre variabili, può capitare che tale variabile non è selezionata dalla procedura di forward selection Bayesiana. In tal caso il piccolo contributo dato dalla variabile è eliminato per l'inclusione delle risposte impulsive relative alle variabili, alle quali quella determinata variabile non contribuisce. Può quindi essere conveniente rivalutare anche le risposte impulsive non contenute nel relativo set delle variabili selezionate dalla forward selection Bayesiana; tale procedura di selezione delle variabili è stata quindi anticipata alla all-best selection per selezionare le singole risposte impulsive.

Con le procedure FORWARD-AllBest-Input e FORWARD-AllBest-Output si è cercato di includere nella procedura di selezione un impostazione fisica. Pensando che i dati impiegati per l'identificazione del modello siano state misurate da qualche sensore (come nel caso dell'identificazione di modelli termodinamici di un edificio), la forward selection Bayesiana identifica prima i sensori che danno informazioni alla predizione di certe misure, definite come uscite del sistema. Queste

si usano come punto iniziale delle procedure di selezione delle risposte impulsive. La procedura di FORWARD-AllBest-Input (Input nel senso di ingressi per la predizione delle uscite di un sistema, dati dagli effettivi ingressi e dai passati delle uscite di tale sistema) valuta per ogni sensore, impiegando una procedura all-best selection, il suo contributo alla predizione di quale sensore definito come uscita. Si noti che si valutano anche le variabili non selezionate dalla forward selection Bayesiana. La procedura di FORWARD-AllBest-Output (Output per le variabili di uscita del sistema, che si cercano di predire) invece impiega diverse procedura all-best selection per valutare sequenzialmente per ogni singolo sensore definito come uscita, quale tra tutti i sensori contribuisce alla predizione del relativo sensore di uscita.

Capitolo 4

Risultati degli studi Monte Carlo

Si presentano in questo capitolo i risultati ottenuti da diversi studi Monte Carlo, utilizzati per validare il metodo di identificazione SSEH e per confrontarlo con il metodo classico PEM, da lungo lo stato dell'arte nell'ambito dell'identificazione. Nella prima parte di questo capitolo si analizzano le influenze della lunghezza del set di dati N e della lunghezza delle risposte impulsive (dato che per questioni computazionali devono essere troncati a una certa lunghezza t_0) sulle prestazioni di SSEH e anche sui metodi PEM. Successivamente, per diversi N e t_0 fissati, si confrontano diversi metodi di selezione per l'algoritmo SSEH. Infine, per i medesimi N e t_0 fissati, si confrontano le diverse versioni dell'approccio SSEH e dell'approccio PEM.

L'indice FIT_k , definito in capitolo 3, dà una valutazione della bontà dell'approssimazione di un andamento predetto a k passi a un andamento reale e si calcola in relazione a una specifica uscita. In questo capitolo è stata fatta la scelta arbitraria di riportare solo i FIT_k relativi all'uscita $y^{[2]}$ (quella intermedia, dato che i modelli sono stati generati con tre uscite), poiché per l'alto numero di esperimenti fatti in ogni studio Monte Carlo, i FIT_k relativi alle tre uscite sono distribuiti in maniera equilibrata. Un'alternativa avrebbe potuto essere ovviamente riportare i FIT_k relativi a una qualunque altra uscita oppure riportare la media dei FIT_k sulle tre uscite.

Di seguito, per confrontare i risultati ottenuti dai diversi metodi di identificazione in termini di FIT_k , si useranno dei "box-plot", utili in generale per illustrare la distribuzione statistica di una qualche popolazione di dati. Questi sono stati generati impiegando il comando `boxplot.m` di *MATLAB*[®] e rappresentano il mediano, un rettangolo delimitato dal primo e quarto quartile e i segmenti ("whisker") che individuano i valori massimi e minimi non compresi gli "outlier" [56]. Anche gli outlier sono a loro volta evidenziati nel box-plot e, nel caso in cui questi

si trovano al di fuori del limite del plot, si usa la notazione “Plus xx ”, dove xx indica il numero di outlier all'esterno del limite.

4.1 Influenze delle lunghezze del set di dati e delle risposte impulsive

In questa sezione si analizzano le influenze della lunghezza del set di dati N e della lunghezza delle risposte impulsive t_0 alle prestazioni del metodo SSEH. Si considera a tal fine la versione SSEH sub-ottima che, come spiegato in sezione 3.3, impiega il modello Bayesiano ridotto ed è quindi computazionalmente più efficiente rispetto alle versioni ottime, che presentano prestazioni leggermente migliori, come si vedrà in sezione 4.3. Come lunghezze del set di dati nelle simulazioni sono state considerate $N \in \{150, 200, 300, 500, 800\}$ e come lunghezze delle risposte impulsive $p \in \{5, 10, 30, 40\}$. Come lunghezza delle risposte impulsive è stata considerata inoltre la lunghezza adeguata, stimata mediante un'identificazione PEM di modelli ARX, impiegando l'indice AIC come criterio di selezione del modello (PEM/ARX+AIC).

Per avere un confronto con l'approccio PEM, in questa sezione si considerano anche la versione PEM+Oracolo+set e la versione PEM+BIC. Inoltre è stato calcolato, per ogni esperimento, il FIT_k ottenuto dalla predizione impiegando il modello di riferimento, che presenta ovviamente il limite superiore raggiungibile da un modello identificato dai dati e che qui permette un confronto generale sulla bontà dei diversi metodi di identificazione.

I diversi metodi di identificazione sono stati applicati a dati, ottenuti da modelli con livello di sparsità differente, utilizzando ingressi non correlati e ingressi correlati, generati come descritto in sezione 3.1.

4.1.1 Ingressi non correlati e modelli con livello di sparsità ridotto

Si riportano qui i risultati degli studi Monte Carlo eseguiti per analizzare l'influenza di N e di t_0 sulle prestazioni dei diversi metodi di identificazione, nei quali sono stati generati ingressi scorrelati e modelli con livello di sparsità ridotto così da avere qualche ingresso che non contribuisce in alcun modo all'evoluzione del

sistema.

Tipicamente l'indice FIT_k decresce all'aumentare dei passi di predizione k e, dopo un certo numero di passi, tende a stabilizzarsi. Questo è dovuto al fatto che, all'aumentare di k , il passato dell'uscita contribuisce sempre di meno alla predizione. In Figura 4.1 sono riportati gli andamenti di mediani dell'indice FIT_k in funzione dei passi di predizione k di uno studio Monte Carlo con una lunghezza del set di dati $N = 300$, per il modello di riferimento con livello di sparsità ridotto e per modelli identificati impiegando diversi metodi.

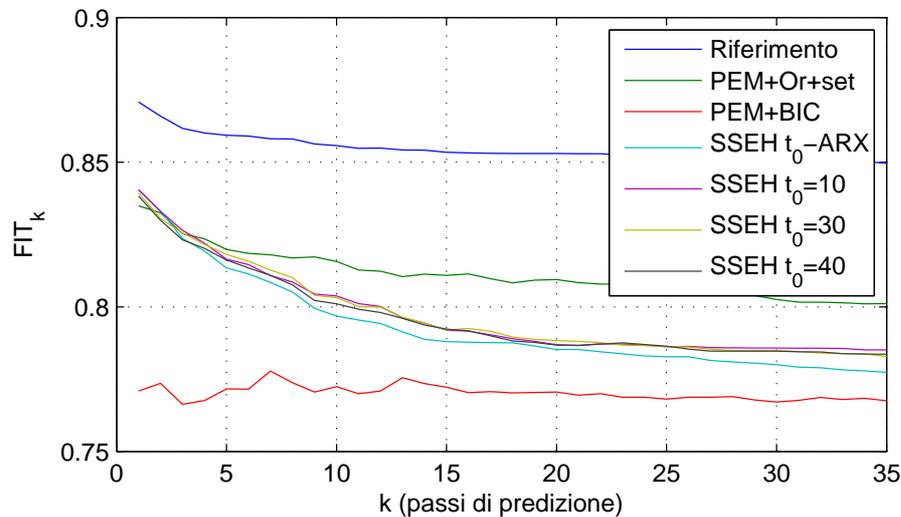


Figura 4.1 – Andamento dei mediani dei FIT_k al cambiare del passo di predizione k di uno studio Monte Carlo con la lunghezza del set di dati $N = 300$. Sono riportati gli andamenti relativi al modello di riferimento sparso e ai modelli identificati mediante i metodi PEM+Oracolo+set, PEM+BIC e SSEH subottimo con diverse lunghezze delle risposte impulsive ($t_0 \in \{10, 30, 40\}$ e t_0 stimato da PEM/ARX+AIC).

La Figura 4.1 mostra che, come c'è da aspettarsi, all'aumentare dei passi di predizione di solito l'indice FIT_k peggiora. Non è così invece per i modelli identificati impiegando PEM+BIC. Il motivo sta nel fatto che questa versione di PEM, per un set di dati N ridotto, ha difficoltà soprattutto a identificare il modello di rumore, cioè i termini g_k in equazione (2.26). Infatti tale modello ha effetto positivo nella predizione solo a pochi passi e peggiora quindi il FIT_k per piccoli k . Come si vedrà di seguito, all'aumentare di N , PEM+BIC riesce a identificare meglio anche il modello d'errore e quindi questa versione presenta un andamento del FIT_k che decresce all'aumentare dei passi di predizione k , similmente a quan-

to visto in Figura 4.1 per le altre versioni di identificazione.

Siccome gli andamenti riportati in Figura 4.1 non contengono molta informazione statistica, si è scelto di non riportare tutti gli andamenti per tutti i diversi N considerati. Nella presentazione degli risultati seguente si è scelto invece di limitarsi agli indici FIT_1 e FIT_{20} (cioè i FIT relativi alle predizioni a un passo e a venti passi di predizione) poiché con tali due valori ci si può fare un'idea dell'andamento dell'indice FIT_k in funzione di k .

In Tabella 4.1 sono riportati i mediani del FIT_1 e del FIT_{20} (riportato in tabella tra parentesi quadre) per studi Monte Carlo considerando diverse lunghezze del set di dati N , dove per ciascun studio sono stati considerati 200 esperimenti. I valori di tale tabella sono riportati graficamente nei due grafici di Figura 4.2.

Tabella 4.1 – Mediani degli esperimenti degli studi Monte Carlo di FIT_1 e di FIT_{20} (riportati tra parentesi quadre), considerando diverse lunghezze del set di dati N , impiegando $PEM+Oracolo+set$, $PEM+BIC$ e $SSEH$ sub-ottimo con diverse lunghezze delle risposte impulsive t_0 e avendo usato modelli di riferimento con livello di sparsità ridotto per generare i dati. Gli ingressi sono stati generati in modo da essere scorrelati. Come lunghezze delle risposte impulsive di $SSEH$ sub-ottimo sono state usate $t_0 \in \{10, 30, 40\}$ e le lunghezze stimate mediante $PEM/ARX+AIC$, le cui medie per le diverse N sono riportate nella seconda riga della presente tabella.

FIT_1 [FIT_{20}]	$N = 150$	$N = 200$	$N = 300$	$N = 500$	$N = 800$
\bar{t}_0 -ARX	4.63	5.38	6.52	7.72	9.28
Riferimento	0.872 [0.857]	0.873 [0.855]	0.874 [0.856]	0.871 [0.858]	0.872 [0.854]
$PEM+Or+set$	0.793 [0.779]	0.817 [0.795]	0.839 [0.81]	0.849 [0.82]	0.853 [0.829]
$PEM+BIC$	0.53 [0.675]	0.632 [0.723]	0.766 [0.774]	0.824 [0.801]	0.843 [0.819]
$SSEH$ t_0 -ARX	0.807 [0.739]	0.833 [0.75]	0.843 [0.786]	0.854 [0.808]	0.858 [0.825]
$SSEH$ $t_0 = 10$	0.807 [0.746]	0.835 [0.76]	0.844 [0.791]	0.854 [0.812]	0.858 [0.827]
$SSEH$ $t_0 = 30$	0.79 [0.733]	0.825 [0.756]	0.842 [0.788]	0.852 [0.812]	0.857 [0.827]
$SSEH$ $t_0 = 40$	0.78 [0.725]	0.819 [0.754]	0.841 [0.787]	0.853 [0.812]	0.858 [0.827]

Da Tabella 4.1 e dalla relativa Figura 4.2 si vede che, all'aumentare della lunghezza del set di dati N impiegati per l'identificazione, migliorano le prestazioni di tutti i metodi di identificazione impiegati in termini di FIT_k e si avvicinano alla prestazione in predizione del modello di riferimento sparso, usato per generare i dati.

Per quanto riguarda le prestazioni, dalla Figura 4.2 (soprattutto quella relativa

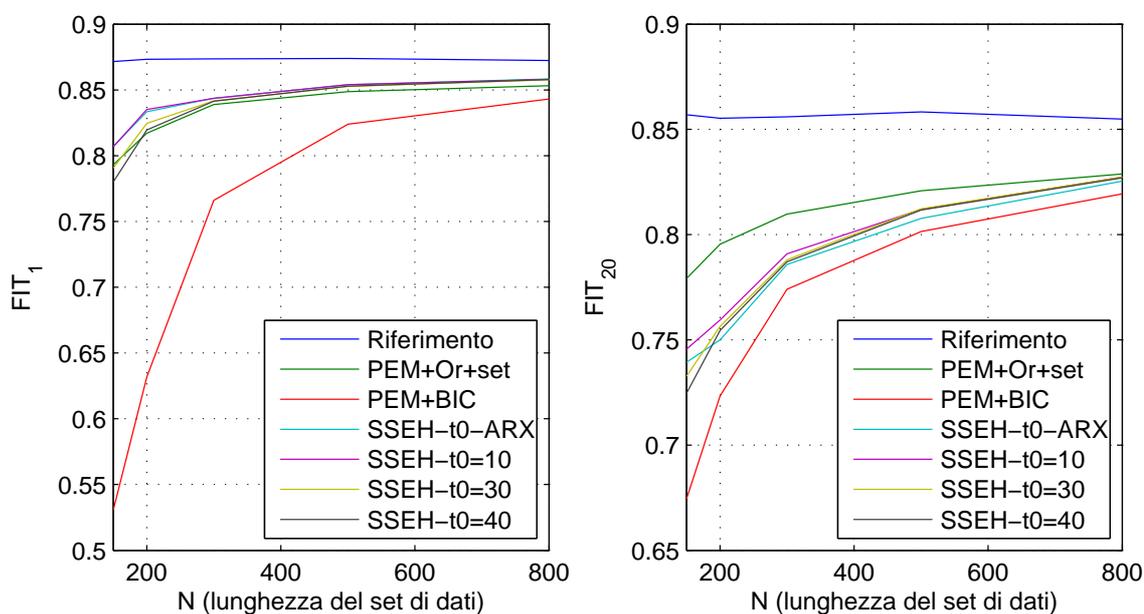


Figura 4.2 – Mediani di FIT_1 e FIT_{20} di uno studio Monte Carlo in funzione alla lunghezza del set di dati N , relativo a Tabella 4.1.

al FIT_1 , sinistra) è immediatamente evidente che l'approccio PEM, nel caso in cui deve effettuare una selezione dell'ordine e non conosce il set degli ingressi che non contribuiscono all'evoluzione del sistema (la versione PEM+BIC riportata nell'immagine), è poco efficiente quando si ha un numero ridotto di dati a disposizione. All'aumentare dei dati invece, le prestazioni di tale versione si avvicinano a quella della versione PEM che ha accesso ai dati di test per la selezione dell'ordine e conosce il set di ingressi che non contano nell'evoluzione del sistema (PEM+Oracolo+set). Le prestazioni di PEM+Oracolo+set sono confrontabili a quelle delle versioni SSEH sub-ottima.

Dalla Tabella 4.1 si verifica anche quanto accennato prima, relativamente all'andamento dell'indice FIT_k in funzione dei passi di predizione k della versione PEM+BIC. Per N ridotti, il modello di rumore non è identificato in maniera soddisfacente. Infatti fino a $N = 300$ incluso, i FIT_k a un passo di predizione sono inferiori di quelli a venti passi. Per $N = 500$ e $N = 800$ invece, anche questi modelli presentano prestazioni migliori a pochi passi di predizione.

Nella seconda riga di Tabella 4.1 è riportata la media delle lunghezze delle risposte impulsive t_0 nei singoli esperimenti degli studi Monte Carlo di SSEH, stimate mediante il metodo PEM/ARX+AIC. Le prestazioni di SSEH, che impiega tale stima della lunghezza t_0 , sono riportate nella sesta riga. Dalla Tabella 4.1 e dalla relativa Figura 4.2 si vede che le prestazioni del metodo SSEH sub-ottimo non

variano di tanto al cambiare della lunghezza delle risposte impulsive. Si nota però che, per piccole dimensioni del set di dati N , non conviene scegliere lunghezze delle risposte impulsive troppo lunghe. Questo è dovuto al fatto che in realtà la lunghezza del set di dati usato per l'identificazione è dato da $N - t_0$ (si guardi sezione 2.2) e l'aumento della lunghezza delle risposte impulsive t_0 comporta una riduzione dei dati utilizzabili per l'identificazione. Questo si evidenzia comunque solo per lunghezze del set di dati N ridotti.

Per valutare l'influenza della lunghezza del set dei dati N e delle risposte impulsive identificate con SSEH sono stati riportati i box-plot, rappresentando per questione di spazio solo quelle relative alle lunghezza dei dati massima ($N = 800$) e minima ($N = 150$) considerate; in Figura 4.3 è riportato il box-plot per lo studio Monte Carlo relativo a $N = 150$ e in Figura 4.4 quello relativo a $N = 800$.

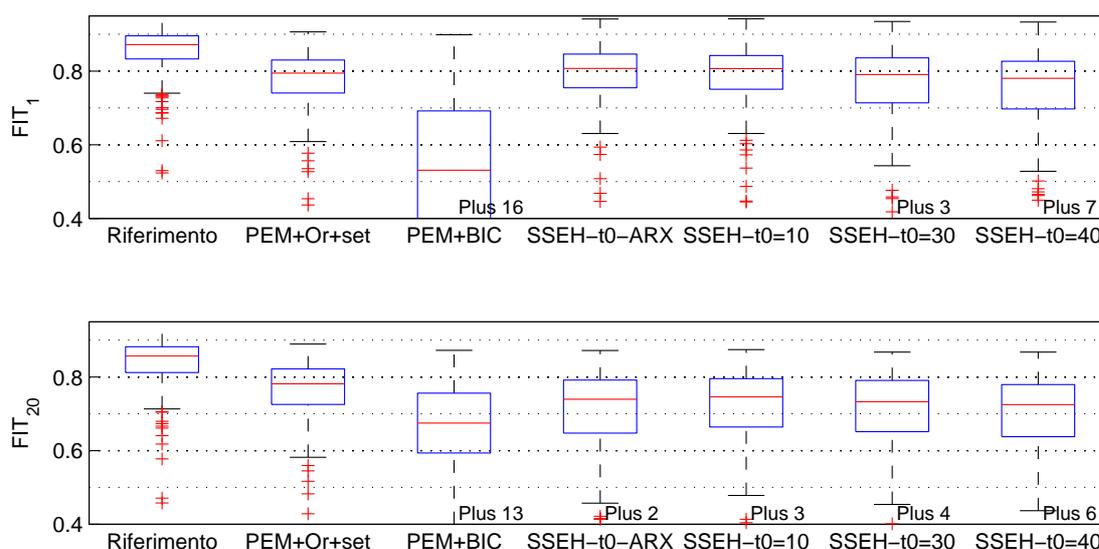


Figura 4.3 – Box-plot di FIT_1 e FIT_{20} di uno studio Monte Carlo con la lunghezza del set dei dati $N = 150$, relativo alla seconda colonna di Tabella 4.1.

Esaminando il box-plot relativo a $N = 150$ di Figura 4.3 e quello relativo a $N = 800$ di Figura 4.4, si confermano le considerazioni fatte a proposito di Tabella 4.1 e Figura 4.2: all'aumentare della lunghezza del set di dati migliorano le prestazioni dei modelli identificati e i FIT_k si avvicinano a quelli ottenuti con il modello di riferimento sparso.

Dalle Figure è immediatamente evidente che, per $N = 150$, PEM+BIC ha delle

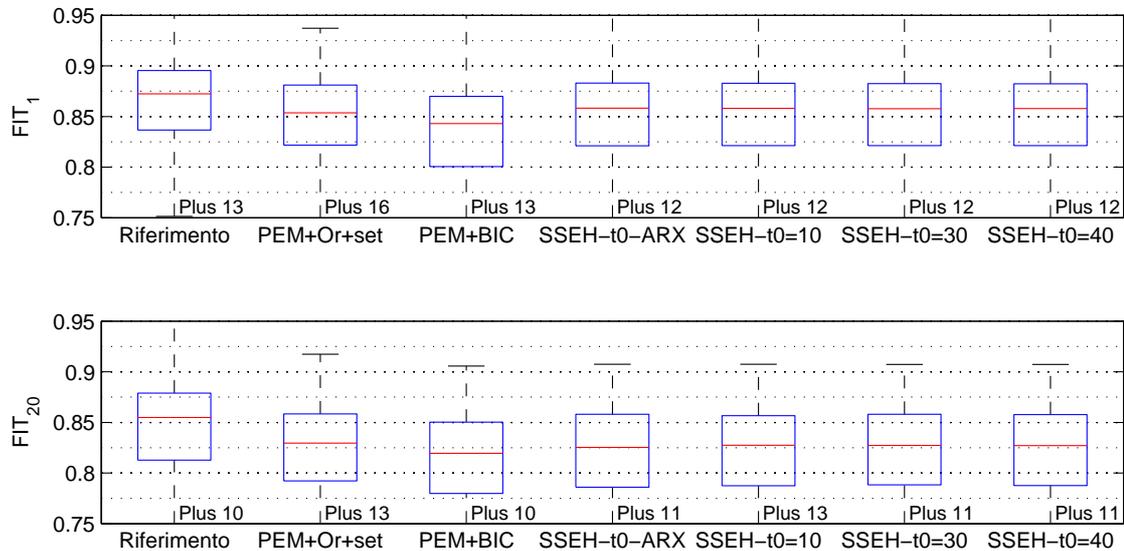


Figura 4.4 – Box-plot di FIT_1 e FIT_{20} di uno studio Monte Carlo con la lunghezza del set dei dati $N = 800$, relativo all'ultima colonna di Tabella 4.1.

prestazioni molto inferiori rispetto agli altri metodi di identificazione considerati, conducibile alle necessità sia di dover selezionare un modello adeguato dai pochi dati, sia di non conoscere le risposte impulsive nulle. Da Figura 4.3 si vede che questo comporta, per il metodo PEM+BIC, non solo che il mediano dei FIT_k sia inferiore, ma anche che la varianza degli indici sia più alta. Per $N = 800$ invece le prestazioni di PEM+BIC si avvicinano a quelle degli altri metodi e ai FIT_k del modello di riferimento.

Dalla Figura 4.3 si vede ancora che, per lunghezze di dati ridotte, conviene usare lunghezze delle risposte impulsive dei modelli identificati tramite SSEH non troppo lunghe.

Le percentuali delle risposte impulsive considerate in maniera corretta dal metodo SSEH sub-ottimo sono riportate in Tabella 4.2, separando le risposte impulsive relative agli ingressi da quelle relative alle uscite (riportate tra parentesi quadre). Si ricorda che in questa sezione, per questioni di tempo di calcolo, del metodo SSEH si considera solo la versione SSEH sub-ottima che impiega il modello Bayesiano ridotto, e la procedura di selezione delle variabili forward selection Bayesiano.

Come si vede da Tabella 4.2 la procedura di selezione nel caso di modello

Tabella 4.2 – Percentuali delle risposte impulsive relative agli ingressi e alle uscite (riportate in tabella tra parentesi quadre) considerate in maniera corretta per i diversi studi Monte Carlo, impiegando modelli con livello di sparsità ridotto e ingressi scorrelati. Nella seconda riga sono riportate le medie delle t_0 stimate mediante PEM/ARX+AIC per i relativi N .

%-in [%-out]	$N = 150$	$N = 200$	$N = 300$	$N = 500$	$N = 800$
\bar{t}_0 -ARX	4.63	5.38	6.52	7.72	9.28
SSEH t_0 -ARX	99.45 [94]	99.5 [96.16]	99.8 [97.67]	99.9 [98.67]	100 [99.17]
SSEH $t_0 = 10$	98.85 [91.5]	99.3 [94.5]	99.7 [96.83]	99.95 [98.5]	100 [99]
SSEH $t_0 = 30$	97.65 [91]	99.1 [93.83]	99.6 [96.67]	99.95 [98.3]	99.95 [99.5]
SSEH $t_0 = 40$	97.5 [90.5]	98.95 [93.93]	99.55 [96.1]	99.9 [98.3]	99.95 [99.33]

con livello di sparsità ridotto e ingressi non correlati funziona bene, soprattutto per determinare le risposte impulsive nulle relative agli ingressi (cioè le $h^{[i,j]}$ per $i = 1, \dots, p$ e $j = 1, \dots, m$). La prestazione della selezione delle $h^{[i,j]}$ relative alle uscite (cioè per $i = 1, \dots, p$ e $j = m + 1, \dots, m + p$) sono un po' inferiori, dato che le uscite sono correlate tra di loro e anche con gli ingressi. Inoltre dalla Tabella 4.2 si nota quanto già visto prima: non conviene usare lunghezze delle risposte impulsive t_0 troppo lunghe, soprattutto per lunghezze del set di dati ridotte, dato che all'aumentare di t_0 si riduce il set di dati effettivamente impiegato per l'identificazione.

4.1.2 Ingressi correlati e modelli con livello di sparsità ridotto

Si riportano qui i risultati ottenuti generando il modello ancora con livello di sparsità ridotto, in modo che solo qualche ingresso non contribuisca in alcun modo nell'evoluzione del sistema (qualche $h^{[i,j]} = 0$ per $j \in \{1, \dots, m\}$ e per ogni $i = 1, \dots, p$). Qui però gli ingressi sono stati generati in modo da essere tra di loro correlati, come spiegato in sezione 3.1.

In Tabella 4.3 sono riportati i mediani degli indici FIT_1 e FIT_{20} (riportati tra parentesi quadre) di uno studio Monte Carlo di 200 esperimenti, avendo generato gli ingressi in modo correlato. Sono riportati gli indici relativi al modello di riferimento e a modelli ottenuti impiegando i metodi di identificazione SSEH sub-ottimo con diverse lunghezza delle risposte impulsive t_0 , PEM+Oracolo+set

e PEM+BIC. In Figura 4.5 tali valori sono illustrati graficamente, riportando a sinistra l'andamento di FIT_1 e a destra l'andamento di FIT_{20} in funzione della lunghezza del set di dati N .

Tabella 4.3 – Mediani per gli studi Monte Carlo di FIT_1 e di FIT_{20} (riportati tra parentesi quadre), considerando diverse lunghezze del set di dati N , impiegando PEM+Oracolo+set, PEM+BIC e SSEH sub-ottimo con diverse lunghezze delle risposte impulsive t_0 e avendo usato modelli di riferimento con livello di sparsità ridotto per generare i dati. Gli ingressi sono stati generati in modo da essere correlati. Come lunghezze delle risposte impulsive di SSEH sub-ottimo sono state usate $t_0 \in \{5, 10, 30, 40\}$ e le lunghezze stimate mediante PEM/ARX+AIC, le cui medie per le diverse N sono riportate nella seconda riga della presente tabella.

FIT_1 [FIT_{20}]	$N = 150$	$N = 200$	$N = 300$	$N = 500$	$N = 800$
\bar{t}_0 -ARX	4.47	5.07	6.21	7.25	8.63
Riferimento	0.873 [0.86]	0.874 [0.852]	0.875 [0.859]	0.872 [0.858]	0.872 [0.852]
PEM+Or+set	0.689 [0.658]	0.723 [0.694]	0.771 [0.722]	0.797 [0.766]	0.819 [0.784]
PEM+BIC	-0.51 [0.238]	-0.08 [0.401]	0.43 [0.538]	0.676 [0.655]	0.751 [0.708]
SSEH t_0 -ARX	0.683 [0.547]	0.704 [0.623]	0.761 [0.678]	0.782 [0.73]	0.808 [0.761]
SSEH $t_0 = 10$	0.666 [0.556]	0.71 [0.616]	0.765 [0.693]	0.779 [0.735]	0.808 [0.76]
SSEH $t_0 = 30$	0.599 [0.484]	0.686 [0.589]	0.751 [0.678]	0.778 [0.732]	0.807 [0.763]
SSEH $t_0 = 40$	0.574 [0.466]	0.67 [0.573]	0.751 [0.677]	0.776 [0.728]	0.808 [0.764]

Da Tabella 4.3 e relativa Figura 4.5 si possono trarre conclusioni simili a quelle del caso di ingressi scorrelati: all'aumentare della lunghezza del set di dati N migliorano le prestazioni di tutti i metodi di identificazione impiegati. Le prestazioni di PEM+BIC sono molto scarsi per N ridotti, dato che, oltre a identificare i parametri, deve scegliere un'ordine del modello adeguato e non conosce gli ingressi che non contribuiscono all'evoluzione del sistema. PEM+Oracolo+set, che dispone dell'informazione degli ingressi superflui e impiega i dati di test per la selezione dell'ordine, ha invece prestazioni confrontabili con quelle di SSEH sub-ottimo. Si osserva inoltre che, all'aumentare della lunghezza delle risposte impulsive t_0 di SSEH, le prestazioni di tale metodo di identificazione peggiorano. Come accennato precedentemente questo è dovuto al fatto che la dimensione del set di dati utile per l'identificazione SSEH si riduce all'aumentare di t_0 (si guardi sezione 2.2).

In confronto al caso di ingressi scorrelati, le prestazioni usando ingressi correlati sono peggiori. Si nota però che, all'aumentare di N , le prestazioni dei due casi si

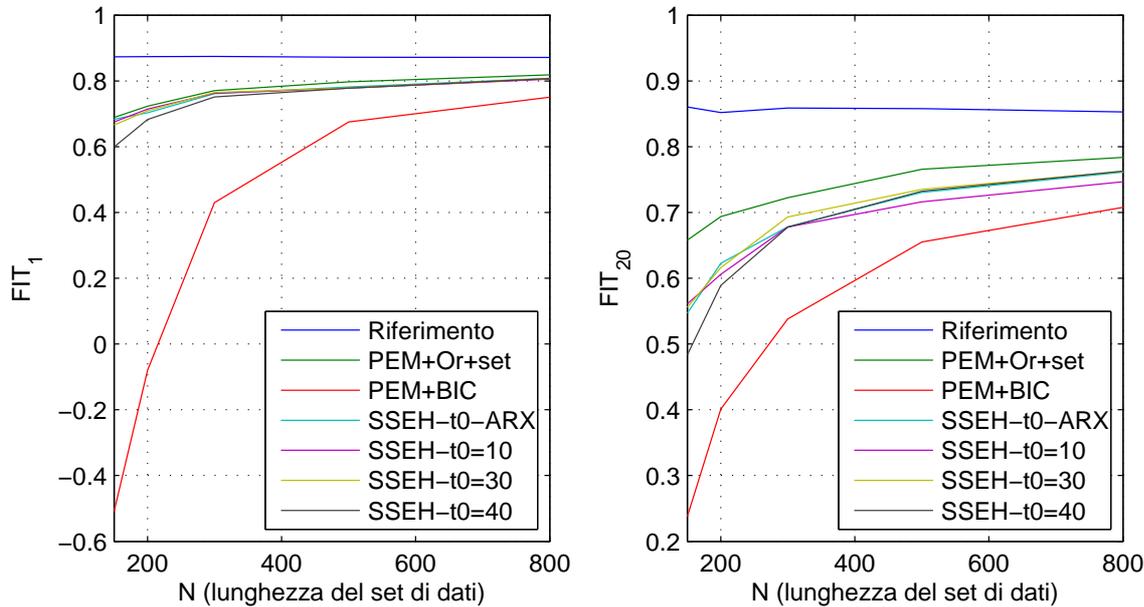


Figura 4.5 – Mediani di FIT_1 e FIT_{20} di uno studio Monte Carlo in funzione alla lunghezza del set di dati N , relativo a Tabella 4.3.

avvicinano sempre di più per il fatto che sia con ingressi correlati sia con ingressi scorrelati, i FIT si avvicinano sempre di più a quelli ottimali del modello di riferimento. Questo allineamento delle prestazioni è molto evidente per l'identificazione PEM+BIC, ma si riconosce anche per gli altri metodi di identificazione impiegati.

In Figura 4.6 e in Figura 4.7 sono riportati i box-plot dei mediani degli indici FIT_1 e FIT_{20} , rispettivamente per $N = 150$ e per $N = 800$ per studi di Monte Carlo, avendo generato modelli con livello di sparsità ridotto e ingressi correlati.

Dalle Figure 4.6 e 4.7 si riaffermano le considerazioni fatte in precedenza sulle prestazioni dei diversi metodi di identificazione al variare di N e di t_0 . In confronto agli ingressi scorrelati, nel caso di ingressi correlati, si vede che oltre alla diminuzione dei mediani, anche la varianza degli FIT aumenta. Questo si evidenzia soprattutto per l'identificazione PEM+BIC, dove i FIT , anche per $N = 800$, presentano una varianza nettamente maggiore di quelle di altri modelli.

Le prestazioni in termini di percentuale di ingressi e uscite correttamente considerate dalla procedura di selezione (forward selection Bayesiano) di SSEH sono riportati in Tabella 4.4.

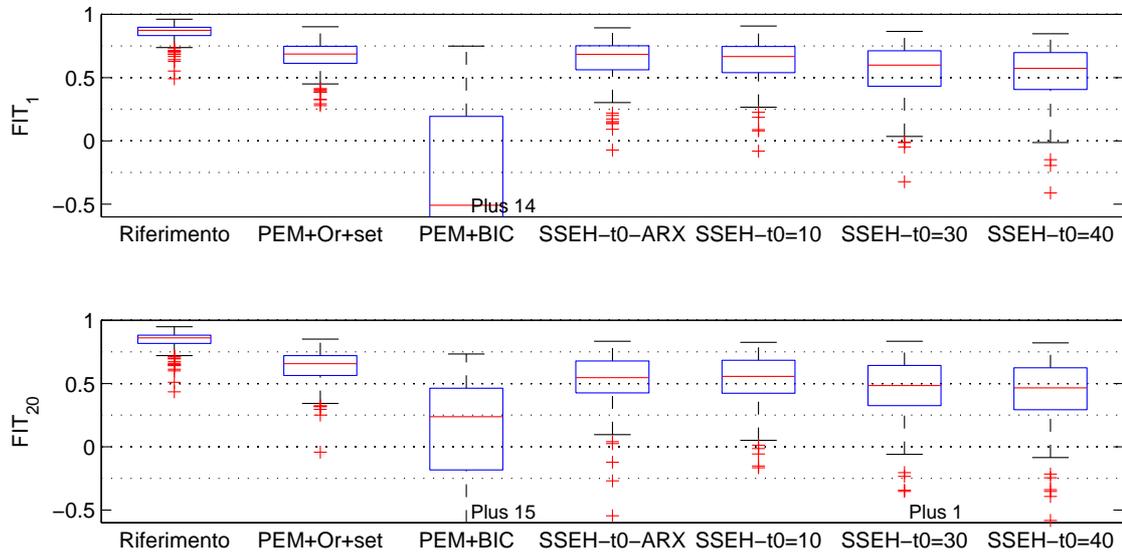


Figura 4.6 – Box-plot di FIT_1 e FIT_{20} di uno studio Monte Carlo con la lunghezza del set dei dati $N = 150$, relativo alla seconda colonna di Tabella 4.3.

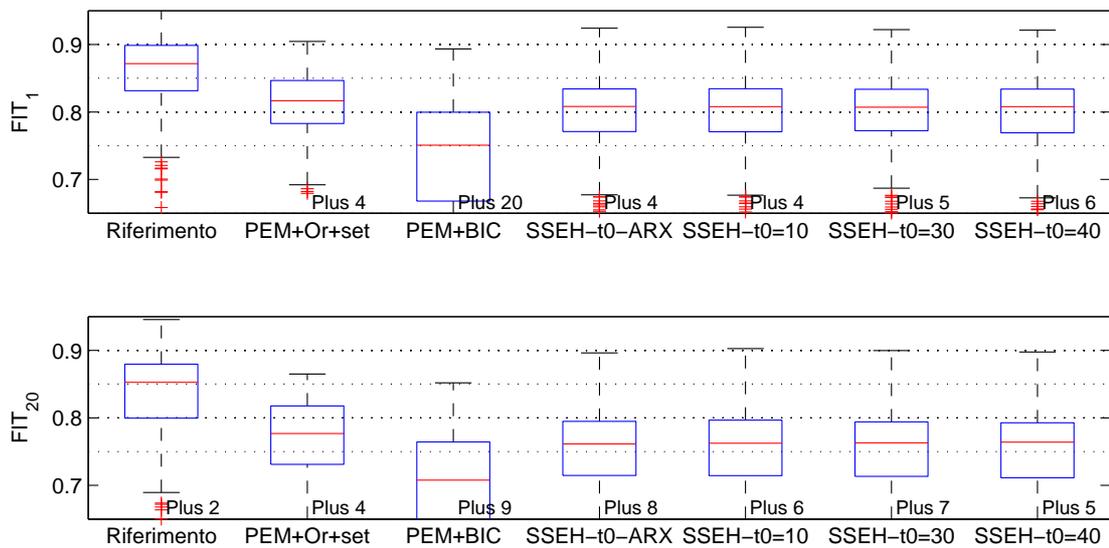


Figura 4.7 – Box-plot di FIT_1 e FIT_{20} di uno studio Monte Carlo con la lunghezza del set dei dati $N = 800$, relativo all'ultima colonna di Tabella 4.3.

Tabella 4.4 – Percentuali delle risposte impulsive relative agli ingressi e alle uscite (riportate tra parentesi quadre) considerate in maniera corretta per i diversi studi Monte Carlo, impiegando modelli con livello di sparsità ridotto e ingressi correlati. Nella seconda riga sono riportate le medie delle t_0 stimate mediante PEM/ARX+AIC per i relativi N .

%-in [%-out]	$N = 150$	$N = 200$	$N = 300$	$N = 500$	$N = 800$
\bar{t}_0 -ARX	4.47	5.07	6.21	7.25	8.63
SSEH t_0 -ARX	93.9 [95]	94.75 [96.5]	97.05 [97.83]	97.9 [98.5]	98.4 [99.33]
SSEH $t_0 = 10$	92.4 [92.3]	94.9 [95.17]	97.15 [97.17]	97.8 [99]	98.5 [98.83]
SSEH $t_0 = 30$	90.2 [91.5]	93.85 [93.5]	96.9 [96.33]	97.8 [98.67]	98.5 [98.5]
SSEH $t_0 = 40$	88.85 [90.33]	93.35 [93.5]	96.95 [96.33]	97.45 [98.67]	98.45 [98.5]

Anche nel caso di ingressi correlati, la procedura di selezione mantiene delle prestazioni buone, seppure leggermente inferiori del caso di ingressi scorrelati. Come si vede da Tabella 4.4 i risultati migliorano all'aumentare della lunghezza del set di dati N , mentre per N ridotti peggiorano anche all'aumentare della lunghezza delle risposte impulsive t_0 . Diversamente da quanto visto per ingressi scorrelati, la percentuale di uscite considerate in maniera corretta è in generale leggermente più alta. Analizzando i dati si è visto che, per ingressi non correlati le uscite hanno coefficienti di correlazione nettamente superiori agli ingressi, mentre in questo caso di ingressi correlati le uscite hanno coefficienti di correlazione inferiori rispetto a quello degli ingressi. Si presume che la percentuale più alta di uscite considerate in maniera corretta in questo caso sia dovuta proprio al fatto che le uscite risultano meno correlate degli ingressi.

4.1.3 Ingressi non correlati e modelli con livello di sparsità aumentato

Si riportano qui i risultati dei studi Monte Carlo avendo generato ingressi scorrelati e modelli con livello di sparsità aumentato (si è annullato qualche $h^{[i,j]}$ per ogni $i = 1, \dots, p$ e in aggiunta qualche $h^{[i,j]}$ per qualche $i \in \{1, \dots, p\}$), come spiegato in sezione 3.1.

Nei paragrafi precedenti, nella presentazione dei risultati di studi Monte Carlo dove sono stati utilizzati modelli con livello di sparsità ridotto, si è visto come l'aumento della lunghezza delle risposte impulsive per SSEH t_0 influenza le prestazioni. Si è scelto per questo di considerare $t_0 = 5$ al posto di $t_0 = 40$, per valutare

le prestazioni di SSEH impiegando una lunghezza delle risposte impulsive ridotta, spesso inferiore a quella stimata da PEM/ARX+AIC. Nella presentazione dei risultati impiegando modelli con livello di sparsità aumentato si considerano inoltre entrambe le versioni PEM che hanno accesso al set dei dati di test per effettuare la selezione dell'ordine, PEM+Oracolo e PEM+Oracolo+set.

Si ricorda che, come spiegato in sezione 3.1, non è chiaro come sia possibile usare l'informazione delle $h^{[i,j]}$ nulle nel modello di riferimento per qualche (non per tutte) $i \in \{1, \dots, p\}$ nell'identificazione PEM. Per questo tale informazione non è considerata nella versione PEM+Oracolo+set, ma è utilizzata solo l'informazione delle $h^{[i,j]} = 0$ per ogni $i = 1, \dots, p$. Anche la versione SSEH sub-ottima con la procedura forward selection Bayesiana presentata in sezione 2.2.5 e considerata qui, prevede la sola possibilità di annullare tutte le $h^{[i,j]}$ per un determinato j . Il confronto tra SSEH sub-ottimo e PEM+Oracolo+set, impiegando modelli con livello di sparsità aumentato, è quindi equo.

In Tabella 4.5 si riportano i mediani degli indici FIT_1 e FIT_{20} (riportato tra parentesi quadre) di diversi studi Monte Carlo per valutare l'influenza di N e di t_0 , impiegando ingressi scorrelati e modelli con livello di sparsità aumentato. Questi valori sono riportati graficamente in Figura 4.8, riportando a sinistra i FIT_1 e a destra i FIT_{20} . Nella seconda riga della tabella sono riportati inoltre le medie per i diversi N dell'ordine stimato mediante PEM/ARX+AIC nei diversi studi Monte Carlo.

Come nel caso di modelli di riferimento con livello di sparsità ridotto, anche per modelli di riferimento con livello di sparsità aumentato si verifica che, all'aumentare della lunghezza del set di dati N , migliorano le prestazioni dei diversi metodi di identificazione, come si vede da Tabella 4.5 e dalla relativa Figura 4.8. Per N ridotto impiegando la versione SSEH sub-ottima sono stati ottenuti prestazioni migliori che impiegando i metodi PEM, anche rispetto alla versione PEM+Oracolo+set. All'aumentare della lunghezza del set di dati N le prestazioni di tutti i metodi si avvicinano e tendono al FIT_k ottenuto impiegando il modello di riferimento.

Per quanto riguarda l'influenza della lunghezza delle risposte impulsive di SSEH t_0 sulle proprie prestazioni, vale ancora che, per N ridotti, non conviene scegliere t_0 troppo alto, dato che si riduce la lunghezza del set di dati effettivamente utilizzabile per l'identificazione. Da Tabella 4.5 e dalla relativa Figura 4.8 si osserva però anche che, per t_0 troppo piccolo, aumentando N , le prestazioni dopo un

Tabella 4.5 – Mediani per gli studi Monte Carlo di FIT_1 e di FIT_{20} (riportati tra parentesi quadre) considerando diverse lunghezze del set di dati N , impiegando PEM+Oracolo+set, PEM+BIC e SSEH sub-ottimo con diverse lunghezze delle risposte impulsive t_0 e avendo usato modello di riferimento con livello di sparsità aumentato per generare i dati. Gli ingressi sono stati generati in modo da essere scorrelati. Come lunghezze delle risposte impulsive di SSEH sub-ottimo sono state usate $t_0 \in \{5, 10, 30\}$ e le lunghezze stimate mediante PEM/ARX+AIC, le cui medie per le diverse N sono riportate nella seconda riga della presente tabella.

FIT_1 [FIT_{20}]	$N = 150$	$N = 200$	$N = 300$	$N = 500$	$N = 800$
\bar{t}_0 -ARX	6.71	8.95	10.58	12.36	13.83
Riferimento	0.985 [0.976]	0.984 [0.975]	0.984 [0.975]	0.985 [0.976]	0.983 [0.975]
PEM+Or+set	0.862 [0.813]	0.916 [0.89]	0.967 [0.952]	0.975 [0.962]	0.978 [0.968]
PEM+Or	0.787 [0.783]	0.897 [0.872]	0.958 [0.943]	0.972 [0.956]	0.976 [0.967]
PEM+BIC	0.532 [0.717]	0.781 [0.8]	0.871 [0.886]	0.944 [0.92]	0.965 [0.947]
SSEH t_0 -ARX	0.955 [0.895]	0.965 [0.927]	0.973 [0.95]	0.978 [0.961]	0.978 [0.964]
SSEH $t_0 = 5$	0.937 [0.849]	0.941 [0.851]	0.943 [0.876]	0.948 [0.879]	0.949 [0.891]
SSEH $t_0 = 10$	0.956 [0.904]	0.966 [0.927]	0.971 [0.945]	0.976 [0.956]	0.975 [0.956]
SSEH $t_0 = 30$	0.946 [0.876]	0.961 [0.924]	0.973 [0.953]	0.979 [0.964]	0.979 [0.967]

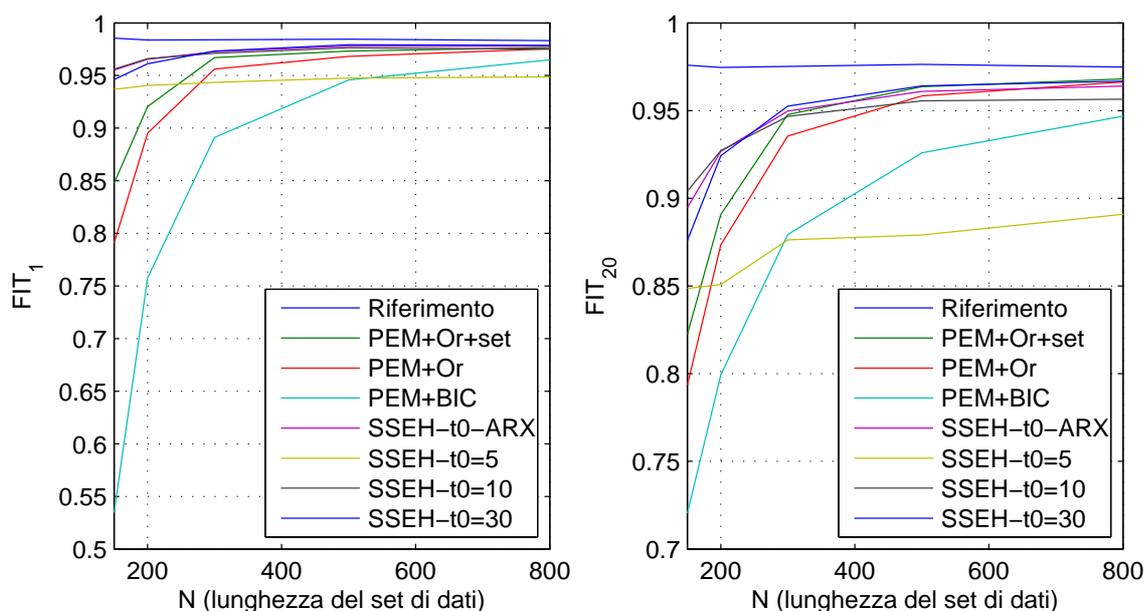


Figura 4.8 – Mediani di FIT_1 e FIT_{20} di uno studio Monte Carlo in funzione alla lunghezza del set di dati N , relativo a Tabella 4.5.

certo istante non migliorano praticamente più. Il motivo è che dopo un certo N i singoli campioni sono identificati in modo quasi ottimale, senza lasciare grandi incertezze, e quindi, aumentando ulteriormente N , alle risposte impulsive manca la flessibilità per poter spiegare informazioni aggiuntive.

Confrontando le tre versioni del metodo PEM considerate: PEM+Oracolo+set, PEM+Oracolo e PEM+BIC, per N piccolo sembra che sia più importante selezionare l'ordine adeguato del modello che conoscere le variabili che effettivamente servono a predire l'evoluzione del sistema. All'aumentare di N però entrambe le informazioni si ricavano dai dati e le diverse versioni convergono alle stesse prestazioni.

Di seguito sono riportati i box-plot per gli N più estremi considerati negli studi Monte Carlo, dove sono stati generati ingressi scorrelati e modelli di riferimento con livello di sparsità aumentato: in Figura 4.9 è riportato il box-plot per $N = 150$ e in Figura 4.10 quello per $N = 800$.

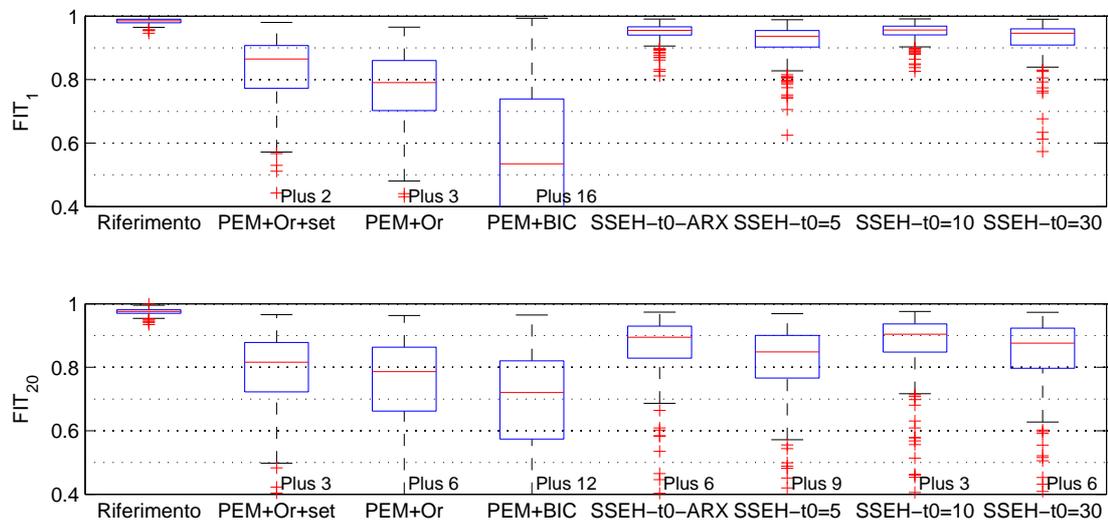


Figura 4.9 – Box-plot di FIT_1 e FIT_{20} di uno studio Monte Carlo con la lunghezza del set dei dati $N = 150$, relativo alla seconda colonna di Tabella 4.5.

Valutando i box-plot di Figura 4.10 e di Figura 4.9 si verificano le considerazioni appena fatte sia in termini di mediani, che in termini di varianza degli risultati.

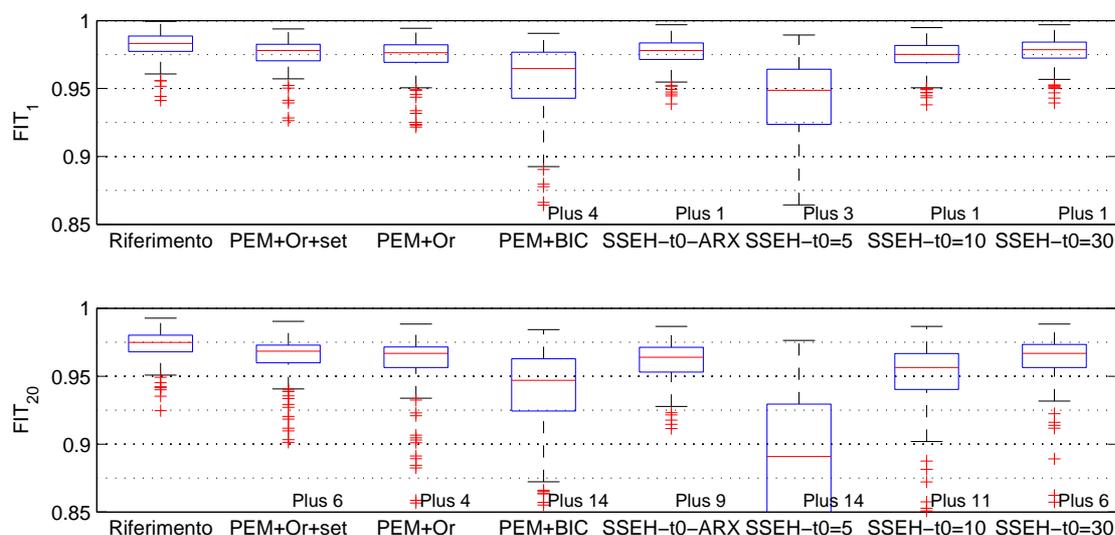


Figura 4.10 – Box-plot di FIT_1 e FIT_{20} di uno studio Monte Carlo con la lunghezza del set dei dati $N = 800$, relativo all'ultima colonna di Tabella 4.5.

In Tabella 4.6 sono presentati le percentuali degli ingressi e delle uscite (riportati tra parentesi quadre) considerate in maniera corretta nei diversi studi Monte Carlo, avendo generato ingressi scorrelati e modelli di riferimento con livello di sparsità aumentato.

Tabella 4.6 – Percentuali degli ingressi e delle uscite (riportate in tabella tra parentesi quadre) considerate in maniera corretta per diversi studi Monte Carlo impiegando modelli con livello di sparsità aumentato e ingressi scorrelati. Nella seconda riga sono riportate le medie delle t_0 stimate mediante PEM/ARX+AIC per i relativi N .

%-in [%-out]	$N = 150$	$N = 200$	$N = 300$	$N = 500$	$N = 800$
\bar{t}_0 -ARX	6.71	8.95	10.58	12.36	13.83
SSEH t_0 -ARX	99.9 [74.83]	100 [72.83]	100 [71.83]	100 [71.67]	100 [70.5]
SSEH $t_0 = 5$	99.75 [74.83]	99.95 [72.83]	100 [71.83]	99.8 [71.67]	100 [70.5]
SSEH $t_0 = 10$	99.9 [74.83]	100 [72.83]	100 [71.83]	100 [71.83]	100 [70.5]
SSEH $t_0 = 30$	99.9 [75]	100 [72.83]	100 [71.83]	100 [72]	100 [70.5]

La Tabella 4.6 non è così espressiva come le relative tabelle nel caso di modelli di riferimento con livello di sparsità ridotto. Mentre la procedura di selezione funziona molto bene per gli ingressi, la percentuale di uscite considerate in maniera

corretta è molto inferiore. Controllando i risultati si è visto che, quasi in ogni esperimento dello studio Monte Carlo, tutte le uscite sono state considerate dalla procedura di selezione, anche se nel modello di riferimento c'erano delle uscite, il cui passato non contribuiva alla predizione. Non è ben chiaro il motivo per cui queste percentuali di uscite considerate correttamente sono più scarse di quelle del caso di modelli di riferimento con livello di sparsità ridotto. Un motivo sta nel fatto che le prestazioni in termini di FIT_k sono già molto buone e potrebbero migliorare quindi di poco se la selezione fosse fatta in maniera corretta. Un'altra ragione sta nel fatto che le uscite non sono molto correlate tra di loro e quindi, dato che l'informazione di una certa uscita non è contenuta nelle altre, sono poche quelle che possono essere eliminate. Analizzando i dati infatti, si è trovato che la media della correlazione lineare delle uscite è intorno a 0.5, un valore molto ridotto rispetto ai 0.985 che si verifica per gli ingressi se sono generati in maniera correlata.

Esaminando solo le percentuali di ingressi considerati in maniera corretta di Tabella 4.6, le tendenze per N piccolo e $t_0 = 5$ evidenziano come la procedura funzioni peggio che negli altri casi. La procedura di selezione migliora all'aumentare di t_0 , fissato N , anche per le percentuali di uscite considerate in maniera corretta. Queste tendenze sono in accordo con quanto detto in precedenza.

Dalla Tabella 4.6 sembra inoltre che le percentuali di uscite considerate in maniera corretta peggiorino all'aumentare di N . Controllando però i risultati si è visto che questa decrescita si è sviluppata casualmente annullando le $h^{[i,j]}$ per $j \in \{m+1, \dots, m+p\}$ (le risposte impulsive del predittore relative alle uscite) negli studi Monte Carlo.

4.1.4 Ingressi correlati e modelli con livello di sparsità aumentato

Si presentano qui i risultati degli studi Monte Carlo con lo scopo di valutare l'influenza di N e di t_0 sui diversi metodi di identificazione, generando ingressi correlati e modelli con livello di sparsità aumentato.

In Tabella 4.7 sono riportati i mediani di FIT_1 e di FIT_{20} per i diversi studi Monte Carlo. In Figura 4.11 tali mediani sono riportati graficamente.

Valgono le stesse considerazioni qualitative fatte nel caso di modelli di riferimento con livello di sparsità aumentato e ingressi scorrelati. Come si vede da

Tabella 4.7 – Mediani per gli studi Monte Carlo di FIT_1 e di FIT_{20} (riportati tra parentesi quadre), considerando diverse lunghezze del set di dati N , impiegando PEM+Oracolo+set, PEM+Oracolo, PEM+BIC e SSEH sub-ottimo con diverse lunghezze delle risposte impulsive t_0 e avendo usato modello di riferimento con livello di sparsità aumentato per generare i dati. Gli ingressi sono stati generati in modo da essere correlati. Come lunghezze delle risposte impulsive di SSEH sub-ottimo sono state usate $t_0 \in \{5, 10, 30\}$ e le lunghezze stimate mediante PEM/ARX+AIC, le cui medie per le diverse N sono riportate nella seconda riga della presente tabella.

FIT_1 [FIT_{20}]	$N = 150$	$N = 200$	$N = 300$	$N = 500$	$N = 800$
\bar{t}_0 -ARX	6.4	7.92	9.23	10.9	12.19
Riferimento	0.983 [0.975]	0.985 [0.974]	0.984 [0.975]	0.983 [0.976]	0.984 [0.976]
PEM+Or+set	0.818 [0.672]	0.876 [0.807]	0.935 [0.89]	0.952 [0.927]	0.958 [0.94]
PEM+Or	0.591 [0.487]	0.749 [0.676]	0.887 [0.833]	0.929 [0.89]	0.946 [0.92]
PEM+BIC	-0.563 [0.282]	0.05 [0.472]	0.709 [0.723]	0.841 [0.8]	0.911 [0.854]
SSEH t_0 -ARX	0.882 [0.76]	0.917 [0.827]	0.928 [0.86]	0.95 [0.912]	0.963 [0.932]
SSEH $t_0 = 5$	0.858 [0.703]	0.881 [0.731]	0.896 [0.761]	0.911 [0.756]	0.914 [0.803]
SSEH $t_0 = 10$	0.887 [0.78]	0.921 [0.836]	0.928 [0.877]	0.95 [0.909]	0.959 [0.923]
SSEH $t_0 = 30$	0.849 [0.728]	0.906 [0.826]	0.937 [0.887]	0.95 [0.918]	0.963 [0.938]

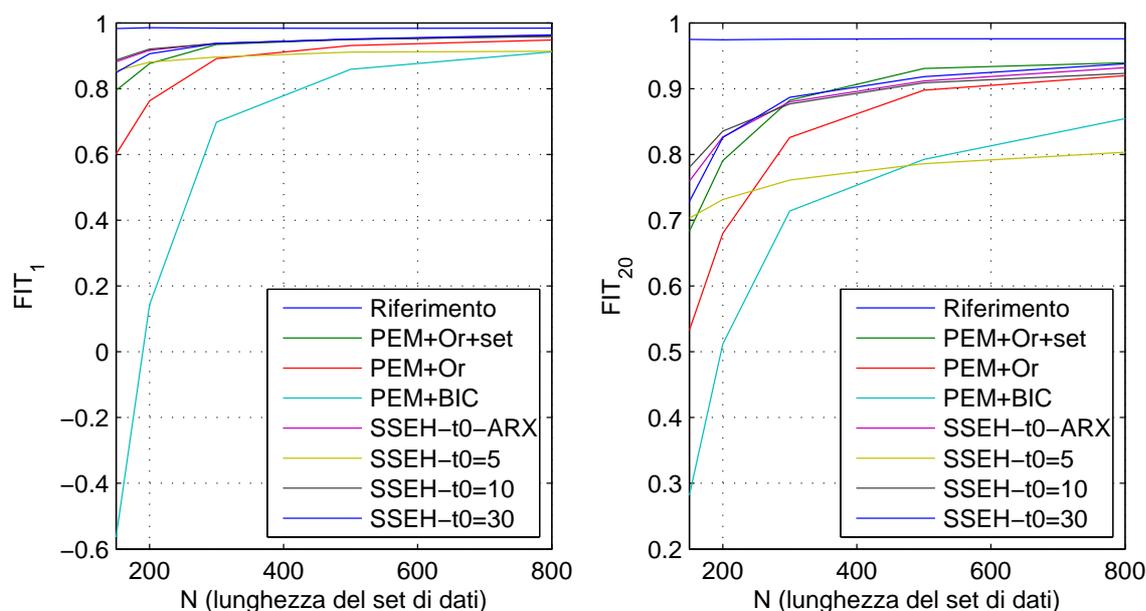


Figura 4.11 – Mediani di FIT_1 e FIT_{20} di uno studio Monte Carlo in funzione alla lunghezza del set di dati N , relativo a Tabella 4.7.

Tabella 4.7 e dalla relativa Figura 4.11, oltre al fatto che le prestazioni migliorano all'aumento di N , per lunghezze del set di dati N ridotte, i modelli identificati mediante SSEH sub-ottimo producono risultati migliori dei metodi PEM.

Per i modelli identificati mediante SSEH si verifica ancora che, per N ridotti, non conviene scegliere la lunghezza delle risposte impulsive t_0 eccessiva e che d'altro canto, per riuscire a sfruttare tutti i dati all'aumentare di N , servono t_0 maggiori. Confrontando i metodi PEM, si verifica ancora che soprattutto per N ridotti l'informazione dell'adeguato ordine del modello incide molto sulle prestazioni. Anche l'informazione sulle variabili effettivamente impiegate per predire il modello comporta un miglioramento delle prestazioni. Aumentando N si osserva comunque che le prestazioni delle diverse versioni PEM convergono.

Quantitativamente, in confronto al caso di modelli di riferimenti con livello di sparsità aumentato e ingressi scorrelati, le prestazioni delle identificazioni sono inferiori, come attesa a causa delle correlazioni degli ingressi.

Le considerazioni appena fatte si osservano anche dai box-plot riportati in Figura 4.12 per $N = 150$ e in Figura 4.13 per $N = 800$. Tali considerazioni possono essere tratte in relazione alle varianze dei FIT_1 e FIT_{20} , oltre che ai mediani.

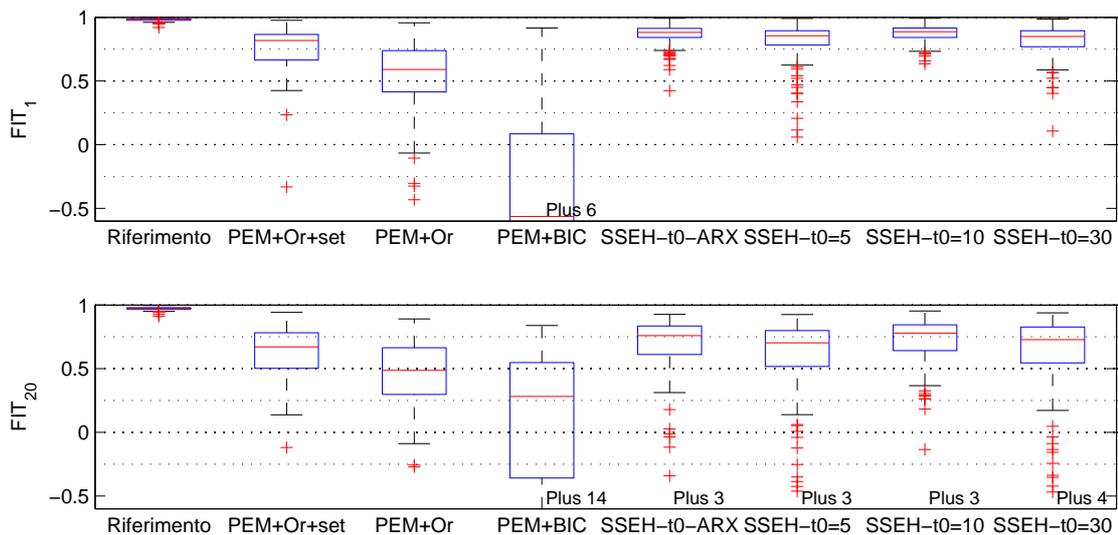


Figura 4.12 – Box-plot di FIT_1 e FIT_{20} di uno studio Monte Carlo con la lunghezza del set dei dati $N = 150$, relativo alla seconda colonna di Tabella 4.7.

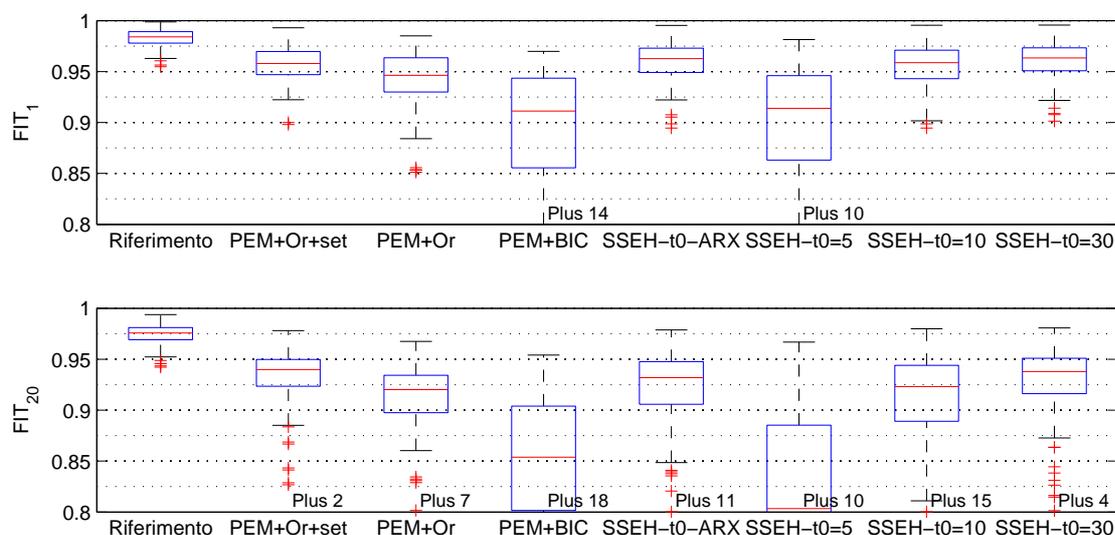


Figura 4.13 – Box-plot di FIT_1 e FIT_{20} di uno studio Monte Carlo con la lunghezza del set dei dati $N = 800$, relativo all'ultima colonna di Tabella 4.7.

In Tabella 4.8 sono riportati le percentuali degli ingressi e delle uscite (riportati tra parentesi quadre) considerati in maniera corretta dalla procedura di forward selection Bayesiana di SSEH sub-ottimo.

Tabella 4.8 – Percentuali degli ingressi e delle uscite (riportate in tabella tra parentesi quadre) considerate in maniera corretta per diversi studi Monte Carlo impiegando modelli con livello di sparsità aumentato e ingressi correlati. Nella seconda riga sono riportate le medie delle t_0 stimate mediante PEM/ARX+AIC per i relativi N .

%-in [%-out]	$N = 150$	$N = 200$	$N = 300$	$N = 500$	$N = 800$
\bar{t}_0 -ARX	6.4	7.92	9.23	10.9	12.19
SSEH t_0 -ARX	96.3 [71.33]	97.9 [72.5]	98.4 [73.33]	98.9 [71.83]	99.35 [72.67]
SSEH $t_0 = 5$	94.9 [71.5]	96.45 [72.5]	97.75 [73.33]	98.3 [71.83]	99.35 [72.67]
SSEH $t_0 = 10$	96.35 [71.67]	98.05 [72.67]	98.45 [73.5]	98.95 [72]	99.45 [73]
SSEH $t_0 = 30$	95.7 [71.5]	97.85 [72.67]	98.45 [73.5]	98.9 [72.33]	99.25 [73]

Le percentuali di uscite considerate in maniera corretta sono ancora scarse, come nel caso di modelli con livello di sparsità aumentato e ingressi scorrelati. In quel caso, come atteso e come si vede da Tabella 4.8, le percentuali di ingressi considerati correttamente era superiore, dato che la selezione è più onerosa per

ingressi correlati. Dalla Tabella si osserva anche chiaramente che tale percentuale migliora all'aumentare della lunghezza del set di dati N e che, per lunghezze delle risposte impulsive inadeguate (troppo corte o troppo lunghe), la percentuale diminuisce.

4.2 Diverse procedure di selezione di SSEH

In questa sezione si riportano i risultati di studi Monte Carlo svolti per confrontare diverse procedure di selezione delle variabili per SSEH.

Per questioni di efficienza computazionale, nelle simulazioni eseguite in questa sezione, è stata considerata solo la versione SSEH sub-ottima. L'influenza delle lunghezze del set di dati N e delle risposte impulsive t_0 sulle prestazioni dei modelli identificati con SSEH sub-ottimo è stata analizzata nella sezione precedente 4.1. In questa sezione si fa riferimento solo a $N = 300$, lunghezza per la quale, nei diversi casi considerati in precedenza, sia SSEH sia le diverse versioni di PEM considerate, hanno mostrato prestazioni accettabili. Come lunghezza delle risposte impulsive nelle simulazioni è stata impiegata la stima di t_0 adeguata mediante PEM/ARX+AIC, siccome i risultati apparivano buoni conservando comunque delle lunghezze t_0 ridotte.

Come modelli di riferimento nelle simulazioni eseguite in questo capitolo sono stati considerati sia quelli con livello di sparsità ridotto sia quelli con livello di sparsità aumentato. Sono stati utilizzati però solo ingressi correlati, per i quali la selezione è più impegnativa che non impiegando ingressi scorrelati, facendo emergere meglio vantaggi e svantaggi delle diverse procedure di selezione.

Questa sezione è divisa in due parti. Nella prima si considerano solo procedure che effettuano la selezione di variabili, cioè selezionano j , per i quali si annullano tutte le $h^{[i,j]}$ (per ogni $i = 1, \dots, p$). Lo scopo è analizzare le prestazioni delle procedure forward selection, backward elimination e stepwise selection sia nelle versioni sequenziali, sia in quelle Bayesiane, e la procedura all-best selection Bayesiana, presentate in sezione 3.3. Nella seconda parte di questa sezione si illustrano le procedure che possono selezionare le singole risposte impulsive del predittore $h^{[i,j]}$, utili soprattutto per modelli con livello di sparsità aumentato.

4.2.1 Procedure di selezione di variabili

Si presentano qui gli studi Monte Carlo eseguiti con lo scopo di valutare la bontà delle diverse procedure di selezione delle variabili per la predizione di un sistema. Questo significa che le procedure considerate sono in grado di annullare per determinati $j \in \{1, \dots, m + p\}$ le $h^{[i,j]}$ per ogni $i = 1, \dots, p$. Le procedure considerate sono forward selection, backward elimination, stepwise selection e all-best selection, i cui pregi e difetti sono presentati in sezione 3.3. Si considerano inizialmente le versioni delle procedure di selezione sequenziali e poi le versioni Bayesiane.

Versioni sequenziali delle procedure di selezione Si considerano qui le procedure di selezione di variabili sequenziali di forward selection, backward elimination e stepwise selection, per ciascuna sia la versione in avanti sia la versione all'indietro (si guardi 3.3). Si ricordi che le versioni in avanti valutano prima gli ingressi e successivamente le uscite, mentre le versioni all'indietro valutano prima le uscite e poi gli ingressi. Di maggior interesse è in realtà il confronto tra la versione in avanti e all'indietro della procedura forward selection, mentre sono state riportate entrambe le versioni per le altre procedure per dovere di completezza. In Tabella 4.9 sono riportati i mediani di FIT_1 e di FIT_{20} di uno studio Monte Carlo eseguito per confrontare tali procedure, avendo impiegato modelli di riferimento con livello di sparsità ridotto e ingressi scorrelati. Per avere un confronto si è impiegata anche la procedura forward selection Bayesiana e si sono riportati anche i mediani dei FIT in predizione dei modelli di riferimento. In Figura 4.14 è riportato il box-plot degli indici FIT relativi alla Tabella 4.9. In Tabella 4.10 sono riportate inoltre le percentuali degli ingressi e uscite selezionate in maniera corretta delle diverse procedure.

Tabella 4.9 – Mediani di FIT_1 e FIT_{20} di uno studio Monte Carlo generando modelli di riferimento con livello di sparsità ridotto e ingressi correlati, impiegando diverse procedure di selezione sequenziali. La media delle lunghezze delle risposte impulsive stimate da PEM/ARX+AIC e impiegate da SSEH sub-ottimo risulta $\bar{t}_0 = 5.98$.

	RIF	FOR-BAY	FOR-av	FOR-in	BAC-av	BAC-in	STEP-av	STEP-in
FIT_1	0.876	0.754	0.676	0.679	0.764	0.757	0.757	0.752
FIT_{20}	0.857	0.681	0.577	0.587	0.679	0.683	0.679	0.676

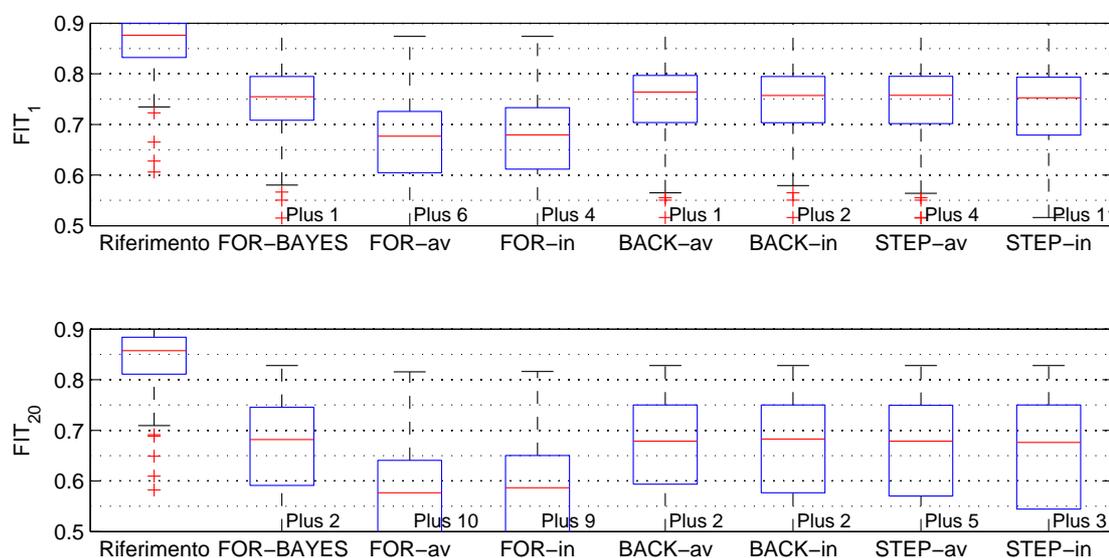


Figura 4.14 – Box-plot di FIT_1 e FIT_{20} di uno studio Monte Carlo eseguito per confrontare diverse procedure di selezione sequenziali del metodo di identificazione SSEH sub-ottimo, avendo generato modelli di riferimento con livello di sparsità ridotto e ingressi correlati.

Tabella 4.10 – Percentuali degli ingressi e delle uscite considerate in maniera corretta da diverse procedure di selezione sequenziali in uno studio Monte Carlo generando modelli di riferimento con livello di sparsità ridotto e ingressi correlati. La media delle lunghezze delle risposte impulsive stimate da PEM/ARX+AIC e impiegate da SSEH sub-ottimo risultava $\bar{t}_0 = 9.26$.

	FOR-av	FOR-in	BAC-av	BAC-in	STEP-av	STEP-in
% – IN	54.1	61.5	98.4	98.35	98.1	97.2
% – OUT	98.33	99.67	97.5	97	97.5	95

Dalla Tabella 4.9 e dalla relativa Figura 4.14 si nota chiaramente lo svantaggio in termini di FIT delle procedure sequenziali di forward selection in confronto alle procedure di backward elimination. In Tabella 4.10 lo svantaggio si mostra anche in termini di percentuale di variabili selezionate in maniera corretta. Come accennato in sezione 3.3 questo è dovuto al fatto che variabili, che non contribuiscono alla predizione del sistema ma che sono valutate prima di variabili con esse correlate che contribuiscono, sono selezionate erratamente. Che il motivo delle prestazioni inferiori sia riconducibile a quanto detto, si mostra nel fatto che la bontà delle versioni in avanti e all'indietro della forward selection sequenziali cambiano significativamente, soprattutto in termini di percentuali di variabili selezionate in maniera corretta, riportate in Tabella 4.10. Il problema è risolto nella procedura stepwise selection sequenziale, che è in pratica una procedura di forward selection equipaggiata con la possibilità di eliminare variabili superflue scelte in precedenza. Infatti, come si nota dalle tabelle e dalla figura, tale procedura ha prestazioni confrontabili con quelle della procedura sequenziale backward elimination.

Le prestazioni delle procedure sequenziali backward elimination sono comunque leggermente superiori a quelle delle procedure sequenziali stepwise selection. Da un'analisi approfondita dei risultati delle simulazioni sembra che questo sia dovuto al fatto che la procedura stepwise selection sequenziale ogni tanto sbaglia, non selezionando variabili che contribuiscono alla predizione del sistema e sommando quindi gli errori della forward selection e della backward elimination. Questo accade soprattutto nel caso in cui si valutino prima le uscite e successivamente gli ingressi e per questo le prestazioni della versione all'indietro sono inferiori a quelle della versione all'avanti.

In Tabella 4.11 sono riportati i mediani di FIT_1 e di FIT_{20} di uno studio Monte Carlo avendo generato ingressi correlati e modelli di riferimento con livello di sparsità aumentato e impiegando le diverse procedure sequenziali. Per avere un confronto si è impiegata anche la procedura forward selection Bayesiana e sono riportati anche i mediani dei FIT in predizione dei modelli di riferimento. In Figura 4.15 è riportato il relativo box-plot degli indici di FIT . In Tabella 4.12 invece sono riportati le percentuali degli ingressi e delle uscite considerate in maniera corretta dalle procedure sequenziali.

Nel caso dello studio Monte Carlo con modelli di riferimento con livello di sparsità aumentato, la differenza tra la procedura sequenziali forward selection e

Tabella 4.11 – Mediani di FIT_1 e FIT_{20} di uno studio Monte Carlo generando modelli di riferimento con livello di sparsità aumentato e ingressi correlati, impiegando diverse procedure di selezione sequenziali. La media delle lunghezze delle risposte impulsive stimate da PEM/ARX+AIC e impiegate da SSEH sub-ottimo risulta $\bar{t}_0 = 9.26$.

	RIF	FOR-BAY	FOR-av	FOR-in	BAC-av	BAC-in	STEP-av	STEP-in
FIT_1	0.983	0.935	0.902	0.906	0.936	0.936	0.936	0.936
FIT_{20}	0.976	0.874	0.816	0.821	0.874	0.874	0.874	0.874

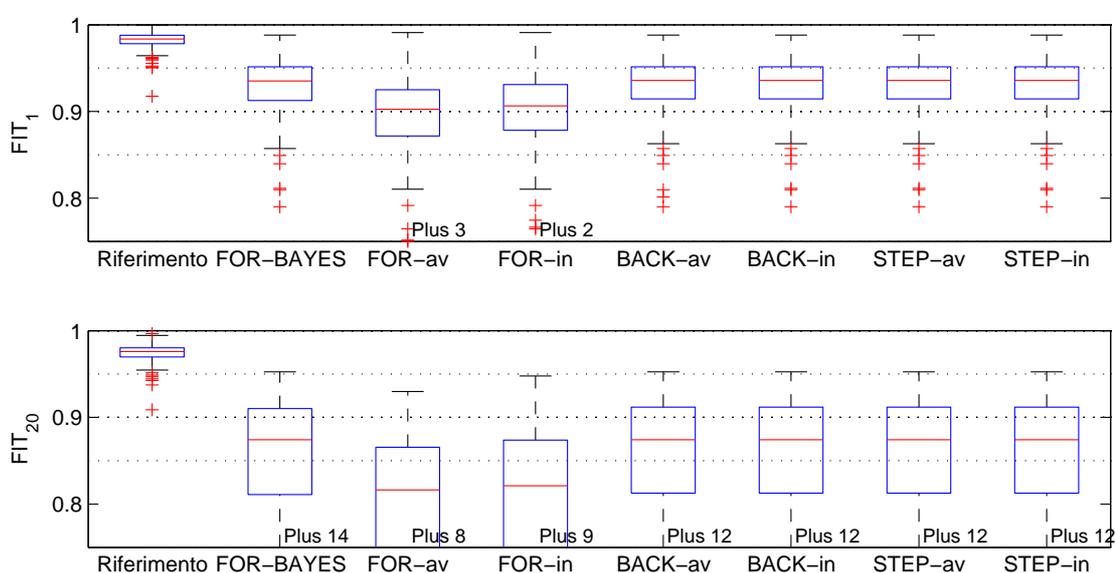


Figura 4.15 – Box-plot di FIT_1 e FIT_{20} di uno studio Monte Carlo eseguito per confrontare diverse procedure di selezione sequenziali del metodo di identificazione SSEH sub-ottimo, avendo generato modelli di riferimento con livello di sparsità aumentato e ingressi correlati.

Tabella 4.12 – Percentuali degli ingressi e delle uscite considerate in maniera corretta da diverse procedure di selezione sequenziali in uno studio Monte Carlo generando modelli di riferimento con livello di sparsità aumentato e ingressi correlati. La media delle lunghezze delle risposte impulsive stimate da PEM/ARX+AIC e impiegate da SSEH sub-ottimo risulta $\bar{t}_0 = 9.26$.

	FOR-av	FOR-in	BAC-av	BAC-in	STEP-av	STEP-in
% – IN	58.85	66.25	99.7	99.7	99.7	99.7
% – OUT	72.83	72.67	72.83	72.83	72.83	72.83

le altre procedure sequenziali si presenta ancora, ma in modo meno evidente che nel caso di modelli con livello di sparsità ridotto. Questo si nota dalle Tabelle 4.11 e 4.12 e dalla Figura 4.15. A differenza del caso di modelli con livello di sparsità ridotto si vede inoltre che le prestazioni delle procedure sequenziali backward elimination e stepwise selection sono identiche.

Versioni Bayesiane delle procedure di selezione Si presentano qui i risultati delle simulazioni eseguite per confrontare le procedure Bayesiane di selezione delle variabili: forward selection Bayesiana (la procedura standard impiegata già in precedenza e presentata in dettaglio in sezione 2.2.5), backward elimination Bayesiana, stepwise selection Bayesiana e all-best selection Bayesiana. In Tabella 4.13 sono riportati i mediani dei FIT_1 e FIT_{20} di uno studio Monte Carlo avendo considerato le diverse procedure Bayesiane e avendo generato modelli di riferimenti con livello di sparsità ridotto e ingressi correlati. Il relativo box-plot è riportato in Figura 4.16. In Tabella 4.14 sono riportate le percentuali degli ingressi e delle uscite considerati in maniera corretta dalle differenti procedure Bayesiane.

Tabella 4.13 – Mediani di FIT_1 e FIT_{20} di uno studio Monte Carlo generando modelli di riferimento con livello di sparsità ridotto e ingressi correlati, impiegando diverse procedure di selezione Bayesiane. La media delle lunghezze delle risposte impulsive stimate da PEM/ARX+AIC e impiegate da SSEH sub-ottimo risulta $\bar{t}_0 = 9.26$.

	RIFERIMENTO	FORWARD	BACKWARD	STEPWISE	ALL-BEST
FIT_1	0.871	0.748	0.753	0.754	0.754
FIT_{20}	0.858	0.662	0.667	0.672	0.672

Nel caso delle procedure Bayesiane non si presenta più una grande differenza tra la forward selection, la backward elimination e la stepwise selection, come visto per le procedure sequenziali. Le prestazioni sono infatti solo leggermente inferiori per la forward selection Bayesiana sia in termini di FIT (come si vede da Tabella 4.13 e dalla relativa Figura 4.16) sia in termini di percentuali di variabili considerate in maniera corretta (si guardi Figura 4.14). Dalle tabelle e dalla figura si vede inoltre che la procedura all-best selection Bayesiana, la cui versione sequenziale non è stata considerata in precedenza, ha prestazioni identiche alla procedura stepwise selection Bayesiana. L'onere computazionale maggiore della

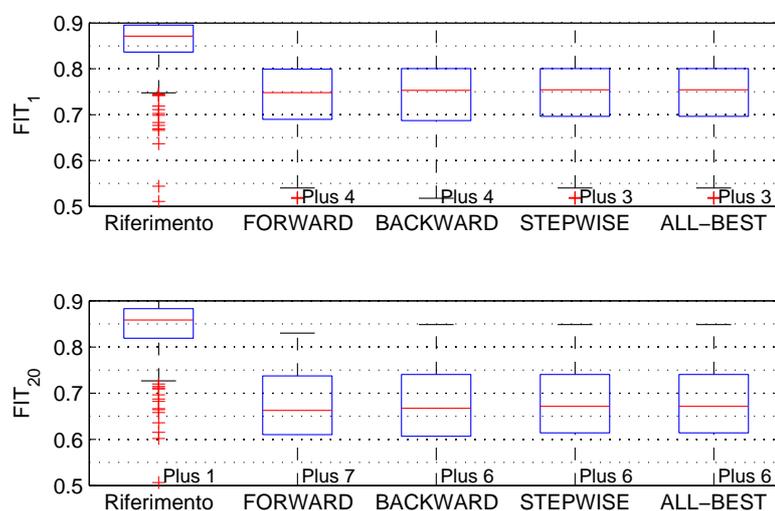


Figura 4.16 – Box-plot di FIT_1 e FIT_{20} di uno studio Monte Carlo eseguito per confrontare diverse procedure di selezione Bayesiane del metodo di identificazione SSEH sub-ottimo, avendo generato modelli di riferimento con livello di sparsità ridotto e ingressi correlati.

Tabella 4.14 – Percentuali degli ingressi e delle uscite considerate in maniera corretta da diverse procedure di selezione Bayesiane in uno studio Monte Carlo generando modelli di riferimento con livello di sparsità ridotto e ingressi correlati. La media delle lunghezze delle risposte impulsive stimate da PEM/ARX+AIC e impiegate da SSEH sub-ottimo risulta $\bar{t}_0 = 9.26$.

	FORWARD	BACKWARD	STEPWISE	ALL-BEST
% – IN	96.8	97.8	98.6	98.6
% – OUT	98	98	98	98

all-best selection non ha quindi condotto al vantaggio che ci si aspettava inizialmente.

Considerazioni simili a quelle appena fatte per gli studi Monte Carlo in cui sono stati usati modelli con livello di riferimento ridotto, sono valide anche per studi che considerando modelli di riferimento con livello di sparsità aumentato. Questo sia per quanto riguarda la differenza tra le procedure Bayesiane di forward selection e le procedure di backward elimination e stepwise selection, che nel caso sequenziale era molto più evidente, sia per quanto riguarda le prestazioni della procedura all-best selection Bayesiana, che malgrado l'onere computazionale maggiore ha le stesse prestazioni della stepwise selection Bayesiana.

I mediani dei FIT_1 e FIT_{20} delle simulazioni relative alle procedure Bayesiane e ai modelli di riferimento con livello di sparsità aumentato sono riportati in Tabella 4.15. In Figura 4.17 è illustrato il relativo box-plot e in Tabella 4.16 sono riportate le percentuali delle variabili considerate in maniera corretta dalle differenti procedure di selezione Bayesiane.

Tabella 4.15 – Mediani di FIT_1 e FIT_{20} di uno studio Monte Carlo generando modelli di riferimento con livello di sparsità aumentato e ingressi correlati, impiegando diverse procedure di selezione Bayesiane. La media delle lunghezze delle risposte impulsive stimate da PEM/ARX+AIC e impiegate da SSEH sub-ottimo risulta $\bar{t}_0 = 9.26$.

	RIFERIMENTO	FORWARD	BACKWARD	STEPWISE	ALL-BEST
FIT_1	0.983	0.935	0.935	0.935	0.935
FIT_{20}	0.974	0.88	0.881	0.881	0.881

Tabella 4.16 – Percentuali degli ingressi e delle uscite considerate in maniera corretta da diverse procedure di selezione Bayesiane in uno studio Monte Carlo generando modelli di riferimento con livello di sparsità aumentato e ingressi correlati. La media delle lunghezze delle risposte impulsive stimate da PEM/ARX+AIC e impiegate da SSEH sub-ottimo risulta $\bar{t}_0 = 9.26$.

	FORWARD	BACKWARD	STEPWISE	ALL-BEST
% – IN	99.1	99.9	99.9	99.9
% – OUT	70.5	70.5	70.5	70.5

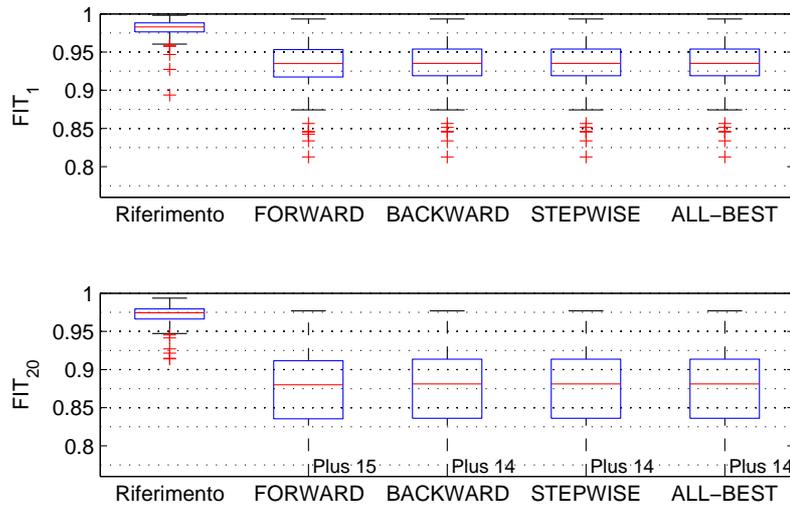


Figura 4.17 – Box-plot di FIT_1 e FIT_{20} di uno studio Monte Carlo eseguito per confrontare diverse procedure di selezione Bayesiane del metodo di identificazione SSEH sub-ottimo, avendo generato modelli di riferimento con livello di sparsità aumentato e ingressi correlati.

4.2.2 Procedure di selezione di singole risposte impulsive

Qui si riportano i risultati dei studi Monte Carlo eseguiti per confrontare diverse procedure di selezione delle risposte impulsive. Mentre le procedure di selezione di variabili possono annullare per ogni $i = 1, \dots, p$ le sole risposte impulsive del predittore $h^{[i,j]}$ per $j \in \{1, \dots, m+p\}$ relative alle variabili non selezionate, quelle di selezione delle risposte impulsive possono annullare singolarmente ogni $h^{[i,j]}$ per $i \in \{1, \dots, p\}$ e $j \in \{1, \dots, m+p\}$.

Non è stata fatta un'analisi così dettagliata come per le procedure di selezione delle variabili, perché le considerazioni fatte in quel caso si possono trasferire alle procedure di selezione delle risposte impulsive. Si considerano qui la procedura di selezione delle risposte impulsive all best selection e estensioni della procedura di selezione delle variabili forward selection Bayesiana, presentate in sezione 3.3. Le procedure sono state applicate solo a dati generati con livello di sparsità aumentato, perché nel caso di dati generati da modelli con livello di sparsità ridotto le procedure di selezione delle variabili hanno ovviamente prestazioni maggiori (dato che dispongono dell'informazione che le risposte impulsive relative a una certa variabile contano congiuntamente).

In Tabella 4.17 sono riportati i mediani di FIT_1 e di FIT_{20} per lo studio Monte Carlo, avendo generato modelli con livello di sparsità aumentato e ingressi correlati e avendo utilizzato le diverse procedure di selezione delle risposte impulsive presentate in sezione 3.3. In Figura 4.18 sono riportati i box-plot degli indici FIT relativi a tale studio Monte Carlo e in Tabella 4.18 le percentuali di risposte impulsive considerate in maniera corretta. Nelle tabelle e in figura sono riportati anche i relativi valori per il modello di riferimento sparso e per il modello identificato impiegando la procedura forward selection Bayesiana standard.

Tabella 4.17 – Mediani di FIT_1 e FIT_{20} di uno studio Monte Carlo generando modelli di riferimento con livello di sparsità aumentato e ingressi correlati, impiegando diverse procedure di selezione di risposte impulsive. La media delle lunghezze delle risposte impulsive stimate da PEM/ARX+AIC risulta $\bar{t}_0 = 9.23$.

	Rif	FOR	ALL-Best	FOR-AB	FOR-AB-In	FOR-AB-Out	FOR-bac
FIT_1	0.983	0.935	0.937	0.936	0.933	0.936	0.937
FIT_{20}	0.975	0.872	0.873	0.875	0.871	0.875	0.876

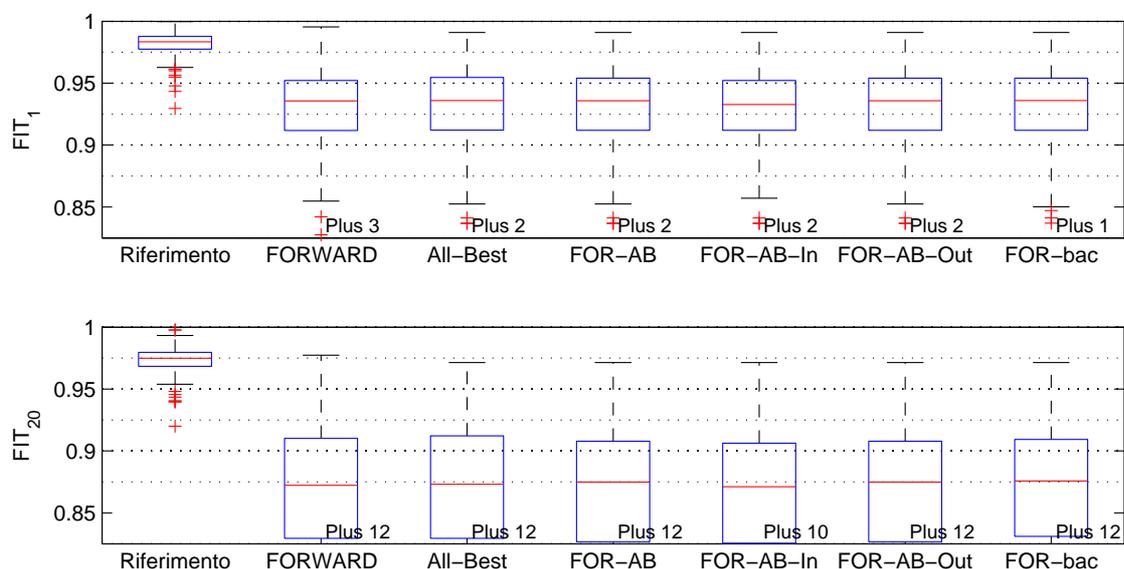


Figura 4.18 – Box-plot di FIT_1 e FIT_{20} di uno studio Monte Carlo eseguito per confrontare diverse procedure di selezione di risposte impulsive del metodo di identificazione SSEH sub-ottimo, avendo generato modelli di riferimento con livello di sparsità aumentato e ingressi correlati.

Tabella 4.18 – Percentuali delle risposte impulsive relative agli ingressi e alle uscite considerate in maniera corretta da diverse procedure di selezione di risposte impulsive in uno studio Monte Carlo generando modelli di riferimento con livello di sparsità aumentato e ingressi correlati. La media delle lunghezze delle risposte impulsive stimate da PEM/ARX+AIC risulta $\bar{t}_0 = 9.23$.

	FOR	ALL-Best	FOR-AB	FOR-AB-In	FOR-AB-Out	FOR-bac
% – IN	74.22	91.76	89.43	89.08	89.43	89.58
% – OUT	51.78	59.17	57.89	56.39	57.89	57.89

Purtroppo, come anche negli altri casi in cui sono stati considerati modelli con livello di sparsità aumentato, anche in questo caso le differenze tra le diverse procedure in termini di *FIT* sono piccole e non è quindi immediato confrontare le diverse procedure di selezione delle risposte impulsive. Le differenze tra le varie procedure di selezione si evidenziano in maniera più chiara in termini di percentuali di risposte impulsive considerate correttamente, riportate in Tabella 4.18. Si vede comunque anche da Tabella 4.17 e da Figura 4.18 che le prestazioni migliorano estendendo la procedura forward selection Bayesiana. Tra queste estensioni la procedura FORWARD-backward ha dato i risultati migliori ed è tra l'altro la versione computazionalmente più efficiente tra tutte le procedure di selezione delle variabili considerate. Anche la procedura all-best selection, che non impiega come condizione iniziale le variabili scelte dalla forward selection Bayesiana, ma parte da un'insieme vuoto, ha dato delle prestazioni buone, come si evidenzia soprattutto dalla percentuale di risposte impulsive considerate in maniera corretta.

4.3 Confronto tra diversi versioni di SSEH e PEM

Si presentano in questa sezione i risultati di studi Monte Carlo eseguiti per confrontare le diverse versioni di identificazione di SSEH e di PEM citati in sezione 3.3: di SSEH sono state considerate la versione sub-ottima, la versione ottima e la versione ottima con preselezione. Del metodo PEM invece sono state considerate due versioni di Oracolo (PEM+Oracolo+set conosce le variabili che danno contributo all'evoluzione del sistema, PEM+Oracolo invece no), due versioni che impiegano l'indice BIC (PEM+BIC+preselezione usa solo le variabili selezionate da SSEH sub-ottimo, PEM+BIC le usa tutte), e la versione che impiega l'indice

AICc per selezionare l'ordine del modello.

Negli studi Monte Carlo come lunghezza del set di dati per l'identificazione si è usato $N = 300$, perché in sezione 4.1 si è visto che tutti i metodi di identificazione per tale lunghezza danno risultati accettabili e perché l'onere computazionale è molto ridotto rispetto a $N = 500$ o a $N = 800$. Come lunghezza delle risposte impulsive per SSEH è stata usata la lunghezza stimata da PEM/ARX+AIC, che in precedenza ha dato buone prestazioni. Gli ingressi nelle simulazioni sono stati generati solo in modo da essere correlati, dato che tale situazione è più vicina alle condizioni reali e che riportare le simulazioni per entrambi i tipi di ingressi sarebbe superfluo.

Nelle simulazioni presentate qui, come procedura di selezione delle diverse versioni di SSEH è stata impegnata ancora la procedura standard di selezione delle variabili forward selection Bayesiana. Questa scelta è stata fatta per essere coerente con gli articoli [20–22] anche se nella precedente sezione si è visto che ha dato prestazioni leggermente inferiori alle altre procedure Bayesiane considerate. Soprattutto in termini di *FIT* la differenza è però trascurabile.

Modelli di riferimento con livello di sparsità ridotto In Tabella 4.19 sono riportati i mediani di FIT_1 e di FIT_{20} dello studio Monte Carlo eseguito per confrontare le prestazioni delle diverse versioni SSEH e delle diverse versioni PEM, dove sono stati impegnati modelli di riferimento con livello di sparsità ridotto. In Figura 4.19 sono riportati i box-plot relativi a questo studio Monte Carlo, dove per questione di leggibilità, sono illustrati i risultati relativi alle identificazioni SSEH nel grafico sopra e quegli relativi alle identificazioni PEM nel grafico sotto. Per avere un miglior confronto, nel grafico relativo al metodo SSEH sono riportati anche i box-plot relativi alla versione PEM+Oracolo+set e ai modelli di riferimento, e nel grafico relativo al metodo PEM è riportato il box-plot relativo ai modelli di riferimento. In Tabella 4.20 sono riportati invece le percentuali di variabili (ingressi e uscite rispettivamente) considerati in maniera corretta dalla procedura di selezione forward selection, relativi alle diverse versioni di SSEH. Per le diverse versioni di SSEH la media della lunghezza delle risposte impulsive stimata da PEM/ARX+AIC risulta $\bar{t}_0 = 6.05$.

Per quanto riguarda il confronto tra i metodi di identificazione PEM e SSEH, da Tabella 4.19 e da Figura 4.19 si possono trarre conclusioni simili a quanto visto in sezione 4.1: solo nel caso in cui il metodo PEM ha accesso agli dati di test per

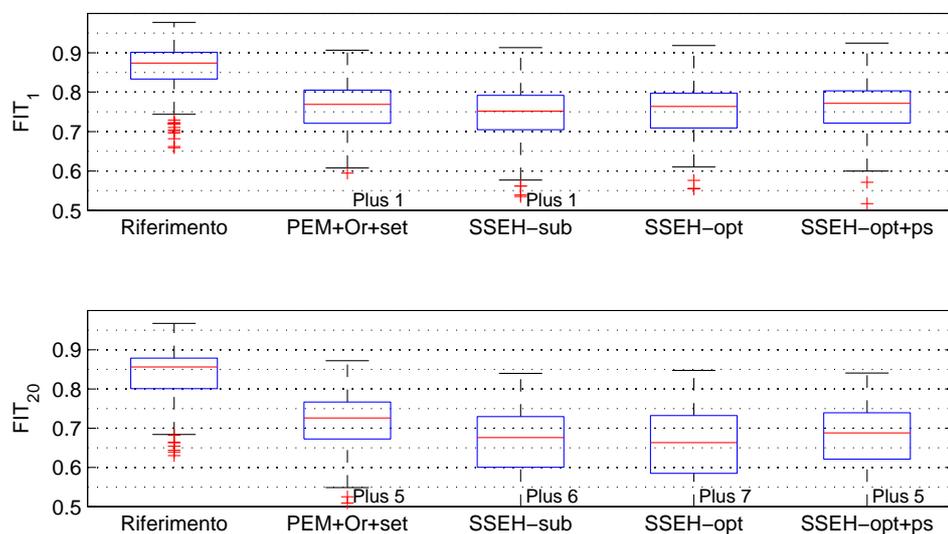
Tabella 4.19 – Mediani di FIT_1 e FIT_{20} di uno studio Monte Carlo generando modelli di riferimento con livello di sparsità ridotto e ingressi correlati, impiegando diverse versioni dei metodi di identificazione SSEH (sopra) e PEM (sotto). La media delle lunghezze delle risposte impulsive stimate da PEM/ARX+AIC risulta $\bar{t}_0 = 6.05$.

	Riferimento	SSEH-sub	SSEH-opt	SSEH-opt+ps
FIT_1	0.874	0.752	0.764	0.772
FIT_{20}	0.856	0.676	0.663	0.688

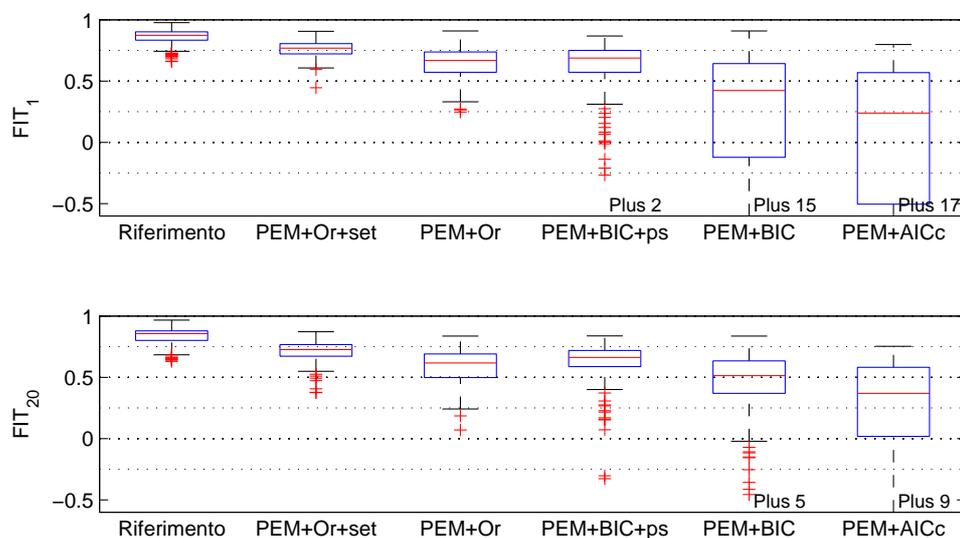
	PEM+Or+set	PEM+Or	PEM+BIC+ps	PEM+BIC	PEM+AICc
FIT_1	0.769	0.668	0.688	0.423	0.238
FIT_{20}	0.726	0.618	0.663	0.513	0.369

Tabella 4.20 – Percentuali degli ingressi e delle uscite rispettivamente considerate in maniera corretta da diverse versioni del metodo SSEH in uno studio Monte Carlo generando modelli di riferimento con livello di sparsità aumentato e ingressi correlati. La media delle lunghezze delle risposte impulsive stimate da PEM/ARX+AIC risulta $\bar{t}_0 = 6.05$.

	SSEH-sub	SSEH-opt	SSEH-opt+ps
% – IN	97.2	91.1	98.55
% – OUT	97.67	95.83	95.1



(a) Identificazione SSEH



(b) Identificazione PEM

Figura 4.19 – Box-plot di FIT_1 e FIT_{20} di uno studio Monte Carlo eseguito per confrontare impiegando diverse versioni dei metodi di identificazione SSEH (sopra) e PEM (sotto), avendo generato modelli di riferimento con livello di sparsità ridotto e ingressi correlati.

selezionare l'ordine adeguato del modello e conosce il set degli ingressi che hanno effettivamente influenza alla predizione del sistema (versione PEM+Oracolo+set), PEM ha prestazioni leggermente migliori del metodo SSEH. Per tutte le altre versioni PEM, le versioni SSEH hanno prestazioni nettamente superiori.

Confrontando le diverse versioni PEM si vede che PEM+AICc e PEM+BIC ha dato le prestazioni più scarse. Questo risultato era atteso, siccome le due versioni non dispongono dell'informazione di quali variabili contribuiscono effettivamente alla predizione del sistema e devono selezionare l'ordine del modello dai dati. Equipaggiando il metodo PEM+BIC con l'informazione delle variabili selezionate dalla versione SSEH sub-ottima (PEM+BIC+preselezione) le prestazioni migliorano nettamente e diventano confrontabili con la versione PEM+Oracolo. Infatti, come si vede bene da Figura 4.19, i mediani dei *FIT* di PEM+BIC+preselezione sono leggermente superiori a quelli di PEM+Oracolo, ma presentano un numero maggiore di outlier. La versione PEM+Oracolo+set presenta ovviamente le prestazioni migliori di tutte le versioni PEM.

Per quanto riguarda il confronto tra le diverse versioni di SSEH si può dire che le prestazioni sono buone per ogni versione e quindi molto simili. Partendo dalla versione SSEH sub-ottima, come previsto, migliorano le prestazioni in termini di *FIT* considerando il modello Bayesiano intero (versione SSEH ottimo). Le prestazioni migliorano ulteriormente impiegando solo le variabili selezionate dalla versione SSEH sub-ottima (versione SSEH ottimo con preselezione), dato che non deve più considerare le variabili scartate prima.

La procedura SSEH ottima ha prestazioni inferiori in termini di percentuale di variabili selezionate correttamente rispetto alla versione SSEH sub-ottima. Questo si verifica perché, anche se i fattori di scala λ_{ij} inferiori a e^{-10} nelle versioni SSEH ottime sono eliminate a posteriori, può capitare che la procedura di ottimizzazione riesca a fissare qualche λ_{ij} , per i quali la marginale decresce, e che quindi le relative variabili non sono scartate dalla procedura di selezione. È da aspettarsi che all'aumentare della lunghezza del set di dati N tale situazione si verifichi raramente. Questo non è però stato controllato per l'onere computazionale notevole delle versioni SSEH ottime. Si noti che per la versione SSEH ottima con preselezione tale difetto si somma alle variabili non selezionate nella fase di inizializzazione. La percentuale di ingressi selezionati in maniera corretta di tale procedura presenta invece i risultati migliori, anche rispetto alla procedura sub-ottima. Questo perché nella fase di inizializzazione la procedura di forward selection Bayesiana tende a selezionare troppe variabili, cioè seleziona anche variabili che in realtà sono superflue, le quali dopo l'ottimizzazione di SSEH ottima

con preselezione possono essere a volte determinate e quindi scartate dalla procedura di forward selection Bayesiana. Inoltre non si considerano più le variabili già scartate correttamente nella fase di inizializzazione.

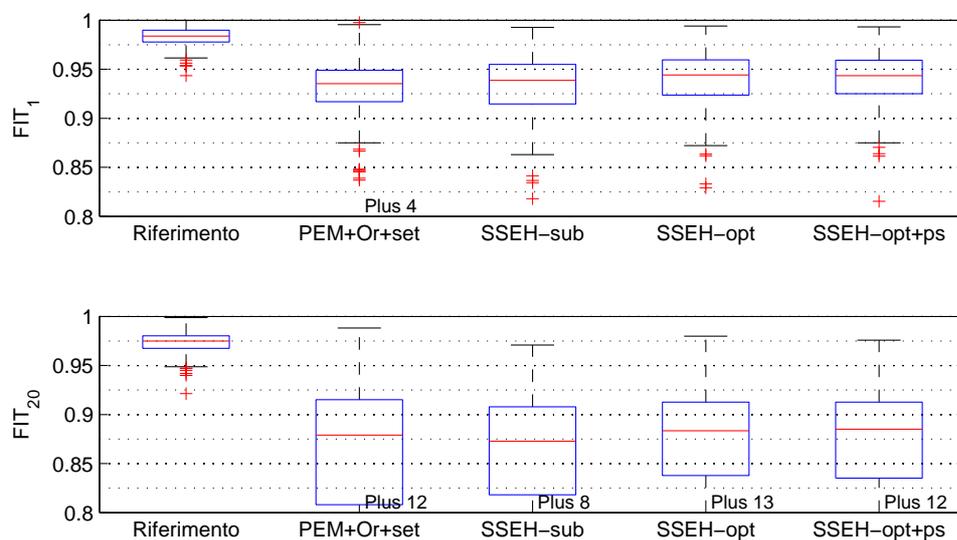
Modelli di riferimento con livello di sparsità aumentato I mediani di FIT_1 e di FIT_{20} dello studio Monte Carlo eseguito per confrontare le prestazioni delle diverse versioni SSEH e delle diverse versioni PEM, dove sono stati impegnati modelli di riferimento con livello di sparsità aumentato, sono riportati in Tabella 4.19. In Figura 4.19 sono riportati i relativi box-plot degli indici FIT e in Tabella 4.20 sono riportate invece le percentuali di ingressi e uscite rispettivamente considerati in maniera corretta dalla procedura di selezione forward selection. Per le diverse versioni di SSEH la media della lunghezza delle risposte impulsive stimata da PEM/ARX+AIC risulta $\bar{t}_0 = 9.5$.

Tabella 4.21 – Mediani di FIT_1 e FIT_{20} di uno studio Monte Carlo generando modelli di riferimento con livello di sparsità ridotto e ingressi correlati, impiegando diverse versioni dei metodi di identificazione SSEH (sopra) e PEM (sotto). La media delle lunghezze delle risposte impulsive stimate da PEM/ARX+AIC risulta $\bar{t}_0 = 9.5$.

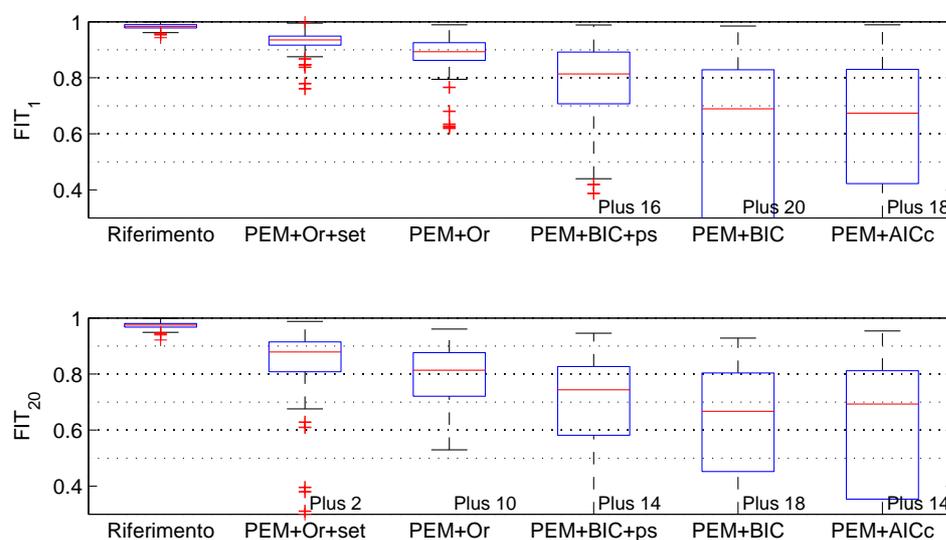
	Riferimento	SSEH-sub	SSEH-opt	SSEH-opt+ps
FIT_1	0.984	0.939	0.944	0.943
FIT_{20}	0.975	0.873	0.883	0.885

	PEM+Or+set	PEM+Or	PEM+BIC+ps	PEM+BIC	PEM+AICc
FIT_1	0.935	0.893	0.814	0.689	0.674
FIT_{20}	0.879	0.814	0.744	0.667	0.693

Quanto succedeva per modelli di riferimento con livello di sparsità ridotto, si verifica anche in questo caso. Come si vede da Tabella 4.21 e Figura 4.20, le diverse versioni del metodo SSEH hanno prestazioni confrontabili alla versione PEM+Oracolo+set mentre le altre versioni PEM hanno prestazioni inferiori. Per questo tipo di modelli di riferimento, PEM+AICc ha dato prestazioni confrontabili a PEM+BIC. La differenza tra le prestazioni di queste versioni e le altre versioni del metodo PEM considerate è inferiore di quanto si è verificato per i modelli di riferimento con livello di sparsità ridotto. In questo caso la versione



(a) Identificazione SSEH



(b) Identificazione PEM

Figura 4.20 – Box-plot di FIT_1 e FIT_{20} di uno studio Monte Carlo eseguito per confrontare impiegando diverse versioni dei metodi di identificazione SSEH (sopra) e PEM (sotto), avendo generato modelli di riferimento con livello di sparsità ridotto e ingressi correlati.

Tabella 4.22 – Percentuali degli ingressi e delle uscite rispettivamente considerate in maniera corretta da diverse versioni del metodo SSEH in uno studio Monte Carlo generando modelli di riferimento con livello di sparsità aumentato e ingressi correlati. La media delle lunghezze delle risposte impulsive stimate da PEM/ARX+AIC risulta $\bar{t}_0 = 9.5$.

	SSEH-sub	SSEH-opt	SSEH-opt+ps
% – IN	99	89.45	99.6
% – OUT	72.83	73	73

PEM+BIC+preselezione presenta mediani dei *FIT* inferiori a quelli della versione PEM+Oracolo. La versione PEM+Oracolo+preselezione ha dato ovviamente le prestazioni migliore tra i diversi metodi PEM.

Per quanto riguarda le versioni SSEH, si può dire ancora che le prestazioni in termini di *FIT* migliorano considerando il modello Bayesiano intero invece che quello ridotto, e migliorano ulteriormente ottimizzando solo le λ_{ij} relative alle variabili selezionate in una fase di inizializzazione (effettuata da parte della versione SSEH sub-ottima). Valgono anche le considerazioni fatte in precedenza relativamente alle percentuali di ingressi considerati in maniera corretta come si vede da Tabella 4.22: mentre la percentuale degli ingressi considerati correttamente di SSEH ottimo peggiora rispetto a quella di SSEH sub-ottimo per una decrescita errata della marginale, quella di SSEH ottima con preselezione migliora perché le variabili già scartati correttamente nella fase di inizializzazione non sono più considerate. Considerando invece le percentuali di uscite considerate in maniera corretta non si riesce a trarre delle informazioni (come tra l'altro è successo anche in altre situazioni in cui sono stati considerati i modelli con livello di sparsità aumentato), perché sono state scelte quasi sempre tutte le uscite, anche se in qualche caso qualcuna non influenzava l'evoluzione del sistema.

Capitolo 5

Risultati dell'identificazione di modelli termodinamici

Per testare il metodo non-parametrico SSEH e confrontarlo con il metodo classico PEM nell'identificazione di modelli termodinamici, questi sono stati applicati su dati climatici ottenuti dal monitoraggio di un edificio esistente (descritto in sezione 3.2). Se non diversamente specificato, come uscite del modello sono stati definiti tutti gli andamenti dei sensori di temperatura sulle schede multisensore, come ingressi invece quelli dei sensori di umidità e di irraggiamento e di tutti i sensori della stazione meteo. In questo capitolo si presentano i risultati delle identificazioni.

In questo capitolo, per questioni di complessità computazionale, del metodo SSEH è stata usata solo la versione sub-ottima. Come configurazione di riferimento nelle simulazioni si sono definiti una lunghezza del set di dati $N = 500$ e un tempo di campionamento $T = 6$ minuti. Il metodo PEM/ARX+AICc, usato per stimare la lunghezza delle risposte impulsive t_0 di SSEH, a differenza di quanto visto nel capitolo precedente, non ha prodotto risultati ragionevoli quando applicato a dati reali, probabilmente a causa di N troppo piccolo. Per questo di seguito sono stati riportati solo i risultati relativi a 5 passi e a 30 passi, che equivalgono a $t_0 = 5 \cdot 6 = 30$ minuti e a $t_0 = 30 \cdot 6 = 180$ minuti.

Nella prima parte del capitolo si presentano i risultati delle simulazioni effettuate per confrontare i metodi SSEH e PEM. In essa, per SSEH, è stata usata solo la procedura di selezione delle variabili forward selection Bayesiana, che scarta quindi i sensori che non contengono informazioni utile per la predizione delle variabili definite come uscite. Del metodo PEM sono state considerate le versioni PEM+Oracolo+preselezione, PEM+Oracolo, PEM+BIC+preselezione e PEM+BIC. Per intuire l'influenza della lunghezza del set di dati N e del tempo

di campionamento T è stato usato anche $N = 250$ con $T = 6$ e $T = 12$ minuti. Inoltre per singole uscite è stata effettuata anche una selezione “manuale” dei sensori a cui sono stati applicati i metodi SSEH e PEM.

Nella seconda parte del capitolo sono riportati i risultati delle simulazioni, nelle quali sono state usate procedure di selezione delle singole risposte impulsive del metodo SSEH. In questo modo, oltre a scartare i sensori che non contengono alcuna informazione utile per la predizione di nessuna uscita, per ogni uscita sono selezionati solo i sensori utili a predirla. Si evidenzia quindi la struttura del sistema, chiamata causalità di Granger, simile a quella riportata in Figura 2.5. Delle procedure di selezione delle risposte impulsive considerate in sezione 4.2, qui sono usate solo le procedure FORWARD-backward e FORWARD-AllBest-Output, che rispetto alle altre sono computazionalmente più efficienti e hanno dato buoni risultati.

5.1 Confronto tra SSEH e PEM

In questa sezione si riportano i risultati delle diverse simulazioni effettuate per confrontare i metodi PEM e SSEH per l'identificazione di modelli termodinamici. Come condizione di riferimento sono stati scelti una lunghezza del set di dati $N = 500$ e un tempo di campionamento di $T = 6$ minuti. È stata considerata anche la lunghezza ridotta $N = 250$ con tempi di campionamento di $T = 6$ e di $T = 12$ minuti. Alla fine si presentano i risultati ottenuti considerando per l'identificazione solo un sottoinsieme di sensori.

5.1.1 Configurazione di riferimento: $N = 500$ e $T = 6$ minuti

Si riportano qui i risultati delle simulazioni effettuate per confrontare i metodi SSEH e PEM usando i dati monitorati in un edificio esistente. Del metodo SSEH è stata usata la versione sub-ottima con la procedura di selezione delle variabili forward selection Bayesiana e del metodo PEM le versioni PEM+Oracolo+preselezione, PEM+Oracolo, PEM+BIC+preselezione e PEM+BIC. Come configurazione di riferimento nelle simulazioni è stata usata la lunghezza del set di dati $N = 500$ con un tempo di campionamento di $T = 6$ minuti. Per l'identificazione SSEH sub-ottima sono state considerate le lunghezze delle rispo-

ste impulsive $t_0 = 30$ minuti e $t_0 = 180$ minuti.

I risultati delle simulazioni per tutti i sensori di temperatura definiti come uscite con $N = 500$ e con $T = 6$ minuti in termini di FIT_1 (6 minuti in avanti) e di FIT_{20} (120 minuti in avanti) sono riportati in Figura 5.1.

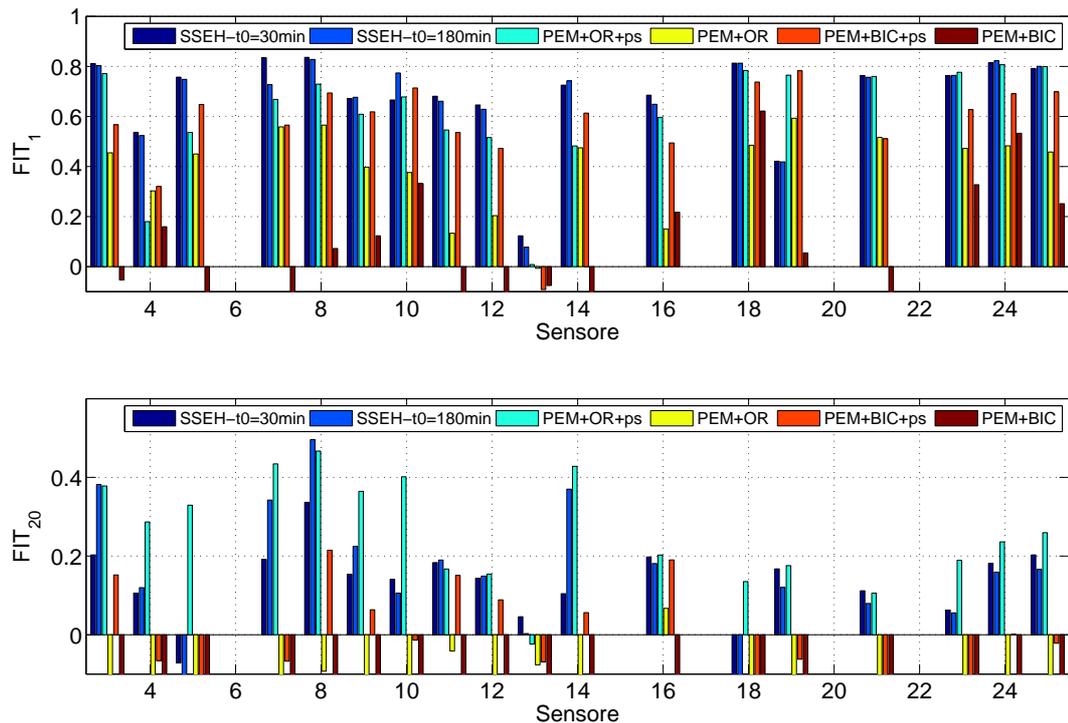


Figura 5.1 – FIT_1 (6 minuti in avanti; sopra) e FIT_{20} (120 minuti in avanti; sotto) per i diversi metodi di identificazione, con $N = 500$ e $T = 6$ minuti e avendo impiegato tutti i sensori di temperatura delle schede multisensore come uscite. L'asse delle ascisse è relativa ai nomi delle schede multisensore distribuite nell'edificio e indicati in Figura 5.4.

Da Figura 5.1 si vede che la versione non-parametrica SSEH sub-ottima ha dato delle prestazioni molto buone rispetto al metodo PEM. Soprattutto a 6 minuti di predizione gli indici FIT ottenuti da SSEH sub-ottimo sono in generale superiori a tutti quelli ottenuti dalle diverse versioni PEM. Anche a 120 minuti le versioni PEM hanno fornito prestazioni inferiori, ad eccezione delle versioni PEM+Oracolo+preselezione che ha ottenuto prestazioni superiori.

In generale le versioni PEM che usano l'informazione della selezione delle variabili

effettuata dalla procedura di selezione di SSEH sub-ottimo hanno dato prestazioni nettamente superiori alle versioni PEM, che usano tutti i sensori per predire l'evoluzione del sistema. Si nota poi che le versioni PEM con Oracolo hanno dato prestazioni superiori a quelle che effettuano la selezione impiegando l'indice BIC. Per quanto riguarda la lunghezza delle risposte impulsive t_0 di SSEH, da Figura 5.1 si vede che in generale, a un passo di predizione, la lunghezza maggiore ($t_0 = 120$ minuti) assicura una prestazione migliore, mentre a venti passi di predizione la prestazione relativa alla lunghezza $t_0 = 30$ minuti è superiore. In ogni caso i risultati corrispondenti ai due t_0 erano confrontabili.

Considerazioni simili si possono trarre dalla Figura 5.2, nella quale sono stati riportati gli andamenti dell'indice FIT_k relativi a due sensori di uscita in funzione dei passi di predizione k per i diversi metodi di identificazione. È stato scelto di riportare gli indici FIT_k relativi ai sensori $S10$ e $S25$, posti al centro dei rispettivi locali (si guardi Figura 5.4 oppure Figura 3.3) e non troppo vicini a ventilconvettori, termosifoni o finestre.

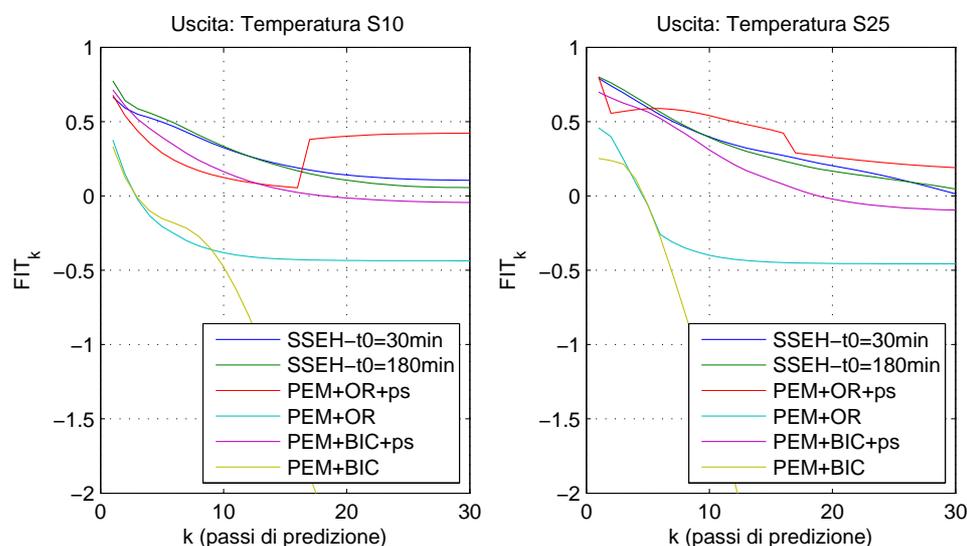


Figura 5.2 – Andamento dell'indice FIT in funzione dei passi di predizione k per ciascuno dei metodi di identificazione testati, con $N = 500$ e $T = 6$ minuti e usando come uscite i sensori di temperatura delle schede multisensore $S10$ (sinistra) e $S25$ (destra).

Come si vede anche da Figura 5.2 la versione SSEH ha dato prestazioni buone in confronto alle versioni PEM. Solo la versione PEM+Oracolo+preselezione ha fornito prestazioni che in determinati casi sono rivelanti. Per le due diverse

lunghezze di t_0 , le prestazioni di SSEH sono confrontabili. Si noti che in generale l'indice FIT_k decresce all'aumentare dei passi di predizione, ma negli andamenti ottenuti dalle versioni PEM con Oracolo si presentano delle irregolarità. Questo è dovuto al fatto che per ogni passo di predizione le versioni Oracolo selezionano l'ordine del modello più adatto, che può quindi variare. Analizzando i dati ottenuti dalle simulazioni si è visto che le irregolarità negli andamenti dei FIT_k si presentano al cambiamento di ordine del modello selezionato dall'Oracolo.

Considerando la singola uscita, le prestazioni dei modelli ottenuti dalle versioni PEM con Oracolo sono di norma inferiori di quelle ottenute dalle relative versioni PEM che impiegano l'indice BIC per effettuare la selezione. Questo è dovuto al fatto che l'Oracolo seleziona l'ordine migliore per il modello complessivo, e può quindi capitare che, per singole uscite, l'ordine scelto dia prestazioni inferiori a quelle ottenibili impiegando un qualche altro ordine.

In Figura 5.3 sono confrontati gli andamenti reali delle temperature dei sensori $S10$ (sinistra) e $S25$ (destra) con le predizioni a 6 e a 120 minuti impiegando il modello identificato con SSEH sub-ottimo e $t_0 = 180$ minuti. Questo confronto fornisce un'idea generale del legame tra FIT e bontà dell'approssimazione della predizione all'andamento reale.

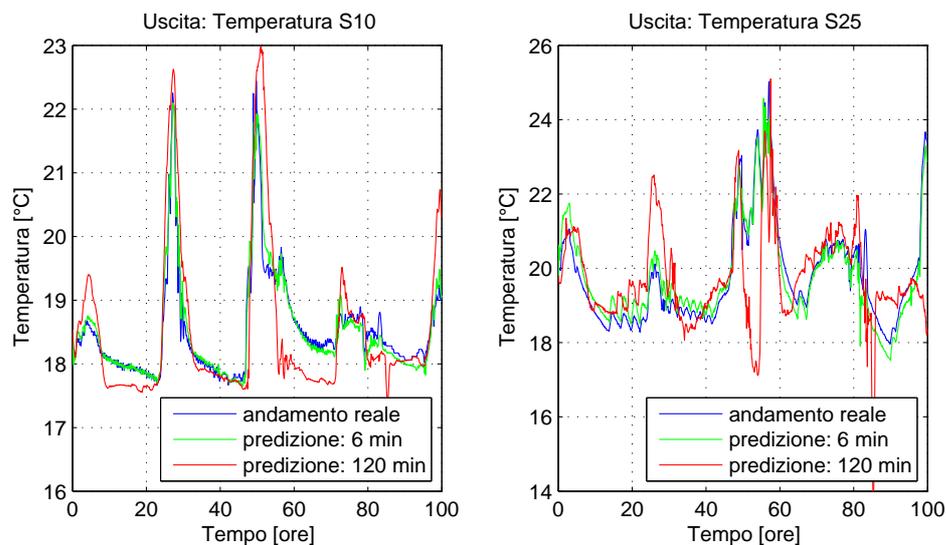


Figura 5.3 – Confronto tra gli andamenti reali di temperatura e le predizioni a un passo e a venti passi per i sensori di temperatura sulle schede multisensore $S10$ (sinistra) e $S25$ (destra) impiegando i modelli identificati con SSEH sub-ottimo con $t_0 = 130$ minuti e con $N = 500$ e $T = 6$ minuti.

L'approssimazione della predizione a 6 minuti all'andamento reale è molto buona, come si vede da Figura 5.3. In tal caso il FIT_1 per i due sensori risulta tra 0.7 e 0.8. All'aumentare dei passi di predizione il FIT decresce raggiungendo valori tra 0.1 e 0.2 per 120 minuti, e di conseguenza l'approssimazione peggiora. Comunque, anche per 120 minuti, l'errore tra predizione e andamento reale è in generale inferiore a $2^\circ C$.

Si noti che, per eseguire le identificazioni, tutti gli andamenti sono stati normalizzati in modo tali da avere media nulla e varianza unitaria. Per ottenere le predizioni riportate in Figura 5.3, scalate in maniera corretta, per prima cosa il modello deve essere denormalizzato utilizzando la varianza dei dati del set di training e successivamente, le uscite dello stesso devono essere sommate alla media dei dati di test, ottenuta in una fase di inizializzazione.

È stato scritto un piccolo codice per rappresentare graficamente, per una determinata uscita, i sensori selezionati da una qualunque delle procedure di selezione del metodo SSEH. Nella rappresentazione sono evidenziati i sensori selezionati per predire almeno una delle uscite e i sensori scartati completamente dalla procedura di selezione. Inoltre sono rappresentati diversamente i sensori che nel modello raffigurano le uscite dai sensori che raffigurano gli ingressi. In Figura 5.4 è riportato il grafico generato dal codice per l'uscita $S10$. Si noti che in questo caso la selezione ottenuta per le altre uscite è identica, siccome è stata impiegata la procedura di selezione delle variabili forward selection Bayesiana, che scarta solo i sensori superflui per la predizione dell'intero sistema e non effettuando la selezione riguardante le singole uscite.

Da Figura 5.4 si vede che per predire il sistema, dalla procedura di selezione forward selection Bayesiana di SSEH sono state selezionate quasi tutte le uscite, ovvero quasi tutti i sensori di temperatura delle schede multisensore. Sono stati selezionati alcuni sensori di irraggiamento e un solo sensore di umidità relativa, mentre nessuna sensore della stazione meteo è stato ritenuto utile per la predizione del sistema.

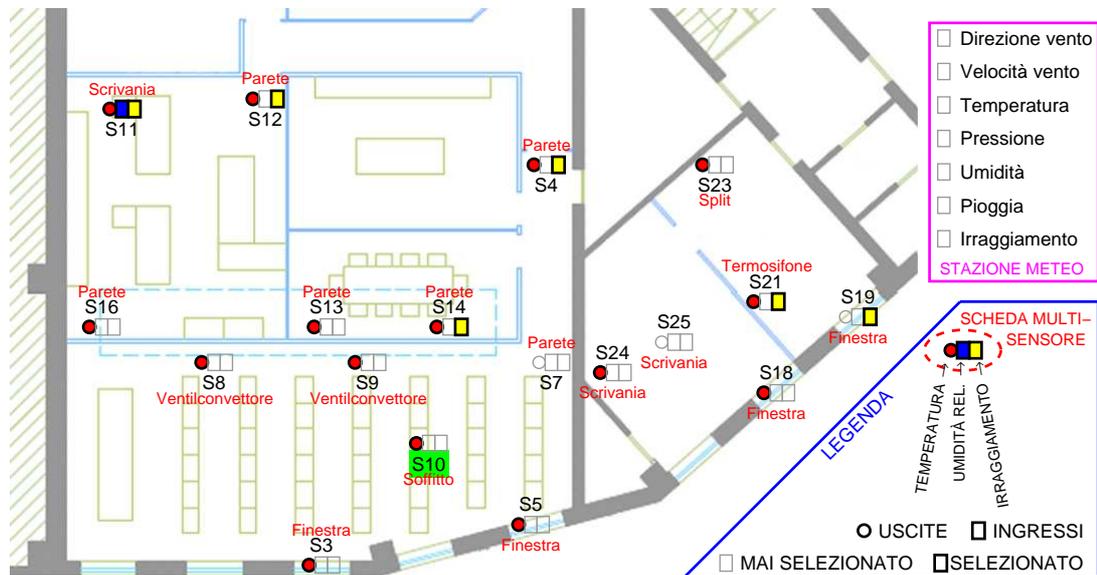


Figura 5.4 – Selezione dei sensori effettuata dalla procedura di selezione delle variabili forward selection Bayesiana di SSEH sub-ottimo per il sensore S10. Tutti i sensori rappresentati in Figura come cerchi sono uscite del modello, quelli rappresentati con un rettangolo sono invece gli ingressi. Il sensore, evidenziato in verde, è l'uscita alla quale si riferisce la Figura. I sensori colorati sono quelli selezionati per la predizione di tale uscita, quelli con la cornice nera e spessa sono i sensori scelti per predire almeno un'uscita e quelli con la cornice grigia e sottile sono invece i sensori completamente scartati dalla procedura di selezione.

5.1.2 Influenza della lunghezza del set di dati e del tempo di campionamento

Si presentano di seguito i risultati ottenuti impiegando due set di dati con una lunghezza di $N = 250$ e con tempi di campionamento rispettivamente di $T = 6$ e di $T = 12$ minuti.

Lunghezza ridotta del set di dati Per analizzare l'influenza della lunghezza del set di dati, i diversi metodi di identificazione sono stati applicati a dati con $N = 250$, mantenendo il tempo di campionamento a $T = 6$ minuti. Come uscite sono state considerate ancora tutti i sensori di temperatura delle schede multi-sensore e sono state impiegate le stesse versioni dei metodi di identificazione di prima. In Figura 5.5 sono riportati i risultati in termini di FIT_1 (6 minuti in avanti) e di FIT_{20} (120 minuti in avanti) delle simulazioni per tutti i sensori d'uscita.

Si sono tralasciate le corrispettive immagini delle Figure 5.2 e 5.3 poiché non vi sono contenute informazioni utili al confronto dell'influenza della lunghezza del set di dati utilizzati. Ugualmente non si riporta la selezione dei sensori poiché la procedura forward selection Bayesiana per $N = 250$ seleziona un set di sensori del tutto simile al caso $N = 500$.

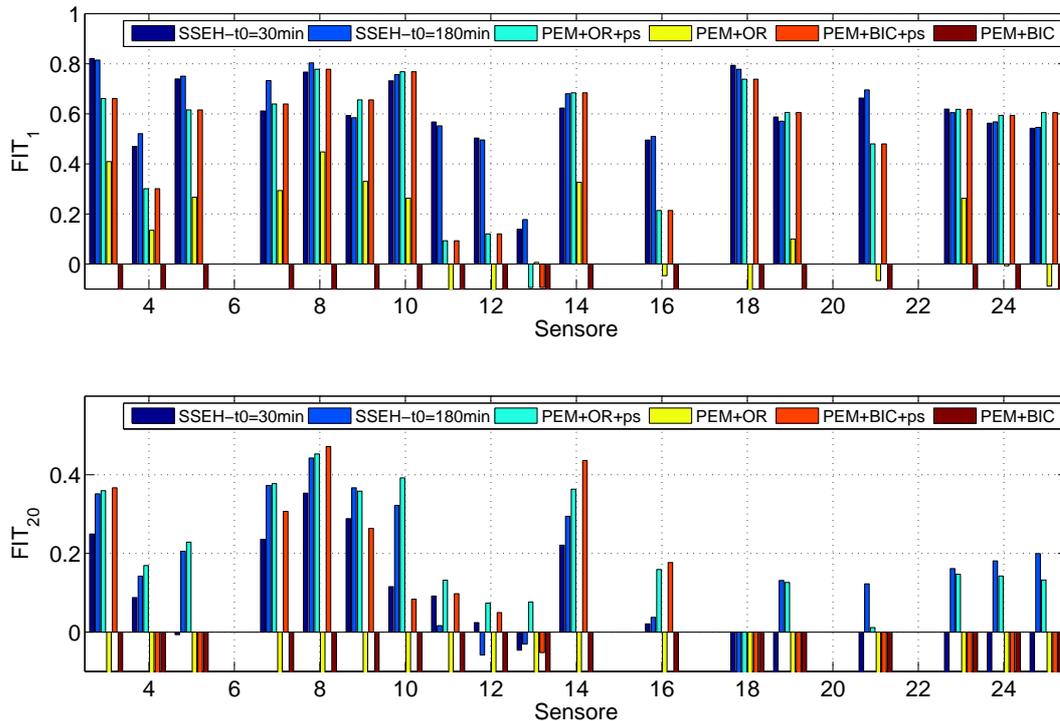


Figura 5.5 – FIT_1 (6 minuti in avanti; sopra) e FIT_{20} (120 minuti in avanti; sotto) per i diversi metodi di identificazione, con $N = 250$ e $T = 6$ minuti avendo impiegato tutti i sensori di temperatura delle schede multisensore come uscite. L'asse delle ascisse è relativa ai nomi delle schede multisensore distribuite nell'edificio e indicati in Figura 5.4.

Anche per il set di dati di lunghezza $N = 250$ vale che, in generale, aumentando i passi di predizione peggiorano le prestazioni in termini di FIT_k , come si vede da Figura 5.5. Il metodo SSEH fornisce prestazioni buone, confrontabili con la versione PEM+Oracolo+preselezione. A un passo di predizione, le versioni PEM+BIC+preselezione e PEM+Oracolo+preselezione hanno raggiunto le stesse prestazioni. Analizzando i dati si è notato che le due versioni di PEM hanno scelto lo stesso ordine del modello. A più passi di predizione invece,

la versione PEM+BIC+preselezione ha dato prestazioni in generale inferiori a PEM+Oracolo+preselezione.

Mentre le prestazioni del metodo SSEH e delle due versioni PEM che effettuano una preselezione sono solo leggermente inferiori a quanto ottenuto per $N = 500$, le prestazioni ottenute dalle versioni PEM, che impiegano tutti i dati per identificare il modello, peggiorano considerevolmente. Questo comportamento è evidente dal confronto di Figura 5.1 con Figura 5.5.

Si noti che, mantenendo il tempo di campionamento $T = 6$ minuti, anche per $N = 250$ sono state ottenute prestazioni confrontabili per le due lunghezze delle risposte impulsive t_0 di SSEH sub-ottimo considerate. Infatti, come si vede da Figura 5.5, per le diverse uscite, alcuni FIT_k sono migliori impiegando $t_0 = 30$ minuti, altri sono migliori impiegando $t_0 = 18$ minuti.

Tempo di campionamento T aumentato Per analizzare l'influenza del tempo di campionamento sull'identificazione di modelli termodinamici del set di dati di lunghezza $N = 500$ è stata eliminata ogni seconda misura, ottenendo così un set di dati di lunghezza $N = 250$ sullo stesso intervallo temporale, ma con tempo di campionamento $T = 12$ minuti. I risultati in termini di FIT_1 (12 minuti in avanti) e di FIT_{10} (120 minuti in avanti) per i diversi sensori di uscita sono rappresentati in Figura 5.6.

L'aumento del tempo di campionamento da $T = 6$ minuti a $T = 12$ minuti ha messo in difficoltà un po' tutte le versioni di identificazione considerate, come si vede da Figura 5.6. A un passo di predizione (12 minuti) la versione SSEH sub-ottima con $t_0 = 60$ minuti ($t_0 = 5$ passi) ha dato prestazioni buone, ma aumentando i passi ha dato prestazioni in generale inferiori al caso di $N = 250$ e $T = 6$ minuti (si guardi Figura 5.5). SSEH sub-ottimo con $t_0 = 360$ minuti ($t_0 = 30$ passi) ha fornito prestazioni pessime, così come anche le versioni di PEM, che impiegano l'indice BIC per selezionare l'ordine del modello. La versione PEM+Oracolo, che impiega tutti i dati per identificare il modello, ha dato prestazioni buone in confronto alle altre versioni considerate, soprattutto per il FIT_{10} .

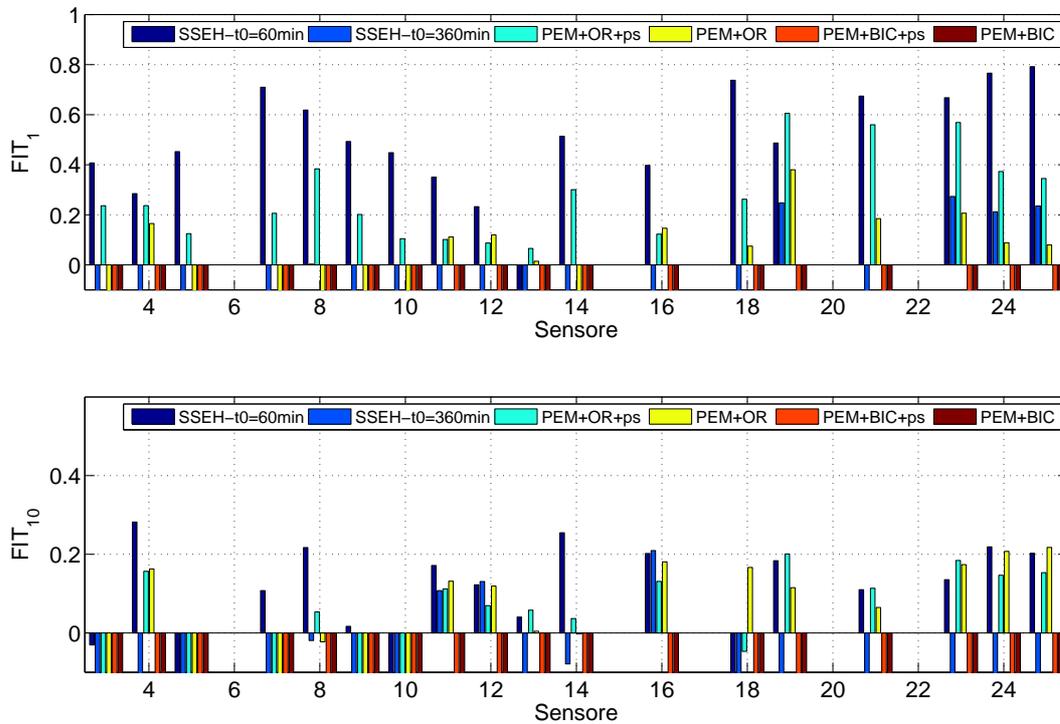


Figura 5.6 – FIT_1 (12 minuti in avanti; sopra) e FIT_{10} (120 minuti in avanti; sotto) per i diversi metodi di identificazione, con $N = 250$ e $T = 12$ minuti avendo impiegato tutti i sensori di temperatura delle schede multisensore come uscite. L'asse delle ascisse è relativa ai nomi delle schede multisensore distribuite nell'edificio e indicati in Figura 5.4.

5.1.3 Selezione manuale di sensori

Per non tutti i sensori, sono state fatte delle identificazioni selezionandone un sottoinsieme. Questo perché, come accennato più volte (si guardi sezione 2.1), il metodo PEM non dà risultati affidabili nel caso in cui il set di dati a disposizione è ridotto in confronto al numero di parametri da identificare.

In queste simulazioni, per l'identificazione si è considerato un set di dati con $N = 500$ e $T = 6$ minuti. Sono stati considerati due set di dati: uno relativo alla grande sala riunioni composto dalle schede multisensore $S3$, $S5$, $S7$, $S8$, $S9$ e $S10$ e l'altro relativo ai due uffici affiancati composto dalle schede multisensore $S18$, $S19$, $S21$, $S23$, $S24$ e $S25$ (si guardi Figura 5.4). Come uscite del modello sono stati fissati tutti gli andamenti di temperatura di tale schede. Come ingressi invece sono stati scelti tutti gli andamenti di umidità relativa e irraggiamento delle schede multisensore e gli andamenti di temperatura, umidità relativa e ir-

raggiamento misurati dalla stazione meteo.

In Figura 5.7 sono riportati i risultati in termini di FIT_1 (6 minuti in avanti) e di FIT_{20} (120 minuti in avanti) delle simulazioni impiegando i due set di dati rispettivamente.

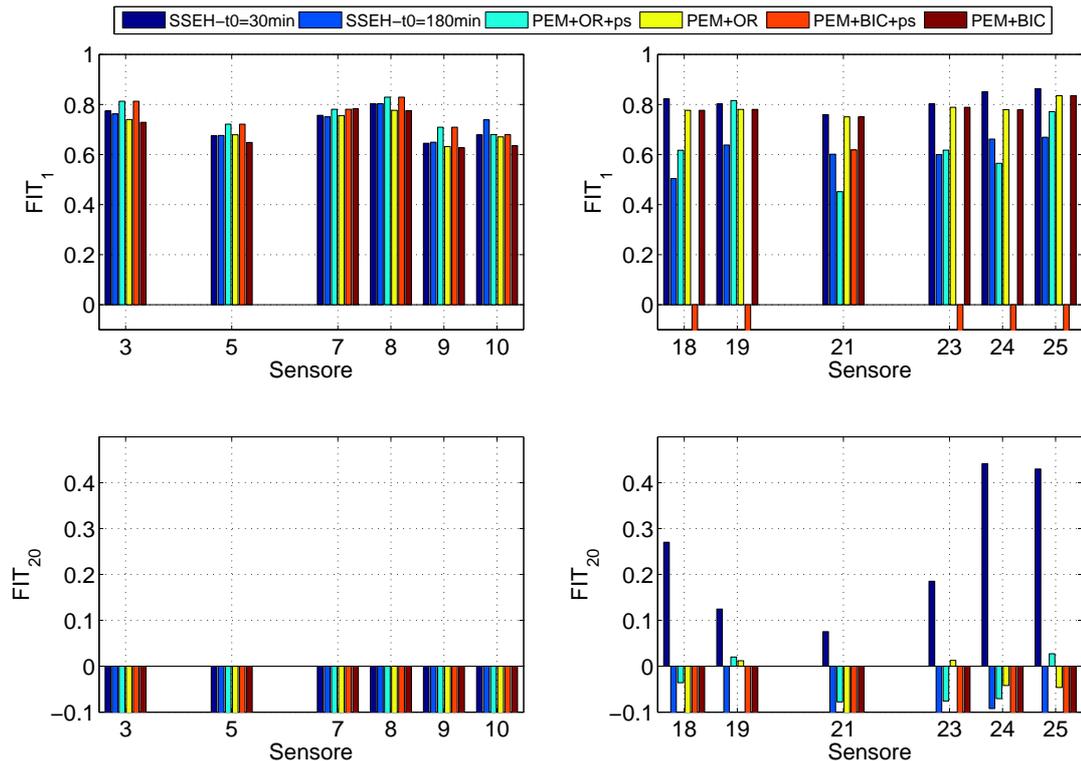


Figura 5.7 – FIT_1 (6 minuti in avanti; grafici sopra) e FIT_{20} (120 minuti in avanti; grafici sotto) relativi al set di schede multisensore S_3, S_5, S_7, S_8, S_9 e S_{10} (grafici a sinistra) e al set di schede multisensore $S_{18}, S_{19}, S_{21}, S_{23}, S_{24}$ e S_{25} (grafici a destra) rispettivamente, con $N = 500$ e $T = 6$ minuti. L'asse delle ascisse è relativa ai nomi delle schede multisensore indicati in Figura 5.4.

Dalla Figura 5.7 si vede che a un passo di predizione (6 minuti in avanti) le diverse versioni di identificazione forniscono prestazioni buone. Le prestazioni delle versioni del metodo PEM che non impiegano la selezione effettuata da SSEH sub-ottimo sono in generale superiori a quelle ottenute dalla configurazione di riferimento (con $N = 500$, $T = 6$ minuti e impiegando per l'identificazione tutti i sensori disponibili) vista in precedenza. La versione SSEH sub-ottima ha prestazioni in termini di FIT_1 leggermente inferiori rispetto la configurazione di riferimento. Le versioni PEM che impiegano la preselezione invece generano FIT_1

leggermente superiori per il set di schede multisensore $S3, S5, S7, S8, S9$ e $S10$, ma soprattutto la versione PEM+BIC+Oracolo ha prestazioni molto inferiori per il set di schede multisensore $S18, S19, S21, S23, S24$ e $S25$.

Aumentando i passi di predizione, il FIT cala fortemente e, come si vede dal confronto di Figura 5.7 con Figura 5.1, il FIT_{20} è in generale inferiore considerando un sottoinsieme di tutti i sensori. Solo la versione SSEH sub-ottima con $t_0 = 30$ minuti ha prestazioni in termini di FIT_{20} , per il set di schede multisensore $S18, S19, S21, S23, S24$ e $S25$, superiori alla condizione di riferimento.

5.2 Confronto tra diverse procedure di selezione

In questa sezione si riportano i risultati delle simulazioni eseguite per testare procedure di selezione di singole risposte impulsive. È stata considerata la configurazione di riferimento con $N = 500$, $T = 6$ minuti e impiegando tutti i sensori, usando quelli di temperature delle schede multisensore come uscite e gli altri, inclusi quelli della stazione meteo, come ingressi. Di SSEH è stata impiegata solo la versione sub-ottima con $t_0 = 180$ minuti, che in sezione 5.1 ha mostrato avere buone prestazioni. Come procedure di selezione delle singole risposte impulsive sono state usate le procedure FORWARD-backward e FORWARD-AllBest-Output presentate in sezione 3.3 e impiegate anche negli studi Monte Carlo. Le altre procedure di selezione non sono invece state considerate a causa del loro onere computazionale. A causa del gran numero di variabili infatti, già la procedura FORWARD-backward ha impiegato qualche giorno per calcolare il risultato: è legittimo attendersi un tempo di calcolo ancora superiore delle altre, in particolare dalla procedura all-best selection.

In Figura 5.8 sono confrontati i FIT_1 (6 minuti in avanti) e i FIT_{20} (120 minuti in avanti) per tutte le uscite, impiegando le procedure di selezione delle singole risposte impulsive FORWARD-backward e FORWARD-AllBest-Output. Nella Figura 5.8 sono riportati anche i FIT_1 e i FIT_{20} impiegando la procedura di selezione delle variabili forward selection Bayesiana, già viste in Figura 5.1.

Da Figura 5.8 si vede che le tre procedure di selezione considerate hanno fornito prestazioni confrontabili. A conferma di ciò, Figura 5.9 riporta gli andamenti di FIT_k in funzione dei passi di predizione k (un passo equivale alla predizione 6 minuti in avanti) per i sensori di uscita $S10$ e $S25$: non si può stabilire quale pro-

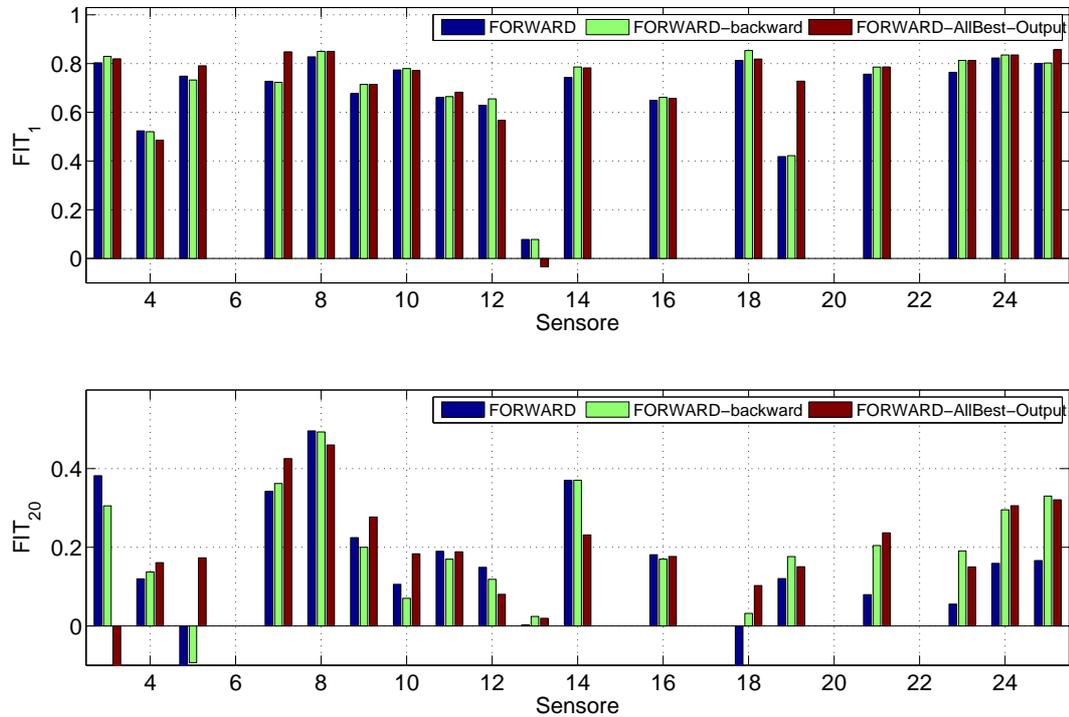


Figura 5.8 – FIT_1 (6 minuti in avanti; sopra) e FIT_{20} (120 minuti in avanti; sotto) per i diversi metodi di identificazione, con $N = 500$ e $T = 6$ minuti avendo impiegato tutti i sensori di temperatura delle schede multisensore come uscite. L'asse delle ascisse è relativa ai nomi delle schede multisensore distribuite nell'edificio e indicati in Figura 5.4.

cedura tra FORWARD-backward e FORWARD-AllBest-Output sia la migliore.

In ogni caso, per poter classificare le diverse procedure, in Tabella 5.1 sono riportate le medie di FIT_1 , FIT_{15} e FIT_{30} relative alle diverse procedure di selezione impiegate.

Da Tabella 5.1 si evidenzia quanto si vede in Figura 5.9: le procedure che effettuano la selezione delle singole risposte impulsive hanno prestazioni migliori della procedura di selezione delle variabili forward selection Bayesiana. Delle procedure delle singole risposte impulsive la FORWARD-AllBest-Output ha le prestazioni migliori a un passo di predizione (6 minuti in avanti) e a trenta passi (180 minuti in avanti) mentre la FORWARD-backward ha le prestazioni migliori ai passi intermedi. Si noti comunque che le medie riportate in tabella non sono

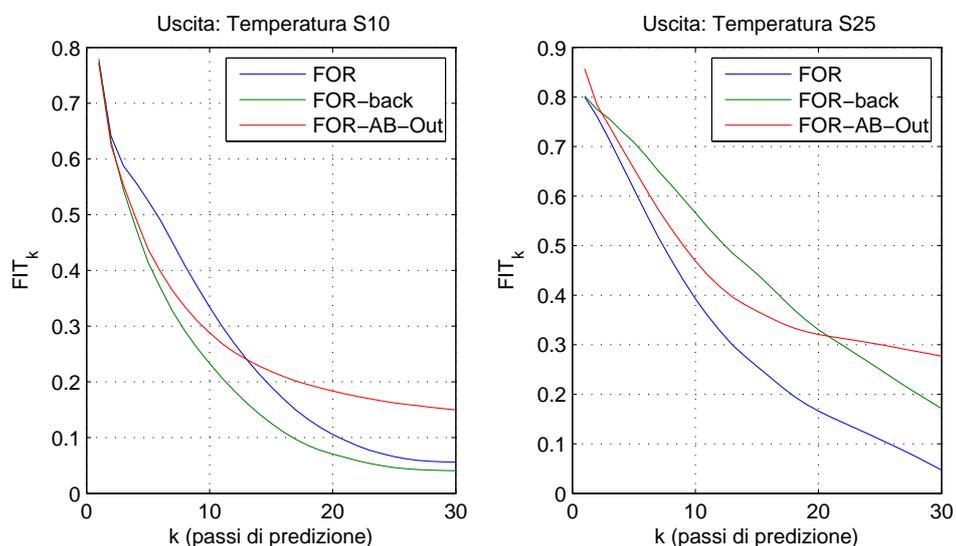


Figura 5.9 – Andamenti dell'indice FIT in funzione dei passi di predizione k per ciascuna delle procedure di selezione di SSEH, con $N = 500$ e $T = 6$ minuti e usando come uscite i sensori di temperatura delle schede multisensore S10 (sinistra) e S25 (destra).

Tabella 5.1 – Medie sulle uscite di FIT_1 , FIT_{15} e di FIT_{30} , differenziate per le procedure di selezione forward selection Bayesiana, FORWARD-backward e FORWARD-AllBest-Output.

	FOR	FOR-back	FOR-AB-Out
FIT_1	0.678	0.695	0.709
FIT_{15}	0.206	0.255	0.228
FIT_{30}	0.1	0.127	0.131

significative da un punto di vista statistico, perché si riferiscono solo alle 18 uscite di questo specifico monitoraggio.

Ovviamente, mentre il set di sensori scelti della procedura forward selection Bayesiana è uguale per ogni uscita (si guardi Figura 5.4), per le procedure di selezione delle singole risposte impulsive il set impiegato per effettuare la predizione cambia per ogni uscita. In Figura 5.10 sono riportate le selezioni relative ai sensori S_{10} e S_{25} effettuate della procedura FORWARD-backward, in Figura 5.11 invece sono riportate le selezioni relative ai sensori S_{10} e S_{25} effettuate della procedura FORWARD-AllBest-Output.

Confrontando le Figura 5.10 e 5.11 si vede che la scelta di sensori effettuata dalle due procedure di selezione delle singole risposte impulsive è molto simile. Per predire la relativa uscita in entrambi i casi è stato scelto solo un piccolo set di sensori, in generale spazialmente non molto distanti dall'uscita relativa. Si noti comunque che il numero totale di sensori selezionati per predire il sistema dalla procedura FORWARD-AllBest-Output è molto superiore a quello della procedura FORWARD-backward. Questo è dovuto ovviamente al fatto che la procedura FORWARD-AllBest-Output, partendo dal set di risposte impulsive selezionate dalla procedura di selezione delle variabili forward selection Bayesiana, può aggiungere, se ritenuto utile, singole risposte impulsive. La procedura FORWARD-backward invece, partendo dal set selezionato dalla procedura forward selection Bayesiana, può solo eliminare delle risposte impulsive.

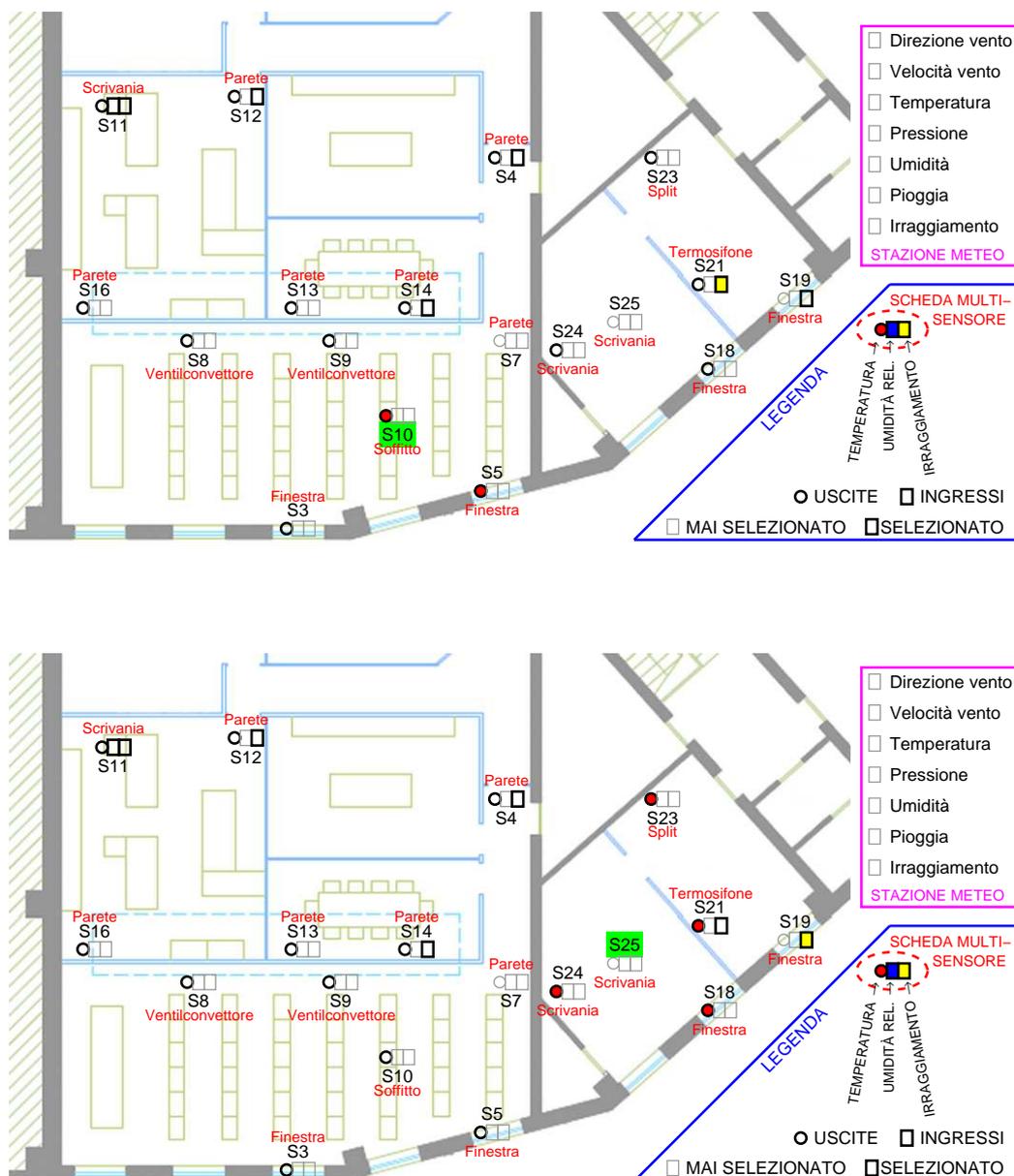


Figura 5.10 – Selezione dei sensori effettuata dalla procedura di selezione delle singole risposte impulsive FORWARD-backward di SSEH sub-ottimo per il sensore S10 (sopra) e per il sensore S25 (sotto). Tutti i sensori rappresentati in Figura come cerchi sono uscite del modello, quelli rappresentati con un rettangolo sono invece gli ingressi. Il sensore, evidenziato in verde, è l'uscita alla quale si riferisce il grafico rispettivo. I sensori colorati sono quelli selezionati per la predizione di tale uscita, quelli con la cornice nera e spessa sono i sensori scelti per predire almeno un'uscita e quelli con la cornice grigia e sottile sono invece i sensori completamente scartati dalla procedura di selezione.

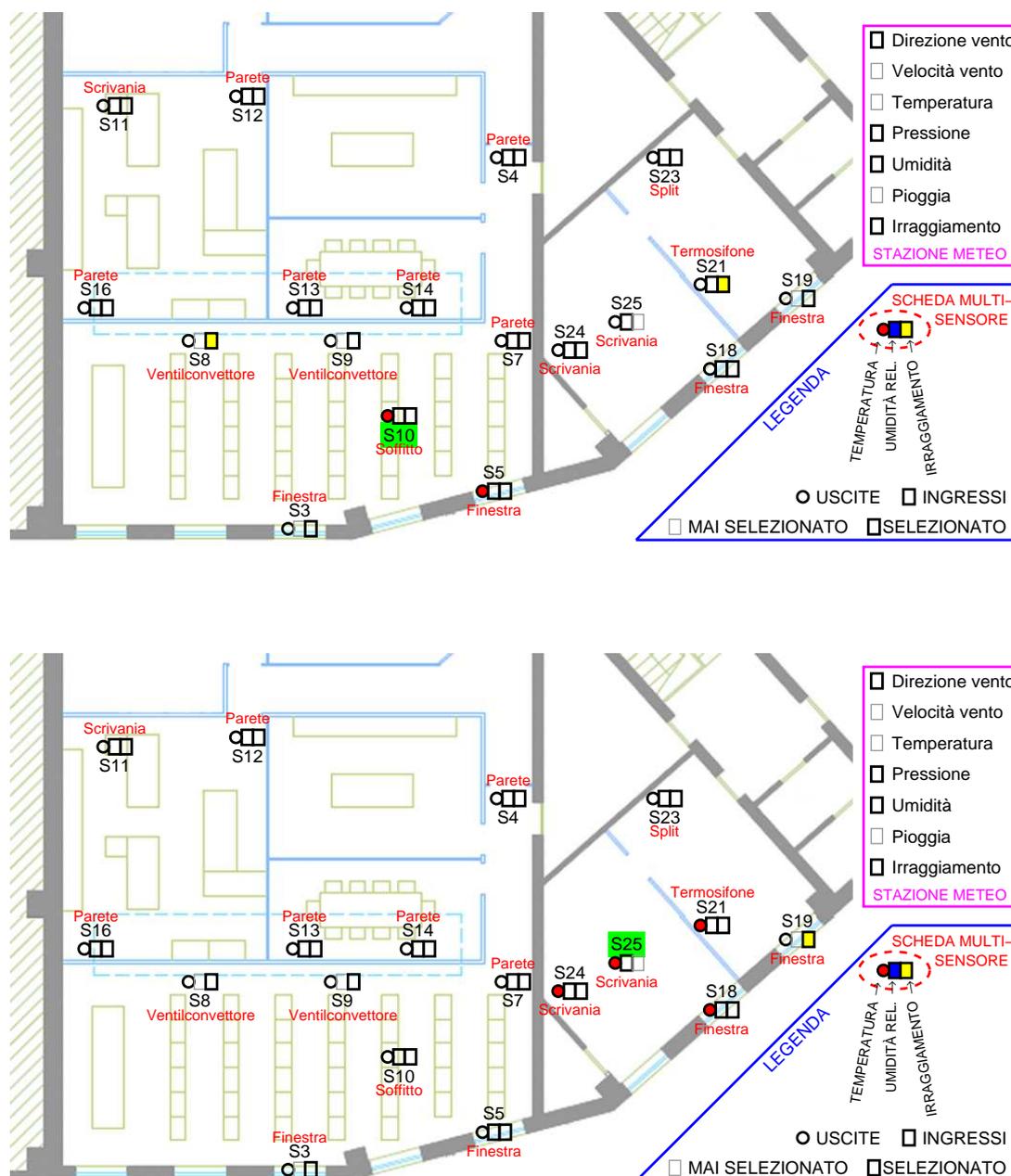


Figura 5.11 – Selezione dei sensori effettuata dalla procedura di selezione delle singole risposte impulsive FORWARD-AllBest-Output di SSEH sub-ottimo per il sensore S10 (sopra) e per il sensore S25 (sotto). Tutti i sensori rappresentati in Figura come cerchi sono uscite del modello, quelli rappresentati con un rettangolo sono invece gli ingressi. Il sensore, evidenziato in verde, è l'uscita alla quale si riferisce il grafico rispettivo. I sensori colorati sono quelli selezionati per la predizione di tale uscita, quelli con la cornice nera e spessa sono i sensori scelti per predire almeno un'uscita e quelli con la cornice grigia e sottile sono invece i sensori completamente scartati dalla procedura di selezione.

Capitolo 6

Discussione e confronto dei risultati

In questo capitolo si discutono i risultati presentati nelle sezioni precedenti. Si analizzano prima i risultati relativi ai diversi studi Monte Carlo considerati e poi quelli relativi alle identificazioni effettuate sui dati monitorati in un edificio esistente.

Nelle sezioni precedenti non sono stati indicati i tempi di calcolo impiegati dai diversi algoritmi di identificazione, perché nella programmazione degli algoritmi che realizzano le diverse versioni SSEH è stato tenuto in considerazione solo l'aspetto funzionale, non ottimizzando l'aspetto di onere computazionale. Siccome l'efficienza computazionale è però una caratteristica molto importante per valutare l'impiego di un algoritmo in applicazioni reali, nella parte iniziale di questo capitolo si faranno alcune considerazioni a questo riguardo.

6.1 Considerazioni relative all'efficienza computazionale

Nella presentazione dei risultati non sono stati riportati i tempi di calcolo dei vari algoritmi, perché questo dipende ovviamente da molti fattori. Per esempio, per eseguire le simulazioni sono stati utilizzati due differenti computer, uno con un processore "intel centrino due T5600" e l'altro con un processore "intel core i5-2540M": il primo computer ha impiegato più del doppio del tempo per sviluppare calcoli confrontabili. Dato che il tempo di calcolo è un criterio molto importante per la scelta di impiego di un certo algoritmo piuttosto di un altro, qui si riporta-

no alcuni confronti tra le diverse versioni di metodi di identificazione impiegati. Si noti comunque che i diversi algoritmi SSEH non sono stati ottimizzati in termini di efficienza computazionale e sono quindi migliorabili. Si noti inoltre che le complessità computazionali in termini di $O(\cdot)$ delle diverse procedure di selezione di SSEH sono riportati in sezione 3.3. I tempi di calcolo riportati qui sono relativi al computer con un processore “intel core i5-2540M”, con 4GB di memoria RAM e avendo impiegato come programma di calcolo *MATLAB*[®] - versione R2011a.

Studi di Monte Carlo Negli studi Monte Carlo sono stati impiegati $m = 10$ ingressi e $p = 3$ uscite. Il tempo di calcolo dell’identificazione SSEH sub-ottima dipende fortemente alla lunghezza del set di dati N . Per $N = 150$ il tempo impiegato è di 12 – 18 secondi e crescendo all’aumenta di N , ha impiegato per $N = 300$ circa 30 – 40 secondi e per $N = 800$ circa 2 – 3 minuti. Al contrario, il tempo di calcolo delle diverse versioni di PEM considerate non dipende così fortemente da N . Infatti, nel caso in cui l’ordine massimo da considerare era fissato uguale a 12, si ottenevano tempi di calcolo pari a 30 – 35 secondi per $N = 150$, 35 – 40 secondi per $N = 300$ e 60 – 70 secondi per $N = 800$. Si noti però che il tempo di calcolo dei metodi PEM cresce fortemente aumentando l’ordine massimo da considerare. Prendendo 30 come ordine massimo, per $N = 300$ il tempo di calcolo delle versioni PEM è cresciuto fino a 10 – 15 minuti.

La lunghezza delle risposte impulsive t_0 di SSEH invece non ha un grosso impatto sul tempo di calcolo. Si noti però, che aumentando t_0 , aumenta lo spazio di memoria necessario per eseguire i calcoli.

Le versioni SSEH ottime sono state considerate solo per $N = 300$ e hanno impiegato un tempo di calcolo nettamente superiore alla versione sub-ottima. Mentre la versione SSEH ottima impiegava circa 10 – 15 minuti, la versione con preselezione è stata più veloce, dipendendo dal numero di variabili eliminate dalla identificazione sub-ottima di inizializzazione e quindi non più considerate nella fase di ottimizzazione. Se solo poche variabili sono state selezionate, il tempo di calcolo si è ridotto anche a 2 – 3 minuti.

Negli studi Monte Carlo, il tempo impiegato dalla procedura di selezione nelle identificazioni SSEH è solo una piccola quota del tempo complessivo impiegato dall’identificazione SSEH. Questo vale soprattutto per le versioni sequenziali di forward selection e backward elimination, che nel caso $N = 300$ impiegavano so-

lo 1 – 2 secondi. Le versioni Bayesiane di queste procedure impiegavano invece 5 – 7 secondi. Sia la versione Bayesiana sia la versione sequenziale della procedura di stepwise selection e sia la versione Bayesiana della procedura all-best selection hanno richiesto circa 7 – 9 secondi. Si noti che il tempo di calcolo delle due versioni Bayesiane è simile a quello della versione sequenziale stepwise selection anche se internamente le procedure Bayesiane sono molto più complesse, perché queste ultime dispongono di un criterio di arresto e devono quindi considerare molte meno combinazioni di variabili della procedura stepwise selection sequenziale.

Identificazioni impiegando i dati monitorati Nelle simulazioni impiegando i dati monitorati nell'edificio esistente, rispetto agli studi Monte Carlo, sono state usate più variabili di ingressi e di uscite. Infatti, nel monitoraggio sono state impiegate 18 schede multisensore, ciascuna delle quali registrava dati di temperatura, umidità relativa e irraggiamento, e sono stati utilizzati i dati di una stazione meteo che registrava direzione e velocità del vento, temperatura, umidità, pioggia, irraggiamento e pressione. Al posto di 3 uscite e 10 ingressi, in gran parte delle simulazioni impiegando i dati monitorati, tutti i sensori di temperatura della schede multisensore sono stati usati come uscite del modello e come ingressi tutti gli altri, ottenendo quindi in tutto 18 uscite e 43 ingressi. Questo ha fatto aumentare di molto il tempo di calcolo dei diversi metodi di identificazione.

Nelle diverse simulazioni con $N = 250$, la versione SSEH sub-ottima ha impiegato circa 20 minuti, le diverse versioni PEM invece 10 – 15 minuti impostando 12 come ordine massimo da considerare.

Mentre negli studi Monte Carlo la procedura di selezione aveva poca influenza sul tempo totale dell'identificazione SSEH, in questo caso l'efficienza della procedura è il fattore determinante. Questo si è mostrato in particolare per $N = 500$, dove sono stati impiegati la procedura di selezione delle variabili forward selection Bayesiana e le procedure di selezione delle risposte impulsive FORWARD-AllBest-Output e FORWARD-backward. L'identificazione SSEH complessiva impiegando queste tre procedure di selezione ha richiesto circa 67 ore. Di questo tempo la procedura di selezione delle variabili forward selection Bayesiana, usata come procedura di inizializzazione delle due seguenti, ha impiegato circa 50 minuti, la procedura FORWARD-backward circa 50 ore e la procedura FORWARD-AllBest-Output circa 16 ore. Eliminando i tempi impiegati dalle procedure di selezione dal tempo totale, si ricavano circa 25 minuti come tempo impiegato dall'identificazione SSEH per eseguire tutti i rimanenti calcoli (compreso anche il problema

di ottimizzazione 1. nella procedura descritta in sezione 2.2.5).

6.2 Analisi dei risultati degli studi Monte Carlo

Si discutono qui di seguito i risultati ottenuti dai vari studi Monte Carlo, riferendosi alle sezioni di capitolo 4.

Diverse lunghezze del set di dati N e delle risposte impulsive t_0 In sezione 4.1 sono stati riportati i risultati degli studi Monte Carlo, eseguiti per mostrare l'influenza della lunghezza del set di dati N sulle prestazioni dei metodi SSEH e PEM e l'influenza della lunghezza delle risposte impulsive t_0 sulle prestazioni di SSEH. In queste simulazioni sono stati utilizzati modelli di riferimento sia con livello di sparsità ridotto sia con livello di sparsità aumentato e sono stati generati ingressi scorrelati e correlati.

Dalle simulazioni si è evidenziato chiaramente che, all'aumentare di N , le prestazioni dei metodi di identificazione migliorano chiaramente (si guardino per esempio le Tabelle 4.3 e 4.7, e le relative Figure 4.5 e 4.11). Per lunghezze N ridotte, la versione PEM+BIC, che deve stimare l'ordine del modello dai dati e non conosce il set delle variabili che danno effettivamente contributo alla predizione del sistema, ha dato prestazioni scarse nelle simulazioni. La versione PEM+Oracolo+set, cioè la versione di PEM che ha accesso ai dati di test per selezionare l'ordine del modello migliore e che conosce il set di variabili che effettivamente contribuiscono alla predizione del sistema, ha ottenuto prestazioni in generale confrontabili con quelle ottenute impiegando le versioni SSEH sub-ottime. Particolarmente per modelli di riferimento con livello di sparsità aumentato, pure da questa versione PEM i risultati ottenuti per N ridotti sono stati inferiori a quelli ottenuti con SSEH. Con PEM+Oracolo, che non conosce il set di variabili che contano, sono state ottenute prestazioni intermedie tra PEM+BIC e PEM+Oracolo+set. Per N estese si è visto che le prestazioni di tutte le diverse versioni PEM e SSEH si avvicinano e convergono al FIT_k ottenuto dalla predizione impiegando il modello di riferimento.

Si sottolinea che la versione PEM+BIC è utilizzabile in applicazioni reali. Le due versioni PEM Oracolo invece di solito non lo sono, dato che in applicazioni reali di solito non si ha un set di dati di test a disposizione e non si conosce il set delle variabili che contribuiscono alla predizione del sistema. Questo vale in particolare se il tempo di acquisizione dei dati è ridotto, il che rappresenta la situazione di

interesse di questa tesi. Le versioni con Oracolo infatti sono state considerate con lo scopo di mostrare le prestazioni migliori ottenibili impiegando metodi PEM. Per le considerazioni appena fatte si può dire che SSEH è un metodo molto valido in confronto al metodo PEM nel caso in cui la lunghezza del set di dati N è ridotta, proprietà richiesta in questa tesi. Si sottolinea poi che, anche aumentando N , dall'identificazione SSEH sono state ottenute prestazioni confrontabili con PEM.

Come si vede per esempio dalle Tabelle 4.4 e 4.8, nei diversi studi Monte Carlo in generale la procedura di selezione delle variabili forward selection Bayesiana del metodo SSEH ha portato a risultati molto buoni, che migliorano leggermente all'aumentare di N . In confronto alle altre, la percentuale di uscite considerate in maniera corretta, riportata in Tabella 4.8 e ottenuta dalle simulazioni impiegando modelli con livello di sparsità aumentato è notevolmente inferiore. Questo fenomeno è emerso anche nelle sezioni 4.2 e 4.3. Analizzando i risultati si è visto che impiegando modelli con livello di sparsità aumentato di norma tutte le uscite sono state selezionate dalla procedura, anche se spesso il passato di qualche uscita non ha dato contributo all'evoluzione del sistema. Non è ben chiaro perché si presenti questo fenomeno; un motivo è sicuramente che le uscite del modello di riferimento non risultano essere molto correlate e ogni uscita contiene quindi delle informazioni non esprimibili dalle altre. Infatti, analizzando i dati delle simulazioni, si è visto che il coefficiente di correlazione lineare risultava tra 0.2 e 0.9, valori piccoli in confronto al 0.98 ottenuto per gli ingressi generati in modo da essere correlati.

Si ritiene che un altro motivo delle basse percentuali di uscite considerate in maniera corretta sia che i risultati in termini di FIT_k , sia delle predizioni con i modelli di riferimento sia con i modelli identificati, ottenuti impiegando modelli con livello di sparsità aumentato, sono nettamente migliori di quelli ottenuti impiegando modelli con livello di sparsità ridotto. Questo si osserva per esempio confrontando Tabelle 4.3 con Tabella 4.7. Il fatto di considerare una qualche variabile di uscita in realtà superflua non ha quindi in pratica alcun impatto sulle prestazioni. Le prestazioni migliori sono dovute sicuramente al fatto che modelli con livello di sparsità aumentato hanno molte risposte impulsive del predittore nulle e la predizione di una specifica uscita dipende quindi solo da poche altre variabili. Questo facilita la predizione dell'uscita del modello di riferimento. Probabilmente anche il modo con cui si genera il modello di riferimento con livello di sparsità aumentato contribuisce alle prestazioni alte in termini di FIT_k : si genera un modello di riferimento random stabile del quale si calcolano le risposte

impulsive del predittore. Si annullano un certo numero di risposte impulsive del predittore e si ricalcola il modello che può però risultare instabile. In tale caso si ripete questa procedura. Può essere che tale procedimento elimini proprio i modelli di riferimento particolarmente difficili da predire.

Come accennato, nelle simulazioni sono stati considerati sia ingressi scorrelati, sia ingressi correlati. Confrontando i risultati in termini di FIT_k si vede che, come atteso, le prestazioni peggiorano impiegando ingressi correlati. Questo si evidenzia, per esempio, dal confronto tra Tabella 4.1 e Tabella 4.3. Anche le prestazioni della selezione delle variabili di SSEH peggiora impiegando ingressi correlati come può essere visto chiaramente dal confronto di Tabella 4.6 con Tabella 4.8. Del resto, quanto detto prima relativamente al confronto delle prestazioni dei diversi metodi, vale sia impiegando ingressi correlati sia impiegando ingressi scorrelati. Per questo motivo e dato che ingressi scorrelati sono meno impegnativi per le procedure di identificazione che ingressi correlati, negli studi Monte Carlo successivi, eseguiti per analizzare le influenze di N e t_0 , non sono più stati considerati ingressi scorrelati.

Per quanto riguarda la lunghezza delle risposte impulsive t_0 dei modelli identificati con SSEH, dalle diverse simulazioni si è visto che, per N ridotto, conviene sceglierla non troppo lunga (si guardi per esempio Figura 4.5 e Figura 4.11). Questo è dovuto al fatto che il set dei dati effettivamente usato per eseguire le identificazioni è di lunghezza $N - t_0$ e quindi per N non molto esteso un aumento di t_0 si fa sentire. D'altronde, se N è abbastanza esteso, un aumento delle risposte impulsive comporta una maggior flessibilità del modello nello spiegare i dati e quindi si migliorano le prestazioni in termini di FIT_k . Per scegliere la lunghezza t_0 adeguata, l'impiego del metodo PEM/ARX+AICc ha ottenuto buone prestazioni nelle simulazioni (si guardino per esempio le Figure 4.5 e 4.11). Per tale motivo, negli studi Monte Carlo seguente quelli eseguiti per analizzare l'influenza di N e t_0 , quello è il metodo usato per scegliere t_0 .

Si attende che l'indice FIT_k peggiori all'aumentare dei passi di predizione k impiegando i modelli ottenuti mediante i diversi metodi di identificazione, dato che all'aumentare di k l'influenza sul sistema del passato delle uscite diminuisce. Ad eccezione di PEM+BIC, e solo per N ridotti, questo si è verificato per tutti i diversi metodi, come si vede per esempio da Tabella 4.1. Il fatto che per PEM+BIC in certi casi il FIT_k sia peggiore per piccoli k si spiega notando che

tale metodo ha difficoltà a identificare il modello di rumore per N ridotti. Dato che la predizione dipende dall'uscita per via del modello di rumore e dato che questa dipendenza diminuisce all'aumentare dei passi, il FIT_k è inferiore per piccoli k . All'aumentare di N il risultato dell'identificazione del modello di rumore migliora e con questo i FIT_k , presentando l'andamento in funzione dei passi di predizione k atteso.

Confronto tra le diverse procedure di selezione di SSEH In sezione 4.2 si sono confrontate diverse procedure di selezione delle variabili e delle risposte impulsive di SSEH, impiegando per la loro maggior efficienza la versione SSEH sub-ottima. Sono stati usati solo ingressi correlati, dato che con tali si formano situazioni più impegnative per le procedure di selezione e si evidenziano quindi meglio le differenze tra le diverse procedure. Per confrontare le procedure di selezione delle variabili, sono stati utilizzati modelli di riferimento sia con livello di sparsità ridotto sia quelli con livello di sparsità aumentato, mentre per confrontare le procedure di selezione delle risposte impulsive sono stati utilizzati solo modelli di riferimento con livello di sparsità aumentato.

Inizialmente, in sezione 4.2, si sono riportati i risultati relativi alle procedure di selezione delle variabili sequenziali. Come visto in sezione 3.3 in termini di $O(\cdot)$ e in sezione 6.1 riportando i tempi di calcolo impiegati effettivamente nelle simulazioni, rispetto alle procedure Bayesiane queste hanno in generale il grosso vantaggio di essere computazionalmente molto più efficienti.

Per quanto riguarda il confronto tra le diverse procedure di selezione delle variabili sequenziali, come atteso, con la forward selection sequenziale sono state ottenute prestazioni inferiori sia in termini di FIT_k , sia in termini di percentuale di variabili considerate in maniera corretta (si guardino per esempio le Tabelle 4.9 e 4.10). Come accennato precedentemente, questo è dovuto al fatto che variabili che non danno un contributo alla predizione del sistema, ma che sono considerate prima di variabili correlate a esse che ne contribuiscono, sono selezionate per sbaglio e non possono più essere tolte eliminate. Dato che per le procedure sequenziali backwise elimination e stepwise selection tale problema non si presenta, le prestazioni ottenute con queste procedure sono migliori. Si noti però che la procedura stepwise selection è computazionalmente più onerosa rispetto alle altre due, poiché deve valutare ogni variabile per più volte e la procedura backwise elimination sequenziale è quindi preferibile.

Un altro modo per aggirare il problema della procedura forward selection sequenziale è non valutare le variabili sequenzialmente, ma determinare un'ordine con il quale si valutano le variabili impiegando un criterio Bayesiano. Confrontando infatti la procedura forward selection Bayesiana con le altre procedure Bayesiane, per esempio nelle Tabelle 4.13 e 4.14, si vede che le prestazioni sono molto simili. Anche se le prestazioni delle procedure Bayesiane stepwise selection e all-best selection sono leggermente migliori rispetto alle altre due, come procedura standard è stata considerata comunque la procedura forward selection Bayesiana. Questo perché tale procedura è computazionalmente la più efficiente nel caso in cui poche variabili servono a predire un sistema, situazione che si è presentata impiegando i dati monitorati nell'edificio, e per essere coerente con gli articoli [20–22], nelle quali è stata usata la versione MIMO di questa procedura. Si noti comunque che le procedure di selezione sequenziali, tranne la procedura forward selection, hanno presentato prestazioni confrontabili a quelle della procedura Bayesiane, anche se sono computazionalmente molto più onerose. Questo si vede per esempio dal confronto tra Tabella 4.9 e Tabella 4.13.

Per quanto riguarda le procedure di selezione delle risposte impulsive, non è stato fatto un confronto così dettagliato come per le procedure di selezione delle variabili, dato che i risultati ottenuti per la selezione delle variabili sono trasferibili a questo caso. Come pura procedura di selezione delle risposte impulsive è stata considerata solo la all-best selection poiché da quella si attendono i risultati migliori. Le altre procedure considerate sono combinazioni della procedura di selezione delle variabili forward selection Bayesiana con diverse procedure di selezione delle risposte impulsive.

Da Tabella 4.17 e da Tabella 4.18 si vede che i risultati migliori sono stati ottenuti dalla pura procedura di selezione delle risposte impulsive all-best selection Bayesiana, che impiega come stato iniziale il set vuoto di risposte impulsive selezionate. Più sorprendente è il fatto che prestazioni inferiori sono stati ottenuti dalla FORWARD-AllBest, equivalente alla procedura di selezione delle risposte impulsive all-best selection Bayesiana se non per l'impiego del set selezionato dalla procedura di selezione delle variabili forward selection Bayesiana come condizione iniziale. La procedura FORWARD-backward, che parte ancora da tale set e valuta se togliere successivamente singole risposte impulsive, ha ottenuto risultati intermedi tra le due precedenti.

Le prestazioni della FORWARD-AllBest-Input si sono mostrate essere legger-

mente peggiori di quelli della FORWARD-AllBest-Output, motivo per il quale quest'ultima è stata usata nelle simulazioni con i dati monitorati nell'edificio. I risultati ottenuti con queste due procedure sono comunque confrontabili con le altre procedure che usano come condizione iniziale la selezione effettuata dalla forward selection Bayesiana delle risposte impulsive. Si noti però che queste due procedure sono computazionalmente più efficienti delle altre versioni all-best, dato che devono considerare meno variabili alla volta.

Si noti che impiegando i modelli di riferimento con livello di sparsità aumentato si è presentato ancora il problema delle percentuali scarse di uscite considerate in maniera corretta. Questo problema è già stato discusso in precedenza, analizzando le influenze di N e di t_0 .

Confronto tra le diverse versioni SSEH e PEM Si discutono qui i risultati degli studi Monte Carlo riportati in sezione 4.3, eseguiti per confrontare le versioni SSEH sub-ottima, ottima e ottima+preselezione (versione ottima, che ottimizza solo gli iperparametri selezionati dall'identificazione sub-ottima) e le versioni PEM+Oracolo+set, PEM+Oracolo, PEM+BIC+preselezione (versione PEM+BIC che impiega solo gli ingressi selezionati dall'identificazione SSEH sub-ottima) PEM+BIC, e PEM+AICC. Per limitare l'onere computazionale, in precedenza è stata considerata solo una selezione e in particolare del metodo SSEH non sono state considerate le versioni ottime. Come lunghezza del set di dati in questi studi Monte Carlo è stato usato $N = 300$, sono stati impiegati modelli di riferimento con livello di sparsità aumentato e con livello di sparsità ridotto e gli ingressi sono stati generati in modo da essere correlati. La lunghezza delle risposte impulsive impiegata per le diverse versioni SSEH è stata stimata impiegando l'identificazione PEM/ARX+AICC ed è stata usata la procedura di selezione delle variabili forward selection Bayesiana.

Dalle Tabelle 4.19 e 4.21 e dalle Figure 4.19 e 4.20 si vede che negli studi Monte Carlo, solo la versione PEM+Oracolo+set ha dato risultati confrontabili alle diverse versioni SSEH. Come atteso, del metodo SSEH la versione SSEH ottima+preselezione ha dato le prestazioni in termini di FIT_k migliori. Infatti, solo nel caso in cui la selezione effettuata nella fase di inizializzazione fosse sbagliata, da tale versione sarebbero da attendersi prestazioni in termini di FIT_k inferiori a SSEH ottimo, ma in ogni caso prestazioni superiori alla versione SSEH

sub-ottima. La versione SSEH sub-ottima ha dato le prestazioni inferiori e la versione SSEH ottima ha dato prestazioni intermedie tra SSEH sub-ottima e SSEH ottima+preselezione. Comunque la differenza tra i FIT_k risultanti delle diverse versioni SSEH è molto piccola in confronto ai risultati ottenuti dalle versioni PEM. Dato l'onere computazionale molto maggiore delle versioni ottime di SSEH, occorre quindi valutare la convenienza dell'impegno di una di tale versioni o la versione sub-ottima. Nella presente tesi di solito la scelta è stata fatta in base all'onere computazionale minore, usando quindi la versione sub-ottima.

La procedura di selezione delle variabili forward selection Bayesiana impiegata nelle diverse versioni SSEH ha dato in generale buoni risultati, come si vede dalle Tabelle 4.20 e 4.22. L'eccezione è ancora la percentuale di uscite considerate in maniera corretta avendo impiegato modelli di riferimento con livello di sparsità aumentato già discusso in precedenza. Dalle Tabelle si vede inoltre che per la versione SSEH ottima, la selezione ha dato prestazioni inferiori rispetto alle altre due. Analizzando i risultati si è visto che, in confronto alle altre versioni SSEH, per quella ottima la procedura di selezione non è riuscita a scartare certe variabili. Il motivo è che, pur annullando i fattori di scala λ_{ij} che dopo la ottimizzazione risultano essere inferiori a e^{-10} , nella fase di ottimizzazione sono stati trovati certi λ_{ij} , per i quali la marginale decresce anche se le relative variabili non contano in realtà nella predizione del sistema.

Per quanto riguarda le prestazioni del metodo PEM, anche in questi studi di Monte Carlo si è verificato che le prestazioni peggiorano sia non conoscendo l'ordine del modello corretto sia non conoscendo il set di variabili che danno contributo alla predizione del sistema. Come si vede infatti dalle Figure 4.19 e 4.20, con PEM+Oracolo+set sono stati ottenuti le prestazioni migliori, seguito da PEM+Oracolo e PEM+BIC+preselezione. PEM+BIC+preselezione è migliore per modelli di riferimento con livello di sparsità ridotto, PEM+Oracolo invece per modelli di riferimento con livello di sparsità aumentato. Si noti però che PEM+BIC+preselezione, che esegue la selezione delle variabili impiegando l'algoritmo SSEH sub-ottimo, è applicabile in condizioni reali mentre le versioni con Oracolo di solito non lo sono, perché non si ha a disposizione un set di dati di test.

Da PEM+BIC e PEM+AICc sono stati ottenuti i risultati peggiori di tutte le versioni PEM considerate. Mentre le prestazioni di PEM+BIC sono migliori rispetto a PEM+AICc impiegando modelli di riferimento con livello di sparsità ridotto, con uso di modelli di riferimento con livello di sparsità aumentato le

due versioni hanno ottenuto risultati confrontabili. L'identificazione PEM riesce a spiegare modelli di ordine superiore per modelli di riferimento con livello di sparsità aumentato che per modelli di riferimento con livello di sparsità ridotto. Infatti l'ordine selezionato da PEM+Oracolo valeva in media 19 per modelli di riferimento con livello di sparsità ridotto e 6 per modelli di riferimento con livello di sparsità aumentato. Questo, connesso al fatto che l'impiego dell'indice AICc tende a stimare ordini del modello superiori che l'impiego dell'indice BIC [36], è il motivo delle prestazioni differenti di PEM+BIC e PEM+AICc, per diversi tipi di modelli di riferimento.

6.3 Analisi dei risultati delle identificazioni su dati reali

Si discutono in questa sezione i risultati riportati in capitolo 5, ottenuti identificando modelli termodinamici con l'impiego di diverse versioni di SSEH e di PEM. Si noti che in queste simulazioni come condizione di riferimento è stata usata una lunghezza del set di dati $N = 500$ e un tempo di campionamento $T = 6$ minuti, considerando tutti i sensori di temperatura delle schede multisensore come uscite e tutti gli altri sensori, inclusi quelli della stazione meteo, come ingressi del modello.

Confronto tra le diverse versioni SSEH e PEM Si discutono qui i risultati delle simulazioni eseguite per confrontare i metodi SSEH e PEM. Oltre alla configurazione di riferimento, sono stati usati due set di dati con lunghezza ridotta $N = 250$ e tempo di campionamento rispettivamente $T = 6$ e $T = 12$ minuti. Nel primo caso, rispetto alla configurazione di riferimento, i dati considerati si riferiscono solo alla metà dell'intervallo temporale. Nel secondo caso si riferiscono allo stesso intervallo temporale della configurazione di riferimento, ma presentano un tempo di campionamento doppio.

Nelle simulazioni, del metodo SSEH è stata impiegata solo la versione SSEH subottima con $t_0 = 5$ passi (per $T = 6$ minuti $t_0 = 30$ minuti) e $t_0 = 30$ passi (per $T = 6$ minuti $t_0 = 180$ minuti), per questioni di onere computazionale, e del metodo PEM sono state considerate le versioni PEM+Oracolo+preselezione, PEM+Oracolo, PEM+BIC+preselezione e PEM+BIC.

Da Figura 5.1, ma anche dalle Figure 5.5 e 5.6 si nota che le prestazioni a 20 passi di predizione sono nettamente inferiori di quelle a un passo. La differenza è superiore a quanto si è verificato per gli studi Monte Carlo (si guardi per esempio Figura 4.5). La ragione è riconducibile al fatto che i predittori dei modelli termodinamici dipendono fortemente dalle uscite del modello, mentre gli ingressi hanno un ruolo secondario. A conferma di ciò, per esempio nella configurazione di riferimento, dalla procedura di selezione sono stati selezionati molti più sensori di uscita che sensori di ingresso (si guardi Figura 5.4).

Impiegando un tempo di campionamento $T = 6$ minuti, dalle Figura 5.1 e 5.5 si vede che il metodo SSEH ha dato prestazioni molto buone in confronto al metodo PEM. Soprattutto a un passo di predizione, i FIT_1 ottenuti dalle versioni SSEH sub-ottima con diversi t_0 sono migliori di quelli ottenuti dai diversi metodi PEM. Gli indici FIT_{20} , cioè per la predizione a 180 minuti, ottenuti dalla versione PEM+Oracolo+preselezione risultano spesso superiori rispetto a quelli ottenuti da SSEH sub-ottimo. Si ricorda però che le versioni PEM con Oracolo non sono di solito applicabili in situazioni reali, specialmente se il set di dati è a dimensione ridotta, siccome non si ha a disposizione un set di dati di test.

In generale del metodo PEM le versioni che effettuano la preselezione impiegando i risultati di SSEH sub-ottimo raggiungono prestazioni migliori delle versioni che impiegano tutti i sensori. Sempre per $T = 6$ minuti, la differenza tra le prestazioni è più evidente per $N = 250$, ma si nota chiaramente anche per $N = 500$ (si confrontino Figura 5.1 con Figura 5.5). Considerando i risultati degli studi Monte Carlo, tale risultato è dovuto alla lunghezza del set di dati N comunque ridotta in entrambi i casi, in confronto al numero di parametri da stimare dalle versioni PEM che non impiegano la preselezione. È legittimo attendersi che aumentando N , le prestazioni delle versioni con e senza preselezioni convergano, assunzione non verificata per questioni di onere computazionale.

Dalle Figura 5.1 e 5.5 si vede inoltre che, per singole uscite, le prestazioni ottenute delle versioni PEM, che impiegano l'indice BIC per selezionare l'ordine del modello, sono superiori a quelle delle versioni PEM con Oracolo. In apparenza questo può sembrare strano poiché l'Oracolo ha accesso ai dati di test e seleziona per ogni passo di predizione l'ordine che dà i risultati migliori. In realtà, siccome la selezione è effettuata considerando i risultati dell'intero sistema, e può capitare che esista un'ordine migliore per singole uscite diverso da quello globale.

Per un tempo di campionamento di $T = 12$ minuti le prestazioni ottenute

sono molto inferiori a quelle ottenute con $T = 6$ minuti, anche confrontando i risultati con la stessa lunghezza di dati ($N = 250$; si confrontino Figura 5.5 con Figura 5.6). Le prestazioni di qualche uscita della versione PEM+Oracolo sono migliori per $T = 12$ minuti che per $T = 6$ minuti, ma per tutte le altre versioni le prestazioni ottenute per $T = 12$ minuti sono molto inferiori. Uno dei motivi, ma non quello principale, è che l'indice FIT_1 per $T = 6$ minuti si riferisce alla predizione 6 minuti in avanti, mentre per $T = 12$ minuti si riferisce ovviamente alla predizione 12 minuti in avanti, cioè predice il doppio del tempo di prima. Per questo in Figura 5.6 invece di riportare il FIT_{20} è stato riportato il FIT_{10} , in modo che questo si riferisca alla predizione 120 minuti in avanti, come nelle relative Figure con $T = 6$ minuti.

Si assume che il motivo principale delle prestazioni peggiori dei metodi di identificazione per $T = 12$ minuti di quelle per $T = 6$ minuti risieda nel fatto che le dinamiche degli andamenti termici non siano più modellabili da modelli con un tempo di campionamento di $T = 12$ minuti.

Come si vede da Figura 5.6 le prestazioni di SSEH sub-ottimo con $t_0 = 60$ minuti ($t_0 = 5$ passi) sono nettamente migliori a quelle con $t_0 = 360$ minuti ($t_0 = 30$ passi). Negli studi Monte Carlo si è visto che conviene scegliere lunghezze delle risposte impulsive t_0 non troppo estese per lunghezze del set di dati N ridotti dato che aumentando t_0 si riduce la lunghezza del set di dati effettivamente utilizzati per l'identificazione. In questo caso la ragione deve essere diversa poiché altrimenti le prestazioni inferiori per $t_0 = 30$ passi dovrebbero verificarsi anche nel caso $N = 250$ e $T = 6$ minuti. In questo caso, con l'impiego di t_0 troppo esteso, si introducono degli errori, dovuti soprattutto a incertezze nell'identificazione dei campioni delle risposte impulsive sopra una certa lunghezza, che comportano un peggioramento delle prestazioni in termini di FIT .

Confrontando Figura 5.1 con Figura 5.7 si vede, soprattutto a un passo di predizione, che impiegando una selezione manuale dei sensori le prestazioni ottenute da PEM senza preselezione migliorano considerevolmente rispetto a quelle relative all'impiego di tutti i sensori. Le prestazioni delle versioni PEM che impiegano la preselezione e della versione SSEH sub-ottimo invece sono migliori nel caso in cui si utilizzano tutti i sensori, come si evidenzia soprattutto confrontando i FIT_{20} delle due Figure. Questo mostra che la selezione delle variabili effettuata dalla procedura forward selection Bayesiana del metodo SSEH ha delle buone proprietà anche in applicazioni reali.

Confronto tra procedure di selezione di singole risposte impulsive In sezione 5.2 sono stati riportati i risultati delle simulazioni impiegando la versione SSEH sub-ottima con $t_0 = 180$ minuti. Come procedure di selezione di SSEH sono state impiegate le procedure di selezione delle singole risposte impulsive FORWARD-backward e FORWARD-AllBest-Output, poi confrontate con la procedura di selezione delle variabili forward selection Bayesiana. Nelle simulazioni è stata adottata la configurazione di riferimento con $N = 500$, $T = 6$ minuti impiegando tutti i sensori di temperatura delle schede multisensore come uscite del modello e i rimanenti sensori, inclusi quelli della stazione meteo, come ingressi. Non sono state usate le altre procedure di selezione per questioni di onere computazionale. Qui si discutono i risultati ottenuti da tali simulazioni.

Come già accennato in precedenza e dalle Figura 5.8 e 5.9 è difficile determinare quale tra le tre procedure di selezione considerate dia la miglior prestazioni in termini di *FIT*. Da Tabella 5.1, nella quale sono riportate le medie dei *FIT* riportati in Figura 5.8, si può dire che le procedure di selezione delle singole risposte impulsive forniscono prestazioni migliori della procedura di selezione delle variabili. Questo è dovuto al fatto che tutte le variabili selezionate dalla forward selection Bayesiana sono impiegate per predire ogni singola uscita, anche se spesso alcuni andamenti sono superflui per predire determinate uscite e introducono quindi degli errori nella predizione. Il vantaggio della procedura di selezione delle variabili è comunque la maggiore efficienza computazionale, come visto in sezione 6.1.

Da Tabella 5.1 si vede che la FORWARD-AllBest-Output ha dato prestazioni migliori a un passo di predizione e a molti passi (30), mentre dalla FORWARD-backward sono state ottenute prestazioni migliori a un numero di passi intermedio (15). Si ricorda che le medie riportate in tale Tabella non sono significative da un punto di vista statistico, dato che si riferiscono solo alle 18 uscite e sono disponibili solo i dati di un monitoraggio. Ci si può comunque attendere che, in generale, FORWARD-AllBest-Output raggiunga prestazioni migliori di FORWARD-backward. Entrambe usano come condizione iniziale il set ottenuto dalla procedura di selezione delle variabili forward selection Bayesiana. Mentre però la FORWARD-AllBest-Output valuta per ogni uscita quale sensore, anche tra quelli non selezionati in precedenza, può contribuire alla predizione di tale uscita, la FORWARD-backward non può aggiungere delle singole risposte impulsive non selezionate in precedenza. La procedura FORWARD-backward non può quindi recuperare informazioni utili, che nella fase di inizializzazioni sono state

considerate superflue per il sistema nella sua globalità e non per la singola uscita. Quanto atteso dalle prestazioni delle diverse procedure di selezione si è verificato nelle simulazioni, per esempio per le uscite *S10* e *S25* (si confrontino le Figure 5.4, 5.10 e 5.11). Considerando la selezione del sensore *S25*, dalla procedura di selezione delle variabili, usata come inizializzazione delle altre due, sono stati eliminati tutti i sensori della scheda multisensore *S25*. Mentre questi sensori non possono quindi essere selezionati dalla FORWARD-backward, proprio il sensore di temperatura della scheda *S25*, che in questo caso è l'uscita stessa, è stato aggiunto dalla procedura FORWARD-AllBest-Output per predire se stesso. Considerando la selezione del sensore *S10* invece, al set di partenza dalla FORWARD-AllBest-Output è stata aggiunta la misura di irraggiamento della scheda *S8*. Per questo motivo, come si vede da Figura 5.9, impiegando la procedura FORWARD-AllBest-Output sono stati ottenuti risultati in termini di *FIT* spesso migliori che non impiegando la procedura FORWARD-backward.

Guardando le Figure 5.10 e 5.11 si può dire in generale che dalle procedure di selezione delle singole risposte impulsive sono state ottenute selezioni fisicamente ben interpretabili. Infatti, per predire gli andamenti di temperature, sono state scelte poche altre misure di temperature, spazialmente vicine all'uscita sotto esame. Inoltre, analizzando i dati, si è visto che spesso è stato anche considerato qualche sensore di irraggiamento e di umidità. Confrontando le due Figure si vede che il set di sensori selezionato è simile, un'altra dimostrazione del fatto che le procedure di selezione hanno buone proprietà. Lo svantaggio della procedura FORWARD-AllBest-Output è il numero di sensori selezionati molto maggiore rispetto alle altre procedure, come si vede ancora dal confronto delle Figure 5.4, 5.10 e 5.11. Questo aspetto è importante soprattutto nel caso in cui si voglia effettuare una selezione di sensori wireless per ottenere indicazioni sul posizionamento più favorevoli per l'installazione di una rete di sensori cablati. Si noti che a tale proposito potrebbe essere utile che la procedura di selezione in qualche maniera tenesse conto del fatto che i tre sensori sulle schede multisensore sono fisicamente legati. Questo sarebbe facilmente implementabile effettuando una selezione preliminare che, similmente all'inizializzazione fatta per la FORWARD-backward e la FORWARD-AllBest-Output, valuta per ogni scheda multisensore se selezionare i tre sensori o meno.

Conclusioni

Nel settore edile, modelli matematici sono impiegati in diverse applicazioni. Si è visto che la modellazione basata su dati, chiamata anche identificazione dei modelli, offre diversi vantaggi rispetto la modellazione in avanti. Anche i metodi che costituiscono lo stato dell'arte dell'identificazione di sistemi spesso non raggiungono risultati accettabili, in particolare se si ha un gran numero di sensori in confronto alla lunghezza di dati. Questa situazione si verifica per esempio eseguendo un monitoraggio di breve durata con un gran numero di sensori wireless. L'obiettivo di questa tesi è stato quindi testare metodi di identificazione non-parametrica, dai quali sono da attendersi prestazioni migliori al verificarsi delle predette condizioni, e confrontabili con metodi PEM che costituiscono lo stato dell'arte attuale dell'identificazione e sono spesso impiegati nel settore edile.

In questa tesi, nella parte iniziale sono stati presentati brevemente i diversi approcci di modellazione applicati al settore edile. Successivamente è stata presentata la teoria dei metodi di identificazione usati nelle simulazioni: sono stati descritti dapprima i metodi PEM, considerando in particolare il problema di selezione dell'ordine del modello, e successivamente è stato descritto il metodo non-parametrico SSEH. Tale metodo è già stato sviluppato per sistemi MISO e nella tesi è stato spiegato anche come possa essere generalizzato al caso di sistemi MIMO. Con questo si è conclusa la parte teorica della tesi.

Nelle simulazioni sono state impiegate diverse versioni sia del metodo SSEH, con diverse procedure di selezione, sia del metodo PEM. Questi metodi sono stati confrontati dapprima in diversi studi Monte Carlo e successivamente identificando modelli da dati monitorati in un edificio esistente. Nella parte relativa alle simulazioni della presente tesi, inizialmente sono state descritte le diverse versioni dei metodi di identificazione e il setup degli studi Monte Carlo e del monitoraggio effettuato nell'edificio. Poi sono stati riportati i risultati dei studi Monte Carlo e successivamente i risultati ottenuti dalle identificazioni impiegando i dati moni-

torati. Infine sono stati discussi i risultati ottenuti dalle varie simulazioni.

Sia negli studi Monte Carlo sia dalle identificazioni impiegando i dati reali sono stati ottenuti risultati molto buoni dalle diverse versioni del metodo SSEH in confronto a quelli ottenuti dal metodo PEM. Infatti, nelle simulazioni eseguite, per lunghezze ridotte del set di dati, solo le versioni PEM che dispongono dell'informazione sulle variabili che servono a predire il sistema e che selezionano l'ordine migliore del modello accedendo al set dei dati di test hanno ottenuto risultati confrontabili alle versioni SSEH. Si noti che tali versioni PEM non sono applicabili in situazioni reali poiché di solito non si dispone di un set di dati di test e non si conosce il sottoinsieme delle variabili che contribuiscono alla predizione del sistema. Le altre versioni PEM hanno raggiunto prestazioni spesso molto peggiori. Aumentando la lunghezza del set di dati, le simulazioni effettuate hanno mostrato risultati equiparabili per le diverse versioni di SSEH e PEM. In particolare, le diverse procedure di selezione del metodo SSEH impiegate hanno mostrato di avere delle buone proprietà. Negli studi Monte Carlo la percentuale di variabili o risposte impulsive considerate in maniera corretta è risultata molto alta. Nelle identificazioni che utilizzano i dati reali, la selezione dei sensori si è rivelata fisicamente ben interpretabile.

Come accennato nella tesi, sviluppando gli algoritmi che implementano le diverse versioni SSEH si è considerato il tempo di calcolo. Questo è un aspetto importante da valutare in prossimi studi, siccome il tempo richiesto è stato considerevole e uno dei criteri per la scelta di un metodo di identificazione è anche l'onere computazionale.

Si ritiene che gli studi Monte Carlo riportati nella presente tesi siano abbastanza completi, mentre il lavoro relativo alle identificazioni impiegando dati reali potrebbe essere approfondito. A tale proposito, sarebbe molto interessante testare gli algoritmi impiegando dati ottenuti da altri monitoraggi, includendo per esempio anche dati sui consumi energetici, o dati relativi a settori diversi da quello edile. Al fine di confrontare gli approcci di modellazione basata su dati e modellazione in avanti nel settore edile, inizialmente c'era l'idea di confrontare i modelli ottenuti dai metodi di identificazione con modelli ottenuti impiegando programmi di simulazione energetica. Questo aspetto non è stato purtroppo sviluppato per questioni organizzative e rimane un lavoro per il futuro.

Bibliografia

- [1] (2010) Energy-efficient buildings ppp: Multi-annual roadmap and longer term strategy. EUROPEAN COMMISSION. [Online]. Available: www.ecfp.org/cws/params/ecfp/download_files/36D1191v1_EeB_Roadmap.pdf
- [2] E. Commission, Ed., *Statistical pocketbook*. Publications Office of the European Union, 2010.
- [3] C. Petersdorff, T. Boermans, and J. Harnisch, “Mitigation of co2 emissions from the eu-15 building stock. beyond the eu directive on the energy performance of buildings,” *Environmental Science and Pollution Research*, vol. 13, pp. 350–358, 2006.
- [4] C. Petersdorff, T. Boermans, J. Harnisch, S. Joosen, and F. Wouters, “The contribution of mineral wool an other thermal insulation materials to energy-saving and climate protection in europe,” 2002, ecofys.
- [5] S. Wang and Z. Ma, “Supervisory and optimal control of building hvac systems: A review,” *HVAC&R Research*, vol. 14, no. 1, pp. 3–32, 2008. [Online]. Available: <http://www.tandfonline.com/doi/abs/10.1080/10789669.2008.10390991>
- [6] Y. Ma, F. Borrelli, B. Hancey, B. Coffey, S. Benghea, and P. Haves, “Model predictive control for the operation of building cooling systems,” in *American Control Conference (ACC), 2010*, 30 2010-july 2 2010, pp. 5106 –5111. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5530468&tag=1
- [7] *Decreto del presidente della repubblica (DPR) 2 aprile 2009, n. 59*, Repubblica Italiana Std.
- [8] “11300 - 1: Determinazione del fabbisogno di energia termica dell’edificio per la climatizzazione estiva e invernale,” UNI - TS, Tech. Rep., 2008.

- [9] Energyplus. U.S. Department of Energy. [Online]. Available: http://apps1.eere.energy.gov/buildings/tools_directory/alpha_list.cfm
- [10] ASHRAE, Ed., *2005 ASHRAE handbook: fundamentals*, 2005.
- [11] J. Bloem, Ed., *System identification applied to building performance data*. Office for Official Publications of the European Communities, 1994.
- [12] S. Katipamula and M. R. Brambley, “Review article: Methods for fault detection, diagnostics, and prognostics for building systems: A review, part i,” *HVACandR Research*, vol. 11, no. 1, pp. 3–25, 2005. [Online]. Available: <http://www.tandfonline.com/doi/abs/10.1080/10789669.2005.10391123>
- [13] —, “Review article: Methods for fault detection, diagnostics, and prognostics for building systems: A review, part ii,” *HVACandR Research*, vol. 11, no. 2, pp. 169–187, 2005. [Online]. Available: <http://www.tandfonline.com/doi/abs/10.1080/10789669.2005.10391133>
- [14] A. Cenedese, A. Chiuso, and L. Schenato, “Simea - identificazione termodinamica e controllo del comfort,” 11 2010.
- [15] J. Kreider and J. Haberl, “Predicting hourly building energy use: the great energy predictor shootout- overview and discussion of results,” *ASHRAE Transactions*, vol. 100, no. 2, pp. 1104–1118, 1994.
- [16] R. Dodier and G. Henze, “Statistical analysis of neural networks as applied to building energy prediction,” *Journal of solar energy engineering*, vol. 126, p. 592, 2004.
- [17] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, pp. 267–288, 1996.
- [18] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, “Least angle regression,” *The Annals of statistics*, vol. 32, no. 2, pp. 407–499, 2004.
- [19] M. Yuan and Y. Lin, “Model selection and estimation in regression with grouped variables,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.
- [20] A. Chiuso and G. Pillonetto, “Learning sparse dynamic linear systems using stable spline kernels and exponential hyperpriors,” in *Proc. of NIPS 2010*, accepted, December 2010.

- [21] —, “A bayesian approach to sparse dynamic network identification,” *Automatica - submitted*, 20XX.
- [22] A. Aravkin, J. Burke, A. Chiuso, and G. Pillonetto, “Convex vs nonconvex approaches for sparse estimation: Lasso, multiple kernel learning and hyperparameter lasso,” in *IEEE CDC 2011 (accepted)*, 2011.
- [23] G. Pillonetto and G. De Nicolao, “A new kernel-based approach for linear system identification,” *Automatica*, vol. 46, no. 1, pp. 81–93, 2010.
- [24] N. Aronszajn, *Theory of reproducing kernels*. Harvard University, 1951.
- [25] J. Neymark and R. Judkoff, “International energy agency building energy simulation test and diagnostic method for heating, ventilating, and air-conditioning equipment models (hvac bestest); volume 1: Cases e100-e200,” National Renewable Energy Lab., Golden, CO.(US), Tech. Rep., 2002.
- [26] P. Baker and H. van Dijk, “Paslink and dynamic outdoor testing of building components,” *Building and Environment*, vol. 43, no. 2, pp. 143–151, 2008.
- [27] *Getting started with EnergyPlus - Basic concepts manual*, US Department of Energy, 10 2011.
- [28] Energyplus. U.S. Department of Energy. [Online]. Available: <http://apps1.eere.energy.gov/buildings/energyplus/>
- [29] G. Picci, *Metodi statistici per l'identificazione di sistemi lineari*, 2011.
- [30] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Wiley Online Library, 1999.
- [31] T. Söderström and P. Stoica, *System identification*. Prentice-Hall, 1989.
- [32] H. Akaike, “A new look at the statistical model identification,” *Automatic Control, IEEE Transactions on*, vol. 19, no. 6, pp. 716–723, 1974.
- [33] G. Schwarz, “Estimating the dimension of a model,” *The annals of statistics*, vol. 6, no. 2, pp. 461–464, 1978.
- [34] N. Sugiura, “Further analysts of the data by akaike’s information criterion and the finite corrections,” *Communications in Statistics-Theory and Methods*, vol. 7, no. 1, pp. 13–26, 1978.

-
- [35] C. Hurvich and C. Tsai, “Regression and time series model selection in small samples,” *Biometrika*, vol. 76, no. 2, pp. 297–307, 1989.
- [36] K. Burnham and D. Anderson, *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Verlag, 2002.
- [37] R. Alberton, M. Ausserer, A. Barazzuol, A. Beghi, A. Cenedese, A. Chiuso, and G. Pillonetto, “Sparse dynamical models for thermodynamic identification of buildings,” University of Padova, Tech. Rep., 2011.
- [38] S. Weisberg, *Applied linear regression*. Wiley, 2005, vol. 528.
- [39] T. Hastie, R. Tibshirani, and J. J. H. Friedman, *The elements of statistical learning*. Springer, 2009.
- [40] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” *The Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.
- [41] S. Bakin, “Adaptive regression and model selection in data mining problems,” Ph.D. dissertation, Australian National University, 1999.
- [42] M. Tipping, “Sparse bayesian learning and the relevance vector machine,” *The Journal of Machine Learning Research*, vol. 1, pp. 211–244, 2001.
- [43] J. Maritz and T. Lwin, *Empirical bayes methods*. Chapman and Hall London, 1989.
- [44] F. Bach, “Consistency of the group lasso and multiple kernel learning,” *The Journal of Machine Learning Research*, vol. 9, pp. 1179–1225, 2008.
- [45] C. Rasmussen and C. Williams, “Gaussian processes for machine learning,” 2006.
- [46] G. Pillonetto, A. Chiuso, and G. De Nicolao, “Prediction error identification of linear systems: A nonparametric gaussian regression approach,” *Automatica*, 2010.
- [47] G. Goodwin, M. Gevers, and B. Ninness, “Quantifying the error in estimated transfer functions with application to model order selection,” *Automatic Control, IEEE Transactions on*, vol. 37, no. 7, pp. 913–928, 1992.
- [48] F. Dinuzzo, “Kernel machines with two layers and multiple kernel learning,” *Arxiv preprint arXiv:1001.2709*, 2010.

- [49] K. Kira and L. Rendell, "The feature selection problem: Traditional methods and a new algorithm," in *Proceedings of the National Conference on Artificial Intelligence*. John Wiley & Sons Ltd, 1992, pp. 129–129.
- [50] A. Jain, R. Duin, and J. Mao, "Statistical pattern recognition: a review," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 1, pp. 4–37, jan 2000.
- [51] P. Devijver and J. Kittler, *Pattern recognition: A statistical approach*. Prentice/Hall International, 1982.
- [52] Simea: Sistema integrato/distribuito di monitoraggio energetico ed ambientale. [Online]. Available: <http://automatica.dei.unipd.it/people/cenedese/research/simea.html>
- [53] A. Cenedese, "Simea - definizione ambiente dimostrativo," 11 2010.
- [54] L. Cossaro, "Dimensionamento della rete per il supporto al sistema simea," 06 2010.
- [55] ARPAV. [Online]. Available: http://www.arpa.veneto.it/upload_teolo/dati_rete/vicenza.html
- [56] *System Identification Toolbox V7.4 for MATLAB*, MathWorks, 2010.
- [57] E. George, "The variable selection problem," *Journal of the American Statistical Association*, vol. 95, no. 452, pp. 1304–1308, 2000.
- [58] A. Blum and P. Langley, "Selection of relevant features and examples in machine learning," *Artificial intelligence*, vol. 97, no. 1-2, pp. 245–271, 1997.

Ringraziamenti

Un ringraziamento va innanzitutto al mio relatore, professor Alessandro Chiuso, per aver seguito questo lavoro e per il tempo speso con i miei resoconti periodici. Nonostante sia molto occupato è sempre riuscito a darmi consigli e indicazioni utili.

Desidero ringraziare il team di Fraunhofer Italia, in particolare il direttore professor Dominik Matt e i miei tutori/correlatori Ing. Irene Paradisi e Dipl.-Ing. Lorenz Vorhauser, per aver reso possibile lo svolgimento del lavoro di tesi presso la loro azienda. Specialmente ringrazio l'Ing. Irene Paradisi poiché, nonostante sia costantemente piena di lavoro, ha sempre trovato il tempo per revisionare la mia tesi in maniera molto attenta e precisa, dandomi suggerimenti e consigli molto utili.

Ringrazio il professor Angelo Cenedese e il dottorando Michele Pasqualetto per avermi dato accesso ai dati monitorati nell'edificio di RiCert nel corso del progetto SIMEA.

Infine, un ringraziamento sentito va a tutta la mia famiglia e alla mia ragazza Magdalena per la pazienza, il supporto costante, l'affetto e la fiducia che hanno avuto nei miei confronti, non solo durante lo svolgimento della tesi, ma in tutti gli anni di studio.