POLITECNICO DI MILANO
School of Industrial and Information Engineering
Master of Science in Nuclear Engineering

POLITECNICO
MILANO 1863

ENSEMBLE OF ECHO STATE NETWORKS FOR PREDICTING THE ENERGY PRODUCTION OF POWER PLANTS

Thesis advisor:
Prof. Piero Baraldi

Co-advisors:
Prof. Enrico Zio
Dr. Sameer Al-Dahidi

Candidate:
Eleonora Nigro
ID number: 852745

Academic year: 2017/2018

A Bìbì

Ringraziamenti

Un primo sincero ringraziamento va al mio relatore Piero Baraldi, per avermi guidato in questo percoso.

Grazie al prof. Zio per avermi accolta nel gruppo Lasar e per la fiducia dimostrata nell'affidarmi questo lavoro.

A heartfelt thanks to Sameer, always ready to encourage me and to give me precious advices.

A tutta la ciurma della sala tesisti: Chiara, Masche, i due Ste, Fede, Pero, Fuma, Miriam, Pincho e Gio, grazie per aver reso la tesi un'avventura, per aver condiviso risate e consigli, ma soprattutto per l'oroscòpo quotidiano.

Grazie a Marina, donna super impegnata ma sempre presente quando ho avuto bisogno, decisa e sincera nel consigliarmi e sempre pronta per un abbraccio affettuoso.

Grazie a Sere, Teo, Busino, Saretta, Pela e Nina, per l'affetto sincero che mi avete costantemente dimostrato in questi anni.

Un grazie speciale va a tutta la mia famiglia, supporto fondamentale in questi anni e in queste mille avventure in cui mi siete sempre stati vicini.

Alla mia mamma e al mio papà, che non potrò mai ringraziare abbastanza. Grazie per aver reso possibile questo traguardo, che in alcuni momenti è sembrato quasi irraggiungibile. Grazie per avermi sostenuto nelle situazioni più dure, grazie per quando

avrei voluto mollare e invece mi avete incoraggiato. Mi avete insegnto a perseverare e a puntare sempre un po' più in alto, e senza di voi non sarei la donna che sono.

A Ludo, tu che mi fai sempre arrabbiare, ma che sei al mio fianco ogni giorno. Da te ho imparato a guardare avanti, sorridendo delle piccole cose belle. Grazie per avermi dato la mia splendida Bibì.

Infine, una dedica speciale alla mia bimba, che mi sorprende ogni giorno con le sue piccole meraviglie. Vederti muovere i tuoi primi passi, sentirti pronunciare il tuo primo "mamma" è quanto di più bello potessi chiedere.

Contents

List of figures

List of Tables

Abstract

Generation portfolios are rapidly changing, with an increasing share of wind and solar electricity generation capacity. The integration of variable resources, whose output is dependent on the environmental conditions, such as wind and solar, is posing the problem of efficiently managing the heterogeneously distributed energy sources in order to meet the user electricity demand, also known as Demand Side Management (DSM). The difficulty of this task is due both to renewable sources stochasticity and to the fluctuating behaviour of the energy market demand. The necessity to integrate and mitigate these two uncertainty sources, requires on one hand to inject flexibility into conventional power plants, such as nuclear power plants, and on the other hand to tackle the problem of unit commitment.

In this context, the objective of the present thesis work is the development of an accurate method for the short-term prediction of the energy production of wind power plants. To this aim, given the difficulty of resorting to physics-based model, we consider Artificial Intelligence (AI) methods. Given the periodicity of the meteorological conditions, we develop a recursive model based on Recurrent Neural Networks (RNN) trained according to the Reservoir Computing (RC) principles. In particular, we have used Echo State Networks (ESNs) given the reduced computational demand of their training and their ability of catching the system dynamics in the states of the reservoir neurons. To further increase the accuracy of the prediction, we have developed an ensemble of ESN models.

The two main novelties of the present thesis work are: 1) the use of different training sets formed by consecutive meteorological and production data for obtaining diverse models and 2) the aggregation of the individual model outputs using a local technique which considers the individual model performance in the last days before the prediction to efficiently dealing with the evolving environment issue.

The proposed method has been verified using real wind plant data and its performance compared with that of feedforward Artificial Neural Networks

Key-words: Echo State Networks (ESNs), Recurrent Neural Networks (RNNs), Prediction Models Ensemble, Wind power forecasting.

Sommario

I portafogli di generazione sono in rapida mutazione, con una significativa crescita della capacità di produzione di energia eolica e solare. L'integrazione di risorse rinnovabili come l'eolico e il solare, la cui capacità produttiva dipende dalle condizioni ambientali, sta ponendo il problema di gestire in modo efficiente le eterogenee fonti energetiche al fine di soddisfare la domanda di elettricità dell'utenza, approcio noto anche come Demand Side Management (DSM).

La difficoltà di questo compito è dovuta sia alla stocasticità delle fonti rinnovabili sia al comportamento fluttuante della domanda del mercato energetico. La necessità di integrare e mitigare queste due fonti di incertezza richiede da una parte di introdurre maggiore flessibilità nelle centrali elettriche convenzionali, come le centrali nucleari, e dall'altra parte di affrontare il problema della potenza impegnata per utenza.

In questo contesto, l'obiettivo del presente lavoro di tesi è lo sviluppo di un metodo accurato per la previsione a breve termine della produzione di energia degli impianti eolici. A tal fine, considerata la difficoltà di ricorrere ad un modello basato sulla fisica, vengono considerati i metodi di Intelligenza Artificiale (AI). Data la periodicità delle condizioni meteorologiche, sviluppiamo un modello ricorsivo basato su reti neurali ricorrenti (RNN) con un training secondo i principi di Reservoir Computing (RC). In particolare, abbiamo usato le cosidette "Echo State Networks" (ESN) in considerazione della ridotta richiesta computazionale per il training e della capacità di apprendere la dinamica del sistema negli stati dei cosidetti "neuroni riserva". Per aumentare

ulteriormente la precisione della previsione è stato successivamente sviluppato un insieme di modelli ESN.

Le due principali novità del presente lavoro di tesi sono: 1) l'uso di diversi set di training ottenuti da dati meteorologici e di produzione consecutivi per ottenere modelli differenti e 2) l'aggregazione dei singoli output del modello utilizzando una tecnica locale che considera le prestazioni del modello individuale negli ultimi giorni prima della previsione per affrontare in modo efficiente il problema delle condizioni in evoluzione.

Il metodo proposto è stato verificato utilizzando dati reali dell'impianto eolico di riferimento e le sue prestazioni confrontate con quelle delle tradizionali Reti Neurali Artificiali (ANN).

Parole chiave: Echo State Networks (ESN), Recurrent Neural Networks (RNN), Ensemble di modelli predittivi, energia eolica.

Estratto

E.1 Introduzione

La rete elettrica moderna si basa su sorgenti energetiche eterogenee, ossia sull'insieme di fonti energetiche convenzionali e rinnovabili. Negli ultimi decenni, la domanda globale di energia è cresciuta in maniera costante, aumentando del 2.1% solo nel 2017, più del doppio rispetto alla crescita stimata nel 2016. Nel 2017, la domanda ha raggiunto un picco di 14 050 milioni di tonnellate di petrolio equivalente (Mtep). Più del 70% della richiesta è stata soddisfatta dai combustibili fossili, il 25% da fonti rinnovabili e la restante parte dal nucleare [30]. Il contributo dei combustibili fossili alla generazione di energia elettrica nel 2017 è rimasto fisso all'81%, nonostante il largo impiego di fonti energetiche rinnovabili.

Nel 2017, la potenza nucleare ammontava al 10% della potenza globale generata, aumentando del 3% rispetto all'anno precedente [30]. Fino ad oggi, la potenza generata dal nucleare è stata utilizzata per soddisfare il carico elettrico di base, in quanto si basa su una tecnologia caratterizzata da un costo fisso molto elevato [31]. Tuttavia, questa situazione sta cambiando, in quanto i moderni impianti nucleari possono essere progettati per produrre energia in modo tale da adattarsi al carico elettrico. Tale caratteristica permette agli impianti nucleari di soddisfare la domanda del mercato energetico che fluttua continuamente anche a causa del largo impiego di sorgenti di energia rinnovabili, che sono intrinsecamente intermittenti [30-34].

Come affermato poco sopra, sorgenti di energia rinnovabile come il sole o il vento non sono disponibili in modo continuo, in quanto dipendono dalle condizioni metereologiche.

In assenza di un sistema di immagazzinamento dell'energia elettrica, le fonti rinnovabili non sono in grado di produrre energia con continuità. Per le ragioni appena esposte, assicurare l'affidabilità del sistema di distribuzione di energia elettrica diventa una questione di primaria importanza, richiedendo la capacità di prevedere la potenza elettrica prodotta da impianti di produzione di elettricità da fonti rinnovabili [1,28,36-39].

In questo contesto di un sistema elettrico con una crescente quota di energia derivante da fonti rinnovabili, la capacità previsionale gioca un ruolo fondamentale.

Per poter ottenere previsioni accurate della produzione di potenza eolica o solare è necessario affrontare diverse difficoltà. Una volta scelto l'algoritmo predittivo, l'analista deve identificare tra le variabili meteorologiche i segnali significativi da utilizzare come input del modello predittivo; ciò significa affrontare un problema di selezione delle feature, con l'obiettivo di ottimizzare l'accuratezza del modello. Per quanto riguarda un impianto eolico, i segnali significativi potrebbero essere le velocità del vento in diverse direzioni e a diverse altezze dell'impianto fornite da vari provider, la temperatura e la pressione atmosferica.

Dal momento che i dati di partenza sono già delle previsioni meteorologiche, è importante sottolineare come queste saranno intrinsecamente affette da errori, i quali si propagheranno poi attraverso il modello predittivo.

L'obiettivo di questa tesi è quello di istituire un modello affidabile per ottenere previsioni accurate della produzione di energia elettrica da impianti eolici, ma potrebbe essere applicata a un gran numero di scopi simili [28,36-39].

In vista dell'obiettivo descritto, il problema è stato affrontato con un approccio basato su due punti chiave:

- Il modello previsionale adottato è di tipo ricorsivo, dal momento che è addestrato tramite dati meteorologici. Affidarsi a una Rete Neurale Ricorrente (RNN) è sembrata la scelta più appropriata dal momento che il sistema in analisi dipende da fenomeni ciclici, cioè il vento.

  Questa classe di modelli di regressione si è dimostrata una buona soluzione al problema di imparare la dinamica del sistema.
- Modelli ricorrenti diversi sono generati e aggregati per costituire un ensemble, la cui accuratezza e robustezza è migliore rispetto a quella di un modello singolo, come dimostrato da diversi lavori di ricerca [19,20,29,48,49].

La novità di questa tesi rispetto alla generazione dei modelli è l'uso di RNN con diversi pesi iniziali o addestrate con diversi dataset, o ancora, con un insieme dei due approcci; ciò permette di creare diversità tra i modelli. Dall'altro lato, la novità rispetto all'aggregazione dei modelli di base dell'ensemble è l'uso di un metodo di Fusione Globale (GF) e di un metodo di Fusione Locale (LF), in cui i modelli migliori da aggregare vengono scelti sulla base della loro accuratezza.

La struttura proposta è applicata al caso industriale riguardante la previsione della produzione energetica di un impianto eolico reale situato in Italia. I dati meteorologici sono forniti da due diversi provider.

La parte restante di questo lavoro è organizzata come segue. Il Capitolo 2 è dedicato alla definizione del problema. Nel Capitolo 3 è illustrato il metodo proposto mentre il Capitolo 4 mostra i risultati finali dell'applicazione del metodo proposto.

E.2 Definizione del problema

L'obiettivo di questa tesi è la previsione della produzione, P di un impianto industriale con un orizzonte temporale Δt , i.e. al tempo presente t vogliamo prevedere la capacità produttiva dell'impianto $P(t + \Delta t)$. La performance futura di un impianto è una quantità incerta che dipende da diversi fattori, come la configurazione dell'impianto, il livello di degrado dei suoi component e le condizioni ambientali cui l'impianto stesso è sottoposto. Questa tesi si focalizza sugli effetti delle condizioni ambientali sulla performance dell'impianto.

Assumiamo di avere a disposizione un modello, W, che, sulla base dei dati in input, $\vec{\imath}(t)$, collezionati al tempo presente t, predica le condizioni ambientali, $\vec{\hat{E}}(t + \Delta t)$ cui l'impianto sarà sottoposto al tempo $t + \Delta t$ (Figura 1).

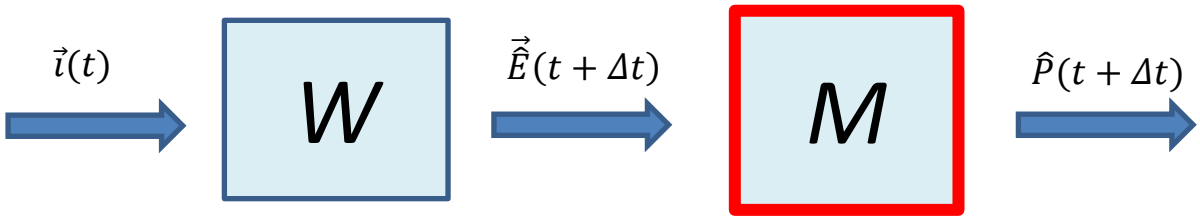

Figura 1: Architettura del problema

Dal momento che il modello W è una rappresentazione imprecisa della realtà, la sua previsione $\vec{\hat{E}}$ è affetta da un errore $\vec{e}$ rispetto alle reali condizioni ambientali, $\vec{E}(t + \Delta t)$, al tempo $t + \Delta t$:

$$\vec{\hat{E}}(t + \Delta t) = \vec{E}(t + \Delta t) + \vec{e} \quad (1)$$

L'obiettivo è quindi quello di costruire un modello M che riceva in input le previsioni delle condizioni ambientali, $\vec{\hat{E}}(t + \Delta t)$, fornite dal modello W, e che fornisca in output la

previsione della performance dell'impianto, $\hat{P}(t + \Delta t)$ (Figura 1 (destra)). Per sviluppare il modello *M* assumiamo di avere a disposizione un dataset contenente le previsioni storiche delle condizioni ambientali e I corrispondenti dati reali delle performance:

$$D = \left\{\vec{\hat{E}}(t + \Delta t), P(t + \Delta t)\right\} \quad (2)$$

E.3 Il metodo

L'analisi predittiva può essere basata su modelli fisici oppure sui dati [1,2]. Nel primo caso si presenta la necessità di disporre di modelli fisici che facciano affidamento su assunzioni e semplificazioni che ne riducono la capacità predittiva rispetto agli approcci basati sui dati [1]. La presente tesi si basa su tecniche di Machine Learning (ML), i quali si basano interamente su modelli statistici che "imparano" andamenti e correlazioni dai dati [2].

Il modello statistico scelto per il nostro scopo è di tipo *dinamico*, in quanto possiede la capacità di cogliere la dinamica del sistema, grazie al fatto che il suo output previsto al tempo t dipende dall'intera storia dell'input. La scelta di un modello *dinamico* è stata fatta sulla base della considerazione che nei dati meteorologici è presente una periodicità, o meglio, una ciclicità.

Le Reti Neurali Ricorrenti (RNN) (rappresentate in Figura 2 (destra)) costituiscono la più diffusa classe di regressori dinamici; esse sono caratterizzate da connessioni cicliche e feedback tra i neuroni, che conferiscono alle RNN la proprietà di memoria, rendendole in grado di gestire la previsione di serie temporali [2].

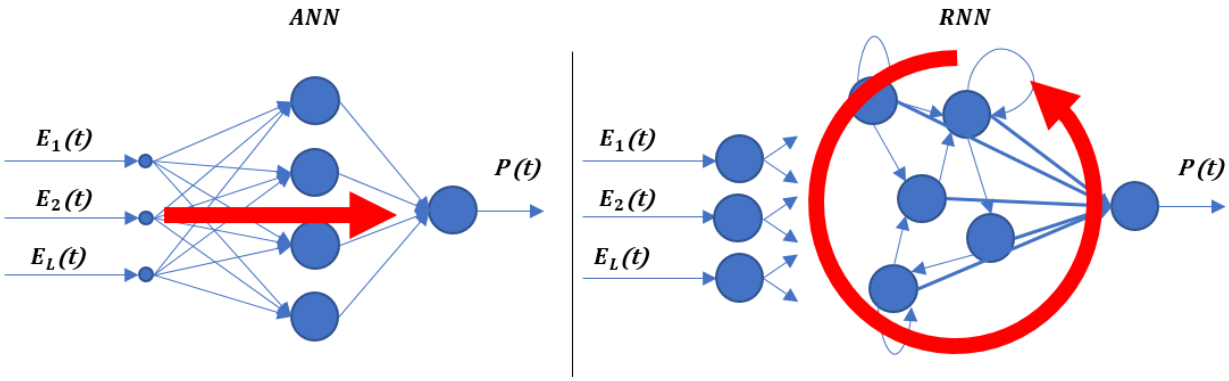

Figura 2: Schema ANN (sinistra) e RNN (destra).

Per il nostro scopo, tra i diversi tipi di RNN disponibili, sono state selezionate le Echo State Networks (ESN) come modello predittivo tra i diversi tipi di RNN disponibili. Una ESN è una RNN addestrata secondo un metodo basato sulla generazione casuale della RNN, detta *reservoir*, la quale rimane invariata durante la fase di addestramento. Dal momento che gli unici pesi della ESN da ottimizzare sono quelli delle connessioni tra gli stati interni della *reservoir*, la sua fase di addestramento risulta più efficiente [26, 27, 37].

E.3.1 – Ensemble di Modelli

Dopo aver creato il singolo modello, i.e. la ESN, abbiamo generato un ensemble di modelli per migliorare l'accuratezza delle previsioni fornite dal singolo modello [2,18,19,20]. L'ensemble è costituito da più modelli di regressione i cui output sono aggregati in un'unica previsione finale. L'idea alla base di questo approccio è quella di creare singoli modelli diversificati, in modo tale da rinforzare le loro caratteristiche positive e da smorzare le loro debolezze [2].

In questa tesi ricorriamo a un ensemble di ESN al fine di accrescere la robustezza e l'accuratezza della previsione.

In quest'ottica, lo sviluppo di un ensemble richiede di *i)* generare singoli modelli diversi (Sezione E.3.1.1) e quindi di *ii)* definire una strategia di aggregazione degli output dei singoli modelli (Sezione E.3.1.2) [27, 37].

E.3.1.1 Diversità dei Modelli

Rispetto a *i)* l'idea fondamentale è quella di generare modelli diversi, caratterizzati da diverse performance e quindi diversi errori di previsione.

La diversità tra modelli può essere ottenuta utilizzando:

a. Modelli addestrati con diversi dataset;
b. Modelli caratterizzati da diverse architetture (in particolare da diversi pesi iniziali);
c. Diverse feature in input al modello, i.e., diverse variabili ambientali.

In questa tesi ricorriamo ad (a) e (b) per generare due diversi ensemble:

- "**Ensemble 1**" è costituito da ESN addestrate con diversi dataset. I dataset di addestramento differiscono sia nel loro punto di inizio (campionato in modo casuale dalla prima metà dell'intero dataset di addestramento), sia nella loro lunghezza (campionata casualmente da una distribuzione uniforme limitata tra metà del dataset di addestramento e l'intero dataset). In questo caso la diversità è generata da (a) e (b).
- "**Ensemble 2**" costituito da ESN addestrate su tutti i dati disponibili per l'addestramento; in questo caso la diversità è generata solo da (b).

E.3.1.2 Aggregazione dei Modelli

Per quanto riguarda lo sviluppo del secondo ensemble, le strategie di aggregazione sono tipicamente classificate in metodi di aggregazione Globale e Locale [1,2,21,22]. Metodi di fusione globali (GF) si basano sull'assunzione che il contributo dei singoli modelli sia lo stesso, qualsiasi sia il pattern in input considerato.

Dall'altro lato, metodi di fusione locali (LF) assumono che il contributo dei singoli modelli al risultato dell'ensemble sia proporzionale alla loro performance locale nell'intorno della sequenza di test analizzata.

La parte restante di questo capitolo è dedicata alla descrizione dei metodi di fusione globale e locale successivamente adottati in questa tesi.

A. Metodi di Fusione Globali (GF)

Assumiamo di avere a disposizione un numero $H = A$ x I di diversi modelli di regressione che ricevono in input un vettore di previsioni meteorologiche, $\vec{E}$, e che forniscono la produzione prevista $\hat{P}^h$, dove $h=1,\ldots,H$. Scegliendo la mediana come metodo di fusione globale, l'output dell'ensemble $\hat{P}^{GF}$, per un generico pattern di test in input, E, si calcola come il valore mediano delle previsioni dei singoli modelli, $\hat{P}^h$:

$$\hat{P}^{GF} = median_{ah=1,\ldots,H}\left(\hat{P}^h\right) \tag{3}$$

Viene scelta proprio la mediana grazie alla sua robustezza rispetto ad altri indicatori statistici, come la media [1].

Una volta che è stata calcolata la mediana, la capacità predittiva è ottenuta come la capacità del "modello mediano" $\hat{P}^{GF}$. La performance predittiva viene calcolata affidandosi alla metrica standard quale l'errore medio assoluto (MAE).

B. Metodi di Fusione Locali (LF)

L'idea alla base dei metodi LF è che, se un modello è accurato nella previsione del risultato nell'intorno della sequenza di test, sarà anche accurato nella previsione della sequenza di test stessa [1, 2, 22].

La strategia adottata in questa tesi per aggregare in modo Locale i singoli modelli può essere riassunta nei seguenti step:

Step 1: Gli I x A modelli disponibili sono testate sui primi s punti di ciascun *j-esimo* periodo del dataset di test ($j=1,\ldots,J$) e quindi per ciascun *j-esimo* periodo vengono scelti gli H modelli migliori. A questo punto, le performance locali degli A x I modelli individuali vengono valutate considerando il risultato del modelli *i,a-esimo*, $\hat{P}_{i,a}^{j}$ ($i=1,\ldots,I$ e $a=1,\ldots,A$), e il corrispondente errore medio assoluto di previsione risulta (MAE):

$$MAE_{i,a}^{j} = \frac{\sum_{i,a=1}^{IxA} \left| \hat{P}_{i,a}^{j} - P_{i,a} \right|}{I \; x \; A} \tag{4}$$

dove $P_{i,a}$ indica il vero output del *i,a-esimo* modello.

Step 2: Una volta scelte le migliori H ESN per ciascun *j-esimo* periodo della sequenza di test, i loro valori output predetti, $\hat{P}_{h}^{j}$ ($h=1,\ldots,H$), vengono aggregati calcolando il valore output mediano:

$$\hat{P}^{LF} = mediana_{h=1,\dots,H}\left(\hat{P}_h^j\right) \quad (5)$$

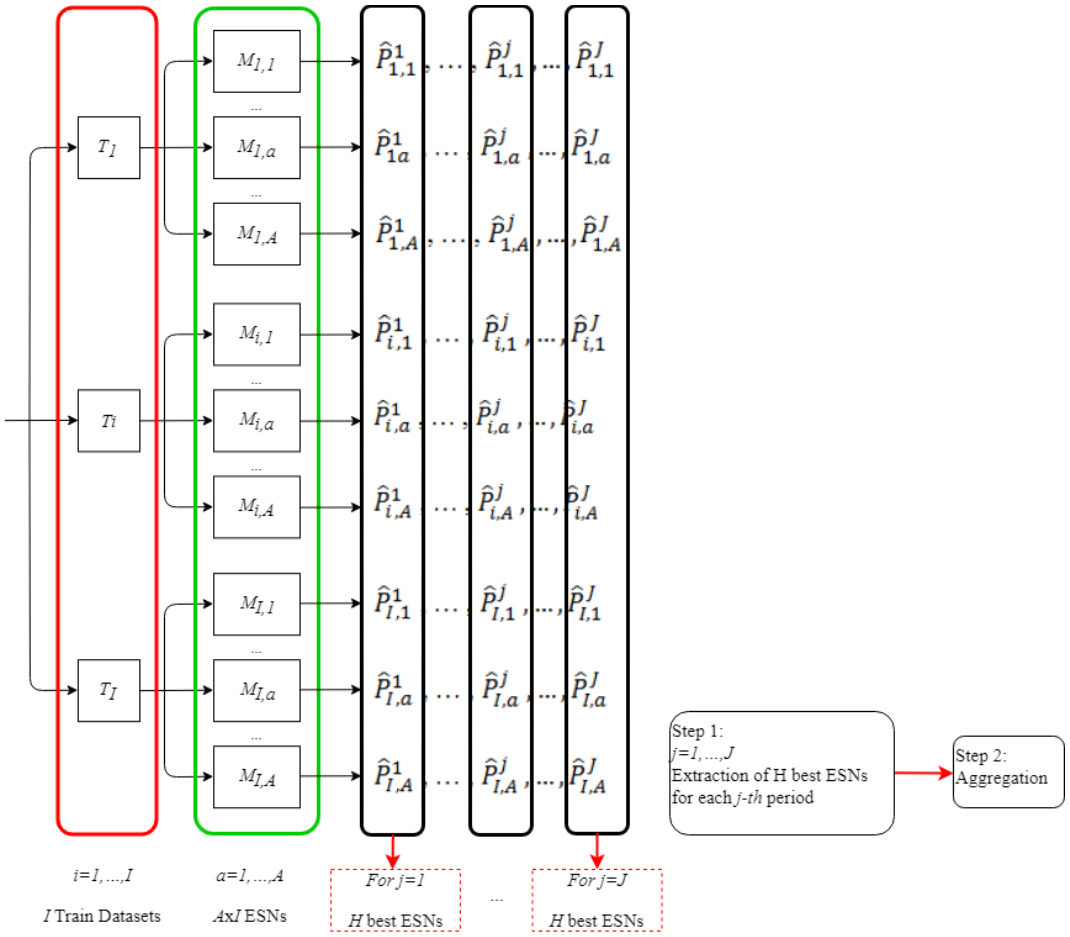

Figura 3: Diagramma di flusso del metodo di Fusione Locale (LF).

E.4. Caso Studio: Previsione della Produzione Energetica di Impianti Eolici

E.4.1 Descrizione del Sistema

In questa sezione viene affrontato il problema della previsione della produzione energetica di un impianto eolico da 34 MW. Questo caso studio si basa su previsioni meteorologiche e I corrispondenti valori di produzione energetica collezionati in un period di tre anni, dal 2011 al 2013 [1]. I dati di previsioni meteorologiche, $\vec{\hat{E}}$, comprendono $F = 19$ features (i.e., numero della settimana, tempo in ore del dato di previsione meteorological, il ritardo della previsione, velocità orizzontale e verticale del vento e la sua direzione a diverse altezze e in diversi punti dell'impianto). L'orizzonte delle previsioni è $\Delta t = 24$ ore, i.e., in un dato istante t_o, è disponibile la previsione meteorologica per le successive $\Delta t = 24$ ore [1].

Il dataset ottenuto dagli anni 2011-2013 è suddiviso come segue:

- I dati del 2011 e 2012 sono utilizzati per sviluppare/addestrare I singoli modelli dell'ensemble e per settare i loro parametri interni;
- I dati del 2013 sono utilizzati per valutare la performance dell'ensemble proposto, in riferimento all'errore medio assoluto (*MAE*).

E.4.2 Settaggio dei Parametri

Una delle maggiori difficoltà nello sviluppo delle ESN è quella di settarne i parametri dell'architettura [2]. Poiché questo compito richiede un certo livello di competenza, le successive due sezioni sono dedicate rispettivamente al settaggio dei parametri dei singoli modelli ESN e dell'ensemble.

E.4.2.1 Modelli Individuali

Per affrontare l'ottimizzazione dell'architettura della singola ESN, abbiamo innanzitutto prioritizzato l'importanza dei parametri [2], come mostrato nella tabella sottostante:

Tabella 1: Prioritizzazione dei parametri dell'architettura della ESN.

Parametri di Primo Ordine (*FOP*)	**Parametri di Secondo Ordine (*SOP*)**
Dimensione della reservoir (*S*) Raggio Spettrale (*SR*) Connettività (*C*)	Input scaling (*IS*) Teacher scaling (*TS*)

Una volta prioritizzati i parametri, è stata utilizzata una ricerca a griglia [50] per misurare l'impatto della variazione dei parametri sulla performance della ESN. La procedura prevede di:

1. Ottimizzare i *FOP* mantenendo i *SOP* fissi a valori scelti dalla letteratura.
2. Fissare i *FOP* ai valori migliori trovati nel punto precedente e ottimizzare i *SOP*.
3. Verificare se si ottengono gli stessi valori per i *FOP* quando i *SOP* sono fissati ai loro valori ottimizzati.

La tabella sottostante riporta i valori dei parametri ottenuti per la ESN migliore:

Tabella 2: Parametri e MAE della ESN migliore risultante dalla ricerca a griglia gerarchica.

Size	Spectral Radius	Connectivity	Input Scaling	Teacher Scaling	MAE
466	0.02	0.13	$10^{-1.44}$	10^{-3}	3.8802

E.4.2.2 – Ensemble

Come descritto precedentemente, lo sviluppo di un ensemble di modelli richiede la generazione di modelli singoli diversi tra loro, caratterizzati da errori di previsione di segno ed entità diversi.

A questo scopo sono state adottate due strategie per generare diversità tra i modelli dei due ensemble proposti: *i)* l'uso di modelli addestrati con diverse sequenze di addestramento; e *ii)* l'uso di modelli caratterizzati da diversi pesi iniziali. La prima strategia è stata utilizzata per costruire il primo ensemble proposto (chiamato Ensemble 1), mentre la seconda è stata utilizzata per costruire il secondo ensemble proposto (chiamato Ensemble 2).

La figura sottostante mostra le previsioni della produzione ottenute dai 500 modelli (ESN) dei due ensemble proposti (linee sottili) (Ensemble 1 a sinistra ed Ensemble 2 a destra), confrontati con le vere produzioni (linee spesse).

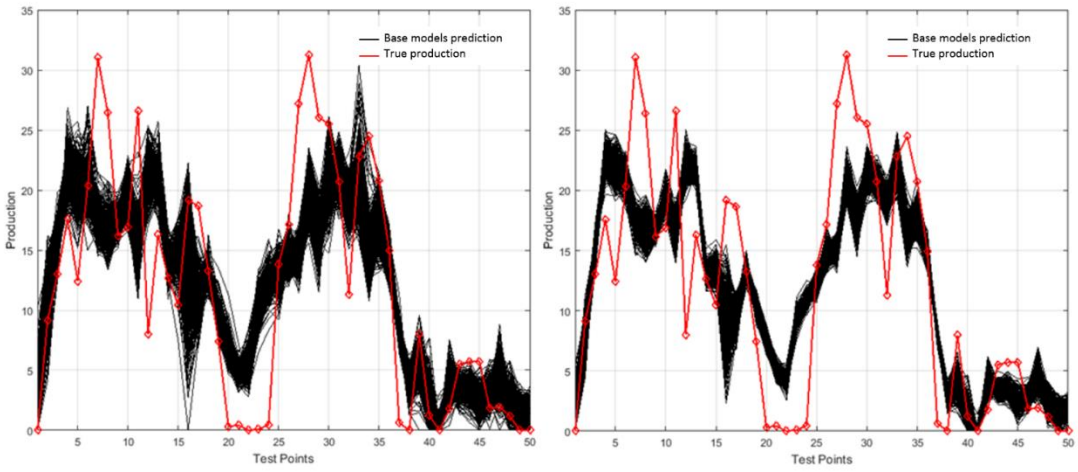

Figura 4: Diversità tra I modelli dell'Ensemble 1 (sinistra) e dell'Ensemble 2 (destra).

Guardando la Figura 4 si notano punti in cui tutte le ESN sottostimano o sovrastimano la produzione energetica reale (per entrambi gli ensemble), e altri punti in cui diversi

modelli ESN commettono diversi errori di previsione. La diversità di cui si è discusso poco sopra si riferisce proprio a questa differenza di errore previsionale tra i diversi modelli.

Dalla stessa Figura 4 si può osservare che i singoli modelli dell'Ensemble 1 mostrano una grande variabilità delle previsioni della produzione (sinistra), mentre i modelli dell'Ensemble 2 (i.e., ugualmente addestrati su tutta la sequenza di addestramento disponibile) (destra) mostrano minore variabilità. Questa maggiore variabilità presente nell'Ensemble 1 può costituire un vantaggio nella riduzione del *MAE* quando i modelli vengono aggregati.

Una volta ottenuti i singoli modelli, questi sono aggregati per ottenere la previsione. Per questo scopo si è utilizzato un metodo di Fusione Locale, che prevede l'aggregazione degli

$H = 50$ modelli migliori tra i 500 generati, scelti sulla base del *MAE*.

E.4.3 – Risultati

E.4.3.1 – Singola ANN e singola ESN

Il primo confronto utile è quello tra la performance di una singola ESN rispetto a una singola ANN, quando entrambe sono addestrate e testate sulle stesse sequenze di addestramento e di test, rispettivamente.

La Figura 5 mostra il *MAE* medio mensile calcolato sulla sequenza di test per la ESN e

per la ANN.

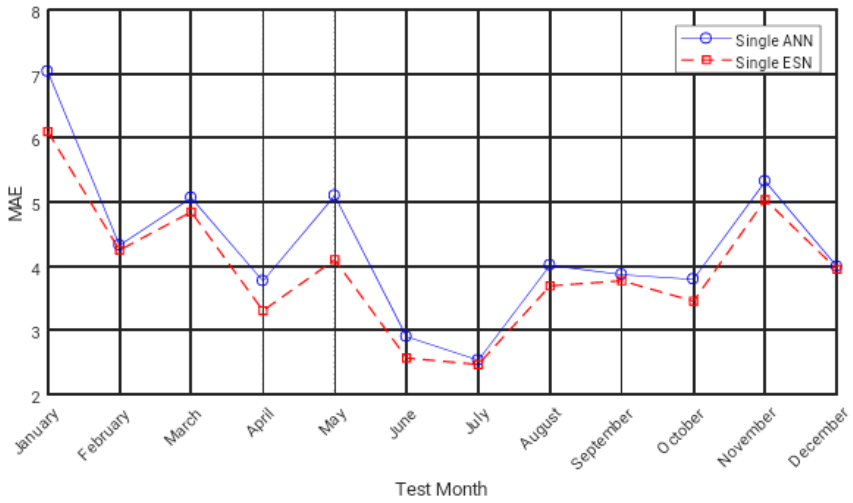

Figura 5: valori di MAE (unità di misura arbitraria) di una singola ANN e di una singola ESN addestrate e testate con gli stessi dataset.

Si può facilmente riconoscere che il trend generale del MAE è piuttosto simile per I due modelli, ma risulta evidente che il MAE è, in media, più basso per la ESN che per la ANN. Questo risultato viene confermato dal calcolo del MAE sull'intera sequenza di test per i due modelli: 4.31 per la ANN e 3.96 per la ESN.

E.4.3.2 – Ensemble di ESN

Nella Tabella 4.5 sono riportati i valori medi di MAE e le loro deviazioni standard ottenuti dal singolo modello ESN e dagli ensemble con Fusione Globale (*GF*) e Fusione Locale (*LF*), composti da 500 e 5000 singoli modelli, addestrati con i due diversi dataset di test.

Tabella 3: valori MAE (unità di misura arbitraria) e deviazioni standard.

	Single ESN	***Ensemble of 500 ESNs***		***Ensemble of 5000 ESNs***	
		GF	***LF***	***GF***	***LF***
Random Train *(Ensemble 1)*	4.1685 ± 0.1082	4.0393 ± 0.0050	3.8878 ± 0.0053	4.0391	3.8098
Same Train (*Ensemble 2*)	3.9584 ± 0.0309	3.9226 ± 0.0010	3.8508 ± 0.0025	3.9225	3.8094

Dalla tabella riportata sopra, si può dedurre che: addestrare il singolo modello ESN sull'intera sequenza di addestramento disponibile costituisce un vantaggio rispetto ad addestrare il singolo modello con un set di punti di inizio e di lunghezze random. Ciò è dovuto al fatto che, utilizzare tutto il dataset di addestramento disponibile, permette alla ESN di essere correttamente addestrata a riconoscere la dinamica del sistema. Inoltre, la deviazione standard del *MAE* più elevata quando viene utilizzata una sequenza random di addestramento, è giustificata dal fatto che i dataset di addestramento sono diversi l'uno dall'altro, portando così a una maggiore diversità tra modelli. Tuttavia, questa differenza nel *MAE* decresce quando viene aggregato un numero di modelli sempre maggiore; è il vantaggio di avere maggiore diversità tra i modelli.

La differenza nella capacità predittiva dei due ensemble si reduce ulteriormente passando dal metodo GF al LF, arrivando ad avere una differenza quasi nulla tra due ensemble Locali costituiti da 5000 modelli singoli.

Tutti i risultati appena mostrati per il metodo LF si riferiscono all'aggregazione dei 50 modelli migliori. Risulta quindi interessante vedere se il numero dei migliori modelli

aggregate nell'ensemble Locale può influenzare il MAE finale; per questo motivo, in Figura 6 è rappresentato l'andamento del MAE dell'ensemble Locale in funzione del numero dei modelli migliori aggregati, nel caso dell'Ensemble 1 e dell'Ensemble 2 (entrambi con 500 e 5000 ESN):

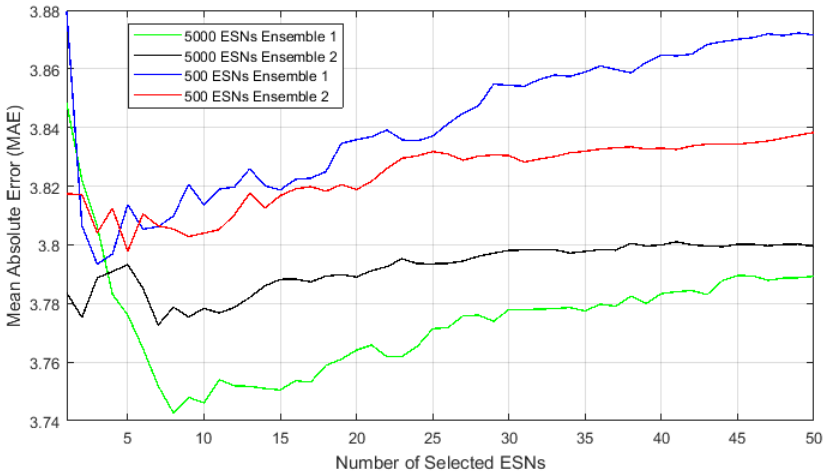

Figura 6: MAE (unità di misura arbitraria) in funzione del numero selezionato di ESNs nel metodo LF, per l'Ensemble 1 e l'Ensemble 2 (entrambi con 500 e 5000 modelli ESN).

Dalla Figura 6 si deduce che aggregare le 50 migliori reti singole non è la scelta migliore per minimizzare il *MAE*. In Tabella 4.6, sono mostrati i numeri migliori di ESN e i corrispondenti *MAE* per gli Ensemble 1 and Ensemble 2 (con 500 e 5000 ESN):

Tabella 4: Combinazione migliore di numero di ESN selezionate e corrispondenti MAE (unità di misura arbitraria) per Ensemble 1 ed Ensemble 2 (composti da 500 e 5000 modelli).

		Numero di ESN selezionate	**MAE**
Ensemble 1 (Train Random)	***LF* 500 ESN**	3	3.793
	***LF* 5000 ESN**	7	3.743
Ensemble 2 (Train uguale)	***LF* 500 ESN**	5	3.798
	***LF* 5000 ESN**	8	3.773

È evidente dalla Tabella 4 che i migliori valori di performance si ottengono utilizzando un ensemble di 5000 reti piuttosto che di 500, sia per l'Ensemble 1 che per l'Ensemble 2. Il valore di *MAE* più basso, pari a 3.74, si ottiene dall'Ensemble 1, generando 5000 ESN e aggregando le migliori 8; ciò indica che la diversità generata da 5000 ESN addestrate in modo random ha un impatto positivo sulla performance quando il numero di modelli singoli viene fatto variare.

1. Introduction

The modern electricity grid relies on heterogeneous energy sources; it is based on a mixture of conventional and renewable energy sources. Over the last decades, the global energy demand has been constantly increasing, raising by 2.1% only in 2017, more than twice the estimated 2016 growth rate [30]. In 2017, global energy demand reached an estimated peak of 14 050 million tonnes of oil equivalent (Mtoe). More than 70% of the rise was met by fossil fuels, 25% by renewables and the remaining part by nuclear [30]. The contribution of fossil fuels in global energy demand in 2017 remained fixed at 81%, a level that has remained stable over the last decades, despite the strong deployment of renewable energy sources.

In 2017, nuclear power accounted for the 10% of global power generation (Figure 1.1), growing by 3% relatively to 2016 [30]. Until today, nuclear power has been mainly used as a base-load source of electricity, since nuclear power plants rely on a technology characterized by a high fixed cost [20]. However, the situation is currently changing, since modern nuclear power plants can be designed to work also as load-following energy source; this characteristic would allow them to match the fluctuating energy market demand [13,20,44,46].

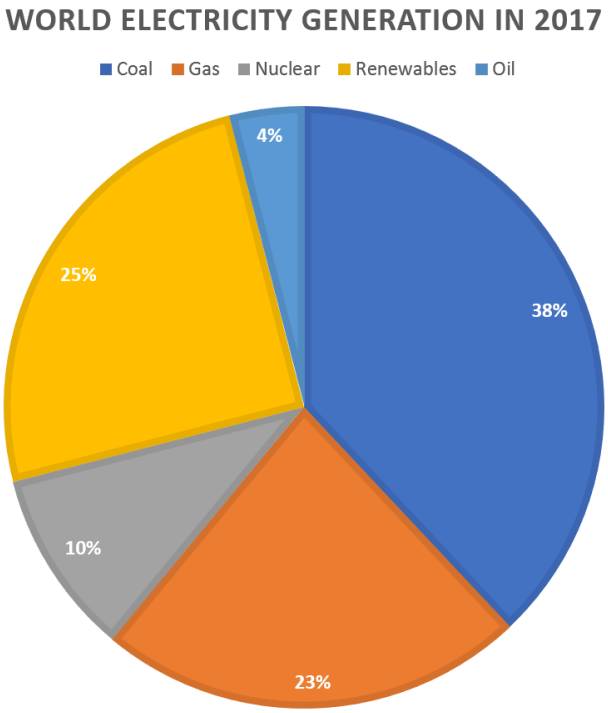

Figure 1.1 World electricity generation in 2017 divided by source. Renewables here include hydroelectric.

This capability of nuclear plants to work as load-following energy sources is becoming more and more important due to the large-scale employment of renewable energy sources, which are intrinsically intermittent [44].

As already stated, renewable energy sources such as wind and solar, are not continuously available, since they depend on weather conditions; their behaviour is characterized by an intrinsic stochasticity.

The issue of efficiently managing the heterogeneously distributed energy sources in order to meet the user electricity demand, is known as Demand Side Management (DSM). The difficulty of this task is due both to renewable sources stocasticity and to the fluctuating behaviour of energy market demand. The necessity to integrate and mitigate these two uncertainty sources, requires on one hand to inject flexibilty into conventional power plants, and on the other hand to tackle the problem of unit commitment [1,14,29,41].

For these reason, ensuring the reliability of electric power distribution becomes a fundamental and complex issue, requiring the capability of predicting electrical power output from plants with intermittent energy sources.

In this context, the present thesis work is dedicated to the problem of short-term forecasting of the energy production from renewable power plants.

It is important to underline that when dealing with production prediction from renewable plants, the input data on which the prediction model is based are mainly constituted by weather forecasting and environmental conditions. An example is constituted by the model used to predict the energy output from a wind plant: input data could be the wind speed, pressure, temperature and other environmental data.

In order to obtain accurate energy production predictions the following issues should be taken into account:

1) The selection of weather forecast signals those to be used as input of the model.
2) The large uncertainty on the weather forecasting data used as input of the prediction model.
3) The choice of the appropriate approach for the development of the predictive model.

Issue 1) is tipically tackled resorting to feature selection techniques which aim at identifying the best subset of the measured signals for the model prediction. Example of feature selection approaches can be found in [49,50]. With respect to issue 2), a method based on an ensemble of Artificial Neural Networks (ANNs) has been proposed in [23,37,38] to obtain robust estimation in case of large input uncertainty and to estimate the consequent output uncertainty. In this work we focus on the development of the

predictive approach. With respect to issue 3), the prediction approaches are typically divided into model-based and data-driven [5,14,38,43]. The deployment of model-based approaches has been limited by the necessity to have physics-based models which rely on assumptions which in turn lead to a performance reduction [38,43]. On the other hand, data-driven approaches rely only on historical data and do not need a physics-based model. Data-driven methods for predicting the production from renewable plants such as Artificial Neural Networks (ANNs) [35,38,43], Support Vector Machines (SVM) [21,31] and K-nearest neighbours (K-nn) regression [29] have been extensively explored during the last years.

In order to reach this goal, the problem has been tackled from two different angles:

- The selected prediction model is recursive, since it is fed with weather data. Resorting to a Recurrent Neural Network (RNN) seems to be an appropriate choice since the analysed system depends on a cyclical phenomenon, i.e. the wind.

RNNs have been widely used in the last years research work, not only to obtain wind power forecasting [1,41,43], but also for different prognostics tasks, for example the estimation of the remaining useful life (RUL) of industrial components [2,9,6,19,32, 42]. This type of regression models have been demonstrated to be a good solution to the problem of learning the system dynamic; since RNNs internal states are characterized by cyclic connections and feedbacks among neurons, they are capable of encapsulating into their neurons a nonlinear transformation of the input history [6,9,17,19,25,27,32,42].

- Different recursive models are generated and aggregated to build an ensemble, whose accuracy and robustness is improved with respect to that of a single model, as demonstrated by many research works [36,39,7,23,3].

The novelty of the present work with respect to the generation of the ensemble base models is the use of RNNs with different initial weights or trained with diverse training datasets, or a mixture of the two approaches; this allows to create diversity among models. On the other hand, the novelty with respect to the aggregation of the ensemble base models is the use of a Median Global Fusion (GF) method and of a Local Fusion method (LF), in which the best models to be aggregated are chosen with respect to their accuracy.

The proposed framework is applied to the industrial case concerning the energy production prediction of a real wind power plant, located in Italy. Weather data are taken from two weather forecast providers.

The remainder of this work is organized as follows. In chapter 2 the problem statement is discussed. In chapter 3 the proposed method is illustrated. Chapter 4 shows the final results of the application of the proposed method.

2. Problem Statement

The objective of the present thesis work is the prediction of the performance, P of an industrial power plant with a time horizon Δt , i.e. at the present time t we want to predict the plant performance $P(t + \Delta t)$. The future performance of a plant is an uncertain quantity which depends on several factors, such as the plant configuration, the degradation of its components and the environmental conditions experienced by the plant. In this thesis work, we focus on the effects of the environmental conditions on plant performance. We assume to have available a model, W, which, on the basis of input data, $\vec{\imath}(t)$, collected at the present time t, predicts the environmental conditions, $\vec{\hat{E}}(t + \Delta t)$ which the plant will experience at time $t + \Delta t$ (Figure 2.1).

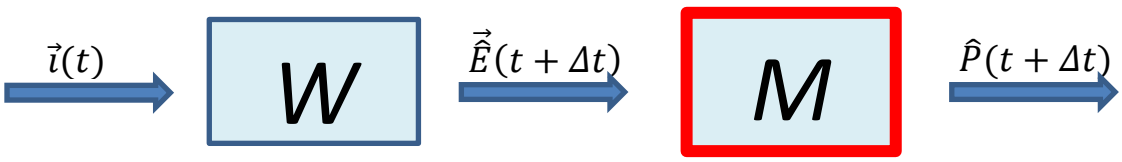

Figure 2.1 Basic architecture of the problem.

Since the model W is an imprecise representation of the reality, its predictions $\vec{\hat{E}}$ is affected by an error $\vec{e}$ with respect to the true environmental conditions, $\vec{E}(t + \Delta t)$, at time $t + \Delta t$:

$$\vec{\hat{E}}(t + \Delta t) = \vec{E}(t + \Delta t) + \vec{e} \qquad (2.1)$$

The objective is, then, to build a model M that receives in input the predicted environmental conditions, $\vec{\hat{E}}(t + \Delta t)$, provided by model W, and provides in output the predictions of the plant performance, $\hat{P}(t + \Delta t)$ (Figure 2.1 (right)). To develop the model M we assume to have available a dataset containing the historical predictions of the environmental conditions and the corresponding real (ground truth) plant performance data:

$$D = \left\{\vec{\hat{E}}(t + \Delta t), P(t + \Delta t)\right\} \tag{2.2}$$

Notice that the use of environmental conditions prediction data to feed empirical models is common in various industrial fields. Some examples of application are:

- the prediction of energy production from renewable sources, which is based on the prediction of weather conditions [1];
- the prediction of walk-in attendance in hospital emergency rooms. This quantity depends from environmental factors and their accurate prediction is of paramount importance for improving service and reducing costs [15];
- prediction of customer demands for products and services [12].

3. Method

Prediction models can be broadly categorized into model-based and data-driven [38,21]. The spread of model-based approaches has been limited by the necessity to rely on assumptions, simplifications and approximations that lead to a lower predictability with respect to data-driven approaches, which, instead, do not require any physical model [38]. For this purpose, the present work is here based on Machine Learning (ML) techniques, which entirely rely on the availability of data.

Data-driven approaches constitute a powerful tool to obtain accurate predictions, especially when a large amount of historical data has been collected. These methods are based on statistical models that "learn" trends from the data [21]. In this regard, a number of different data-driven regression models have been widely developed and applied with success to produce energy production predictions [37,38,21].

Furthermore, these models can be either *static* or *dynamic*. A model is said to be *static* if the predicted output at time t depends solely on the input at the same time t but not on the previous time steps; Artificial Neural Networks (ANNs) (depicted in Figure 3.1 (left)) and the linear regression are just two examples of such models. On the other hand, the main feature of *dynamic* models is their capability to catch the system dynamic, thanks to the fact that their predicted output at time t depends on the whole input history.

For what concerns our goal, it is true that once the input weather data are fixed, they should correspond to a unique output energy production value, thus leading to the choice of a *static* model to make predictions. However, since in weather data there is certainly to be a cyclicality or, at least, a periodicity; a *dynamic* regression model with intrinsic

memory property would be more suitable for this task.

Recurrent Neural Networks (RNNs) constitute the most prominent class of dynamic regressors. Since RNN (depicted in Figure 3.1 (right)) internal states are characterized by cyclic connections and feedbacks among neurons, they are capable of encapsulating into their neurons a nonlinear transformation of the input history; this provides memory properties to RNNs, enabling them to handle sequential tasks, such as time series prediction [21].

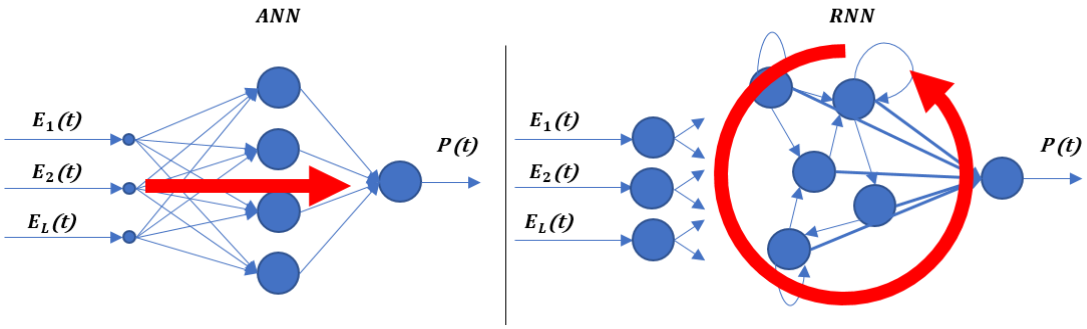

Figure 3.1 ANN (left) and RNN (right) schemes.

In Figure 3.1, the difference between ANNs and ESNs is made clear: while in ANNs the connections go only "from left to right", with no connection cycles, in RNN, the neurons are interconnected and there is at least one connection cycle [32].

The main drawback related to the application of RNNs in various industrial applications is due to the difficulty of optimizing their internal parameters that strongly affect the model performance. In the present thesis work, to overcome the optimization problem, the Reservoir Computing (RC) [25] method is adopted as a computational framework.

The RC paradigm (whose scheme is plotted in Figure 3.2) avoids the "weakness" of gradient-descent RNN training (Figure 3.3 (left)) by randomly creating the RNN (called

the Reservoir) [25]; the reservoir expands the input signals through a non-linear temporal expansion function, and then the readout combine the neuron signals into the desired output target [32]. Since reservoir and readout serve to different purposes, they can be trained separately. The detailed characteristics of RC will be discussed in Section 3.1.

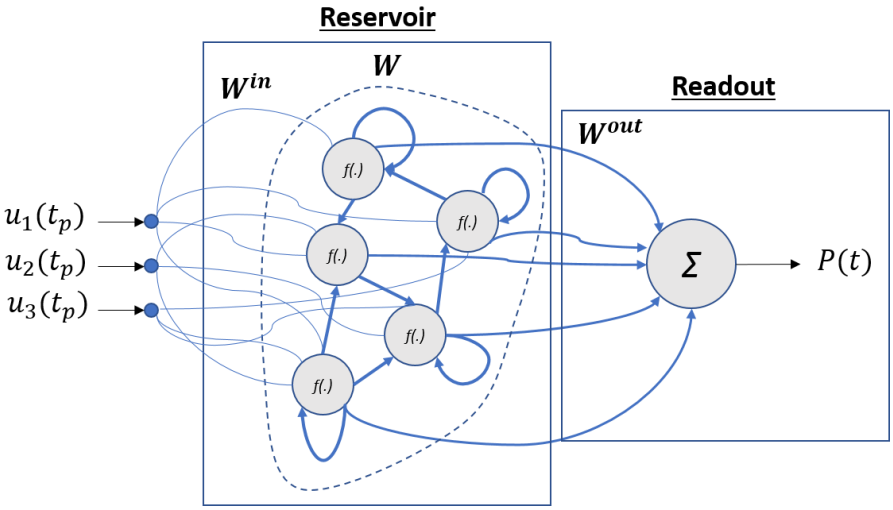

Figure 3.2 Reservoir Computing (RC) scheme.

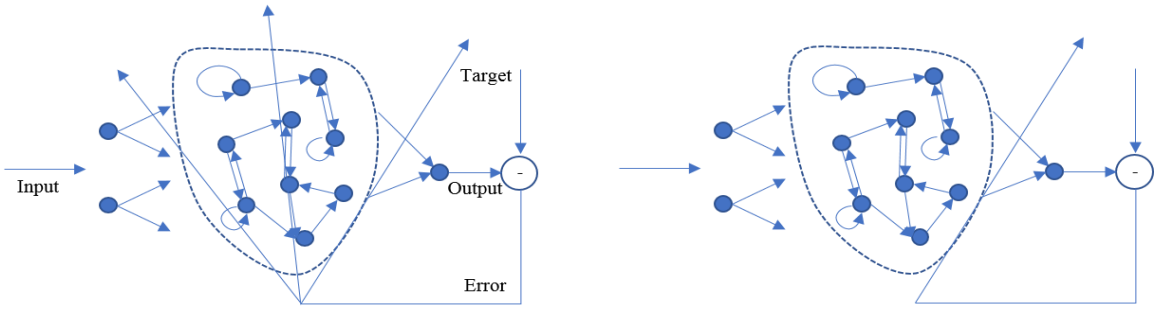

Figure 3.3 : Traditional gradient-descent RNN training adapts all the connection weights (bold arrows) (left). In RC, only RNN-to-output weights are adapted (right) [25].

Liquid State Machines (LSMs) [25], Long Short-Term Memory (LSTM) [40] and Echo State Networks (ESNs) can be mentioned as examples of RNNs.

Echo State Networks (ESNs) are here selected among the different types of RNNs available. An ESN is a RNN trained using a RC method based on the random generation of a RNN that is the reservoir, which remains unchanged during the training phase and is passively excited by the input patterns. Since the only weights of the ESN to be optimized are those of the connections among the reservoir internal states and the output, its training phase is computationally more efficient [21, 25].

3.1 – Echo State Networks (ESNs)

The model considered in this thesis work for the prediction task is the Echo State Network (ESN) due to the following reasons:

- its ability to model the temporal behaviour of dynamic complex systems, and
- its ability to preserve information about the input time history. This is obtained by using feedback connections between the neurons of a layer and those of the preceding layers [2,27]. Thanks to this property, ESNs have intrinsic memory properties, i.e., the system output depends on the observed input time history [2,16], which is a key characteristic when making predictions of weather-dependent quantities, like the electricity energy production in renewable sources plants.

ESNs are a relatively new type of RNNs [16]. The difference from the traditional RNNs lies in the conceptual separation between the reservoir, a randomly created RNN used as nonlinear temporal expansion function, and a linear recurrence-free readout for synthesizing the expansion and producing the desired output (Figure 3.4); notice that this

latter is the only part of the ESN to be trained, which brings significant computational savings [25]. The critical part of the ESN is the reservoir, which is suggested to be generated of large dimension N but with sparse connections among the internal neurons in order to produce a rich set of dynamics [27,16,17] .

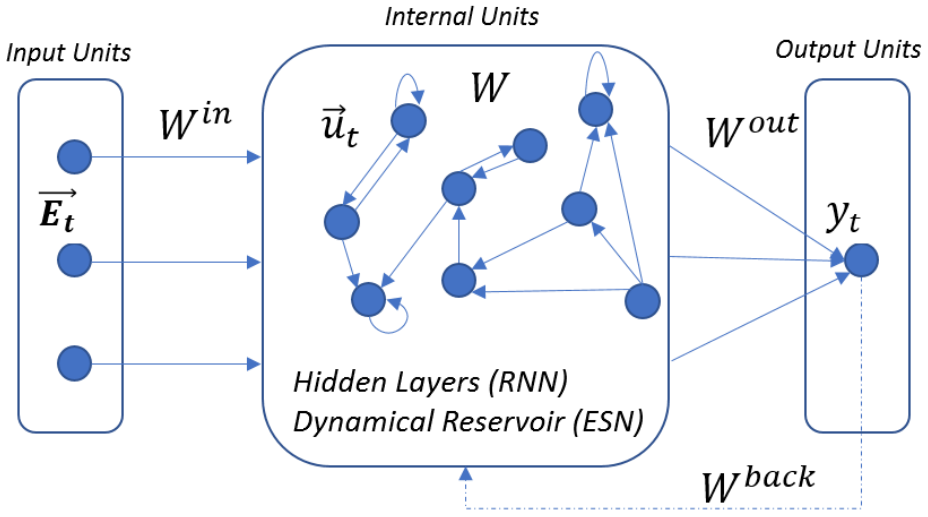

Figure 3.4 Architecture of the ESN.

In this work, we consider an ESN with L input units receiving at time t the weather prediction data, $\vec{E}_t$:

$$\vec{E}_t = [E_t^1 \ E_t^2 \ \dots E_t^L] \tag{3.1}$$

The reservoir is characterized by N internal network units whose internal states are represented by the vector $u_t = [u_t^1 \ u_t^2 \ \dots u_t^N]$ and one output unit producing the output signals, $y_t = P_t$. The activation of internal units u_t at time t is obtained using:

$$\vec{u}_t = \underline{f}\left(W^{in}\vec{E}_t + W\vec{u}_{t-1} + W^{(back)}\, y_{t-1}\right) \tag{3.2}$$

where $\underline{f} = (f_1, \dots, f_N)$ are the internal units activation functions, which are typically sigmoidal, $W^{in} = (w_{ij}^{in})$ is the NxL input weights matrix, $W = (w_{ij})$ is the NxN internal weights matrix, and $W^{back} = (w_{ij}^{back})$ is the NxP output feedback weights matrix, where P represents the output channels (in this case $P = 1$, i.e., the predicted production). The input weights, W^{in}, and the output feedback weights, W^{back}, are randomly generated from a uniform distribution. In order to deal with a specific task, both W^{in} and W^{back} can be scaled: the scaling of W^{in} (IS) and shifting of the input (IF) depend on how much nonlinearity of the processing unit the task needs. If the inputs are close to 0, the sigmoidal neurons tend to operate with activations close to 0, where they are essentially linear, whereas inputs far from 0 tend to drive them more toward saturation, where they exhibit more nonlinearity.

The same idea drives the choice of the output scaling (OS) and shifting (OF), whose values affect the range of the trained W^{out} and might lead to an unstable condition. Finally, the scaling of W^{back} (OFB) is, in practice, limited by a threshold at which the ESN starts to exhibit an unstable behavior, i.e., the output feedback loop starts to amplify the output entering into a diverging generative mode [18] . The output provided by the ESN is given by:

$$y_t = \underline{f_{out}}\left(W^{out}\left(\overrightarrow{E_t}, \overrightarrow{u_t}, y_{t-1}\right)\right) \qquad (3.3)$$

where $\underline{f_{out}} = (f_{out}^1 \dots f_{out}^P)$ are the output units activation functions, which are typically linear, and $W^{out} = (w_{ij}^{out})$ is the $P\times(L+N+P)$ output weights matrix. The ESN training aims at finding optimal values for W^{out} and is performed through a Least Squares linear regression step to minimize the error between the network output and a target signal on a

set of training data. Once the ESN has been trained, it can be used to predict the output, $y_t = P_t$, applying Eqs. (3.2) and (3.3) to the input $\vec{E}_t$ [27].

The ESNs are characterized by the echo state property [16], which states that the effect of a previous state $\vec{u}_t$ and a previous input $\vec{E}_t$ on a future state $\vec{u}_{t+j}$ should vanish gradually as time passes, and not persist or even get amplified. This property is practically assured if the reservoir weight matrix W is scaled so that its spectral radius (SR) $\rho(W)$ (i.e., the largest absolute eigenvalue of W) satisfies $\rho(W) < 1$.

3.2 – Ensemble of Models

Ensembles of models have been used in many different research fields in order to improve the prediction accuracy of the single model [27,34,36]. The ensemble is constituted by multiple regression models (hereafter called base models) whose outcomes are to be aggregated into a final aggregated prediction outcome. The basic idea is to diversify on the base models to leverage their strengths and overcoming their drawbacks [27]. The aggregation of the outcomes provided by the base models of the ensemble has been shown to be more accurate in making predictions compared to each sole base model [27,33,38,36,39]. In the present thesis work, we resort to an ESNs ensemble in order to increase the prediction robustness and accuracy.

Thus, the development of an ensemble requires *i)* to generate diverse base models (Section 3.2.1) and, then, to *ii)* define a strategy for the aggregation of their outcomes (Section 3.2.2) [27,38].

3.2.1 Models Diversity

With respect to the first requirement, the fundamental idea is to generate diverse models, characterized by different performances, i.e., different prediction errors.

Diversity among regression models is generally obtained by using:

d. Models trained with different training datasets;
e. Models characterized by different initial weights;
f. Different models' input features, i.e., different weather variables.

The selected strategy for the present work derives from (a) and (b), and it allows us to create two different ESNs ensembles:

- "**Ensemble 1**" is constituted by ESNs trained with different training datasets. The training datasets are different both *i)* in their starting point (i.e., the starting point is randomly sampled from the first half of the available training dataset) and for *ii)* their length (i.e., the length of the training dataset is randomly sampled from a uniform distribution constrained between half of the available training data and all the available training data). In this case, the diversity is achieved due to both (a) and (b).

- "**Ensemble 2**" is constituted by ESNs trained using all the available training data; in this case the models' diversity is due solely to (b).

For both the ensembles, the training datasets are constituted by consecutive data points. The possibility to use techniques such as bagging [22,38] or boosting to create different

training sets is here excluded since we are dealing with time-series forecasting, and thus, the training dataset needs to contain contiguous data.

Following the described strategy, let us assume to have a number I of different training datasets, $T_i, i = 1, \dots, I$, each one is used to train a number A of ESNs (ESNs are here labelled as $M_{i,a}, a = 1, \dots, A$), thus originating a number AxI of different ESNs (as depicted in Figure 3.5):

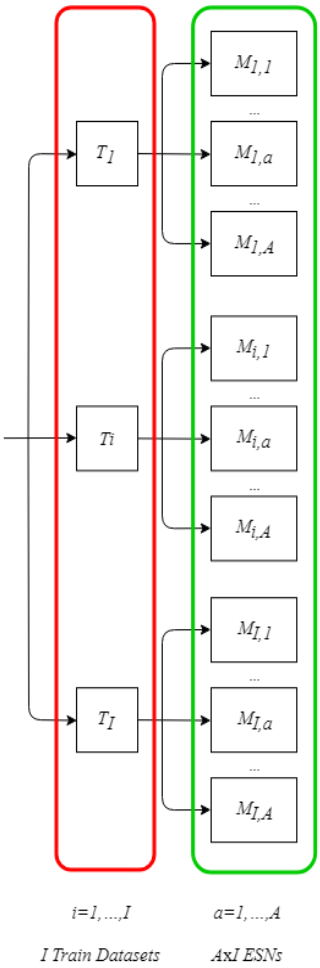

Figure 3.5 Flowchart of generating diverse ESN models.

3.2.2 Models Aggregation

With respect to the second ensemble development requirements, aggregation strategies are typically classified into Global and Local Fusion methods [38,27,11,37]. Global Fusion (GF) methods, such as median and Globally Weighted Average (GWA) are based on the assumption that the contribution of the base models is the same, regardless of the input pattern under study. On the other hand, Local Fusion (LF) methods assume that the contribution of the base models to the ensemble outcome is proportional to their local performance in the neighbourhood of the test pattern under analysis. LF methods, also called dynamic weighting methods, have been successfully applied in various industrial applications such as efficiency and load predictions in power plants [1,11] and remaining useful life predictions of industrial components [27,37].

The remaining of this chapter is dedicated to the description of the GF and LF methods adopted later on in this work.

A. Global Fusion (GF) method

Let us assume to have available $H = A$ x I different regression models which receive in input the vector of weather prediction data, $\vec{E}$, and provide the predicted output $\hat{P}^h$, where $h=1,\ldots,H$. Since we choose the median as GF method, the ensemble outcome $\hat{P}^{GF}$, for a generic test input pattern, E, is computed as the median of the base models predictions, $\hat{P}^h$:

$$\hat{P}^{GF} = median_{h=1,\ldots,H}\left(\hat{P}^h\right) \tag{3.4}$$

The median of the base models predictions is selected because of its increased robustness with respect to other statistical indicators, such as the mean [38].

Once the median has been calculated, the prediction performance is computed as the performance of the "median model" $\hat{P}^{GF}$. The prediction performance is calculated by resorting to standard performance metrics such as the Mean Absolute Error (MAE), as we shall see in the following Chapter.

B. Local Fusion (LF) method

The basic idea of the LF methods is that if a model is accurate in the prediction of the outcomes of the neighbours of the test pattern, it is also expected to be accurate in the prediction of the test pattern [27,37,38].

In this work, the strategy adopted for the individual models aggregation can be summarized in the following steps:

Step 3: Test the *A* x *I* available models on the first *s* points of each *j-th* period of the test dataset (*j=1,…,J*) and then choose for each *j-th* period the best *H* models. In this step, the local performances of the *A* x *I* individual models are evaluated considering the outcome of the *i,a-th* model, $\hat{P}_{i,a}^{j}$ (*i=1,…,I* and *a=1,…,A*), and the corresponding prediction Mean Absolute Error (MAE):

$$MAE_{i,a}^{j} = \frac{\sum_{i,a=1}^{IxA} \left| \hat{P}_{i,a}^{j} - P_{i,a} \right|}{I \; x \; A} \qquad (3.5)$$

where $P_{i,a}$ indicates the true output of the *i,a-th* model.

Step 4: Once the best *H* ESNs for each *j-th* period of the test set have been chosen, their predicted output values $\hat{P}_{h}^{j}$ (*h=1,…,H*) are aggregated by calculating the median output values:

$$\hat{P}^{LF} = median_{h=1,\dots,H}\left(\hat{P}_{h}^{j}\right) \qquad (3.6)$$

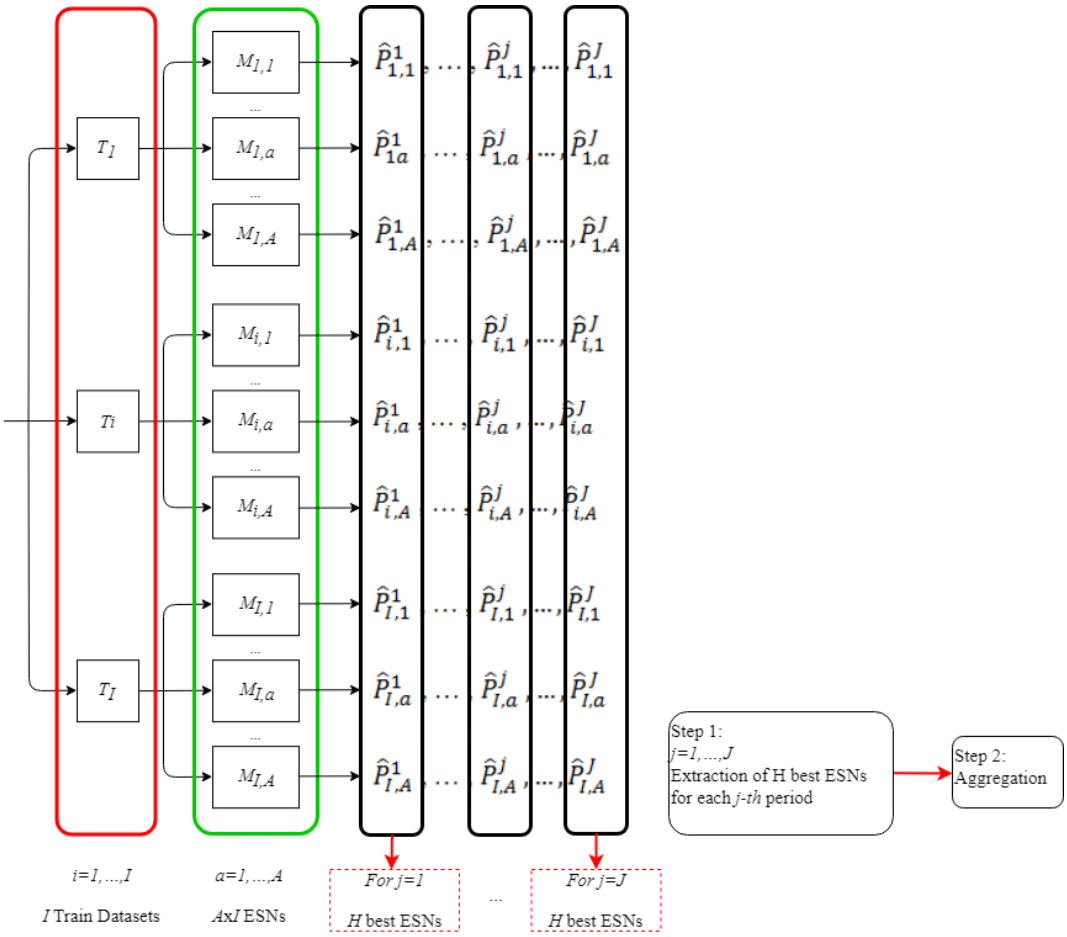

Figure 3.6 Flowchart of the Local Fusion (LF) method.

4. Case Study: Energy Production Prediction in Wind Plants

4.1 System Description

In this Section, the problem of predicting the energy production of a real wind plant is addressed. The present case study is based on weather forecasting and the corresponding energy production values collected in a period of three years from 2011 to 2013 [38]. The weather forecasting data, $\vec{\vec{E}}$, include $F = 19$ features (i.e., week number, time in hours of weather data forecast, the delay of the prediction, horizontal and vertical wind speed and direction at different heights and locations of the plant). The forecast horizon is $\Delta t = 24$ hours, i.e., at a given time t_o, the weather forecasts for the following $\Delta t = 24$ hours being available [38].

For clarification purposes, Figure 4.1 shows the one-day ahead wind speed forecast (top) and the associated energy productions (bottom) of two different days in winter (circles) and summer (squares) conditions. It is interesting to notice the large variability of wind speed in different seasons and at different hours of the day [38].

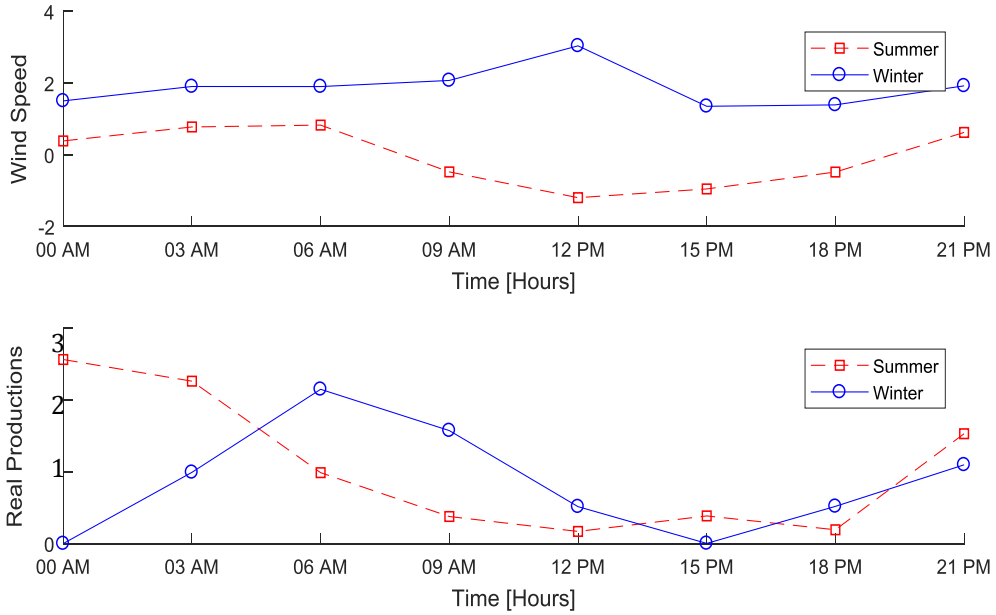

Figure 4.1 One day-ahead wind speed forecasts (top) and associated energy productions (bottom), in winter (circles) and summer (squares) conditions (arbitrary units of measure).

The 2011-2013 dataset is divided as follows:

- Data from 2011 and 2012 are used to develop/train the base models of the proposed ensembles and setting their internal parameters;
- Data from 2013 are used to evaluate the performance of the proposed ensembles with respect to the well-known performance metric, the Mean Absolute Error (*MAE*).

4.2 Model Parameters Setting

One major difficulty when developing an ESN model is the setting of the architecture parameters [27]; this task requires a certain level of expertise. Therefore, Section 4.2.1 and Section 4.2.2 are devoted to the setting of the internal parameters of the ESNs base models and of the ensembles, respectively.

4.2.1 Individual Models

To overcome the optimization of the individual ESN architecture, we first prioritize the importance of parameters, as shown in Table 4.1 [27]:

Table 4.1 Prioritization of ESN architecture parameters.

First Order Parameters (*FOPs*)	**Second Order Parameters (*SOPs*)**
Size of the reservoir (*S*) Spectral radius (*SR*) Connectivity (*C*)	Input scaling (*IS*) Teacher scaling (*TS*)

Once the parameters have been prioritized, a hierarchical grid search approach [10] is used to measure the impact of parameters variation on the ESN performance (Figure 4.2). The procedure is summarized in the following three steps:

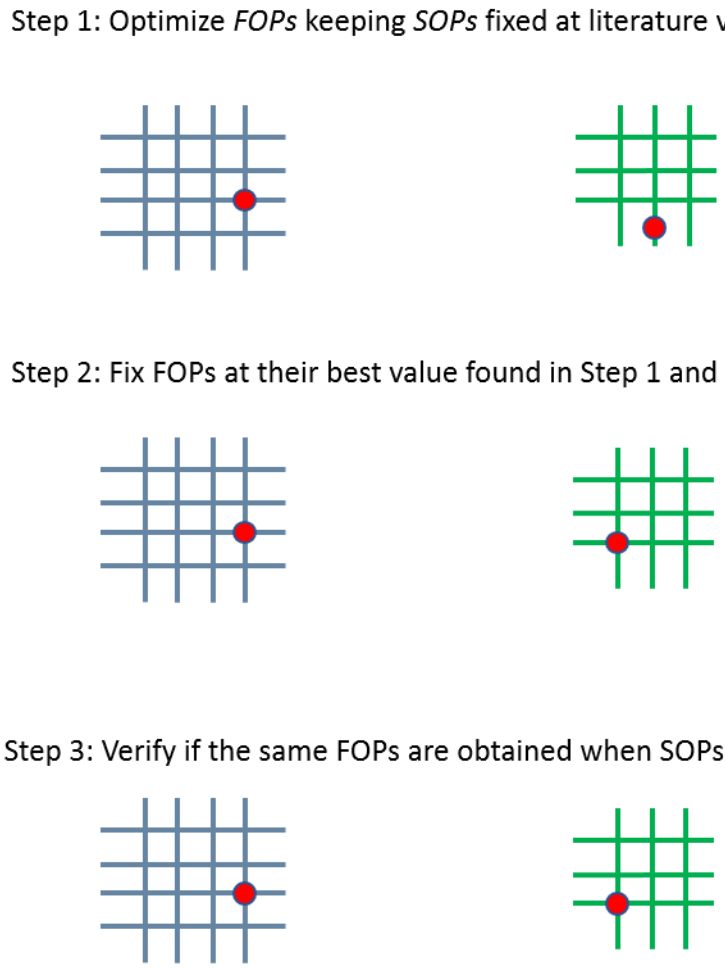

Figure 4.2 Hierarchical grid search Step 1 (top), Step 2 (middle), and Step 3 (bottom).

As performance indicators to be minimized, in this work we consider the Mean Absolute Error (*MAE*):

$$MAE = \frac{\sum_{l=1}^{N_{test}} \left| \hat{P}_l - P_l \right|}{N_{test}}, \qquad (4.1)$$

where N_{test} are the test patterns forming the test dataset, and $\hat{P}_l$ and P_l are the predicted and the real energy production values, respectively.

Figure 4.3 shows the evolution of the *MAE* values with respect to the *SR* (top) and to the *IS* (bottom). In Figure 4.3 (top) one can easily observe the general increasing trend of the *MAE* with the increasing of the *SR*, even though the lowest *MAE* is reached in

correspondence of *SR* = 0.02. In Figure 4.3 (bottom) there is an evident trend of the *MAE*, the minimum value of *MAE* is reached in correspondence of the *IS* = -1.44. It is noteworthy the fact that even if in both Figure 4.3 top and bottom the *MAE* trends are evident, in the first case there is an appreciable difference between the lowest and the highest *MAE* value, while in the second case the *MAE* differences are very small.

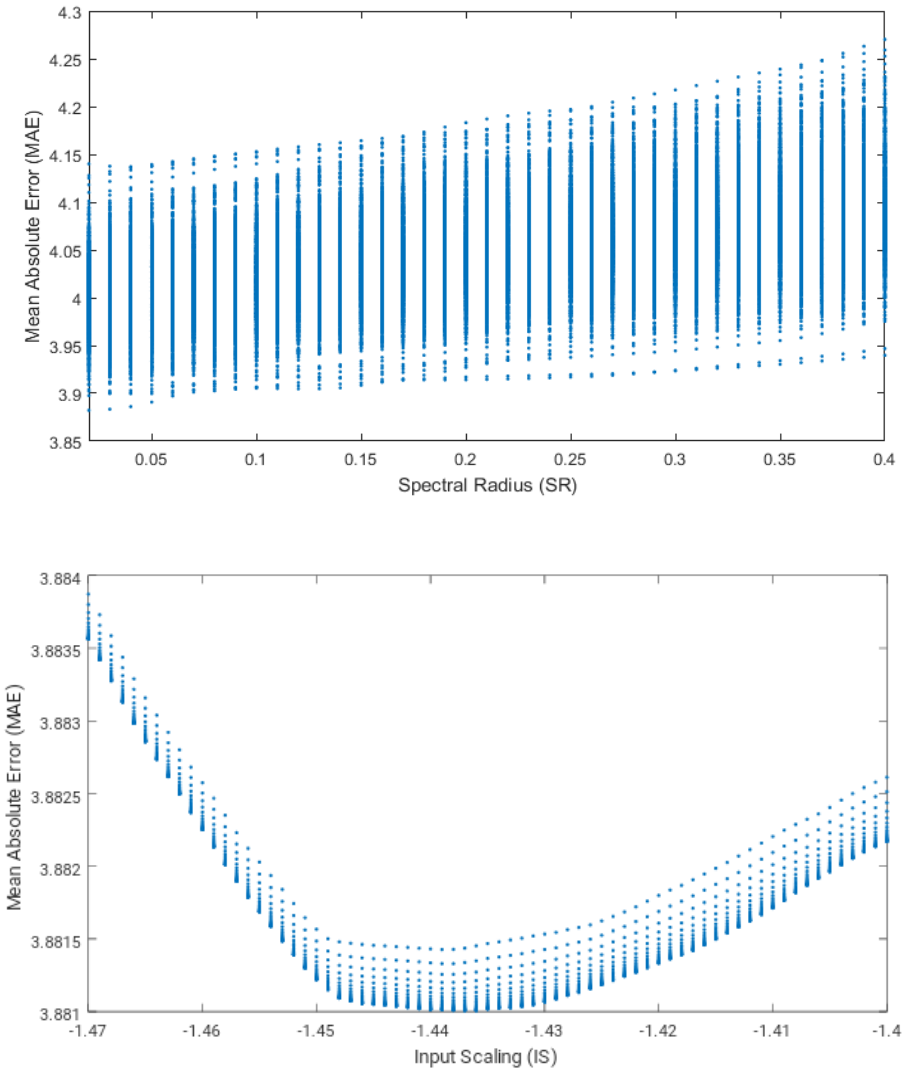

Figure 4.3 Mean Absolute Error (MAE) (arbitrary unit of measure) as function of Spectral Radius (SR) (top) and of Input Scaling (IS) (bottom).

It is worth mentioning that the *MAE* trend with respect to the other parameters, i.e., *C* and *TS*, is less remarkable. In this regard, Table 4.2 reports the best ESN architecture parameters obtained from the hierarchical grid search.

Table 4.2 Parameters and MAE (arbitrary unit of measure) of the best ESN resulting from the hierarchical grid search.

Size	**Spectral Radius**	**Connectivity**	**Input Scaling**	**Teacher Scaling**	**MAE**
466	0.02	0.13	$10^{-1.44}$	10^{-3}	3.8802

4.2.2 – Ensemble

As previously discussed, the development of an ensemble of regression models requires generating diverse individual models, i.e., characterized by prediction errors of different magnitude and sign [38].

To this aim, two strategies have been adopted to inject the diversity among the individual models of the two proposed ensembles: *i)* the use of models trained with different training datasets; and *ii)* the use of models characterized by different initial weights. The former strategy is employed to build the first proposed ensemble (hereafter called Ensemble 1), whereas the latter is employed to build the second proposed ensemble (hereafter called Ensemble 2).

A. The proposed Ensemble 1

For building the Ensemble 1 (refer to Chap. 3.2.1 for more details), $I = 500$ training datasets, $T_i, i = 1, \dots, I$, are built using the data randomly extracted from the period of 2011-2012 (as depicted in Figure 4.4). The test dataset comprises the 2013 data, divided into 12 test months, $J = 12$ (as depicted in Figure 4.5):

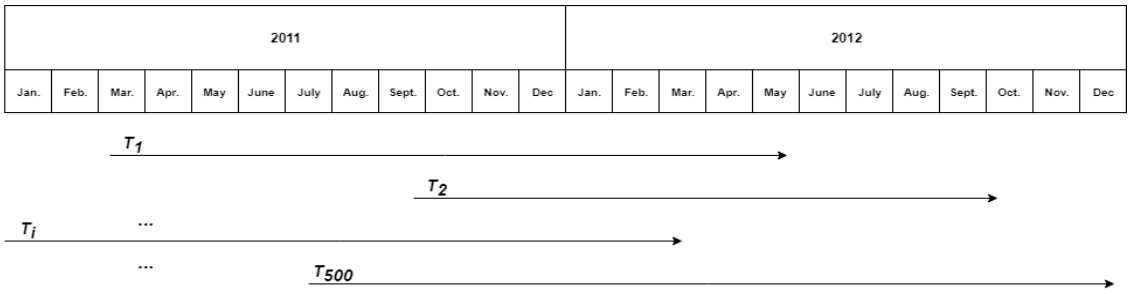

Figure 4.4 The training dataset of the proposed Ensemble 1.

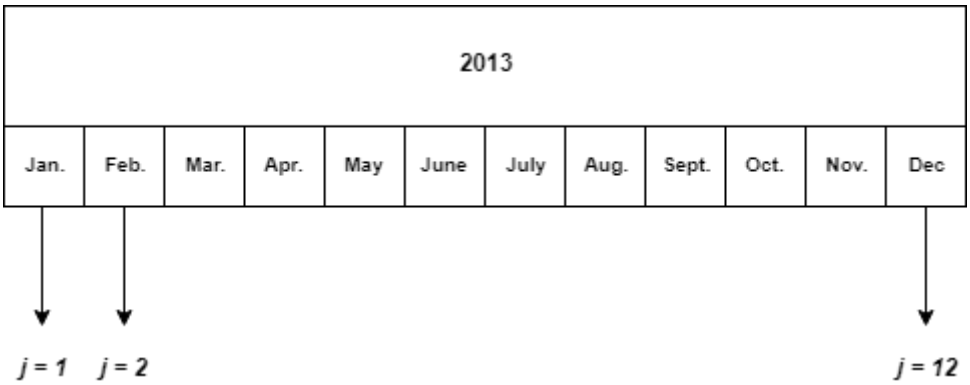

Figure 4.5 The test dataset.

Once the $I = 500$ training datasets, T_i, $i = 1, \dots, 500$, are built, they are used to train a number of $A = 1$ ESN, thus originating $AxI = 500$ different ESNs.

The procedure described to create different training sets, together with the different base models' initial weights, is thus the origin of models diversity in Ensemble 1. This diversity can be put in evidence by plotting the production prediction from the 500 different ESNs and comparing them to the real energy production (Figure 4.7 left).

B. The proposed Ensemble 2

For building the Ensemble 2 (refer to Chapter 3.2.1 for more details), $I = 1$ training dataset, $T_i, i = 1, \dots, I,$ is built using all the available 2011-2012 data (as depicted in Figure 4.6). The same test dataset of the Ensemble 1 subdivided into $J = 12$ test months (Figure 4.5) is used here for evaluating the performance of the Ensemble 2.

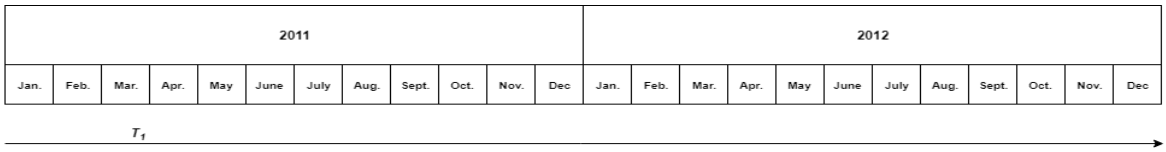

Figure 4.6 The training dataset of the proposed Ensemble 2.

The training dataset, I, is then used to train a number $A = 500$ ESNs, thus originating $AxI = 500$ ESN models. In this case, the base models are different only in their initial weights.

Figure 4.7 shows the production predictions provided by the 500 base models (ESNs) of the two proposed ensembles (thin lines) (Ensemble 1 and Ensemble 2 in Figure 4.7 (left) and (right), respectively), for 50 test data patterns, together compared with the true productions (bold line).

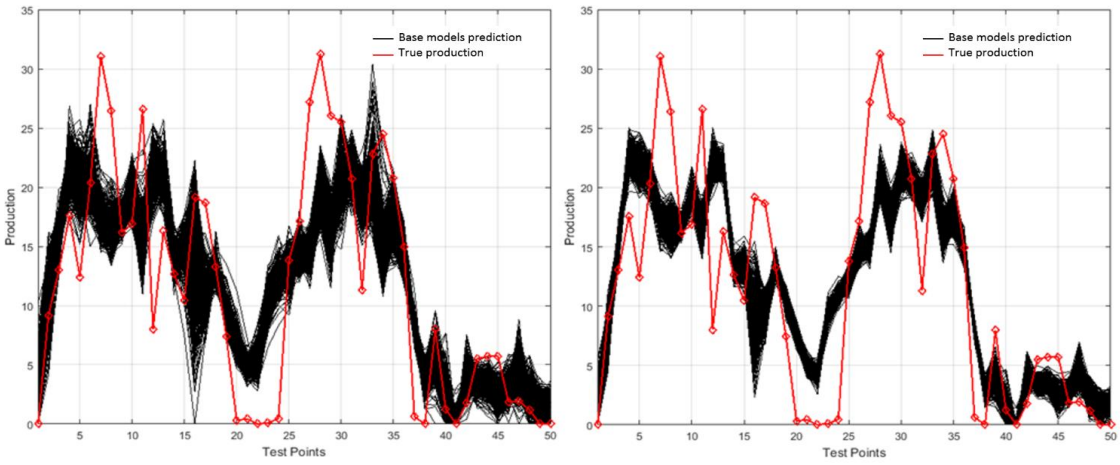

Figure 4.7 Diversity among models of Ensemble 1 (left) and Ensemble 2 (right).

Looking at Figure 4.7, one can recognize that there are points where all the base models (ESNs) (thin lines) underestimate or overestimate the real energy production (bold line) using both of the proposed ensembles, while in other points the base models commit error of different signs. Diversity indeed means that different base models commit errors of different magnitude and sign in predicting the real energy production.

Moreover, by looking at Figure 4.7, it can be observed that base models in Ensemble 1 (i.e., randomly trained) (Figure 4.7 (left)) show large variability of the production predictions, while base models in Ensemble 2 (i.e., equally trained on the entire available training dataset) (Figure 4.7 (right)) show less variability. This higher variability generated in Ensemble 1 could indeed be an advantage in reducing the overall *MAE* when base models are aggregated.

Once the base models' predictions are provided using the two ensembles, a procedure for aggregating the obtained predictions is required. To this aim, the Local Fusion method (as described in Chapter 3.2.2) (whose main two steps are summarized in Chapter 3.2.2 and illustrated in Figure 3.6).

Among the 500 base models of the ensembles, the best $H = 50$ models are selected on the basis of their performances using the *MAE* metric. For both of the ensembles, the number of the base models is $AxI = 500$; each one is tested on the first $s = 35$ points (corresponding to one week of weather previsions) of each *j-th* period, chosen as the test days of the test dataset ($j = 1, \dots, 365$ days). For each *j-th* test day, the best models are selected.

Equations 3.5 and 3.6 are then re-written as:

$$MAE_{i,a}^{j} = \frac{\sum_{i,a=1,1}^{IxA=480} \left| \hat{P}_{i,a}^{j} - P_{i,a} \right|}{IxA} \tag{4.2}$$

where $P_{i,a}$ indicates the true output of the *i,a-th* model.

Once the base models outcomes are aggregated through the computation of their median value, $\hat{P}^{LF}$, the MAE is calculated over the entire test set (from Equation 4.1 with $\hat{P}_l = \hat{P}_l^{LF}$):

$$\hat{P}^{LF} = median_{h=1,\dots,50}\left(\hat{P}_h^{j}\right) \tag{4.3}$$

$$MAE = \frac{\sum_{l=1}^{N_{test}} \left| \hat{P}_l^{LF} - P_l \right|}{N_{test}} \tag{4.4}$$

4.3 – Results

4.3.1 – Single ANN and single ESN

A first comparison is made between the performance of a single ESN with respect to that of a single ANN, when both are trained and tested on the same training and test datasets, respectively.

It is worth mentioning that an ANN model of two hidden layers composed by 9x7 hidden neurons is considered following a trial-and-error procedure [38].

Figure 4.8 shows the average monthly MAE values calculated on the test dataset obtained by the ESN and the ANN single models.

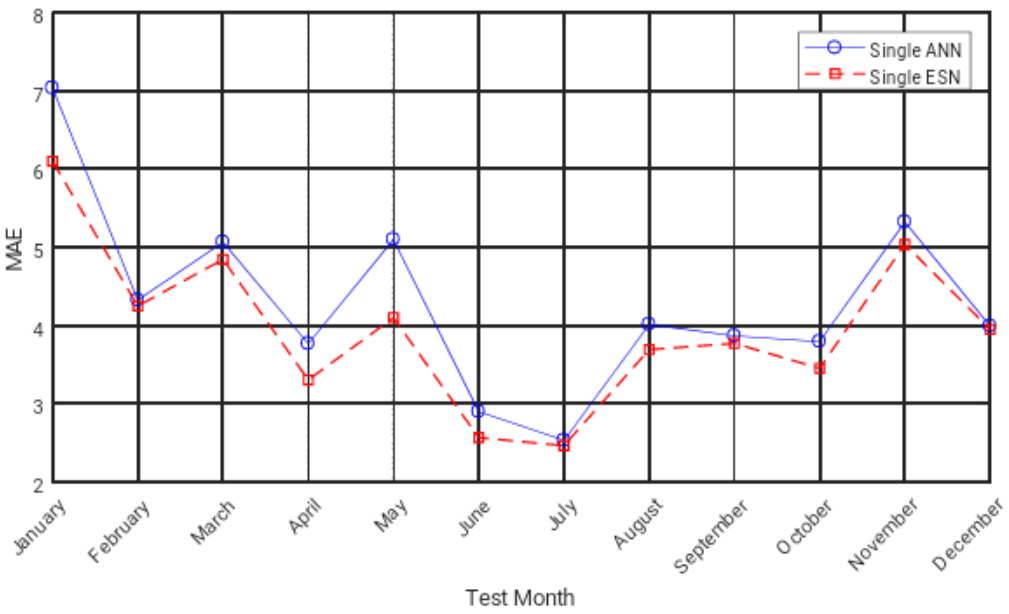

Figure 4.8 MAE values (arbitrary unit of measure) of a single ANN and a single ESN when trained and tested on the same datasets.

One can easily recognize that the overall MAE trend is quite similar for the two models, but it is evident that the MAE is, on average, lower for the ESN than that for the ANN. This finding is confirmed when the MAE is computed over the entire test dataset using the two models as reported in Table 4.3

Table 4.3 Performance (arbitrary unit of measure) of a single ANN and ESN over the entire test dataset.

Prediction Model	MAE
ANN	4.31
ESN	3.96

For clarification purposes, Figure 4.11 shows the productions predictions provided by the single ESN (squares) and the single ANN (circles), together with the real production values (stars). One can recognize that both the ESN and the ANN make a higher-than-average prediction error in regions where the real energy production value suddenly reaches a peak; the higher the peak, the more severe is the underestimation of the energy production.

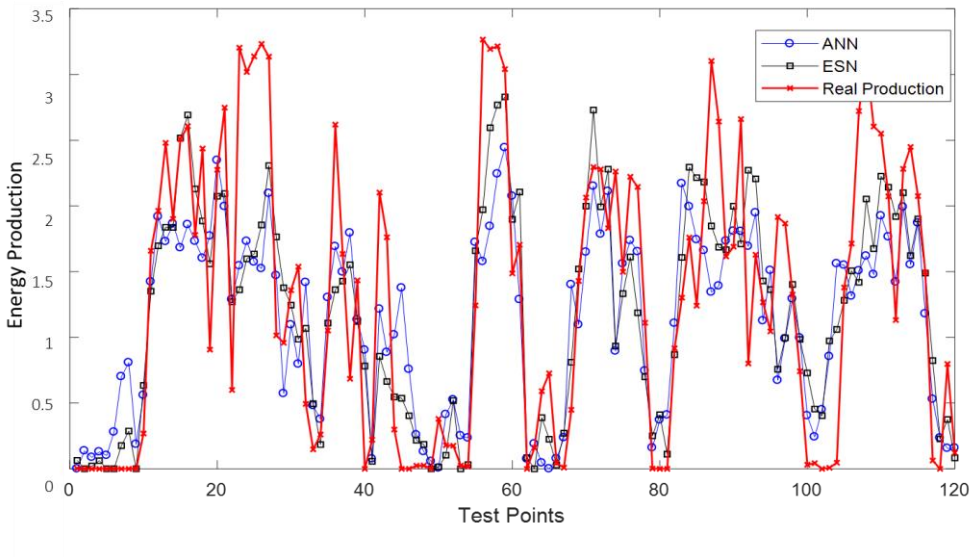

Figure 4.9 Comparison between ANN prediction, ESN prediction and real Production (arbitrary unit of measure).

To have a clear idea of what happens in these cases, it is interesting to investigate the MAE trends of the ANN and ESN (and their standard deviations) when we introduce and progressively increase an energy production threshold (Figure 4.10).

Figure 4.10 shows the average MAE values and their standard deviations obtained by the two models with respect to the increasing energy threshold. The energy production threshold is considered to span the interval [5,30] MW. It is clear that when selecting higher level of energy production (e.g., 30), the prediction error increases, which means that both the ESN and the ANN are less accurate in their predictions as the production peak increases. Moreover, it is possible to observe that the error made by the ANN increases more with the increasing threshold with respect to that made by the ESN.

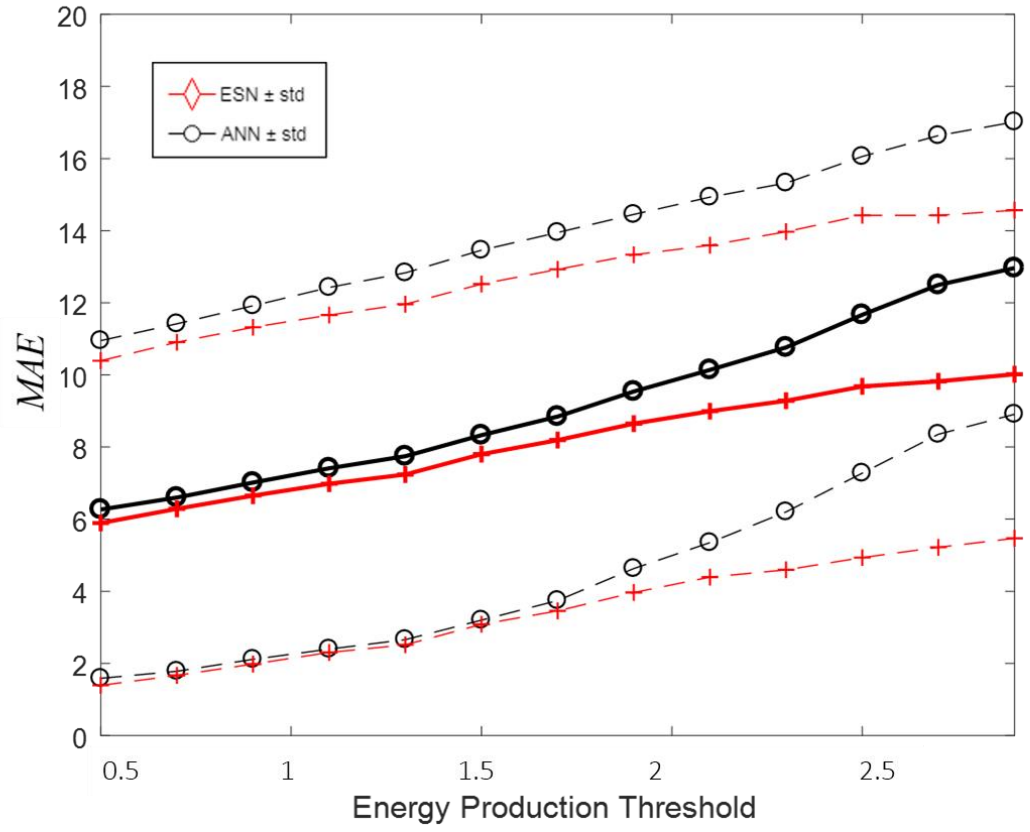

Figure 4.10 Comparison of ANN and ESN prediction errors (arbitrary unit of measure) when the energy production threshold is progressively increased.

In Table 4.4 *MAE* values are reported both for the ANN and the ESN as a function of the energy production threshold; the higher the threshold, the higher the error committed by both the prediction models. Row four in this table provides the difference between the *MAE* of the ANN and the *MAE* of the ESN; this difference is always positive (thus showing that the ANN *MAE* is always higher than ESN *MAE*) and it increases when the energy production threshold is increased.

Table 4.4 MAE (arbitrary unit of measure) of ANN and ESN and their differences as function of the energy production threshold.

Prod. Threshold	5	7	9	11	13	15	17	19	21	23	25	27	29
MAE_{ANN}	6,27	6,60	7,02	7,41	7,75	8,33	8,85	9,54	10,14	10,76	11,67	12,50	12,96
MAE_{ESN}	5,89	6,29	6,65	6,98	7,24	7,80	8,19	8,65	8,99	9,28	9,68	9,82	10,02
$MAE_{ANN} - MAE_{ESN}$	0,38	0,31	0,37	0,43	0,51	0,53	0,66	0,89	1,15	1,48	1,99	2,68	2,94

4.3.2 – ESNs Ensemble

As previously described, all the base models in the ensemble have the same architecture parameters (with the exception of the initial weights), fixed as the best parameters resulting from the hierarchical grid search (see Chapter 4.2.1). As reported in Table 4.2, we choose 466 neurons, with a spectral radius of 0.02, a connectivity of 0.13, an input scaling of $10^{-1.44}$ and a teacher scaling of 10^{-3}.

Table 4.5 reports the average MAE values and their standard deviations obtained by the single ESN model, Global Fusion (*GF*) and Local Fusion (*LF*) ensemble models

composed by 500 and 5000 base models, trained with the two different training datasets, i.e., same and random training datasets. It is worth mentioning that the standard deviations that can be obtained by the GF and LF ensembles when 5000 base models are considered are not reported for computational limitations.

Table 4.5 MAE (arbitrary unit of measure) and their standard deviations.

	Single ESN	***Ensemble of 500 ESNs***		***Ensemble of 5000 ESNs***	
		GF	***LF***	***GF***	***LF***
Random Train *(Ensemble 1)*	4.169 ± 0.108	4.039 ± 0.005	3.872 ± 0.005	4.039	3.789
Same Train (*Ensemble 2*)	3.958 ± 0.031	3.923 ± 0.001	3.838 ± 0.003	3.922	3.799

Looking at Table 4.5, one can easily recognize that: training the single ESN model using the entire available training dataset constitutes an advantage with respect to the training using a set of random starting points and lengths. This is due to the fact that using the entire training dataset can guarantee that the ESN model will be properly trained on the system dynamics. Moreover, the higher standard deviation in the MAE when a random train set is used, is justified by the fact that the training sets are different one from the other, thus causing more diversity among base models. However, this difference in the MAE is inverted, when an increasing number of base models are aggregated; this is the advantage of having more diversity among models. The difference in the performance of the two ensemble is further reduced when resorting to a LF method rather than to a GF method, until Ensemble 1 shows a better performance with respect to Ensemble 2 when

resorting to a LF ensemble of 5000 ESNs.

All the results previously shown for the LF method, are related to the aggregation of the best 50 base models (as described in Chapter 4.2.2). It is then interesting, at this point, to see if the number of the best models aggregated in the Local ensemble influences the finale MAE; for this reason, in Figure 4.11 the MAE of the Local ensemble is plotted as a function of the number of best base models aggregated, for the case of Ensemble 1 and Ensemble 2 (both with 500 and 5000 ESN models):

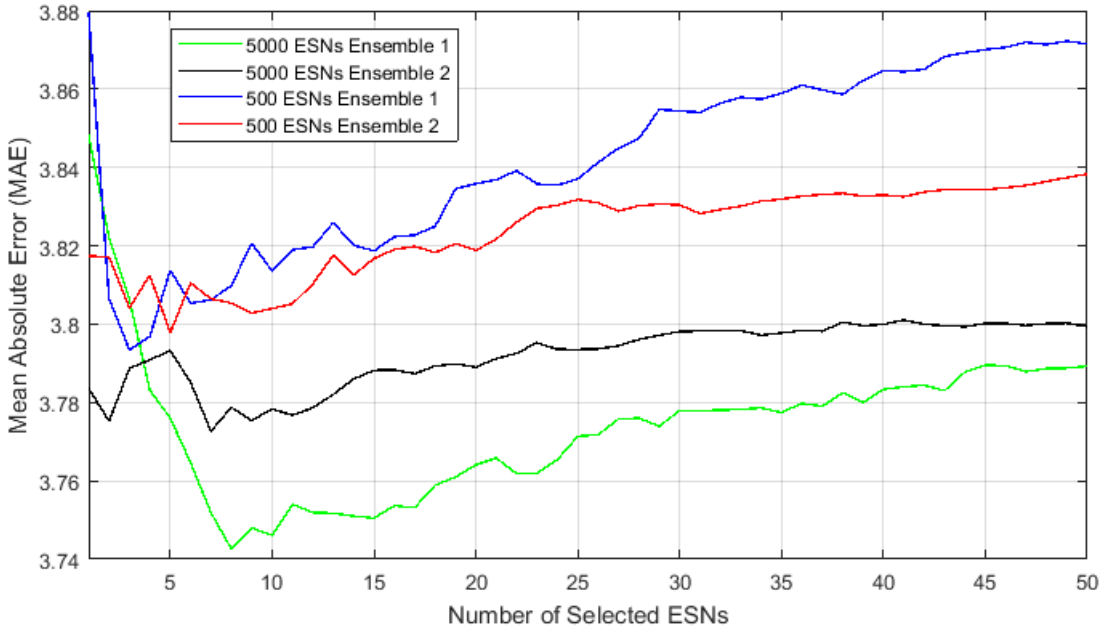

Figure 4.11 MAE (arbitrary unit of measure) as a function of the selected number of ESNs in the LF method, for Ensemble 1 and Ensemble 2 (both with 500 and 5000 ESN models).

From Figure 4.11 it can be deduced that selecting the best 50 base models to be aggregated is not the best choice to minimize the *MAE*. In Table 4.6, the best number of selected ESNs and the correspondent resulting *MAE* is shown for Ensemble 1 and Ensemble 2 (with 500 and 5000 ESNs):

Table 4.6 Best combination of number of selected ESNs and the correspondent MAE (arbitrary unit of measure) for Ensemble 1 and Ensemble 2 (trained with 500 and 5000 base models).

		Number of selected ESNs	**MAE**
Ensemble 1 (Random Train)	***LF* 500 ESNs**	3	3.793
	***LF* 5000 ESNs**	7	3.743
Ensemble 2 (equal Train)	***LF* 500 ESNs**	5	3.798
	***LF* 5000 ESNs**	8	3.773

It is evident from Table 4.6 that the best performance values are obtained when resorting to ensembles of 5000 base models rather than of 500, both for Ensemble 1 and Ensemble 2. The lowest *MAE* value, equal to 3.74, indeed, is obtained from Ensemble 1, by generating 5000 ESNs and then aggregating the best 8; this fact means that the diversity generated by 5000 ESNs randomly trained (Figure 4.12) has a positive impact on the performance when we make the number of base models vary.

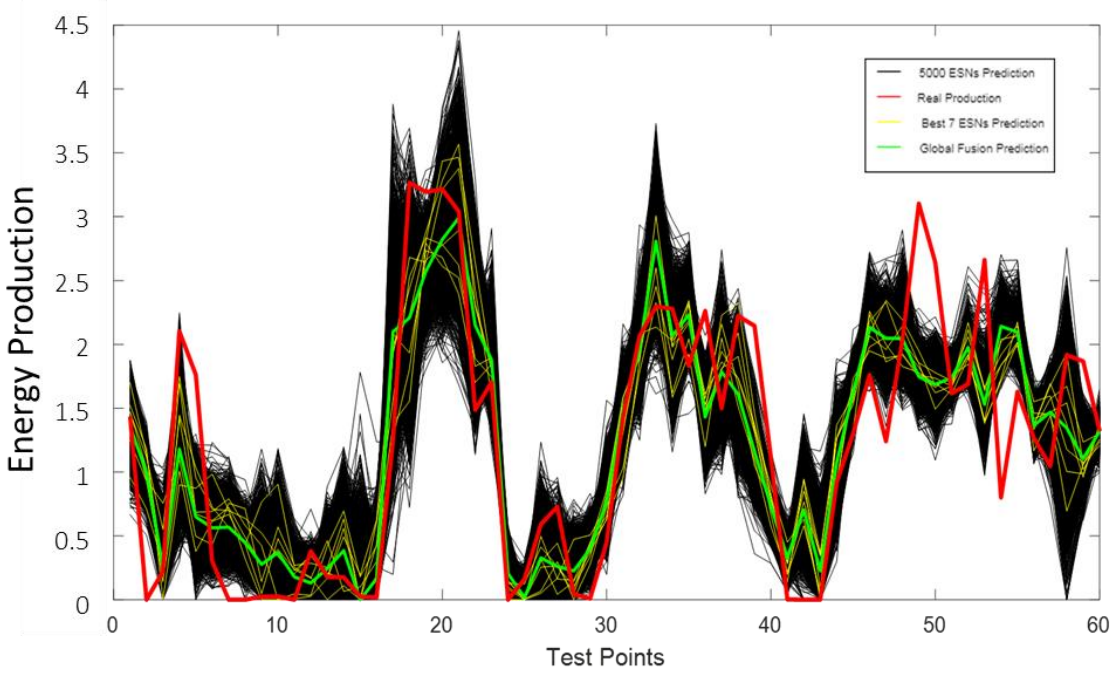

Figure 4.12 Production (arbitrary unit of measure) prediction of Ensemble 1 (5000 base models in black and LF in green), best 7 ESNs in yellow and real production value (red).

4.3.3 – Time horizon

The possibility to conduct a further analysis is given by the availability of more data; this allows us to conduct a sensitivity analysis of the length of the time horizon of the predictions. More specifically, we assume to have available weather forecasts with a forecast horizon increased from the previous value of $\Delta t = 24$ hours to $\Delta t = 96$ hours. This means that at a given time t_0 the weather forecast for the following $\Delta t = 96$ hours are available [38].

When thinking about how to use the more available data, we must discuss separately the case of the ANN and the case of the ESN. With respect to the first, we can try to train the ANN using all the available data with a forecast horizon of $\Delta t = 96$ hours; this certainly increases the noise in the train dataset, since input data are given with a larger time window, but it is also a possible way to provide the system dynamics in input to the ANN.

On the other hand, enlarging the time window in the ESN training sequence is not possible, since Recurrent Neural Networks require input data to flow over time. If weather predictions are made $\Delta t = 96$ hours ahead, the problem of predictions overlapping arises (Figure 4.13).

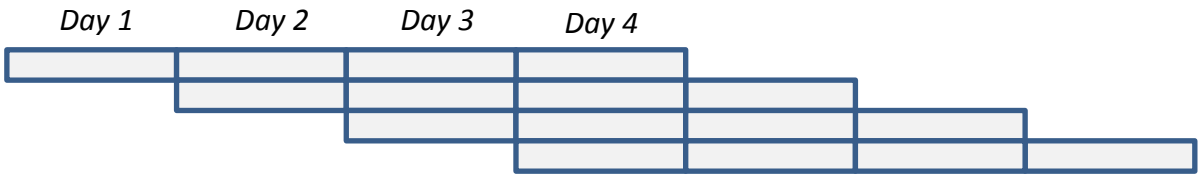

Figure 4.13 Predictions overlapping.

In Figure 4.14 and Table 4.6 the performance of an ANN trained with a time horizon $\Delta t = 24$ hours is compared with that of an ANN trained with a time window of $\Delta t = 96$ hours and with that of an ESN.

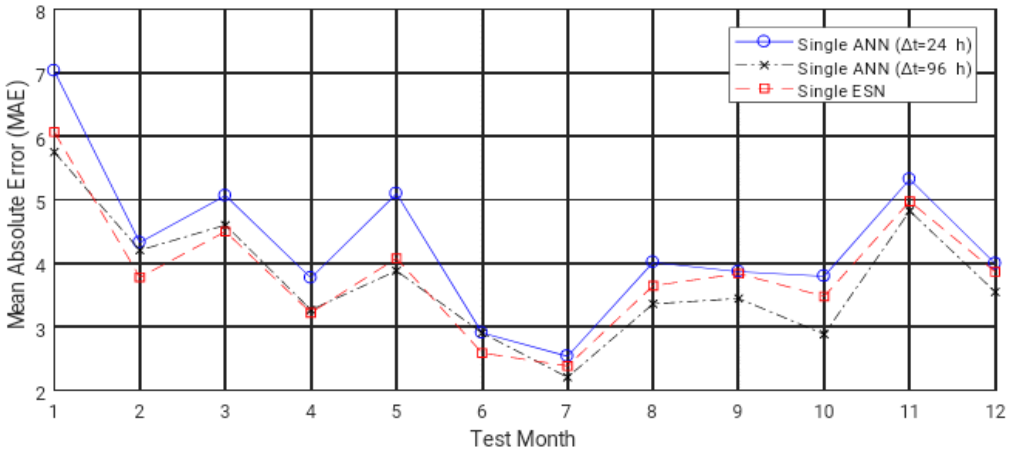

Figure 4.14 Comparison of MAE (arbitrary unit of measure) values of a single ESN, an ANN with $\Delta t = 24$ hours and an ANN with $\Delta t = 96$ hours.

From Figure 4.14 the common trend is evident between the *MAE* for the ESN and for the ANNs, both the one trained with $\Delta t = 24$ hours and the one trained with $\Delta t = 96$ hours. It is also possible to observe that training the ANN with a $\Delta t = 96$ hours allows to obtain a performance improvement, both with respect to a single ESN and to an ANN trained with $\Delta t = 24$ hours.

Performance results over the 12 test months are displayed in Table 4.7 and they allow to quantify a performance improvement of the 9% with respect to the single ESN model and of the 10.5% with respect to an ANN trained with $\Delta t = 24$ hours.

Table 4.7 Performance (arbitrary unit of measure) of a single ESN, of an ANN with $\Delta t = 24$ hours and of an ANN with $\Delta t = 96$ hours.

	MAE
Single ESN	3.9785
ANN $\Delta t = 24$ hours	4.0354
ANN $\Delta t = 96$ hours	3,6144

5. Conclusions

In this thesis work, a methodological framework is developed for the forecasting of the energy production from a wind plant.

The developed method is based on the following two steps:

1. The selection of the optimal regression prediction model,
2. The aggregation of the regression prediction model in an ensemble.

A preliminary study of the different categories of prediction models has been done, revealing that the choice of a recursive prediction model could be appropriate for this task.

Resorting to a Recurrent Neural Network (RNN) seemed an appropriate choice, and among the different available RNNs, the Echo State Network (ESN) has been chosen thanks to his “intrinsic memory” associated to its computational efficiency.

The prediction accuracy of the ESN has been deeply studied; an hierarchical grid search approach has been implemented in order to set the optimal parameters architecture.

With respect to point (2) an ensemble is created, constituted by multiple regression models whose outcomes are aggregated into a final aggregated prediction outcome. The aggregation of the outcomes provided by the base models of the ensemble has been shown to be more accurate and robust in making prediction compared to each sole base model.

The development of an ensemble required to address i) the generation of diverse base models and ii) the definition of a strategy for their outcomes aggregation.

With respect to the first requirement, diversity among regression models has been obtained by using

a. Models trained with different training datasets;

b. Models characterized by different initial weights.

The selected strategy for the present thesis allowed us to create two different ESN ensembles: “Ensemble 1”, whose diversity derives from both (a) and (b), and “Ensemble 2”, whose diversity is due solely to (b). This constitutes the novelty of this thesis work with respect to the single models generation.

With respect to requirement ii), the novelty in the aggregation of the ensemble base models is the use of a Median Global Fusion (GF) method and of a Local Fusion (LF) method, in which the best models to be aggregated are chosen with respect to their prediction accuracy.

As described in chapter 4, the prediction performance study evidenced that the accuracy of an ESN in predicting the wind energy production is higher than that of an ANN when they are equally trained and tested, thus confirming the ESN to be a good choice when predicting cyclical variables of dynamical systems.

Prediction performance quantification also demonstrates that ensembles constitute a proper strategy to address the problem of increasing prediction accuracy, in particular when a certain level of diversity among base models is obtained, and when the correct aggregation method is established.

The proposed methodological framework has been applied to the industrial case concerning the energy production prediction of a 34 MW wind power plant, for which it has been demonstrated to produce a good prediction performance.

List of Acronyms

Artificial Neural Network	ANN	Mean Absolute Erro	MAE
Connectivity	C	Recurrent Neural Network	RNN
Demand Side Management	DSM	Size of the reservoir	S
Echo State Network	ESN	Second Order Parameter	SOP
First Order Parameter	FOP	Spectral Radius	SR
Global Fusion	GF	Teacher Scaling	TS
Hierarchical Grid Search	HGS		
Input Scaling	IS		
Local Fusion	LF		

Bibliography

[1] A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen, and E. Feitosa, “A review on the young history of the wind power short-term prediction,” Renewable and Sustainable Energy Reviews, vol. 12, no. 6. pp. 1725–1744, 2008.

[2] A.I. Moustapha, R.R. Selmic, “Wireless sensor network modeling using modified recurrent neural networks: application to fault detection”, IEEE Trans. Instrum. Meas. 57 (5), 2008, pp. 981-988.

[3] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, “Comprehensive review of neural network-based prediction intervals and new advances.,” IEEE Trans. Neural Netw., vol. 22, no. 9, pp. 1341–1356, 2011.

[4] B. Ernst, B. Oakleaf, M. L. Ahlstrom, M. Lange, C. Moehrlen, B. Lange, U. Focken, and K. Rohrig, “Predicting the wind,” IEEE Power Energy Mag., vol. 5, no. 6, pp. 78–89, 2007

[5] B. Ernst, F. Reyer, J. Vanzetta, “Wind power and photovoltaic prediction tools for balancing grid operation”, in CIGRE/IEEE PES Joint Symposium Integration of Wide-Scale Renewable Resources Into the POWER Delivery System, 2009, pp. 1-9..

[6] E.B. Kosmatopoulos , M.M. Polycarpou , M.A. Christodoulou , P.A. Ioannou , “High-order neural network structures for identification of dynamical systems”, IEEE Trans. Neural Netw. 6 (2) (1995) 422–431.

[7] E. L. R. L H Chiang and R. D. Braatz, “Fault Detection and Diagnosis in Industrial Systems,” Meas. Sci. Technol., vol. 12, no. 10, p. 1745, 2001.

[8] E. Zio, “A study of the bootstrap method for estimating the accuracy of artificial neural networks in predicting nuclear transient processes, IEEE Trans. Nucl. Sci. 53 (3), 2006, pp. 1460-1478.

[9] F. Cadini, E. Zio, N. Pedroni, “Recurrent neural networks for dynamic reliability analysis”, J: Polish Saf. Reliab. Assoc. 1, 2007.

[10] F.J. Pontes et al., “Design of Experiments and Focused Grid Search for Neural Network parameter optimization”, Neurocmputing 186, January 2016.

[11] F. Xue, R. Subbu, P. Bonissone, “Locally Weighted Fusion of Multiple Predictive Models”, in The 2006 IEEE International Joint Conference on Neural Network Proceedings, 2006, pp. 2137-2143.

[12] G.A. Darbellay, M. Slama, “Forecasting the short-term demand for electricity: Do neural networks stand a better chance?”, International Journal of Forecasting,Volume 16, Issue 1, January/March 2000, Pages 71-83.

[13] Glasstone, Sesonske. “Nuclear Reactor Engineering: Reactor Systems Engineering”, Springer, 4th edition, 1994.

[14] G. Sideratos and N. D. Hatziargyriou, “An advanced statistical method for wind power forecasting,” IEEE Trans. Power Syst., vol. 22, no. 1, pp. 258–265, 2007.

[15] Holleman, D.R., Bowling, R.L. & Gathy, C. J Gen Intern Med (1996) 11: 237.

[16] H. Jaeger, “The echo state approach to analyzing and training recurrent neural networks”, German National Research Center for Information Technology, 2001 Technical Report GMD Report 148.

[17] H. Jaeger, “A tutorial on training recurrent neural networks, covering BPTT, RTRL, EFK and the Echo state network approach”, German National Research Center for Information Technology, 2002 Technical Report GMD Report 159.

[18] H. Jaeger, “Short term memory in echo state networks”, GMD-Forschungszentrum Informationstechnik, 2001.

[19] J. Cheng , S. Zhong , Q. Zhong , H. Zhu , Y. Du , “Finite-time boundedness of state estimation for neural networks with time-varying delays”, Neurocomputing 129 (2014) 257–264.

[20] J.R. Lamarsh, A.J. Baratta, “Introduction to Nuclear Engineering”, 3d ed., Prentice-Hall, 2001.

[21] J. Wang, J. Sun, H. Zang, “Short-term wind power forecasting based on support vector machine”, Conference: 2013 5th International Conference on Power Electronics Systems and Applications (PESA) New Energy Conversion for the 21st Century, December 2013, pp. 1-5.

[22] L. Breiman, “Bagging Predictors”, Mach. Learn., vol. 24, n° 421, 1996, pp. 123-140.

[23] L. K. Hansen and P. Salamon, “Neural Network Ensembles,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 12, no. 10, pp. 993–1001, 1990.

[24] M. Lange, U. Focken, “Physical approach to short-term wind power prediction”, Berlin: Springer, 2006.

[25] M. Lukoševičius, H. Jaeger, “Reservoir computing approaches to recurrent neural network training”, Comput. Sci. Rev. 3 (3), 2009, pp.127-149.

[26] M. Rigamonti, P. Baraldi, E.Zio, I. Roychoudhury, K. Goebel, S. Poll, “Echo state network for the remaining useful life prediction of a turbofan engine”, in: PHME 2016 Conference Proceedings, 5, Bilbao, Spain, 2016 July 5th-8th 2016.

[27] M. Rigamonti, P. Baraldi, E. Zio, et at., “Ensemble of optimized echo state networks for remaining useful life prediction”, Neurocomputing, 2017, pp.121-138.

[28] N.A. Shrivastava, B.K. Panigrahi, “Point and prediction interval estimation for electricity markets with machine learning techniques and wavelet transforms”, Neurocomputing 118, 2013, pp. 301-310.

[29] N.A. Treiber, O. Kramer, “Evolutionary feature weighting for wind power prediction with nearest neighbor regression”, Conference: 2015 IEEE Congress on Evolutionary Computation (CEC), May 2015.

[30] OECD/IEA 2018 , “Global Energy & CO_2 Status Report 2017”, March 2018.

[31] O. Kramer, F. Gieseke, “Short-Term Wind Energy Forecasting Using Support Vector Regression,” in Soft Computing Models in Industrial and Environmental Applications, 6th International Conference SOCO 2011, E. Corchado, V. Snášel, J. Sedano, A. E. Hassanien, J. L. Calvo, and D. Ślęzak, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 271–280.

[32] P. Baraldi, “Reservoir computing methods for Prognostics and Health Management (PHM)”, IEEE Seminar series on system reliability, risk and resilience, fall 2017.

[33] P. Baraldi, F. Cadini, F. Mangili, E. Zio, “Model-based and data driven prognostics under different available information”, Probab. Eng. Mech. 32, 2013, pp. 66-79.

[34] P. Baraldi, R. Razavi-Far, E. Zio, “Classifier ensemble incremental-learning procedure for nuclear transient identification at different operational conditions”, Reliab. Eng. Syst. Saf. 96 (4), 2011, pp. 480-488.

[35] P. Ramasamy, S.S. Chandel, A.K. Yadav, “Wind speed prediction in the mountainous region of India using an artificial neural network model,” Renew. Energy, vol. 80, no. March 2014, pp. 338–347, 2015.

[36] R. Polikar, “Ensemble based systems in decision making”, IEEE Circuits Syst. Mag 6 (3), 2006, pp. 21-45.

[37] S. Al-Dahidi, F. Di Maio, P. Baraldi, E. Zio, “A locally adaptive ensemble approach for data-driven prognostics of heterogeneous fleets”, Proc. Inst. Mech. Eng. Part. O J. Risk Reliab., vol. 231, n° 4, pp. 350-363, 2017.

[38] S. Al-Dahidi, P. Baraldi, E. Zio, E. Legnani, “A Dynamic Weighting Ensemble Approach for Wind Energy Production Prediction”, proceeding of 2017 2nd International Conference on System Reliability and Safety, Milan, 2017, pp. 296-302.

[39] S. Han, Y. Liu, J. Yan, “Neural Network ensemble method study for wind power prediction”, in 2011 Asian Pacific Power and Energy Engineering Conference APPEEC 2011, 2011, vol. Wuhan.

[40] S. Hochreiter, J. Schmidhuber, “Long Short-Term Memory”, Neural Computation, Volume 9, Issue 8, 15 November 1997, Pages 1735-1780.

[41] S.S. Soman, H. Zareipour, O. Malik, and P. Mandal, “A review of wind power and wind speed forecasting methods with different time horizons,” North Am. Power Symp. 2010, pp. 1–8, 2010.

[42] S.X. Lun , X.S. Yao , H.Y. Qi , H.F. Hu , “A novel model of leaky integrator echo state network for time-series prediction”, Neurocomputing 159 (2015) 58–66.

[43] T.G. Barbounis, J.B. Theocharis, “Locally recurrent neural networks for long-term wind speed and power prediction”, Neurocomputing, vol. 69, no. 4–6, pp. 466–496, 2006.

[44] U.S. Department of Energy, “Nuclear Physics and Reactor Theory”, DOE Fundamentals Handbook, Volume 1 and 2. January 1993.

[45] V.M. Landassuri-Moreno, J.A. Bullinaria, “Neural network ensembles for time series forecasting”, in: Proceedings of the Eleventh Annual Conference on Genetic and Evolutionary Computation, ACM, July 2009, pp. 1235-1242.

[46] W.M. Stacey, “Nuclear Reactor Physics”, John Wiley & Sons, 2001.

[47] W. Yao, Z. Zeng, C. Lian, H. Tang, “Ensembles of echo state networks for time series prediction”, in: 2013 Sixth International Conference on Advanced Computational Intelligence (ICACI), IEEE, 2013, pp. 299-304.

[48] Y. Peng, H. Wang, D. Liu, X. Peng, “A modified echo state network based remaining useful life estimation approach”, in: PHM 2012-2012 IEEE Int. Conf. on Prognostics and Health Management: Enhancing Safety, Efficiency, Availability, and Effectiveness of Systems Through PHM Technology and Application, Conference Program, 2012 art. No. 6299524.

[48] Z.Yang, P. Baraldi, E. Zio, “A comparison between extreme learning machine and artificial neural network for remaining useful life prediction”, in: Prognostics and System Health Management Conference (PHM-Chengdu), 2016, 2016 October, pp. 1-7.

[49] A. L. Blum, P. Langley, Selection of relevant features and examples in machine learning, Artificial Intelligence, Volume 97, Issues 1–2, December 1997, Pages 245-271.

[50] Bolón-Canedo, V., Sánchez-Maroño, N. and Alonso-Betanzos, A., "A review of feature selection methods on synthetic data", Knowledge and Information Systems, vol. 34, no. 3, 2013, pp. 483-519.