



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

UNIVERSITA' DEGLI STUDI DI PADOVA

**Dipartimento di Ingegneria Industriale DII**

Corso di Laurea Magistrale in Ingegneria dell'Energia Elettrica

Tesi di laurea magistrale

*Sviluppo e valutazione di Key Performance Indicators  
per impianti fotovoltaici*

Relatore:

*Prof. Fabio Bignucolo*

Correlatore:

*Ing. Riccardo Clementi*

Ente: STE Energy

*Giuseppe Maria Piuma*

matricola 1233557

Anno Accademico: 2021/2022







## **Sommario**

Il funzionamento di un impianto fotovoltaico è per molti aspetti descrivibile tramite parametri chiave, detti appunto Key Performance Indicators (KPI in breve): tali valori sono estratti direttamente o indirettamente dalle grandezze fisiche dell'impianto analizzato, e permettono di ottenere un'idea d'insieme sulla situazione della struttura, nonché di paragonarla con altre simili e con il proprio funzionamento trascorso.

Considerati il vasto riconoscimento nel settore fotovoltaico dei KPI principali e la necessità di avere un metro di paragone comune, si è fatto anche un excursus in merito a KPI più inusuali, ma il lavoro di tesi si è focalizzato sul piano pratico, con lo sviluppo e analisi dei parametri davvero rilevanti, basandosi sulle misure di uno specifico impianto.

Si è quindi modellizzato il funzionamento della struttura, optando per l'uso di reti neurali, estraendone la precisione di stima della potenza ed i KPI prodotti dal modello, in modo tale da compararli con quelli di misura ed individuare eventuali funzionamenti subottimali di stringhe, inverter, sensori ecc. In tal modo si è sviluppato un nuovo strumento di monitoraggio, approfondendo le qualità di specifiche reti e la loro funzionalità nella produzione dei KPI di modello.

## Sommario

Introduzione .....	3
1. Modellizzazione dell'impianto: tecniche .....	5
1.1 Single Diode Model – SDM.....	6
1.2 Previsione della generazione di potenza PV .....	7
1.3 Concetti chiave di Machine Learning applicato alla stima di potenza PV .....	8
1.3.1 Suddivisione del dataset .....	8
1.3.2 Pre-processing dei dati in input al modello .....	9
1.3.3 Selezione delle feature (scelta degli input).....	10
1.4 Modelli di previsione (forecast) della produzione energetica PV .....	15
1.5 Modelli di previsione (prediction) della produzione energetica PV .....	16
1.5.1 Reti neurali artificiali (artificial neural network) ANN .....	16
2 Key Performance Indicators KPI .....	21
2.1 Key Performance Indicators KPI legati alla performance dell'impianto PV .....	23
2.1.1 Reference Yield ( $Y_r$ ) .....	23
2.1.2 Array Yield ( $Y_A$ ) , Final Yield ( $Y_f$ ) .....	24
2.1.3 Expected Energy Yield ( $Y_E$ ) – Final Yield atteso .....	24
2.1.4 Performance Ratio ( $PR$ ) .....	25
2.1.4 Performance Ratio corretto in temperatura di riferimento ( $CPRSTC$ ) .....	26
2.1.5 Performance Ratio corretto con la temperatura media annuale ( $CPRT_{avg}$ ).....	28
2.1.6 Fattore di capacità – Capacity Factor (CF).....	28
2.1.7 Rapporto di superficie – Surface Ratio (SR).....	29
2.1.8 Bifacial Performance Ratio (BPR) .....	29
2.1.9 Guadagno bifacciale - Bifacial Gain (BG) .....	29
2.1.10 Guadagno dell'inseguitore - Tracker Gain (TG) .....	30
2.2 KPI legati alla modellizzazione e previsione della potenza PV prodotta .....	31
2.2.1 Errore quadratico medio – Mean Squared Error (MSE).....	31
2.2.2 Scarto quadratico medio – Root Mean Squared Error (RMSE) .....	31
2.2.3 Errore quadratico medio normalizzato (NMSE) .....	32
2.2.4 Coefficiente di determinazione ( $R^2$ ).....	33
2.2.5 Errore medio assoluto – Mean Absolute Error (MAE) .....	33
2.2.5 Errore medio di bias – Mean Bias Error (MBE) .....	33
2.2.6 Indice di prestazione energetica – Energy Performance Index (EPI) .....	34
2.2.7 Disponibilità dei dati .....	34
2.2.7 Qualità dei dati .....	34
3 Impianto fotovoltaico valutato .....	36
4 Sviluppo di modellizzazione e considerazioni sull'impianto .....	44
4.1 Selezione ed estrazione dei dati, scelta del tipo di modello .....	44

4.2	Processo di sviluppo ed ottimizzazione della rete feed-forward .....	47
4.2.1	Pre-processing dei dati d'impianto e ARPAE .....	47
4.2.2	Addestramento della rete con ottimizzazione Bayesiana .....	48
4.3	Applicazione delle reti e conseguenti considerazioni sull'impianto .....	51
4.3.1	Problemi di misura della temperatura di pannello.....	59
4.3.2	Guasto delle stringhe ed impatto sulle stime di energia e potenza .....	61
4.3.3	Calcolo dell'MBE .....	62
4.3.4	KPI di prestazione: PR, CPR .....	64
4.3.5	Confronto con il Performance Ratio dei modelli .....	70
4.3.6	Impiego dati PVGIS nei KPI d'impianto .....	71
4.3.7	Altre opzioni di sviluppo delle reti prese in considerazione .....	73
4.3.8	Modifiche implementabili per potenziare l'efficacia ANN-KPI.....	73
	CONCLUSIONI .....	75
	BIBLIOGRAFIA .....	76

## Introduzione

Nella gestione e supervisione di un impianto fotovoltaico è essenziale avere accesso a strumenti affidabili, ma semplici da implementare, che svolgano funzioni molteplici, tra cui mostrare le performance rispetto all'anno di avvio, ad altre strutture in location differenti, ma anche individuare funzionamenti subottimali o veri e propri guasti.

Tra i vari indicatori chiave, ad avere un ruolo centrale è solitamente il Performance Ratio PR, che lega la radiazione solare giunta nella direzione del piano di array POA con la potenza complessivamente prodotta come conseguenza. L'espressione di tale KPI verrà esposta insieme agli altri nel rispettivo capitolo, ma ciò su cui preme soffermarsi è che, sebbene il PR sia un'ottima rappresentazione sintetica dello stato dell'impianto considerato, sono molteplici i fattori che lo influenzano e che meritano di essere presi in considerazione (quali corretto funzionamento delle stringhe o meno, fenomeni di ombreggiamento ecc.).

Il lavoro svolto non si prefigge di andare a sostituire strumenti specifici per ogni possibile grandezza e circostanza, ma di osservare un quadro d'insieme, anche a lungo termine, sul comportamento della struttura fotovoltaica, concentrandosi sulla modellizzazione della stessa ed il confronto tra i principali KPI ottenuti dalla stima e quelli invece prodotti dalla rete neurale stessa (o reti, come si vedrà più avanti).

In tal senso la ridotta qualità di previsione da parte del modello, se localizzata in specifici mesi con ricorrenza, oppure soltanto in un preciso anno, ed affiancata da una maggior precisione per il resto della vita d'impianto, svolge un ruolo cardine nell'evidenziare anomalie e suggerirne le cause.

Saranno poi l'accortezza ed esperienza di chi opera le analisi a centrare i possibili colpevoli, spingendo al massimo l'impiego di tale strumento.

Tramite l'impiego di reti neurali (Artificial Neural Network ANN) si possono comunque svolgere altre mansioni di monitoraggio, sviluppando gli appositi algoritmi e in alcuni casi l'obiettivo diventa proprio l'individuazione della possibile causa del guasto, come in [1],[2].

Nei paper indicati, invece di modellizzare l'impianto o le stringhe da un punto di vista elettrico, valori come tensione, corrente e potenza sono confrontati direttamente con valori storici o standard del pannello/stringa/impianto e, superati certi delta di differenza, a seconda di tempistiche e grandezze coinvolte, lanciano un algoritmo di tipo ANN adeguato, che cerca di identificare la causa di malfunzionamento più probabile, sulla base dei dati in ingresso. Si tratta in questo caso di reti di classificazione, che a differenza della regressione impiegano il machine learning per distinguere categorie, "labels", piuttosto di cercare di riprodurre quantità.

Sono lo stesso tipo di reti neurali solitamente utilizzate per distinguere la natura delle immagini a partire dai pixel che le compongono, come per tecnologie di riconoscimento facciale ecc.

La potenzialità di tecniche di questo tipo è evidente, anche perché possono proporre sistemi di diagnosi online, come in [3], ma come tutte le reti di classificazione necessitano di una considerevole mole di dati d'allenamento, ed in [3] il sistema impiegato si mostra particolarmente sensibile a rapide o intense variazioni delle condizioni meteorologiche, fenomeno potenzialmente comune a seconda della zona geografica di operatività.

Uno studio simile è stato eseguito [4] applicando a 8 di 12 pannelli le condizioni di guasto da studiare e poi sfruttare la classificazione da parte di una rete neurale di tipo feed forward FFNN per individuare le tipologie di malfunzionamento presenti. Nuovamente con ottimi risultati, si parla però di un setting totalmente sperimentale, con guasti indotti in anticipo, necessità di almeno 8 ore di cielo sereno continuative e sensori di temperatura piazzati in corrispondenza di ogni pannello.

Altri utilizzi di machine learning di questo tipo applicano un tipo di ML più tradizionale all'analisi della curva I-V oppure Macchine a Vettori di Supporto (Support Vector Machines SVM) per risolvere problemi di classificazione dei guasti su immagini termografiche [5].

Tutti questi impieghi di reti neurali, direttamente applicati al monitoraggio di specifici guasti in ambito fotovoltaico, restano comunque ben distinti dalla direzione dell'attività di tesi, ma sono stati riportati per offrire una più ampia panoramica sul crescente utilizzo di Machine Learning di vario genere in



questo settore, con i più svariati approcci, seppure alcuni rimangano relegati ad ora all'ambito della ricerca e sperimentazione.

L'argomento di tesi va ad interagire con vari ambiti, ma nell'ottica di ricavare i KPI, dipendenti da valori ambientali e dalle potenze (continua ed alternata) prodotte, si focalizza sulla stima della produzione energetica fotovoltaica e sull'ottimizzazione del modello impiegato.

La gamma di possibili approcci a tale questione spazia, però, per natura del modello, distanza temporale di stima, dimensioni del sistema indagato ecc.

Le categorie di modelli (passando dai più tradizionali alle svariate tipologie prodotte dal Machine Learning) saranno esplicate e confrontate sulla base di risultati ottenuti in diverse pubblicazioni, con il dovuto approfondimento dedicato alle ANN e le motivazioni della scelta di tale opzione di modellizzazione dell'output di impianto. Dopodiché sarà dedicato ampio respiro ai Key Performance Indicators più usati, sia nella caratterizzazione degli impianti fotovoltaici, ma anche in merito a valutazioni di natura statistica sulle performance dei modelli di stima.

All'atto pratico, verranno descritte le proprietà elettriche e non dell'impianto preso in analisi presso l'azienda STE Energy S.r.l., con cui si è collaborato nello svolgimento del lavoro tesi.

I successivi due capitoli rappresenteranno il risultato dell'applicazione della teoria di ANN e KPI ad un impianto reale, con una prima famiglia di reti neurali per testarne le potenzialità ed individuare eventuali anomalie, completando con una seconda famiglia di reti, allenata con la consapevolezza delle informazioni precedentemente svelate.

## **1. Modellizzazione dell'impianto: tecniche**

Prima di addentrarsi nell'argomento principale con le sue sfaccettature, si sfrutta [6] per un'ultima visione delle metodologie di monitoraggio applicabili direttamente, con e senza modellizzazione, poiché, anche se scala e metodo differiscono da quelli applicati più avanti, permane il concetto del confronto con valori "standard" al fine di notare anomalie.

Tali metodi di individuazione dei guasti possono poi essere real-time oppure su scale temporali maggiori, ma lo scopo rimane nel puntare alla riduzione di tempi e costi di manutenzione, cercando di evitare perdite di energia di qualsivoglia natura, danni agli equipaggiamenti ecc.

Le strategie di monitoraggio in tal senso sono innumerevoli, ma possono complessivamente essere raggruppate in due tipi: model-based e history-based.

I metodi model-based, come suggerisce il nome, sfruttano una modellizzazione di qualche tipo per produrre output stimati, solitamente estraendo dalle misure fisiche i parametri specifici del modello preesistente, dopodiché tali valori sono comparati analiticamente con gli output misurati nel sistema e un guasto si delinea quando la differenza supera una soglia prefissata.

Metodi di questo tipo includono l'approccio basato sul modello a diodo singolo, il filtro Kalman e serie di Fourier. Il modello a diodo singolo (single diode model SDM) è molto diffuso ed una classica rappresentazione del funzionamento elettrico della singola cella o pannello fotovoltaico e può essere impiegato anche come riferimento per lo sviluppo di reti neurali a modellizzazione dei singoli pannelli [7].

Altre tecniche in questo gruppo sono la riflettometria nel dominio del tempo (time domain reflectometry TDR) e thermo-reflectance imaging TR. Queste permettono di identificare eventuali decadimenti nelle performance e guasti alle connessioni di stringa, nonché specifici hotspot nel caso di TR, indipendentemente dalle condizioni ambientali. La debolezza di questi due metodi giace nella necessità di strumentazione costosa e sofisticata, nonché nel dover spegnere l'intero impianto per prendere le misure nel caso di TR.

L'utilità di strategie basate su modelli come quelle appena esposte dipende chiaramente dall'accuratezza di tali modelli, e dalla precisione delle misure che ne hanno permesso la realizzazione.

A differenza delle precedenti strategie, quelle history-based analizzano dataset raccolti dal funzionamento del sistema fotovoltaico per derivarne un modello empirico. Vengono poi applicate procedure di machine learning o di intelligenza computazionale.

In tal modo in letteratura sono state rilevate problematiche specifiche di ombreggiamento parziale, guasti da corto circuito o effetto dell'accumulo di sporco sulle superfici dei pannelli.

Laddove model-based peccano all'applicazione pratica se il modello non ha affidabilità sufficiente, gli history-based richiedono grandi moli di dati e sono per cui poco impiegabili nei primi anni di vita dell'impianto. Di conseguenza [6] mira a potenziare l'efficacia dell'SDM combinandola con schemi di controllo CUSUM (metodo statistico a cumulative sum) per evidenziare anomalie.

Il sistema realizzato in tal modo identifica dapprima i parametri del modello a diodo singolo, poi lo impiega per modellizzare l'intero impianto tramite Matlab/Simulink e PSIM, ne testa la precisione introducendo reali misure ambientali e confrontando i picchi di potenza misurati con quelli simulati dal modello dell'array. Lo schema CUSUM valuta quindi i residuali ottenuti e giunge a conclusioni sullo stato (sano o meno) di operatività dell'array fotovoltaico.

L'approfondimento in merito non si spinge oltre, mentre invece si entrerà nel dettaglio del modello a diodo singolo, tanto è comune ed impiegato nella modellizzazione di celle e pannelli.

## 1.1 Single Diode Model – SDM

Il modello a singolo diodo SDM è il più diffuso e comunemente utilizzato a simulare con una certa precisione il comportamento e la produzione energetica di celle e pannelli fotovoltaici.

Questi ultimi sono modellizzati con un circuito elettrico equivalente (come mostrato in Figura 1.1), contenente un generatore di corrente fotovoltaica  $I_{ph}$  (ovvero la corrente generata nella cella colpita da radiazione per effetto fotovoltaico), un diodo  $D$  a descrivere il fenomeno di ricombinazione interna della cella e due resistenze; la prima è la resistenza serie parassita  $R_s$  dovuta ai reofori di contatto ed agli strati di silicio, la seconda invece è la resistenza shunt  $R_{sh}$ , a rappresentare le perdite per richiusura interna alla cella delle correnti di dispersione  $I_{sh}$ .

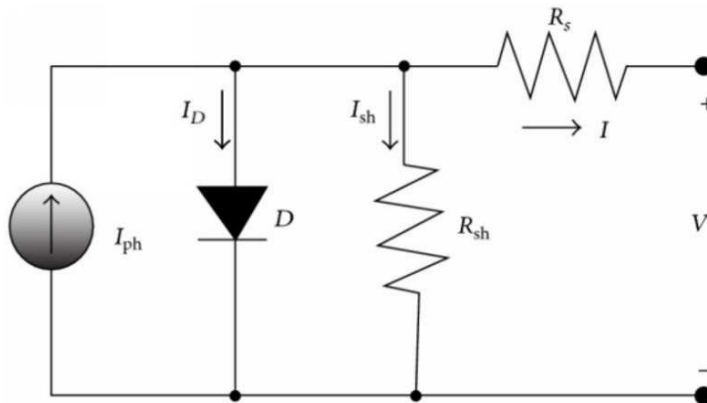


Figura 1.1. Circuito equivalente SDM di celle/pannelli PV

La rappresentazione matematica è ottenuta dalla legge di Kirchoff e dalla composizione del circuito, legando la corrente generata dalla cella  $I$  e la tensione prodotta ai suoi capi  $V$ :

$$I = I_{ph} - I_0 \left[ e^{\frac{q \cdot (V + R_s \cdot I)}{N \cdot k_B \cdot T}} - 1 \right] - \frac{V + R_s \cdot I}{R_{sh}} \quad (1)$$

dove  $I_0$  indica le correnti di saturazione del diodo,  $N$  il fattore di idealità del diodo,  $k_B$  la costante di Boltzmann ( $k_B = 1.3806503 \times 10^{-23} \text{ J/K}$ ),  $q$  la carica dell'elettrone ( $q = 1.60217646 \cdot 10^{-19} \text{ C}$ ) ed infine  $T$  la temperatura della cella.

Essendo poi ogni modulo fotovoltaico composto da numerose celle connesse in parallelo per accrescere la tensione ed in serie per aumentare la corrente, l'equazione complessiva in grado di descrivere elettricamente l'intero array PV nascerà come combinazione dell'equazione SDM appena vista.

Già da (1) si può evincere come sia essenziale estrapolare con la massima precisione il valore dei 5 parametri elettrici determinanti  $I_{ph}, I_0, N, R_s, R_{sh}$ , se si desidera avere un modello davvero in grado di simulare il comportamento delle celle e, di conseguenza, dell'array nella sua totalità.

## 1.2 Previsione della generazione di potenza PV

Si entra ora nel merito di tutte quelle tecniche utilizzate per offrire previsioni di potenza generata dall'impianto modellizzato, in diversi modi e con differenti tempistiche di stima.

Nel dettagliato lavoro di divulgazione ed approfondimento [8] la questione è espansa su più fronti, analizzando le performance di numerosi modelli di previsione della potenza PV, suddivisi secondo diverse tipologie di classificazione, includendo modelli statistici, che impiegano il machine learning, ibridi ecc. con annessi punti di forza e debolezza.

Negli scopi del lavoro di tesi, la produzione dei KPI stimati ed il confronto con quelli reali, nonché la giustapposizione tra potenza prevista/modellizzata e misurata sono il focus, ma gli strumenti che lo permettono hanno anche altri ruoli all'interno della letteratura e dietro al crescente interesse in merito alla previsione della produzione energetica PV. Conviene quindi farne una valutazione.

A livello globale e nazionale, la tecnologia fotovoltaica ha ricevuto grande attenzione nell'ultima decade ed oltre, affiancata alla risorsa eolica come le più accettabili e promettenti nel settore della produzione energetica da risorse rinnovabili. Entrambe queste fonti hanno un elevato rateo di successo della penetrazione nel mercato elettrico, considerati i vasti benefici in termini ambientali (come membri di punta nella spinta alla decarbonizzazione) ma anche economici.

Tuttavia, la generazione di natura PV dipende totalmente dall'andamento di fattori meteorologici incontrollabili e parzialmente imprevedibili, rendendo anche l'output energetico soggetto a forti variazioni, che impattano negativamente su stabilità ed affidabilità degli impianti.

Proprio per le ragioni esposte, un'accurata previsione della generazione PV può ridurre fortemente gli effetti negativi dell'aleatorietà della potenza prodotta, migliorando l'affidabilità del sistema, l'interazione con la rete ed incrementando ulteriormente il livello di penetrazione della tecnologia fotovoltaica.

Le ricerche condotte per produrre modelli di previsione affidabili e con contenuto costo computazionale portano a classificare tali modelli in due macro-categorie: a previsione indiretta o diretta (indirect/direct forecasting).

Nei modelli ad indirect forecasting l'irradianza solare per diverse finestre temporali viene prevista a sua volta tramite stima numerica del meteo (weather numerical prediction WNP), metodi basati su reti neurali a riconoscimento di immagini, di tipo statistico o ibridi, per poi impiegare tali dati in software commerciali di stima della generazione PV (HOMER, PVFORM, ecc.), trasferendo così eventuali errori ed inaffidabilità nel calcolo dell'irradianza al risultato finale di potenza in output.

In modelli a direct forecasting, invece, si usano direttamente dataset contenenti lo storico dell'impianto in quanto a misure, sia meteorologiche che elettriche, portando in media a risultati più precisi rispetto alla tipologia precedente.

Approfondendo la previsione di tipo diretto, si interagisce, come già nell'introduzione, con svariate tecniche e algoritmi di machine learning, con le rispettive peculiarità, operazioni da eseguire ecc.

Si va ora quindi ad espandere sulla breve spiegazione portata in merito nel paper [9], dove tramite machine learning, e ANN nello specifico, è stato impiegato il toolbox ML di Matlab proprio per sviluppare più modelli di previsione della potenza.

## 1.3 Concetti chiave di Machine Learning applicato alla stima di potenza PV

Il Machine Learning ML è un sottoinsieme dell'intelligenza artificiale che sviluppa sistemi in grado di apprendere e migliorarsi sulla base dei dati che sono loro presentati.

Il Machine Learning ha ad oggi un'ampia diffusione anche in molti settori della vita quotidiana, data la sua potenzialità e semplicità concettuale all'origine: si tratta di sistemi che "imparano" in autonomia dai dati analizzati, senza l'intervento di modelli analitici o equazioni predeterminate.

Il riconoscimento di determinati pattern e correlazioni giunge invece dall'apprendimento automatico: con il susseguirsi di esempi (in termini di dati in analisi) il ML migliora le proprie prestazioni in modo adattivo, ripetendo lo svolgimento di determinate azioni o compiti fino al miglioramento degli stessi.

Solitamente vengono somministrate informazioni in input con i rispettivi output, così che l'algoritmo di ML comprenda come correlarli e come rispondere in termini di output a nuovi input (elemento molto rilevante che approfondiremo). Prese in considerazione le numerose applicazioni di natura pratica del ML, vengono spesso impiegate suddivisioni basate proprio su differenti approcci pratici:

Alberi delle decisioni/decision trees: si basano su grafi coi quali realizzano previsioni di conseguenze specifiche in output al variare di determinate decisioni in input

Clustering: modelli matematici con lo scopo di "raggruppare", ovvero associare oggetti ritenuti simili ed identificare specifiche strutture da inserire in suddetti gruppi

Modelli probabilistici: strutturano il processo d'apprendimento sulla base del calcolo delle probabilità, ad esempio collegando in un grafo l'insieme delle variabili aleatorie e le rispettive dipendenze condizionali

Reti neurali artificiali ANN: utilizzano per l'apprendimento algoritmi ispirati alla struttura, al funzionamento ed alle connessioni delle reti neurali biologiche (quelle dell'essere umano). Nel caso delle reti neurali cosiddette multi-strato si entra poi nel campo del Deep Learning (apprendimento profondo), in particolare se il numero di layer che le compongono è elevato.

Un'altra importante distinzione va operata per le due principali famiglie di algoritmi di ML, ovvero tra regressore e classificatore. Entrambi il tipo di algoritmo elaborano dataset composti da un certo numero di esempi per costruire un modello statistico, ma l'output generato come risposta ad un determinato input è un numero reale nel caso della regressione, mentre identifica una classe/categoria di elementi affini nell'altro.

Gli algoritmi di regressione saranno quelli impiegati nella previsione della potenza, mentre quelli di classificazione sono stati studiati ed applicati in vari paper per identificare le tipologie di guasto (vedasi l'introduzione).

### 1.3.1 Suddivisione del dataset

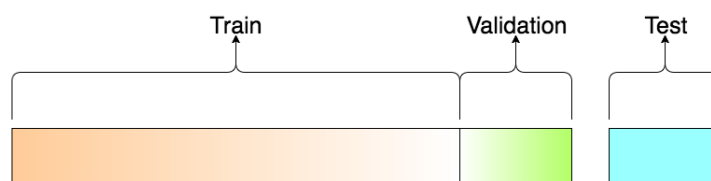
Nel caso in esame il ML attingerà da una raccolta di misure d'impianto, suddividendole dapprima in input (parametri ambientali) ed output (potenze generate) per poi utilizzare tali dati nel training dell'algoritmo. I dati vengono selezionati sulla base di criteri d'importanza temporale, per qualità o pulizia degli stessi, o altre cause e concause specifiche per la situazione in analisi (nel caso di impianti fotovoltaici è prassi impiegare l'intero primo anno di operatività dell'impianto, così da avere una rappresentazione completa a livello mensile e regionale dell'impianto nel suo momento di massima performance).

Durante la fase di training, poi, una parte del totale di dati in input assegnati al training viene tenuta da parte per la successiva convalida del network allenato. La convalida è il principale criterio di verifica della qualità di un network ML, sfrutta l'allenamento eseguito sulla porzione principale dei parametri e tenta di predire gli output della porzione di convalida partendo soltanto dai rispettivi input. Sulla base della convalida, il network viene ulteriormente ottimizzato, incorporando le nuove informazioni nella configurazione del modello. Tali dati di convalida sono "percepiti" dal modello, ma non impara da essi, viene migliorato indirettamente.

La seconda prova di validità eseguita sul ML è testarne la capacità di previsione su dati totalmente estranei al suo network, non facenti parte né della porzione adibita al training né quella usata per la convalida. Questa operazione è svolta sul ML già completo a training terminato, per ottenere riscontro sul possibile futuro funzionamento quando applicato a dati completamente nuovi, senza la possibilità di essere ottimizzato ulteriormente.

Il test set è solitamente selezionato automaticamente durante la suddivisione iniziale dei dati per presentare una certa varietà in ampiezza (o classe), così da offrire al network un primo confronto con dati reali variabili.

La spiegazione sulla suddivisione del dataset è stata adattata all'argomento trattato, ma i concetti si applicano allo stesso modo per tutti i ML, indicativamente come rappresentato in Figura 1.3.1.



**Figura 1.3.1** Suddivisione del dataset in train/allenamento, validation/convalida e test

Lo split ratio, ovvero le percentuali di suddivisione delle tre componenti del dataset adibito al training, non ha regole fisse, ma dipende comunque dal numero totale di campioni a disposizione e dalla natura del modello che si sta allenando.

Se lo specifico modello necessita di tanti dati per il training, si utilizzeranno molti valori proprio per il set di allenamento, mentre se sono presenti molti iperparametri conviene avere un validation set più corposo. Per rapido chiarimento, gli iperparametri sono quei parametri che fissano alcuni elementi del network del ML, come il numero di strati/layer o il numero di nodi per strato nel caso delle ANN.

In sintesi, gli iperparametri sono tutto ciò che caratterizza il modello ML, ma che non cambia durante la fase di training.

Come regola pratica, comunque, il training set dovrebbe valere tra 60% e 80% del totale, con validation e test set di dimensioni comparabili.

### 1.3.2 Pre-processing dei dati in input al modello

Occorre sempre verificare che i dati forniti al network ML per l'addestramento siano formattati correttamente e che i valori anomali (outliers) siano rimossi, trattandosi di valori errati o non affidabili. Il formato deve coincidere con quello richiesto dalla piattaforma utilizzata: nel caso del lavoro di tesi e della ricerca [9] si tratta di Matlab, ma alternative come TensorFlow, RapidMiner o Keras esistono.

Entrando nello specifico dei modelli di previsione della generazione PV, i dati storici impiegati in input e output (valori meteorologici e di potenza) possono presentare picchi anomali, componenti non stazionari dovuti alla grande variabilità delle condizioni meteorologiche, o infine dati mancanti per certi parametri in corrispondenza di dati presenti in altri.

La mancanza di dati storici può essere anche estesa a tutti i parametri, causa disconnessioni volontarie o per guasto, problemi con i sistemi di misura, interruzione della misura per manutenzione ecc., complicando ulteriormente la situazione.

Se lasciati inalterati, questi valori anomali possono impattare negativamente sulla capacità predittiva del modello, peggiorandone il training ed innalzando gli errori di stima, ecco perché il pre-processing dei dati di input è necessario per mantenere elevata l'accuratezza del modello. Vari metodi di pre-processing sono impiegabili: identificazione del ritardo storico, serie temporali trend-free, normalizzazione, trasformazione wavelet e mappe di auto-organizzazione (self-organizing map SOM). In particolar modo nel caso in esame, la normalizzazione è la metodologia più diffusa. Con questa

tecnica si riduce il range dei valori tra 0 e 1 per minimizzare l'errore di regressione e mantenere le corrette correlazioni tra i valori del dataset. Non normalizzare gli input significherebbe avere ampiezze ben diverse tra i vari parametri considerati: questo può facilmente causare in un ML lo spostamento di attenzione sui valori più grandi e trascurare i più contenuti, falsando le correlazioni e dipendenze che il modello stimerebbe. Applicando tale tecnica, invece, ogni tipo di input è scalato rispetto alla propria variabilità ed ottiene la giusta considerazione in fase di training.

La formula di normalizzazione per un certo dato di input  $I$  è la seguente:

$$I_{norm} = \frac{I - I_{min}}{I_{max} - I_{min}} \quad (2)$$

dove  $I_{norm}$  è il dato normalizzato,  $I_{max}$  il massimo pre-normalizzazione e  $I_{min}$  il minimo pre-normalizzazione.

Al momento della futura analisi delle performance del modello di previsione della potenza PV, per tornare alla scala corretta dei valori si devono applicare tecniche di post-processing corrispondenti a quelle di pre-processing impiegate in precedenza. Tra queste si trovano quindi la ricostruzione wavelet e la procedura di anti-normalizzazione. Un elemento di semplicità della normalizzazione è che, nel caso di parametri con minimo nullo (come ad es. l'irradianza), è sufficiente dividere per il valore di picco in pre-processing e moltiplicare per tale valore nuovamente in post-processing.

### 1.3.3 Selezione delle feature (scelta degli input)

Una volta pronti i dati (input e output) per l'addestramento ed il test, è importante selezionare gli input che possono aiutare a prevedere meglio l'output.

Spesso mettere più output a disposizione del ML significa migliorarne le prestazioni ma a volte fornire più input o opzioni per aiutare la previsione può portare a una situazione di overfitting.

L'overfitting è un problema in cui un network ML viene addestrato a funzionare al meglio solo per il set di dati su cui si è addestrato e predice risultati per lo più errati nella fase di test e applicandolo a nuovi dati. Sistemi prони a fenomeni di overfitting sono quelli con molti diversi input, ma pochi valori per ognuno. Anche l'opposto ha generale validità.

Il processo di selezione dei dati di input che possono aumentare l'accuratezza dei test è chiamato selezione delle feature, che consiste nella scelta di un sottoinsieme di dati rilevanti, di alta qualità e non ridondanti. Possibilmente è consigliare filtrare anche dati indirettamente ridondanti: con ciò si intende che se fisicamente, o comunque per ragioni reali e comprovate, un dato di input è strettamente dipendente dalla combinazione di altri due, metterli tutti e tre introdurrebbe a una quota facilmente evitabile di ridondanza.

Nel caso della produzione energetica PV, l'input con la maggiore correlazione con l'output di potenza è l'irradianza solare (l'irraggiamento, a livello istantaneo), ma altri parametri ambientali possono rilevarsi impattanti: temperatura atmosferica, temperatura di pannello, velocità e direzione del vento, umidità. Riconosciuto il numero limitato di parametri, ed il numero possibilmente elevato di misure nel corso di un anno, può non essere necessario operare particolare attenzione alla selezione delle feature, sapendo che il rischio di overfitting è limitato.

Si deve comunque porre attenzione al metodo e strumento con cui ogni input è stato estratto (o in certi casi calcolato a partire da altri): se infatti a crescita dei dati porta mediamente ad un miglioramento del comportamento del modello, ogni misurazione (o stima) giunge con un proprio margine di errore, non sempre facilmente identificabile. Nella fatalità che un parametro abbia grande incertezza o poca affidabilità, può quindi convenire filtrarlo dal training perché l'apporto positivo sarebbe negato dal peggioramento complessivo della stima, avendo introdotto un input con una aleatorietà ingombrante.

Tra le note tecniche di selezione delle feature, in [9] sono state utilizzate e comparate la selezione delle feature di correlazione (Correlation feature selection CFS) e la selezione delle feature di rilievo (Relief feature selection, ReliefF).

La tecnica CFS seleziona sottoinsiemi di caratteristiche in base a una funzione di valutazione euristica basata sulla correlazione, mentre ReliefF è un algoritmo che si adatta alle specifiche istanze, assegnando a ogni feature un peso che ne rappresenta la rilevanza nel modello.

Approfondendo la valutazione della correlazione di determinati parametri di input con la potenza, [8] utilizza questo metodo per selezionare le feature migliori.

La variazione dei parametri meteorologici dipende dalla posizione geografica e dalle condizioni atmosferiche; pertanto, non si verifica un impatto analogo di un parametro meteorologico sulla generazione di energia fotovoltaica in diverse località geografiche.

Di conseguenza, anche la correlazione dei parametri meteorologici e della produzione di energia fotovoltaica non sarà la stessa in diverse località. Eppure, le prestazioni di un modello di previsione dipendono fortemente dalla correlazione tra valori di input e output del modello stesso. In questo caso, occorre (se si presenta la necessità di filtrare gli input) applicare lo studio della correlazione dei diversi input meteorologici (l'irraggiamento solare, la temperatura atmosferica, la temperatura dei moduli, la velocità e il vento) con la produzione di energia fotovoltaica.

Il termine rappresentativo da analizzare è il coefficiente di correlazione  $r$  :

$$r = \frac{cov(X, Y)}{\sigma_x \sigma_y} \quad (3)$$

con due grandezze generiche (nel caso in esame un input e un output)  $X, Y$ , mentre  $cov(X, Y)$  rappresenta la covarianza di campione (4) e  $\sigma_x \sigma_y$  sono rispettivamente la deviazione standard di  $X$  e di  $Y$  (5).

$$cov(X, Y) = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{N - 1} \quad (4)$$

$$\sigma_x = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \mu_x)^2} \quad (5)$$

La deviazione standard rappresenta la dispersione dei dati dalla propria media, mentre la covarianza è una misura di quanto le due grandezze varino assieme (ossia della loro dipendenza reciproca), tuttavia la covarianza non ha specifici limiti, quindi è difficile da interpretare. La versione normalizzata della statistica si calcola appunto dividendo la covarianza per il prodotto delle due deviazioni standard ed esso è il coefficiente di correlazione.

Una correlazione positiva, quando il coefficiente di correlazione è maggiore di 0, significa che le due variabili si muovono nella stessa direzione. Quando  $r = +1$ , significa che le due variabili confrontate hanno una relazione positiva perfetta; quando una variabile cresce o decresce, l'altra si muove nella stessa direzione con la stessa ampiezza relativa.

Più il valore di  $r$  è vicino a +1, più forte è la relazione lineare, mentre un valore di 0 indica che le due grandezze confrontate sono totalmente indipendenti.

Il coefficiente di correlazione è quindi uno strumento potente nella valutazione dei parametri di input da scegliere: le variabili di input fortemente correlate all'output dovrebbero essere poste all'ingresso del modello di previsione, mentre le variabili di input debolmente correlate andrebbero escluse.

Un esempio indicativo, ma dalla lampante efficacia, è portato in [8], tramite il confronto di potenza prodotta da un piccolo impianto PV e l'irraggiamento disponibile nello stesso giorno.

Viene riportato il confronto tra gli andamenti in Figura 1.3.2.



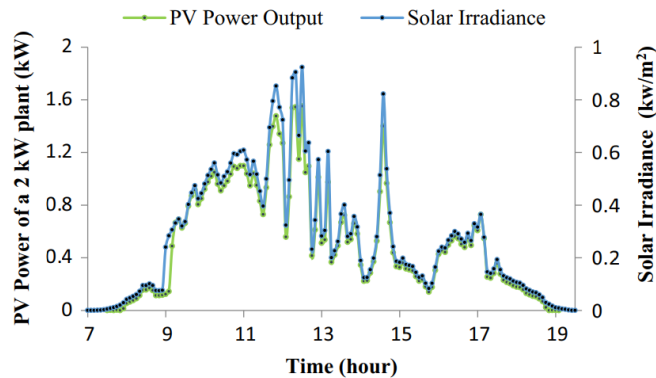


Figura 1.3.2 pattern di irradianza e potenza PV per un giorno

Preso tale giorno in esame, la correlazione risulta evidente (con probabili miglioramenti in giorni con cielo costantemente sereno e peggioramenti in giornate più nuvolose), e viene ulteriormente confermata dalla rappresentazione del coefficiente di correlazione in Figura 1.3.3, reso dalla retta che va a tagliare la distribuzione dei valori di potenza rispetto a quelli d'irradianza, con un valore  $r = 0.989$ , ben prossimo a +1.

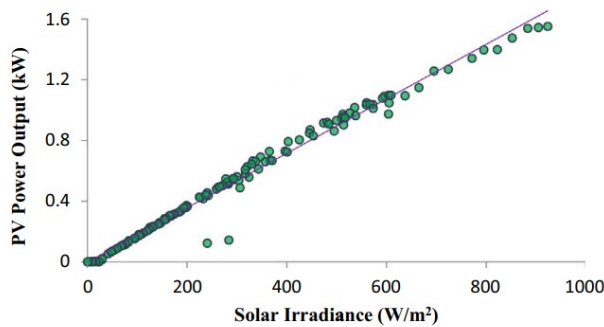


Figura 1.3.3 correlazione tra irradianza e potenza PV

Il risultato non sorprende particolarmente perché l'irradianza è, da un punto di vista fisico e concettuale, il parametro con il più stretto legame con la potenza prodotta dalle tecnologie PV. Altri parametri invece possono avere correlazioni variabili con la località (ad esempio la velocità del vento avrà impatto contenuto nel raffreddare i pannelli e farli produrre a maggiore efficienza in luoghi già naturalmente freddi).

Come precedentemente accennato, in genere l'accuratezza del modello di previsione della produzione di energia fotovoltaica può essere migliorata utilizzando un gran numero di vettori di input. Tuttavia, il costo computazionale e la complessità aumenteranno a causa dell'aggregazione dei parametri di input stessi. Pertanto, progettare un modello di previsione con un numero ottimale di vettori di input basato sulla correlazione è di estrema importanza.

Discusso ora della selezione di feature per creare l'ideale vettore di feature (valori adeguati di input selezionati per l'alta correlazione o con altre strategie), si possono mostrare gli stadi dell'addestramento del Machine Learning, in Figura 1.3.4.

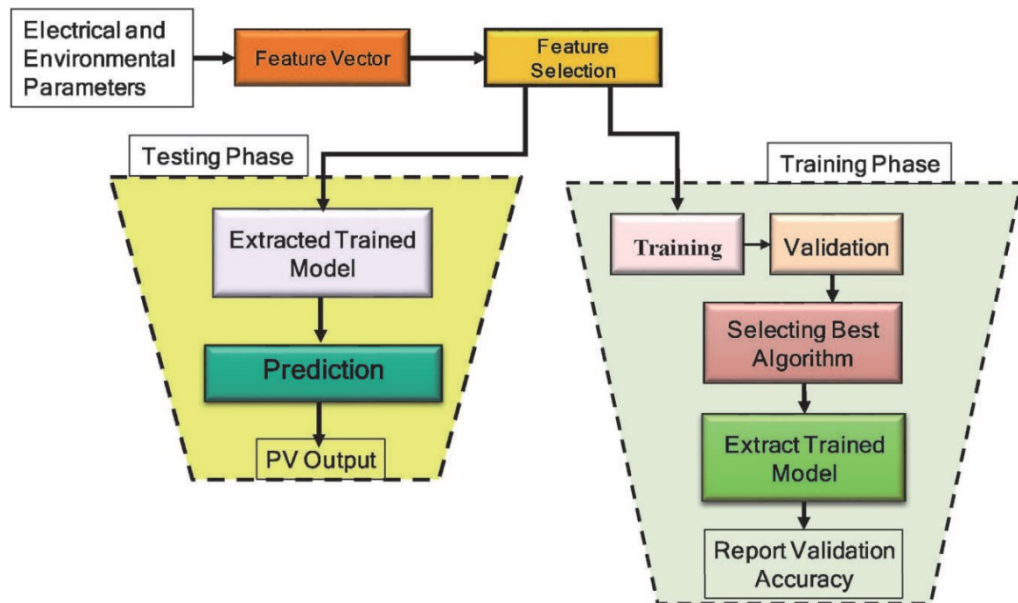


Figura 1.3.4 stadi di cui si compongono addestramento e test del Machine Learning

Ricapitolando con l'ausilio grafico:

- Si esegue la raccolta dei dati elettrici ed ambientali d'impianto (electrical and environmental parameters), si riuniscono in un vettore di feature contenente inizialmente tutti gli input e, prima di iniziare l'addestramento del ML, si esegue un filtro, una feature selection, per scegliere gli input effettivamente meritevoli di essere impiegati
- I dati vengono allora divisi in quelli adibiti all'addestramento (training), alla convalida (validation) ed al test.
- Nella fase d addestramento viene impiegato solo il set di dati corrispondente per quanto riguarda l'apprendimento, dopodiché il ML applica il modello risultante dall'algorithm attuale e lo valuta sul set di dati di convalida, ripetendo l'operazione alla ricerca del minimo errore di convalida (ovvero massima validation accuracy).
- Terminata la fase di training, il modello nella sua versione finale ed ottimizzata (extracted trained model) viene testato sull'ultimo set, svolge una previsione dell'output PV in potenza e così facendo estrae la precisione di stima nel caso di input mai affrontati in fase di training ( il separato dataset della fase d test ).

I temi di preparazione dei dati ed addestramento affrontati fin qui non sono differiti in particolar modo per le due tecniche di previsione della potenza prodotta, ma occorre fare una chiara distinzione per evitare fraintendimenti causa la limitazione linguistica:

- A. La previsione di potenza intesa come *Forecasting* utilizza varie metodologie tra cui il machine learning per anticipare la produzione energetica ed avere una stima prima del tempo (di quanto dipende dal modello impiegato e dalla sua precisione) [10 - 12]
- B. La previsione di potenza intesa come *Prediction* impiega differenti modelli, tra cui quelli ottenuti con Machine Learning, per, confrontati con dati preesistenti di input (o al limite appena misurati), stimare l'output e confrontarlo con quello misurato in cerca di eventuali anomalie e malfunzionamenti

L'obiettivo del lavoro si incentra su Prediction, ma, come visto ripetutamente, molti elementi impiegati in Forecast sono di interesse, in quanto si tratta di impiego di tecniche analoghe nello stesso settore energetico, seppur con scopi differenti. Da qui in avanti, per una questione di chiarezza, considerato che entrambi i termini condividono la traduzione italiana "previsione", la distinzione sarà fatta in inglese.

Le sezioni dedicate al Forecast andranno rapidamente svanendo, ma restano d'interesse per mostrare lo stato dell'arte del ML in campo PV, e soprattutto la pletera di tipologie di ML implementabili nel settore.

## 1.4 Modelli di previsione (forecast) della produzione energetica PV

I ricercatori classificano la previsione della produzione di energia fotovoltaica in diverse categorie in base a diversi fattori, come anticipato; tuttavia, non esistono criteri fissi per classificare la previsione dell'energia fotovoltaica.

La maggior parte delle categorie sono suddivise a seconda dell'orizzonte temporale di previsione, dai dati storici disponibili dell'irraggiamento solare, da altri modelli di dati meteorologici e dei metodi utilizzati per la previsione.

Richiamando quanto spiegato al termine del capitolo precedente, non ci si addenterà oltre nelle varie tipologie di forecast, se non per fare le dovute considerazioni sui tipi di Machine Learning utilizzati, con rispettivi punti di forza e debolezza, in vista dell'impiego successivo per la previsione (prediction).

Gli approcci al forecast sono svariati, e nel caso di modelli basati sul Machine Learning le scelte principali risultano 2:

- Reti neurali - Artificial Neural Network ANN: è il metodo più efficace e diffuso tra i ricercatori, utilizzato in diverse applicazioni, tra cui la prediction ed il forecast della produzione di energia fotovoltaica con un elevato livello di successo. L'ANN è ampiamente utilizzata nella maggior parte delle ricerche, a causa della non linearità dei dati meteorologici, che renderebbe vani altri tipi di modellizzazione.

In genere, infatti, la rete neurale è più adatta rispetto ai metodi statistici quando sussiste un legame non lineare e complicato tra i dati. La struttura delle reti neurali sarà approfondita nel capitolo seguente, ma può essere composta da uno o più strati (layer) interni che permettono la correlazione dei dati. Rivelandosi le reti neurali a singolo layer spesso inadatte a problemi complessi con varie non linearità, altre architetture e procedure di mappatura input-output sono state realizzate ed impiegate.

Tra queste sono comunemente impiegate multilayer perceptron NN (MLPNN), Multilayer feed-forward NN (MLFFNN), radial basis function NN (RBFNN), recurrent NN (RNN), general regression NN (GRNN), e adaptive neuro-fuzzy interface systems (ANFIS). Spiccano la varietà ed adattabilità delle reti neurali, ma possibili problematiche sono la necessità di molti dati in fase di training ed il rischio di overfitting. Tuttavia, come precedentemente evidenziato, il settore PV, con numerose misurazioni ed un numero limitato di parametri, genera un effetto di bilanciamento al riguardo.

- Macchine a vettori di supporto – Support vector machine (SVM): è un metodo di ML supervisionato basato sul principio di principio della minimizzazione del rischio strutturale (structural risk minimization SRM); SRM si occupa di minimizzare un limite superiore del rischio atteso. Pertanto, SVM è in grado di minimizzare l'errore dei dati di addestramento.

SVM è stato originariamente sviluppato da Vapnik per risolvere problemi di classificazione, ma è stato recentemente esteso ai problemi di regressione come il forecast (e la prediction). L'applicazione di SVM alla regressione delle serie temporali è nota come regressione vettoriale di supporto (support vectorial regression SVR).

La previsione della produzione di energia fotovoltaica è un tipico problema di analisi delle serie temporali, pertanto SVR è un metodo adatto in questo caso.

L'algoritmo SVR è un algoritmo di regressione non lineare. Il campione di dati delle serie temporali in ingresso viene mappato in uno spazio di feature a dimensioni superiori mediante una mappatura non lineare, quindi viene eseguita una regressione lineare in quello spazio.

Per come sono strutturati i modelli del tipo SVR, a dominarne le performance sono tre parametri specifici di tale tecnica. L'enorme sensibilità della performance del modello rispetto alla selezione accurata di tali parametri impone una notevole limitazione applicativa.

## 1.5 Modelli di previsione (prediction) della produzione energetica PV

Selezionati i dati (dopo pre-processing e feature selection) da dare in input all'addestramento del ML, si possono scegliere varie tecniche di machine learning, per confrontarle alla ricerca di quella dalla performance migliore. Una ricerca di natura simile è stata svolta in [8] e [13], seppur impiegando tecniche e scelte differenti.

La prima distinzione è eseguita tra le tecniche di machine learning di tipo tradizionale e le reti neurali artificiali ANN, già nominate in più occasioni, che verranno adeguatamente approfondite.

Sono svariati e diffusi i modelli di previsione basati sulla regressione per stimare la potenza PV in output: Regressione Lineare Semplice, Regressione di Processo Gaussiana (Gaussian Process Regression GPR) e l'albero di regressione M5P.

Il modello di regressione lineare semplice ha una relazione lineare tra la risposta in output e i parametri in input, GPR prevede un processo gaussiano che utilizza l'apprendimento "lazy" e una misura della somiglianza dei punti (funzione kernel) per prevedere il valore dai dati di addestramento per un punto non valutato. L'albero di regressione M5P, infine, utilizza un algoritmo che contiene istruzioni *if ed else* diramando una rete decisionale per cui la potenza prevista sarà il risultato di affermazioni del tipo "if ... then ... else ...".

Si entra ora nel merito delle reti neurali.

### 1.5.1 Reti neurali artificiali (artificial neural network) ANN

La rete neurale artificiale (ANN) (Figura 1.5.1) può essere considerata una replica del funzionamento del sistema nervoso umano, semplicemente artificiale nella natura.

La ANN ha tre strati/layer principali: input layer, output layer e hidden layer (strato nascosto).

Va precisato che nella ANN più semplice c'è soltanto un hidden layer (come è realizzata e descritta in [8]), ma gli hidden layer possono essere molteplici (Figura 1.5.1) e dopo i 3 si inizia solitamente a parlare di deep learning (con specifiche di funzionamento differenti che non saranno espanso qui).

L'input layer è costituito dai neuroni artificiali in cui avviene l'apprendimento vero e proprio ed è anche il layer in cui vengono inseriti i parametri di input  $X_1 \dots X_n$ : tale layer contiene un neurone per ogni parametro di input. Ogni neurone di questo strato ha dei pesi specifici  $w_1 \dots w_n$  (Figura 1.5.2), che sono adatti al preciso problema sottoposto alla rete neurale. Questi input, una volta pesati, vengono utilizzati negli hidden layer, dove sono sommati con dei bias  $b_1 \dots b_n$ , attraversano una funzione di trasferimento per hidden layer.

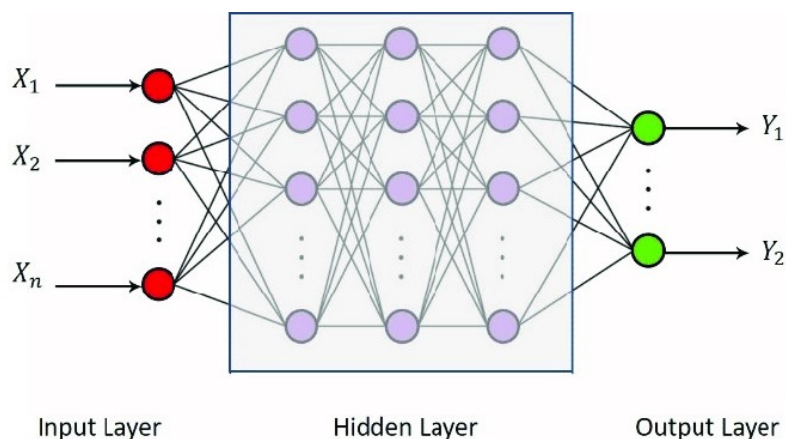


Figura 1.5.1 Struttura di una rete neurale con 3 hidden layers, n input e 2 output

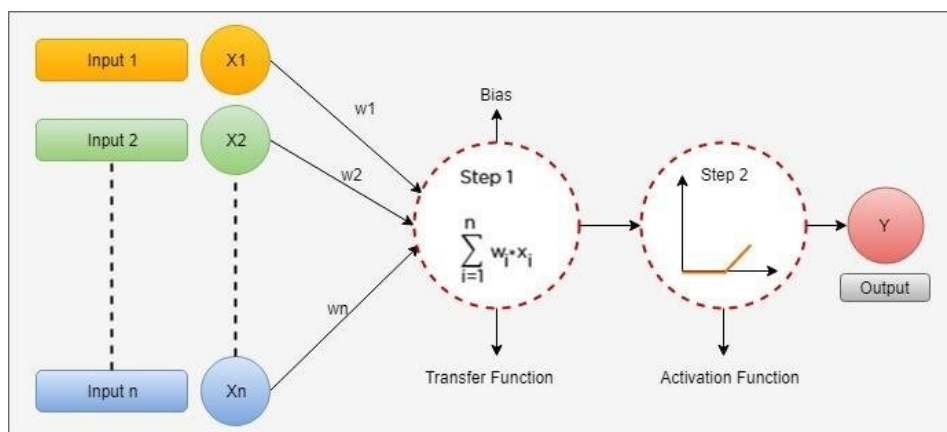


Figura 1.5.2 rappresentazione al dettaglio con indicazione di transfer e activation function

Le funzioni di trasferimento applicano (di consueto) una trasformazione non lineare, per poi divenire gli input degli hidden layer successivi ed infine della funzione di attivazione, che cerca di prevedere l'uscita o fornisce l'errore alla rete come feedback, risultando infine nell'output (Figura 1.5.2).

Il ruolo degli hidden layer e delle funzioni di trasferimento che contengono è proprio di introdurre un elemento essenziale di non linearità, dove la semplice combinazione lineare di input tramite bias e pesi non basterebbe a descrivere il legame complesso che può instaurarsi tra input e output. D'altronde, nel caso di semplici dipendenze di carattere lineare, si potrebbe fare affidamento direttamente a soluzioni più tradizionali come metodi di regressione lineare ecc. che invece risultano inaffidabili nel merito del caso affrontato.

Come si può vedere da un breve scorcio di possibili funzioni di trasferimento (presentate da Matlab) in Figura 1.5.3, la funzione di trasferimento può essere anche soltanto lineare pura, ma in quel caso la non linearità (se richiesta) sarà introdotta dalle funzioni di trasferimento appartenenti ad altri hidden layer.

Resta comunque prassi diffusa l'impiego ripetuto della stessa funzione di trasferimento per ogni hidden layer, con una diversa funzione di attivazione in output.

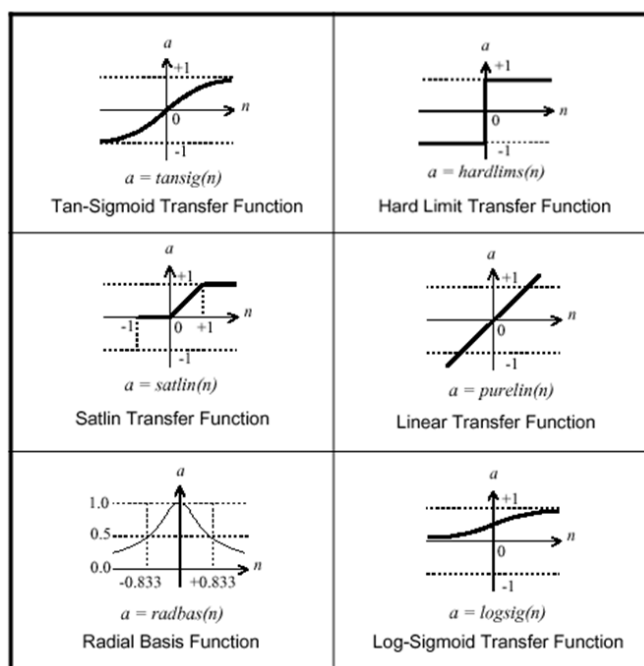


Figura 1.5.3 elenco di comuni funzioni di trasferimento per ANN

Il range in cui operano le funzioni di trasferimento mostrate ( $[0, 1]$ ,  $[-1, 1]$ ) permette di comprendere ulteriormente la comodità e spesso necessità della normalizzazione dei valori in fase di training della rete, poiché tali valori, normalizzati tra -1 e +1, produrranno risultati in uscita dalle fdT degli hidden layer nuovamente inclusi nello stesso range, perfetti per andare in input all'hidden layer successivo trasportando informazioni sull'ampiezza senza sacrificare la scala comune.

In merito allo schema nel suo insieme, il feedback risultante dalla funzione di attivazione funge da apprendimento per i neuroni di input, che tentano nuovamente di fornire input alla funzione di attivazione che aiutino a migliorare la previsione. Questo viene fatto correggendo ed aggiornando bias e pesi, affinché la descrizione complessiva del fenomeno da parte della rete sia la più prossima a quella reale o comunque ad una stretta correlazione tra input e output.

Ci sono diversi algoritmi di addestramento (training algorithms TA) impiegabili e disponibili nell'implementazione Matlab della ANN [8] e ognuno di essi ha i suoi vantaggi e svantaggi e ogni specifica applicazione può avere un TA specifico che fornisca una migliore previsione, a causa della natura dei dati trattati.

È sempre importante esplorare varie combinazioni di numero di hidden layer e di funzioni di addestramento per trovare la combinazione migliore che predica la potenza fotovoltaica con la massima precisione, come mostrato nella Figura 1.5.3. Lì viene mostrato un possibile diagramma di flusso utilizzabile come approccio di ottimizzazione ed è quello impiegato in [8], anche se ha carattere abbastanza generale.

L'algoritmo varia prima gli algoritmi di addestramento TA, poi il numero di hidden layer e quindi compie varie iterazioni. Durante ogni prova l'algoritmo memorizza la rete con le migliori prestazioni per applicarla sul test dataset. La rete finale migliore viene quindi utilizzata per prevedere la potenza fotovoltaica utilizzando le variabili di input.

Al fine di confrontare le varie feature scelte, le tecniche di ML e anche le varie tecniche di selezione delle feature si possono usare vari parametri statistici come il coefficiente di relazione precedentemente definito, ma le varie definizioni degli stessi saranno riportate nel capitolo dei KPI, per affiancare ai parametri di chiave d'impianto anche quelli della valutazione statistica delle performance del modello.

In [8] i parametri a disposizione per il training erano la temperatura ambientale, la temperatura di pannello, l'umidità relativa, l'irraggiamento, l'accumulo di pulviscolo e la velocità del vento in input, con la potenza in output.

In conclusione, la rete neurale ha esibito migliori capacità di prediction della potenza rispetto al modello di regressione lineare ed al GPR (ML di tipologia classica). Inoltre, l'opzione senza l'utilizzo di tecniche di selezione delle feature ha ottenuto errori più contenuti rispetto ad entrambe le tecniche CFS e Relieff.

Questo si dimostra un'ulteriore conferma di come, nello specifico settore di previsione PV in analisi, i parametri a disposizione siano già limitati in numero e legati fisicamente alle performance d'impianto, per cui escluderne alcuni può risultare più deleterio di quanto atteso per una situazione più comune di impiego di ML.

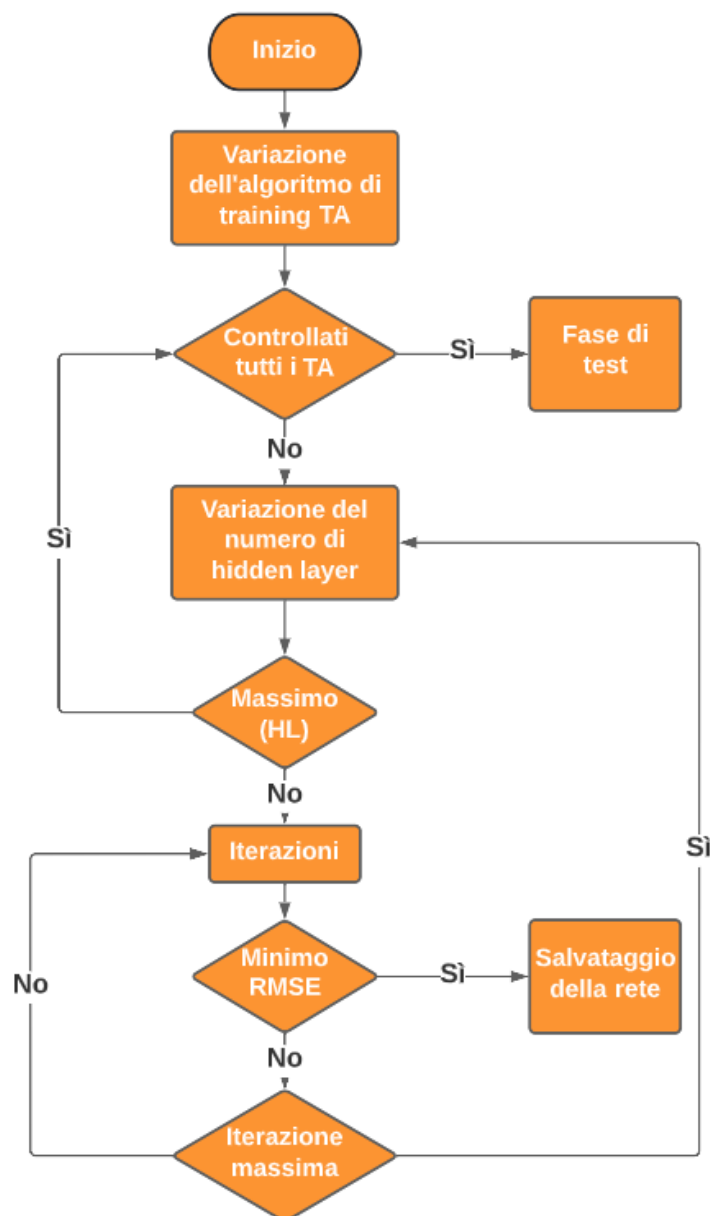


Figura 1.5.4 Metodo di ricerca dell'ANN ottimale per la previsione di potenza PV

La modellizzazione svolta in [13], seppure con tensione e corrente, piuttosto che direttamente la potenza in output, offre spunti interessanti e un'introduzione a un'altra versione di reti neurali dalla particolare struttura, nonché richiamo alla struttura appena vista. Viene innanzitutto evidenziata la ricorrente conferma nella letteratura delle ottime performance di ANN in situazioni di forecast e prediction rispetto ad approcci ML più tradizionali. Le ANN sono riconosciute come i modelli più diffusi anche in virtù di riuscite applicazioni pratiche.

Le reti in questione, impiegate e comparate, sono la rete nonlineare autoregressiva con input esogeni (nonlinear autoregressive network with exogenous inputs NARX) e la rete neurale di tipo deep feed-forward (DFF), o semplicemente feed forward.

Una variabile (nel nostro caso un input) di tipo esogeno è stata ottenuta tramite misura al di fuori del modello e poi inserita nel modello. Si tratta soltanto delle misure dei parametri ambientali legati all'impianto PV.



Una rete di tipo feed-forward è analoga alla struttura generale vista in precedenza: tra le ANN è la più comune e versatile, per cui fa spesso da riferimento. Espandendo i concetti già analizzati in merito, l'obiettivo di una rete feed-forward è quello di approssimare una funzione  $f$  definendo una mappatura  $y = f(x; \theta)$  e apprendendo il valore dei parametri  $\theta$  (bias, pesi) che danno come risultato la migliore approssimazione della funzione, ed esprimono la più precisa correlazione tra input e output.

Questi modelli sono chiamati feed-forward perché l'informazione passa attraverso la funzione valutata da  $x$ , attraverso i calcoli intermedi utilizzati per definire  $f$  e finalmente all'uscita  $y$ , come rappresentato in Figura 1.5.2. Non ci sono connessioni di feedback in cui le uscite del modello vengono reimmesse nello stesso, o la ANN diventerebbe di tipo ricorrente (RNN) o regressivo.

In [13], proprio NARX impiega un feedback pesato tra diversi layer di neuroni, a differenza di DFF. Questo permette al modello di considerare valori passati delle variabili, svolgendo analisi di tipo time series (serie temporali). In tal senso, grazie all'architettura a loop chiuso, l'output stimato in corrispondenza di un certo istante dipende anche da tutti quelli stimati agli istanti precedenti o dagli errori di stima agli istanti precedenti. Si può così migliorare il forecast di valori futuri, ma è necessario un database di dati chiari e consequenziali, nonché l'incremento in complessità rispetto al feed-forward.

Per queste ragioni, nonostante NARX abbia performance migliori nel forecast rispetto a DFF, la richiesta di alta qualità e costanza delle misure e la maggiore difficoltà di realizzazione, fanno preferire DFF almeno per scopi di prediction.

Un ultimo elemento interessante portato in evidenza in [13] è rappresentato da valutazioni in merito alla qualità dei metodi/algoritmi di training (TA) più diffusi.

Questi includono l'approccio di addestramento Gradient Descent, Resilient Back-propagation,

Levenberg-Marquardt (LM) e l'approccio Gauss-Newton. Levenberg-Marquardt è stato scelto per l'addestramento di entrambe le topologie sviluppate (NARX e DFF) poiché meno sensibile alle convergenze locali, fornendo quindi un migliore approccio di apprendimento che offre un compromesso equilibrato tra stabilità e velocità di addestramento. Tale è anche l'algoritmo di default proposto dal toolbox delle reti neurali e deep learning di Matlab.

## **2 Key Performance Indicators KPI**

Lo scopo di questo capitolo è quello di definire alcuni tra i principali parametri che influenzano le prestazioni, l'affidabilità, redditività, impatto ambientale ed integrazione nella rete dei sistemi e dei componenti fotovoltaici. Tali parametri sono gli indicatori di prestazione chiave (KPI) di un impianto fotovoltaico.

Solitamente si tratta delle grandezze più comuni in grado di descrivere il funzionamento d'insieme di una struttura PV, per valutarne il cambiamento di performance negli anni e operare confronti di con impianti nella stessa o altre regioni, anche di taglia diversa.

Nel corrente capitolo, però, saranno anche esposti KPI di impiego più sporadico e specifici per l'accuratezza e incertezza delle stime di produzione PV tramite modellizzazione.

In particolare, i parametri più classici, relativi alle prestazioni, sono riportati nella sezione 2.1, mentre quelli relativi alla modellazione e alla previsione sono riportati nella sezione 2.3.

KPI specifici di monitoraggio, ambito economico ed impatto ambientale non verranno presentati, in quanto hanno rilievo inferiore in merito agli argomenti trattati, ma restano di possibile interesse per ricerche con scopi differenti. Verranno comunque elencate brevemente situazioni nella letteratura scientifica in cui valutazioni di diverso tipo sono direttamente collegate ai KPI tecnici più tradizionali. Nello stato dell'arte del settore, i KPI sono ripetutamente utilizzati per le funzioni più disparate, poiché considerazioni altamente specifiche e strettamente legate a ubicazione, taglia e struttura dell'impianto privano eventuali considerazioni (di carattere tecnico, economico, ecc) della ricercata generalità d'applicazione.

Questi, infatti, permettono di esprimersi sulla situazione economica, presente e futura dell'impianto, come realizzato per una centrale da 30 kWp in Suriname [14], o di evidenziare il collegamento tra costo livellato dell'elettricità (Levelized Cost of Electricity LCOE) e le tendenze del capacity factor di impianti PV distribuiti lungo tutto il territorio statunitense [15].

In situazioni più comuni le considerazioni tratte dagli andamenti di questi indicatori rimangono nell'ambito tecnico [16], con eventuale approfondimento di come concause ambientali ne possano impattare valori ed andamento, con sensibilità variabili a certi fenomeni e grandezze rispetto ad altri. Seguendo tale approccio si può analizzare l'effetto sulle performance d'impianto della sporcizia e polvere accumulate (soiling effect) [17 - 20], di temperatura e velocità del vento in prossimità dell'impianto [21],[22], dello specifico impatto della temperatura sul Performance Ratio (che sarà approfondito a breve ed affrontato con tecniche di correzione) [23] ed infine come i diversi fattori ambientali abbiano un ruolo preciso nel deterioramento delle prestazioni dell'impianto e degli specifici pannelli [24 - 26].

Nello studio svolto in Suriname [14], oltre a previsioni di natura economica sul futuro dell'impianto analizzato, sono monitorate e riportate le condizioni di funzionamento ed un sistema di misurazione e registrazione dei dati fornisce gli input per il calcolo di alcuni indicatori chiave di prestazione (KPI) standard standard, poi confrontati con i valori attesi ottenuti dalla modellazione del sistema utilizzando il software PVSyst.

Nonostante in questo caso la modellizzazione venga realizzata da software esterni invece di realizzare uno specifico modello ad hoc, il concetto di KPI modellizzati da confrontare con quelli di misura è ben presente e sarà al centro dei capitoli successivi.

Come evidenziato nei paper brevemente elencati, l'energia elettrica prodotta da un impianto fotovoltaico dipende da diversi fattori, tra cui il più importante importante è la radiazione solare che raggiunge la superficie dei moduli fotovoltaici, ma è stato ampiamente esposto come le prestazioni di un impianto fotovoltaico dipendano più o meno fortemente anche dalle condizioni climatiche e ambientali locali. In definitiva, la progettazione tecnica di un impianto PV (in particolare del campo dei pannelli PV) determina la correlazione della produzione di energia con le condizioni locali, mentre i KPI ne permettono un'analisi di performance al netto delle stesse.

I fattori che influenzano la produzione energetica dell'impianto si dividono tra quelli intrinsecamente legati alla localizzazione della struttura PV (considerando anche a livello più specifico la forzatura a determinati orientamenti ed inclinazioni per adattarsi alla fisionomia dell'eventuale edificio o struttura d'appoggio) e gli altri, dovuti alle scelte tecniche di componentistica, collegamenti e architettura del sistema in generale.

I fattori dovuti al posizionamento impattano su generazione di potenza, disponibilità e performance del sistema ed introducono elementi di incertezza a causa della variabilità annuale, dei trend a lungo termine difficilmente identificabili e dell'errore di misura o scarto di stima delle grandezze in questione.

Esse si suddividono in:

- Radiazione solare: l'elemento di massimo interesse è l'irradianza globale nel piano d'array (global plane of array irradiation  $G_{POA}$ ), che dipende da orientamento ed inclinazione dei singoli pannelli che compongono l'impianto. Può essere direttamente misurato lungo il POA, o essere calcolato per scomposizione e trasposizione a partire dall'irradianza globale sul piano orizzontale (global horizontal irradiation GHI), la componente normale DNI e la diffusa DIF.
- Elementi meteorologici: temperatura dell'aria, velocità e direzione del vento, valore di umidità
- Situazione ambientale: albedo delle superfici, quantità e frequenza del deposito di neve, polvere, particelle inquinanti

Ragionando invece sui fattori legati alle scelte tecniche, essi possono essere i pannelli PV scelti, le tipologie e taglie degli inverter, il resto della componentistica ecc. Tutto ciò impatta sulla disponibilità e performance dell'intero impianto, nonché sulla sua complessiva affidabilità, tutti e tre elementi che sono suscettibili all'effetto di possibili guasti, problemi di collegamento, distacchi dalla rete ecc. che introducono elementi di incertezza e variabilità al pari dei fattori di posizionamento.

Solitamente le misure dei dati impiegati nei KPI variano per metodologia d'acquisizione con la taglia dell'impianto considerato: con il crescere della stessa si passa dall'impiego di valori risultanti da modelli meteorologici e solari, a modelli dello stesso tipo adattati alle proprietà del sito, fino a giungere a misure eseguite in loco, eventualmente compensate da dati di modello nei momenti di mancanza/gap.

Un'altra categorizzazione dei dati necessari al calcolo dei KPI è lo span temporale di disponibilità dei valori: valori misurati o modellati con maggiore frequenza permetteranno una maggiore libertà nella precisione temporale dei KPI, con eventuale integrazione lungo periodi maggiori se desiderato.

I valori possono quindi spaziare da valori medi sul lungo periodo (annuali o mensili), a serie temporali giornaliere fino a misure con frequenza oraria o addirittura del quarto d'ora o minuto.

## 2.1 Key Performance Indicators KPI legati alla performance dell'impianto PV

Il rendimento di un impianto fotovoltaico è fortemente influenzato da diversi fattori: la risorsa solare disponibile nel sito di installazione, la scelta della tecnologia e l'impatto su questa tecnologia da parte dei vincoli ambientali (come le variazioni di temperatura e il coefficiente di temperatura dei dispositivi interagiscono) e dalle specifiche d'installazione dell'impianto (se è presente un inseguitore o meno, l'entità dell'eventuale inclinazione dei pannelli ecc.)

Nel valutare la produzione di un impianto fotovoltaico, i parametri definiti in questa sezione aiuteranno a differenziare le cause che determinano le differenze nell'energia totale prodotta e nelle sue variazioni nel tempo. I singoli KPI saranno presentati in una forma generalizzata o specifica dal punto di vista della finestra temporale considerata, seppure quasi tutti si applichino a valori giornalieri/mensili/annuali.

I primi 3 termini presi in considerazione sono le Rese (Yields). Per una questione di riconoscibilità e diffusione internazionale di molti dei KPI principali, i nomi saranno spiegati, ma continueranno ad essere riportati in lingua originale nel corso di tutta la sezione.

Gli indicatori di rendimento sono legati all'energia incidente nel piano d'array fotovoltaico. Sono definiti, come la maggior parte degli indicatori chiave, nell'ottica di facilitare il confronto tra diversi impianti fotovoltaici; sono perciò indipendenti dalle dimensioni del campo considerato. Gli Yield possono essere concettualmente ordinati in funzione del flusso di energia che scorre dalla radiazione solare fino alla rete.

Il rendimento di riferimento Reference Yield fornisce informazioni sulla risorsa solare disponibile, mentre il rendimento d'array Array Yield include le perdite di raccolta dell'array ed il rendimento energetico finale Final Yield comprende anche le perdite complessive del sistema ed è strettamente legato al valore dell'energia immessa in rete.

### 2.1.1 Reference Yield ( $Y_r$ )

Il reference yield  $Y_r$ , che rappresenta la risorsa solare disponibile, è definito come il rapporto tra il totale irraggiamento lungo il piano di array  $H$  e l'irradianza di riferimento  $G_{ref}$  in condizioni standard STC:

$$Y_r = \frac{H}{G_{ref}} \quad (6)$$

In condizioni standard  $G_{ref} = 1000 \text{ W/m}^2$ , la massa d'aria vale 1.5AM e la temperatura dell'aria è  $T = 25 \text{ }^\circ\text{C}$ . L'unità di misura dell'irraggiamento  $H$  è  $\text{kWh/m}^2$ , mentre dell'irradianza  $G_{ref}$  è  $\text{kW/m}^2$ , per cui consegue che l'unità di misura di  $Y_r$  è  $\text{kWh/kW}$ , che rappresenta la risorsa di radiazione solare per l'impianto nella forma di ore equivalenti ad irradianza di riferimento in condizioni standard in una certa finestra temporale. L'irraggiamento può essere misurato con piranometri o con celle di riferimento. Le celle di riferimento possono essere scelte con una tecnologia simile a quella dei moduli fotovoltaici dell'impianto, per esibirne un simile comportamento.

Le celle di riferimento, però, posseggono alcuni svantaggi rilevanti: la loro risposta dipende dai parametri ambientali locali (temperatura e spettro solare) ed il segnale di uscita può risultare meno stabile e accurato di conseguenza.

Esistono anche celle di riferimento corrette per la temperatura, che possono essere accurate quanto i piranometri, ma sono molto costose, per cui si tende a preferire piranometri ben mantenuti, per la maggiore accuratezza e stabilità di misura, che favorisce la comparazione tra siti e tecnologie diverse.

### 2.1.2 Array Yield ( $Y_A$ ) , Final Yield ( $Y_f$ )

L'array yield  $Y_A$  è definito come il rapporto tra l'energia DC prodotta per un certo array fotovoltaico  $E_A$  [kWh] e la sua potenza di picco nominale  $P_0$  [kWp] :

$$Y_A = \frac{E_A}{P_0} \quad (7)$$

Il rendimento dell'array include soltanto le perdite di cattura, che possono essere causate da effetti di temperatura, sporcizia, ombreggiamento parziale, scarso inseguimento dell'MPP e altro ancora. Può anche essere interpretato come il numero di ore equivalenti, in un periodo di tempo definito, durante le quali un sistema fotovoltaico dovrebbe funzionare alla sua potenza di picco per fornire l'energia totale DC in uscita.  $P_0$  è uguale alla somma di tutte le potenze di picco in condizioni STC dei pannelli che compongono l'array all'avvio dell'operazione dell'impianto.

Il final yield del sistema PV  $Y_f$  , infine, è dato dal rapporto tra l'energia AC prodotta dall'impianto  $E_{out}$  [kWh] e la potenza di picco nominale dell'array PV installato  $P_0$  [kWp] :

$$Y_f = \frac{E_{out}}{P_0} \quad (8)$$

Esso include anche le perdite di conversione dal lato DC a quello AC dell'impianto.

Gli yield di energia (array e final) sono spesso impiegati per stimare la potenziale produttività di una installazione fotovoltaica e dipendono fortemente dalla radiazione solare e quindi dal luogo di installazione, nonché da anno ad anno ( e da mese a mese), data la variabilità temporale e spaziale della fonte energetica solare.

### 2.1.3 Expected Energy Yield ( $Y_E$ ) – Final Yield atteso

Si propone infine il rendimento energetico atteso  $Y_E$  , il quale descrive il rendimento energetico previsto per un determinato array in corrente alternata (AC), calcolato in base a misurazioni e simulazioni. Questo è il primo KPI direttamente collegato ad uno scenario di stima/modellizzazione dell'impianto in analisi ed esprime l'energia AC che dovrebbe essere prodotta dall'impianto in un determinato arco temporale. Esso si ottiene dalle simulazioni realizzate sulla base delle misurazioni (solitamente) di irradianza e temperatura e si esprime come:

$$Y_E = \frac{E_E}{P_0} \quad (9)$$

$Y_E$  [kWh/kWp] rappresenta l'energia AC specifica stimata e riprende l'equazione (8) del final yield, sostituendo all'energia AC effettivamente prodotta quella attesa  $E_E$  , calcolata sulla base delle specifiche di impianto e con i valori di temperatura ed irradianza in input. Questo KPI non è direttamente rappresentativo del funzionamento dell'impianto, ma dipende principalmente dalla precisione della simulazione eseguita, e dovrebbe pertanto essere utilizzato soltanto come strumento di riconoscimento di anomalie o imperfezioni nella performance.

### 2.1.4 Performance Ratio (PR)

Il KPI più utilizzato per rappresentare l'affidabilità d'impianto rimane il performance ratio, che fornisce un'indicazione delle prestazioni in condizioni reali degli impianti fotovoltaici.

Esso è definito come il rapporto tra il final yield ed il reference yield:

$$PR = \frac{Y_f}{Y_r} = \frac{E_{out}}{P_0} / \frac{H}{G_{ref}} \quad (10)$$

Ricordando che il final yield è uguale all'energia in uscita dall'impianto verso la rete in AC

ponderata per la potenza di picco dell'impianto in corrente continua e che il reference yield è l'irraggiamento totale nel piano dell'array diviso per l'irradianza di riferimento, il PR risulta equivalente al rapporto tra l'energia effettiva prodotta AC, che può essere immessa in rete, e l'energia resa disponibile dalla fonte solare.

Si tratta di un parametro che considera l'impianto nel suo insieme, tenendo conto di tutte le perdite del sistema fotovoltaico, quali la perdita di efficienza di moduli e inverter, perdite resistive nei cavi, effetti della temperatura e dell'irraggiamento, ombreggiamento lontano e vicino, sporcizia, interruzioni dei componenti, ecc.

Il performance ratio, essendo un rapporto tra yield, non ha unità di misura e può essere calcolato (a seconda degli yield da cui viene estratto) su base giornaliera, mensile o annuale.

Come presentato in precedenza, l'informazione sull'irraggiamento, fondamentale per il calcolo del PR, può essere ottenuta tramite diversi percorsi o misurazioni: i piranometri restano complessivamente la tecnologia più precisa ed affidabile, ma necessitano di accurata pulizia e manutenzione; le celle di riferimento devono essere selezionate con tecnologia molto prossima a quella dei pannelli e offrono un certo margine di imprecisione a meno di correzioni di temperatura (che solo le più avanzate includono). Nella maggior parte dei casi anche i valori ottenuti da modelli satellitari sono riconosciuti come sufficientemente affidabili, ed offrono un ottimo compromesso dal punto di vista tecnico ed economico.

Installed	Location	Range of PR	Average PR
1980s	Worldwide	0.50 – 0.75	Individual estimates
1990s	Worldwide (source 1)	0.25 – 0.90	0.66
1990s	Worldwide (source 2)	0.50 – 0.85	0.65 – 0.70
1990s	Germany	0.38 – 0.88	0.67
2000s	France	0.52 – 0.96	0.76
2000s	Belgium	0.52 – 0.93	0.78
2000s	Taiwan	<0.3 - >0.9	0.74
2000s	Germany	0.70 – 0.90	0.84

Figura 2.1.1 Valori medi e range di variazione dei PR per varie installazioni nel tempo

### Yearly integrated Performance Ratio - Europe

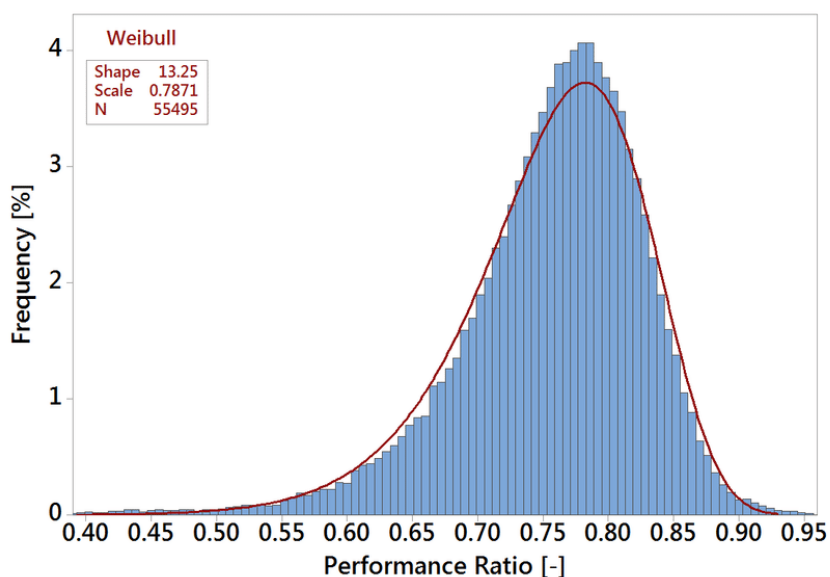


Figura 2.1.2 distribuzione di performance ratio da misurazioni europee

A livello globale i performance ratio annuali hanno esibito una tendenza all'incremento complessivo [27] (Figura 2.1.1), in quanto la ricerca di incremento della produttività degli impianti PV ha portato all'impiego di migliore manutenzione, a maggiore affidabilità dei componenti, al miglioramento della progettazione e ad un maggior numero di impianti fotovoltaici di grandi dimensioni, solitamente di più semplice gestione. La raccolta di valori da tutta Europa in Figura 2.1.2 mostra una distribuzione tipica che varia da circa 0.6 a 0.9, con tutti gli altri valori nel mezzo e 0.78 il più comune.

Localmente, però, i PR tendono a diminuire annualmente, a causa del naturale invecchiamento dei pannelli, in primo luogo, e dell'impianto nel suo insieme. Manutenzione, monitoraggio e pulizia del sistema possono rallentare il fenomeno, mentre guasti e noncuranza possono rapidamente accentuarlo.

Il performance ratio di un sistema è influenzato da vari fattori, ambientali e di specifica tecnica, per cui può essere consigliabile impiegarne versioni modificate per tenere conto di tali parametri o situazioni. La più comune tra queste variazioni è il PR corretto in temperatura per rimuovere (per quanto possibile) l'impatto mensile o annuale della temperatura sulle prestazioni dell'impianto.

#### 2.1.4 Performance Ratio corretto in temperatura di riferimento ( $CPR_{STC}$ )

Per normativa standard, al fine di limitare le variazioni di PR dovute alle variazioni temperatura, si definisce il performance ratio corretto STC, calcolato adeguando i valori di potenza ad ogni misurazione con la rispettiva differenza tra temperatura reale del pannello e temperatura di riferimento STC, ovvero 25 °C. La formula risulta:

$$CPR_{STC} = \frac{Y_{f,corr}}{Y_r} = \frac{E_{out,corr}/P_0}{H/G_{ref}} \quad (11)$$

dove l'energia corretta  $E_{out,corr}$  è stata ottenuta correggendo la potenza per ogni misura e poi eseguendo la sommatoria:

$$E_{out,corr} = \sum P_{out}/C \quad (12)$$

Il fattore di adeguamento della temperatura  $C$  si calcola per ogni misurazione come spiegato nel paragrafo precedente:

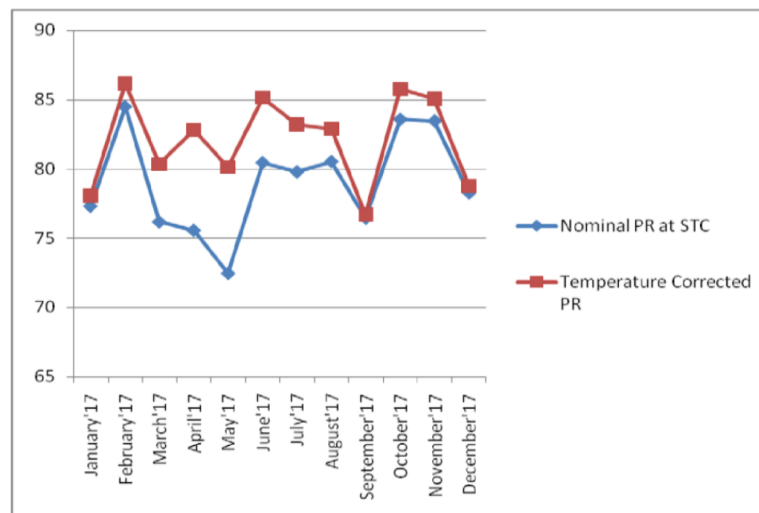
$$C = 1 + \gamma \cdot (T_{mod} - T_{ref}) \quad (13)$$

con:

- $T_{mod}$  la temperatura del pannello in °C
- $T_{ref}$  la temperatura di riferimento STC pari a 25°C
- $\gamma$  il coefficiente di temperatura per la potenza massima (in [1/°C]) ed esprime la perdita di potenza del pannello per ogni grado sopra la temperatura STC

Il performance ratio corretto è spesso indicato anche come Weather Corrected Performance Ratio, dato che viene adeguato agli effetti indiretti della temperatura dell'aria su quella dei pannelli.

Un esempio dell'effetto di tale operazione sull'andamento del KPI in questione è estratto da [29] è mostrato in Figura 2.1.3:



**Figura 2.1.3** Differenti andamenti di PR mensile con e senza correzione di temperatura nella repubblica del Ciad

Concettualmente il CPR è un valore equivalente di PR nello scenario in cui l'impianto avesse funzionato tutto il tempo in condizioni standard, per cui tende solitamente ad essere minore del PR durante i mesi freddi (in cui il contributo di differenza di temperatura rispetto ai 25 °C è negativo e va a ridurre la potenza adeguata, perché elimina gli effetti positivi delle minori temperature sulle prestazioni dei pannelli), ma come si vede in figura, se l'anno o la regione (in questo caso il caldo Ciad) considerata presentano in media elevate temperature, l'innalzamento del CPR può essere più comune tra i mesi.

Un caso più standard, nel Montana, è riportato [30] in Figura 2.1.4, dove la distinzione tra mesi caldi e freddi è più netta e si può apprezzare il positivo effetto di omogeneizzazione dei dati mensili rispetto al PR normale.

Entrambi gli scenari sono stati riportati per evidenziare che, seppur riferendosi al valore standard di temperatura nello svolgere i calcoli, esso non tiene conto della specifica regione e pertanto porta a variazioni anche annuali (con  $CPR_{STC,anno} \neq PR_{anno}$ ). Si può allora scegliere di adottare un'altra formulazione del CPR, come proposta da PVsyst, ovvero valutando rispetto alla temperatura di pannello annuale.



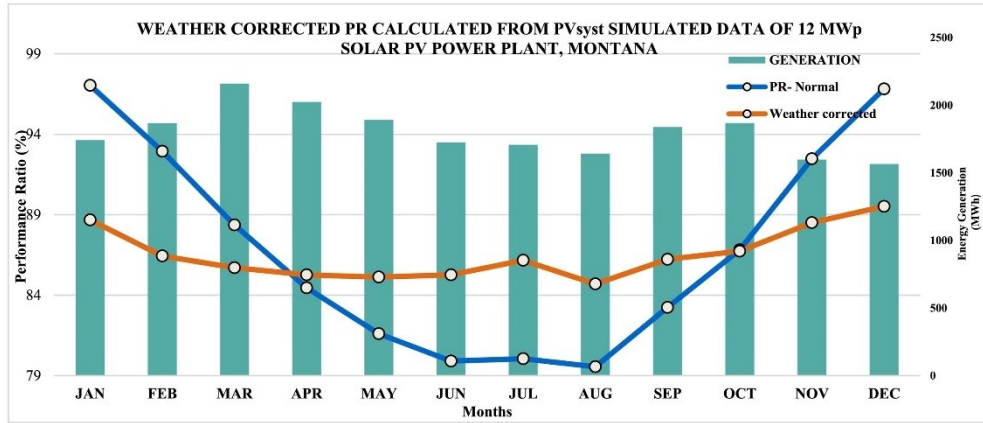


Figura 2.1.4 PR e CPR per impianto in Montana

### 2.1.5 Performance Ratio corretto con la temperatura media annuale ( $CPR_{T_{avg}}$ )

Un altro modo di esprimere il performance ratio corretto con la temperatura, allo scopo di limitare la variabilità stagionale, ma senza modificare il PR annuale, è con la seguente formula:

$$CPR_{T_{avg}} = E_{out} / \left( P_0 \cdot \sum_{hours} \frac{G}{G_{ref}} \cdot (1 + \gamma \cdot (T_{mod} - T_{avg})) \right) \quad (14)$$

$G$  è semplicemente il valore di irraggiamento (in questo caso orario) ed il calcolo è simile al caso precedente, ma la correzione viene applicata all'interno della sommatoria di irradianza piuttosto che nella potenza, e  $T_{avg}$  è la temperatura di pannello annuale media pesata con l'irraggiamento:

$$T_{avg} = \frac{\sum_{hours} (G \cdot T_{mod})}{\sum_{hours} G} \quad (15)$$

### 2.1.6 Fattore di capacità – Capacity Factor (CF)

Il capacity factor (come il  $CPR_{T_{avg}}$ ) non è un KPI standardizzato. Esso è definito come rapporto tra l'energia prodotta annualmente e l'energia che sarebbe stata prodotta assumendo che l'impianto generasse costantemente la potenza massima per tutto l'anno:

$$CF = \frac{E_{out}}{P_0 \cdot 365 \cdot 24} \quad (16)$$

L'utilità nella valutazione dello specifico impianto è limitata, ma il CF può essere impiegato nel confronto con altre fonti per la produzione di energia elettrica, con le rispettive caratteristiche produttive. L'impiego più comune nella letteratura è di paragonare la producibilità PV con le fonti termiche tradizionali.

### 2.1.7 Rapporto di superficie – Surface Ratio (SuR)

Il surface ratio è un altro KPI non standardizzato. Si ricava dalla definizione del PR, sostituendo la potenza nominale con la superficie dei pannelli d'array  $A$ . Si ottiene quindi:

$$SuR = \frac{E_{out}}{A} / \frac{H}{G_{ref}} \quad (17)$$

Tale parametro si misura in  $[W/m^2]$  e può essere utile nel confrontare la superficie coperta da diversi impianti PV, in relazione con l'energia prodotta.

Si introducono ora alcuni KPI della stessa natura di quelli proposti finora, ma adattati a specifiche circostanze.

### 2.1.8 Bifacial Performance Ratio (BPR)

Per moduli bifacciali, che possono produrre elettricità sia dal lato anteriore che da quello posteriore, la produzione totale di energia aumenta grazie alla generazione dal lato posteriore, che dipende dall'irraggiamento posteriore e dalla capacità del modulo di utilizzare questa energia. Pertanto, una misurazione più accurata del PR di un modulo bifacciale comprende la considerazione di un'irradianza totale più elevata e la differenza caratteristica di ciascun tipo di modulo nella produzione di energia tra lato anteriore e posteriore.

Il rapporto tra potenza massima sviluppata dal lato posteriore rispetto all'anteriore viene indicato, in STC, come fattore di bifaccialità  $BF$ .

$$BF = \frac{P_{rear}}{P_{front}} \quad (18)$$

Nel calcolare il PR di un modulo bifacciale si deve quindi pesare l'irraggiamento che raggiunge il lato posteriore con  $BF$  e solo poi aggiungere quel termine all'irraggiamento frontale per ottenere quello complessivo in POA:

$$H_{bifacial} = H_{front} + BF \cdot H_{rear} \quad (19)$$

Dopodiché si può semplicemente applicare la formula standard del PR, sostituendo l'irraggiamento con quello appena calcolato.

### 2.1.9 Guadagno bifacciale - Bifacial Gain (BG)

Il guadagno bifacciale non è ancora un parametro standardizzato. Esso rappresenta il guadagno di energia fornito dal lato posteriore di un modulo o di un sistema bifacciale.

A livello di sistema, è definito come il rapporto tra l'energia posteriore e quella anteriore.

Quando si considera il guadagno bifacciale totale di un sistema  $BG_{sys}$ , è possibile suddividere i diversi aspetti che influenzano il guadagno totale in:

- Guadagno di tipo ottico  $BG_{opt}$  calcolato come il guadagno medio d'irradianza tra retro e fronte del modulo

$$BG_{opt} = G_{rear}/G_{front} \quad (20)$$

dove  $G_{rear}$ ,  $G_{front}$  rappresentano le irradianze medie lungo le due superfici

- Guadagno del modulo  $BG_{mod}$ , anche detto guadagno bifacciale in corrente continua. Viene calcolato come prodotto tra il fattore di bifaccialità ed il guadagno ottico:

$$BG_{mod} = BF \cdot BG_{opt} \quad (21)$$

- Guadagno bifacciale di sistema: è solitamente diverso dal guadagno bifacciale del modulo e può essere calcolato confrontando un modulo bifacciale con uno monofacciale con caratteristiche identiche, operante nelle stesse condizioni di funzionamento (azimut, inclinazione, ecc.), nello stesso luogo e per la stessa durata temporale. Viene espresso in percentuale.

$$BG_{syst} = \frac{E_{rear}}{E_{front}} = \frac{(E_{bifacial} - E_{monofacial})}{E_{monofacial}} \% \quad (22)$$

Il guadagno bifacciale dipende fortemente dalla tecnologia dei moduli, dalla geometria, dalla distanza/inclinazione degli stessi, dall'irraggiamento e dall'albedo del terreno (in quanto molta dell'irradiazione incidente sulla superficie posteriore risulta essere riflessa, soprattutto se l'installazione ha prediletto l'ottimizzazione delle prestazioni del modulo frontale).

### 2.1.10 Guadagno dell'inseguitore - Tracker Gain (TG)

Il guadagno d'inseguimento è un altro KPI non standardizzato (poiché riguarda un tipo preciso di tecnologia e non offre elementi di generalizzazione). Esso quantifica il guadagno energetico risultante dall'inseguimento impiegato in un modulo o intero sistema PV.

Si esprime come il rapporto tra l'energia specifica di un sistema ad inseguimento e quella di un sistema fisso, entrambi operanti con gli stessi pannelli, nello stesso luogo e per la stessa finestra temporale, come fatto per il guadagno bifacciale:

$$TG = \left[ \frac{E_{out,tracking}}{P1} / \frac{E_{out,fixed}}{P2} - 1 \right] \% \quad (23)$$

dove

- TG è espresso in percentuale
- P1 è la potenza nominale installata nel sistema ad inseguimento
- P2 è la potenza nominale installata nel sistema fisso
- $E_{out}$  di entrambi i sistemi è l'energia DC, per escludere l'effetto degli inverter

## 2.2 KPI legati alla modellizzazione e previsione della potenza PV prodotta

Questa sezione descrive i KPI utilizzati per la valutazione dei modelli di previsione (prediction) della potenza fotovoltaica prodotta e per la valutazione di quelli di forecasting della potenza fotovoltaica. L'applicabilità di tali parametri sarà perlopiù comune alle due tipologie di modelli, ma nuovamente ci si focalizzerà sui modelli di previsione della potenza fotovoltaica, base essenziale per valutare l'affidabilità dell'impianto e le sue prestazioni.

I parametri introdotti a seguire valutano l'accordo delle serie temporali di potenza fotovoltaica modellata con le misure di potenza fotovoltaica nel corso della finestra temporale di applicazione dei modelli.

Si ricorda che per potenza fotovoltaica modellata (prediction) si intendono i valori di potenza fotovoltaica calcolati in base ai dati principalmente di irraggiamento solare e temperatura (dei pannelli o dell'aria) ed altri valori ambientali ricavati da osservazioni misurate al suolo o da stime di modelli solari o meteorologici basati su satellite (come PVGIS).

La precisione del modello, qualsiasi ne sia la natura, non potrà mai essere totale, eppure occorre tenere presente che non solo i valori modellati e previsti sono dotati di incertezze: anche le misure della potenza fotovoltaica, utilizzate come riferimento per la convalida, presentano alcune incertezze, e lo stesso vale per le misure o modellizzazioni di dati su cui la previsione si basa. Per tenerne conto, nel contesto in esame, le differenze tra i valori modellati e misurati sono spesso indicate come "deviazioni", piuttosto che come "errori". Il termine "errore" sarà perciò applicato per una ragione di praticità e riconoscibilità, in quanto si tratterà spesso di impiegare comuni tecniche di calcolo dell'errore al sistema.

### 2.2.1 Errore quadratico medio – Mean Squared Error (MSE)

Questo termine indica la discrepanza quadratica tra valori osservati e stimati da modello. Viene solitamente impiegato come valore di passaggio per calcolare lo scarto quadratico medio RMSE, ma possiede applicazioni dirette come termine di riferimento nelle fasi di convalida e test delle tecniche di Machine Learning.

Nello specifico della applicazione come KPI per il fotovoltaico, si esprime come:

$$MSE_p = \frac{\sum_i^N (PP_i - PM_i)^2}{N} \quad (24)$$

con:

- $PP_i$  [kWh] potenza PV stimata allo step temporale  $i$
- $PM_i$  [kWh] potenza PV misurata allo step temporale  $i$
- $i$  indice dello step temporale considerato
- $N$  taglia del campione di valori, ovvero numero totale di step temporali considerati

### 2.2.2 Scarto quadratico medio – Root Mean Squared Error (RMSE)

Generalmente più impiegato dell'errore quadratico medio, anche grazie alla stessa unità di misura dei valori di partenza con cui è più semplice confrontarlo, nasce semplicemente dall'applicazione della radice quadrata all'MSE:

$$RMSE_p = \sqrt{MSE_p} \quad (25)$$

L'RSME misura la dispersione dei valori previsti rispetto a quelli osservati. Un RMSE pari a zero corrisponde a una perfetta corrispondenza del modello con le osservazioni. Valori elevati di RMSE, al contrario, indicano un'elevata incertezza e di conseguenza una bassa accuratezza delle stime. L'RMSE

dà maggior peso ai grandi errori del modello (di previsione) che ai piccoli, pertanto è una metrica adatta quando i piccoli errori sono più tollerabili e gli errori di grandi entità causano problemi sproporzionatamente elevati, come nel caso di molte applicazioni di previsione

Uno svantaggio è che è anche particolarmente sensibile agli outliers, che sono invece valori poco rappresentativi delle grandezze osservate. Chiaramente, le considerazioni appena fatte mantengono invariata la validità quando applicate all'MSE.

Come per tutti i tipi di scarti/errori con una determinata unità di misura, il valore di RMSE ed MSE dipende fortemente dalla grandezza di partenza, e nello specifico dalla potenza nominale d'impianto  $P_0$ , per cui si possono ricavare versioni normalizzate in modi differenti, che sono più facilmente interpretabili, in quanto errori relativi adimensionali.

Normalizzando rispetto alla potenza nominale  $P_0$ , si ottiene:

$$RMSE_{rel1} = \frac{RMSE_p}{P_0} \quad (26)$$

Normalizzando rispetto alla potenza media misurata  $\overline{PM}$ , invece:

$$RMSE_{rel} = \frac{RMSE_p}{\overline{PM}} \quad (27)$$

L'entità di questi scarti diminuisce con l'aumentare dello span temporale di valutazione delle grandezze considerate (raggiungere elevata precisione di modello a livello annuale, ad esempio, è sensibilmente più semplice di riuscire a fare lo stesso per ogni minuto); un forte impatto ha la natura dei dati ambientali che, come per gli altri KPI considerati, influisce introducendo molteplici possibili sorgenti di incertezza.

### 2.2.3 Errore quadratico medio normalizzato (NMSE)

La normalizzazione può assumere altre forme e nel caso dell'NMSE ha una precisa forma dallo scopo preciso, ovvero riportare lo scarto tra stima e misura alla variabilità della misura stessa. Questo viene realizzato dividendo l' $MSE_p$  per la varianza della misura di potenza  $\sigma_p^2$ , così da riportare lo scarto a quanto dispersa sia la variabile misurata.

La varianza (errore quadratico medio rispetto al valore medio) della potenza risulta:

$$\sigma_p^2 = \frac{\sum_i^N (PM_i - \overline{PM})^2}{N} \quad (28)$$

Ne consegue che nel rapporto tra MSE e varianza il numero di elementi N si semplifica ottenendo:

$$NMSE_p = \frac{\sum_i^N (PP_i - PM_i)^2}{\sum_i^N (PM_i - \overline{PM})^2} \quad (29)$$

L'NMSE appena esposto è componente fondamentale del calcolo del coefficiente di determinazione, parametro dall'elevata rilevanza nel valutare la capacità del modello di descrivere i valori osservati.

### 2.2.4 Coefficiente di determinazione ( $R^2$ )

Il coefficiente di determinazione, complementare dell'NMSE, indica quanto la variabilità della potenza misurata sia prevedibile dalle variazioni della potenza stimata; quindi, descrive con quale livello di precisione i valori predetti dal modello replicano quelli reali.

L'espressione è:

$$R_p^2 = 1 - NMSE_p \quad (30)$$

Questo KPI può essere più informativo di MSE ed altri tipi di errore che saranno considerati, poiché è già adimensionale ed esprimibile in percentuale, svincolato dai range arbitrari dei valori delle potenze.

Allo stesso modo del coefficiente di correlazione (3), valori prossimi a +1 implicano una buona precisione del modello (perché NMSE è basso), mentre in prossimità dello 0 anche l'accuratezza del modello è sostanzialmente nulla: le variazioni dalla media delle misure non sono in quel caso spiegabili attraverso variazioni delle potenze di modello.

### 2.2.5 Errore medio assoluto – Mean Absolute Error (MAE)

Anche l'MAE è annoverato tra i KPI per la modellizzazione della potenza PV. Esso è definito come la media degli scarti in valore assoluto:

$$MAE_p = \frac{1}{N} \sum_{i=1}^{i=N} |PP_i - PM_i| \quad (31)$$

Come l'RMSE, l'MEA valuta la dispersione dei valori previsti rispetto a quelli osservati e, se pari a zero, corrisponde a previsioni perfette, mentre valori crescenti corrispondono a un'accuratezza decrescente.

Si differenzia dall'RMSE per l'uguale ponderazione degli errori di previsione piccoli e grandi e per la minore sensibilità agli outlier, e per valori tendenzialmente minori.

Allo stesso modo dell'RMSE, può essere normalizzato dividendolo per la potenza nominale o media d'impianto. Le formule non sono riportate perché sono analoghe a (26), (27).

### 2.2.5 Errore medio di bias – Mean Bias Error (MBE)

L'MBE è la media di tutti gli scarti tra potenza stimata e misurata, ognuno contato con il proprio segno, e quindi coincidente con la differenza delle medie di potenza nel periodo analizzato:

$$MBE_p = \frac{1}{N} \sum_{i=1}^{i=N} (PP_i - PM_i) = \overline{PP} - \overline{PM} \quad (32)$$

Considerazioni sul modello possono essere fatte valutando il segno di MBE: se positivo, indica una sovrastima da parte del modello di previsione, andando ad indicare che in media le potenze previste sono superiori alle osservazioni. Corrispondentemente, un MBE negativo indica una sottostima delle stime/previsioni.

In virtù di tale proprietà, può permettere di identificare deviazioni sistematiche di un certo modello (applicandolo a più siti PV) ed è quindi di particolare importanza nel giudicare la qualità delle previsioni annuali e a lungo periodo, dove eventuali bias sono maggiormente evidenti.

Così come per RMSE ed MAE, anche in questo caso la normalizzazione può essere eseguita nella forma (26) o (27).

I KPI della sezione corrente sulla modellizzazione di potenza possono indicare diversi livelli di qualità del modello impiegato e di accuratezza delle previsioni realizzate dallo stesso, ma il significato di valori elevati o contenuti dei suddetti KPI ha un margine di interpretazione che supera considerazioni sul modello in sé: parte dello scarto giungerà infatti direttamente dalla naturale imprecisione di misura della potenza e dei parametri ambientali impiegati per preparare il modello; un'ulteriore causa di scarto può infine essere dovuta ad una corretta precisione di stima della potenza che va però a cozzare con cambiamenti nel funzionamento dell'impianto PV. In questa ottica, i KPI di previsione svolgono anche un possibile ruolo di monitoraggio.

Per questo motivo si prosegue nella sezione in corso con due KPI di monitoraggio.

### **2.2.6 Indice di prestazione energetica – Energy Performance Index (EPI)**

L'EPI è definito come il rapporto tra final yield  $Y_f$  e l'expected yield  $Y_E$  determinato da un modello PV:

$$EPI = \frac{Y_f}{Y_E} \quad (33)$$

Si raccomanda di calcolarlo regolarmente, utilizzando i dati meteorologici effettivi come input del modello, su base mensile o annuale. In primo luogo, è necessario un periodo di addestramento dei dati (in cui si presume che il sistema funzioni correttamente) per calcolare i coefficienti dell'equazione del modello analitico o per preparare algoritmi di machine learning.

Poi, il modello viene applicato al periodo di test: con un modello perfetto, il suo valore iniziale è pari al 100%. Può essere utilizzato nel monitoraggio per l'identificazione di difetti di prestazione (comportamenti subottimali d'impianto) o guasti nel caso di EPI nettamente minori di 1.

L'affidabilità del suo valore dipende chiaramente dall'accuratezza del modello a causa dell'expected yield al denominatore.

### **2.2.7 Disponibilità dei dati**

La valutazione di questo KPI per un sistema di monitoraggio è abbastanza semplice: è la frazione di tempo in cui il sistema fornisce dati divisa per il tempo del periodo di riferimento per il quale questo KPI è calcolato.

Minore la quantità di valori vuoti registrati dal sistema di monitoraggio, meglio è per il modello basato su tali dati ed il conseguente monitoraggio. Per ogni singola misurazione (ad esempio, i dati di un sensore di irraggiamento in piano d'array) la disponibilità dei dati deve essere registrata separatamente e la disponibilità complessiva dei dati per questa misurazione può essere calcolata moltiplicando la disponibilità del sistema di monitoraggio e la disponibilità dello specifico sensore.

### **2.2.7 Qualità dei dati**

La qualità dei dati si riferisce all'accuratezza dei dati per le analisi successive, per garantire che non ci siano valori falsi o inaffidabili. Per valutare e garantire desiderati livelli di qualità si utilizzano generalmente dei filtri. Pertanto, la qualità dei dati è legata alla quantità di valori filtrati dal sistema di monitoraggio.

Si tratta di misure per cui il valore è stato effettivamente registrato o misurato, ma è chiaramente errato. A tal proposito, questo KPI rende difficile valutare il significato, ad esempio, di un ridotto numero di valori filtrati, perché ciò potrebbe essere conseguenza tanto di una generale bontà de dati quanto dell'impiego di un filtro mal definito.

Tuttavia, una prima stima di questo KPI può essere ottenuta confrontando diversi impianti fotovoltaici che applicano la stessa strategia di filtraggio.

Criteri quantitativi consigliati per questi ultimi due KPI sono variabili e non generalizzabili, poiché le necessità e disponibilità di buone e frequenti misure dipendono dall'ubicazione dell'impianto PV, particolari condizioni meteorologiche ecc.

Come requisiti minimi, però, si possono proporre disponibilità dei dati non inferiore al 90%, con la percentuale di valori filtrati non superiore al 10%.



### **3. Impianto fotovoltaico valutato**

I capitoli precedenti, con riferimento a differenti lavori nella letteratura, hanno gettato le basi teoriche ed introdotto i necessari strumenti per la modellizzazione di un impianto PV tramite l'impiego di ANN, per poi valutarne i KPI più indicativi.

Tutto ciò sarà messo in pratica dal capitolo 4, mentre in questo si svolgerà una panoramica sulle specifiche dell'impianto preso in analisi, ponendo particolare attenzione agli elementi essenziali quali planimetria, pannelli ed inverter.

L'impianto fotovoltaico è installato sulle due falde del tetto di un edificio in Emilia-Romagna. Esso è suddiviso in otto sottocampi fotovoltaici e 392 moduli fotovoltaici (fissi) da 250 Wp ciascuno, produzione SILIKEN, SLK60P6L 255Wp, per una potenza nominale complessiva di 98.00 kWp.

Il collegamento con la rete pubblica è effettuato con sistema trifase con neutro a tensione nominale 400 V - 50 Hz.

L'impianto è dotato di un contatore dell'energia prodotta, e di un contatore dell'energia immessa in rete (fornito dall'Ente distributore dell'energia elettrica) ubicati nei rispettivi quadri elettrici.

I moduli fotovoltaici SLK60P6L 255Wp sono costituiti da 60 celle in silicio policristallino; la tabella 1 riporta le principali specifiche tecniche dei moduli misurate in condizioni standard STC (AM = 1,5; E = 1.000 W/m<sup>2</sup>; T = 25°C).

**Tabella 1 – specifiche tecniche dei moduli fotovoltaici**

<b>Potenza di picco</b>	250 W
<b>Tensione MPP (Maximum Power Point)</b>	30,20 V
<b>Corrente MPP (Maximum Power Point)</b>	8,31 A
<b>Tensione a vuoto</b>	37,70 V
<b>Corrente di corto circuito</b>	8,90 A
<b>Efficienza</b>	15,4 %
<b>Dimensioni</b>	1.640 x 990 x 40 mm
<b>Peso</b>	19 kg

Le stringhe sono suddivise in due tipologie, quelle composte da 16 e da 17 pannelli rispettivamente, vedasi tabella 2.

**Tabella 2 – Caratteristiche tecniche delle stringhe**

<b>Composizione stringa tipo A</b>	
<b>Numero di moduli fotovoltaici</b>	16
<b>Potenza di picco</b>	4.000 W
<b>Tensione MPP (Maximum Power Point)</b>	483,2 V
<b>Corrente MPP (Maximum Power Point)</b>	8,31 A
<b>Tensione a vuoto</b>	603,2 V
<b>Corrente di corto circuito</b>	8,90 A

<b>Composizione stringa tipo B</b>	
<b>Numero di moduli fotovoltaici</b>	17
<b>Potenza di picco</b>	4.250 W
<b>Tensione MPP (Maximum Power Point)</b>	513,4 V
<b>Corrente MPP (Maximum Power Point)</b>	8,31A
<b>Tensione a vuoto</b>	640,9
<b>Corrente di corto circuito</b>	8,90 A

Per integrare i dati elencati nelle tabelle, si riportano sezioni aggiuntive del datasheet dei pannelli fotovoltaici presenti nell'impianto: in figura 3.1 sono presenti i coefficienti di temperatura per potenza, tensione e corrente, mentre in figura 3.2 è mostrato l'invecchiamento massimo garantito dai costruttori, con un valore annuale del -0.7%.

Tensione massima UL / IEC $V_{max}$ (V) UL/IEC	Coefficiente temperatura di $P_{mp}$ $TkP_{mp}$ (%/°C)	Coefficiente temperatura di $V_{oc}$ $TkV_{oc}$ (%/°C)	Coefficiente temperatura di $I_{sc}$ $TkI_{sc}$ (%/°C)
600/1000	-0,43	-0,356	0,062

Figura 3.1 Coefficienti di temperatura

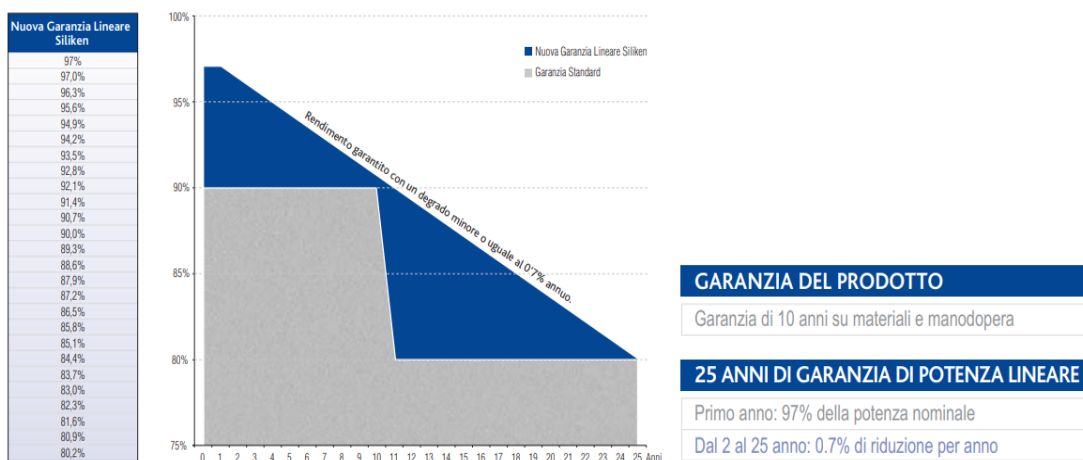


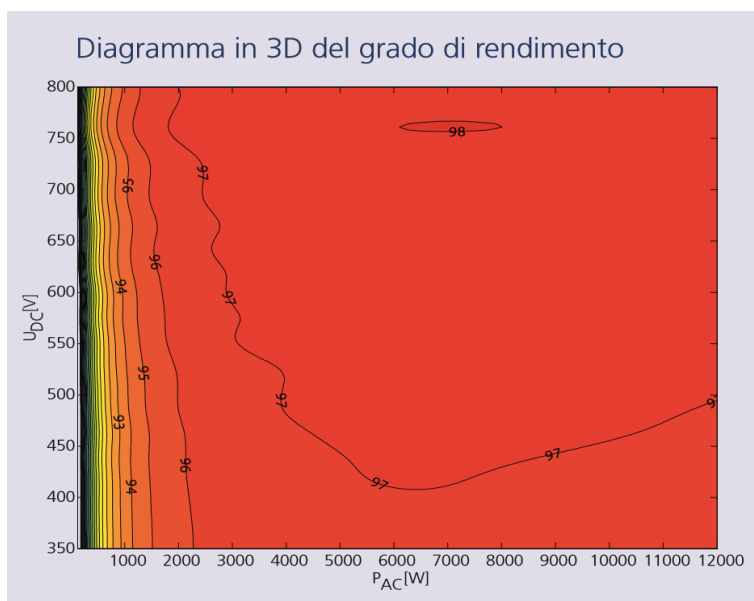
Figura 3.2 Invecchiamento previsto dei pannelli

Sono installati otto inverter trifasi tipo Kaco New Energy Technologie dotati di dispositivo di ricerca MPP (Maximum Power Point), in grado di individuare in ogni istante il punto di massima potenza erogabile dall'impianto. Nella prima parte della tabella 3 sono indicate le principali specifiche tecniche degli inverter lato DC, mentre nella seconda si trovano le specifiche lato AC.

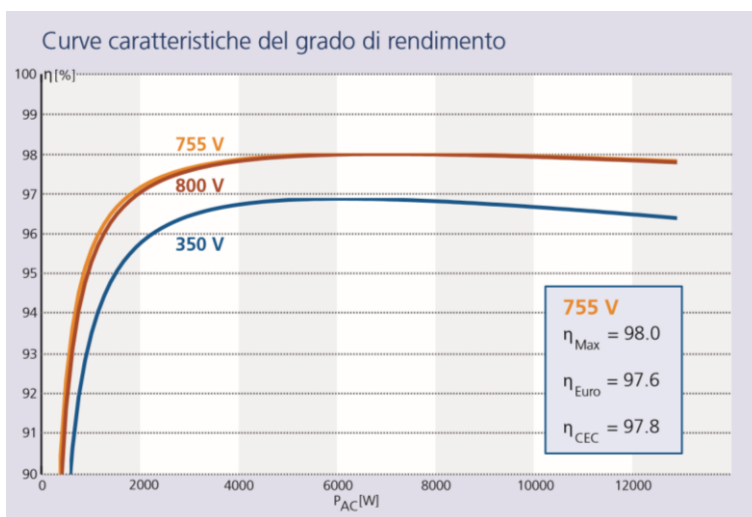
Tabella 3– Caratteristiche tecniche inverter in ingresso

KACO POWADOR 14.0TL 3	
<b>Valori di ingresso</b>	
<b>Potenza</b>	14.000 W
<b>Intervallo MPP</b>	200-800V
<b>Tensione di avvio</b>	250V
<b>Tensione a vuoto</b>	1000 V
<b>Max corrente di impiego</b>	2x18,6A
<b>Numero regolatori MPP</b>	2
<b>Potenza max/regolatore MPP</b>	12,8kW
KACO POWADOR 14.0TL 3	
<b>Valori di uscita</b>	
<b>Potenza di uscita</b>	12.500 VA
<b>Tensione di rete</b>	400/230 (L1-L2-L3- N-PE))
<b>Corrente nominale</b>	3x18,1A
<b>Frequenza nominale</b>	50Hz
<b>Cosφ</b>	1 (0,8 ind.-0,8cap.)
<b>Numero delle fasi di immissione in rete</b>	3
<b>Dati Generali</b>	
<b>Rendimento max</b>	98%
<b>Rendimento europeo</b>	97,6%
<b>Dimensioni</b>	690x420x200mm

Dal datasheet degli inverter sono estratti il diagramma in 3D in figura 3.3 e le curve di rendimento in figura 3.4.



**Figura 3.3** mappa del rendimento d'inverter



**Figura 3.4** curve di rendimento d'inverter

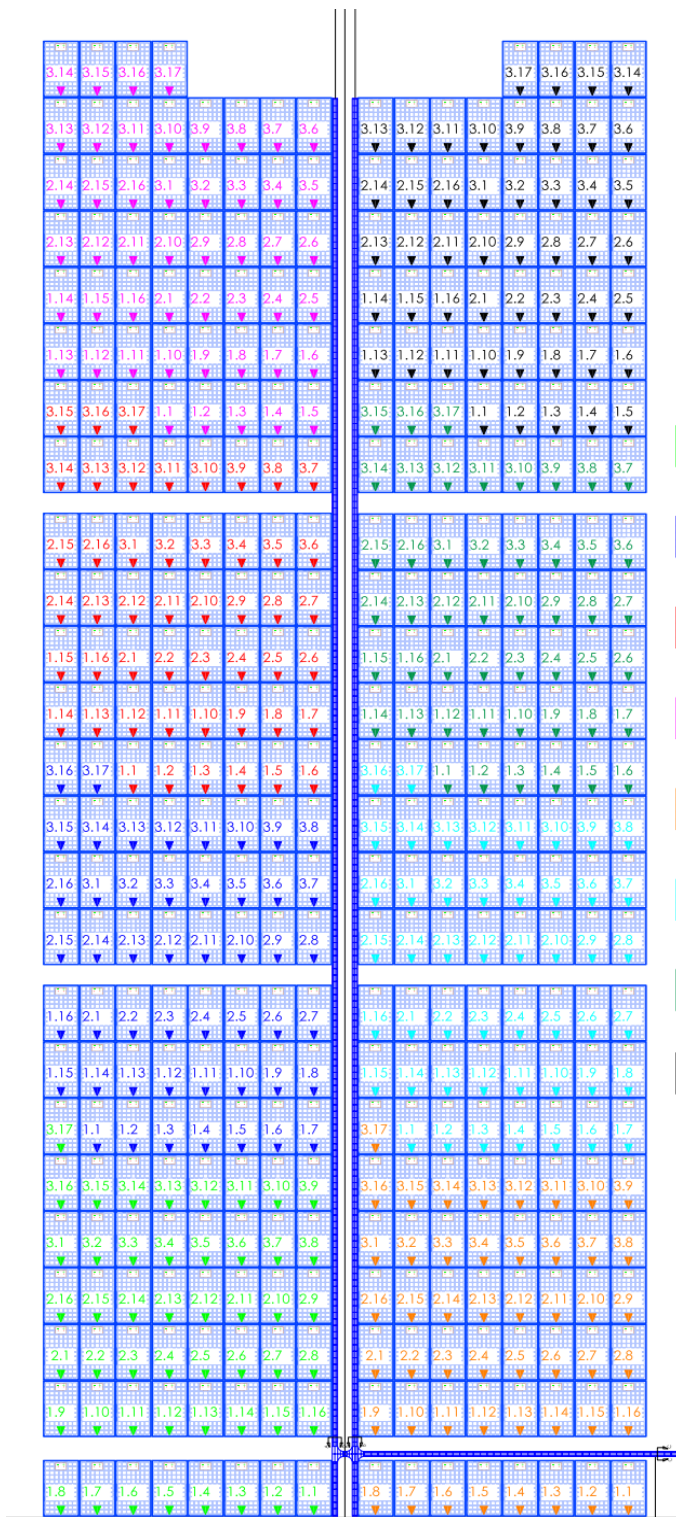


Figura 3.5 planimetria dell'impianto

La planimetria dell'impianto è mostrata nella figura 3.5 qui accanto. L'orientamento della figura pone la falda est a sinistra e la ovest a destra. Le stringhe sono suddivise in 4 sottocampi per falda, ognuno collegato ad un inverter differente, seguendo la distinzione cromatica espressa in figura 3.6.



Figura 3.6 legenda sottocampi e inverter

Nella tabella 4 a seguire sono mostrati due generatori inverter MPPT che in totale combinano i 4 scenari: falda est e falda ovest con orientamenti associati, 2 stringhe da 16 moduli o 1 da 17.

**Tabella 4 diversi scenari lungo le falde**

<b>Generatore INVERTER 1 FALDA EST - MPPT1</b>	
<b>Posizionamento dei moduli</b>	Complanare alle superfici
<b>Struttura di sostegno</b>	Fissa
<b>Inclinazione dei moduli (Tilt)</b>	7°
<b>Orientazione dei moduli (Azimut)</b>	-89°
<b>Estensione totale disponibile</b>	503.75 m <sup>2</sup>
<b>Numero totale moduli</b>	32
<b>Numero di stringhe</b>	2
<b>Numero di moduli per ogni stringa</b>	16
<b>Superficie totale moduli</b>	51.97 m <sup>2</sup>

<b>Generatore INVERTER 5 FALDA OVEST – MPPT2</b>	
<b>Posizionamento dei moduli</b>	Complanare alle superfici
<b>Struttura di sostegno</b>	Fissa
<b>Inclinazione dei moduli (Tilt)</b>	7°
<b>Orientazione dei moduli (Azimut)</b>	91°
<b>Estensione totale disponibile</b>	503.75 m <sup>2</sup>
<b>Numero totale moduli</b>	17
<b>Numero di stringhe</b>	1
<b>Numero di moduli per ogni stringa</b>	16
<b>Superficie totale moduli</b>	51.97 m <sup>2</sup>

Ogni inverter, a qualsiasi falda appartenga, è collegato a due stringhe da 16 pannelli (32 complessivi) ed una da 17. Lo schema multifilare dell'inverter 1 è esibito in figura 3.6., dove i collegamenti risultano più chiari.

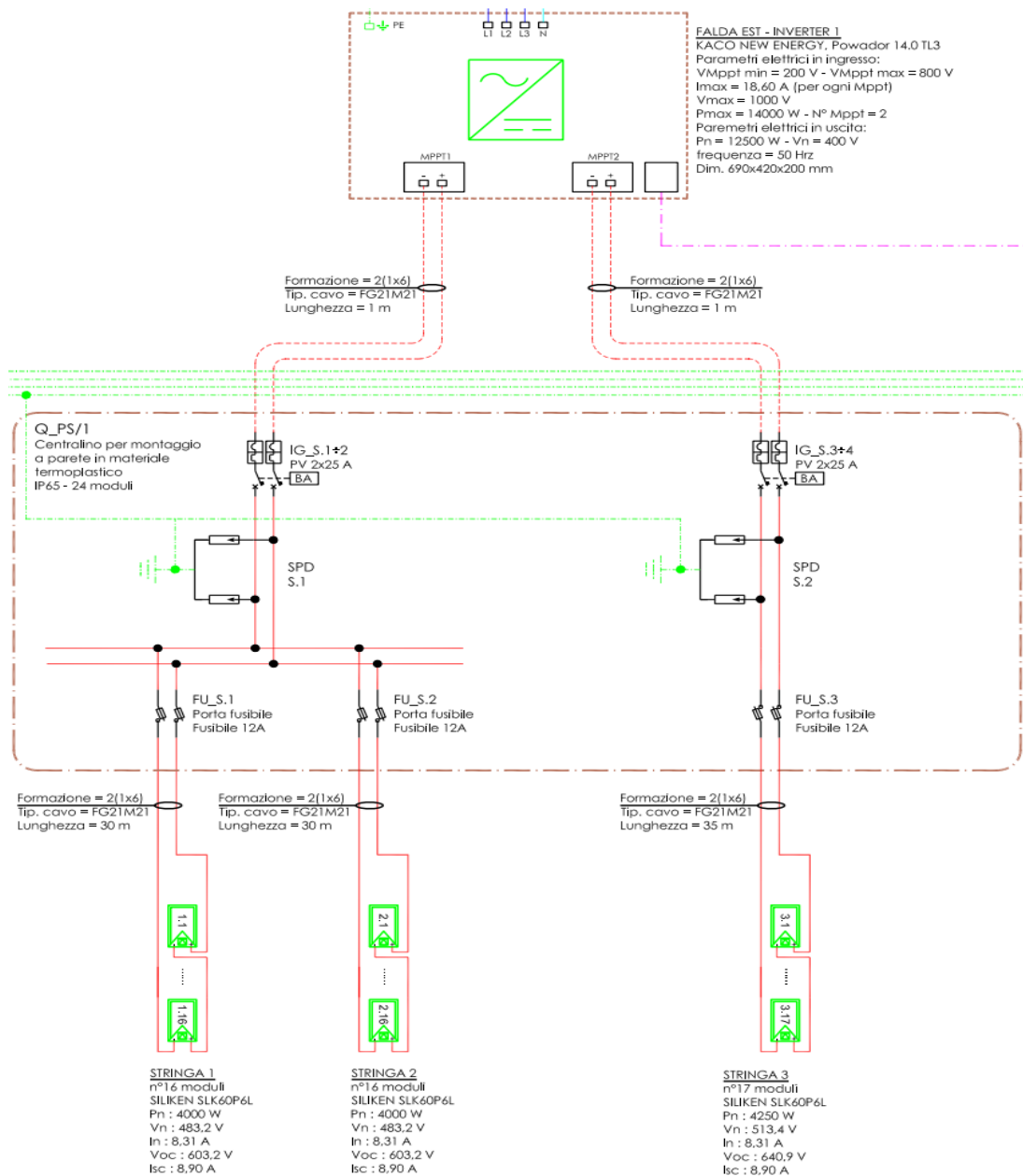


Figura 3.6 schema multifilare dell'inverter 1 sulla falda est



## **4. Sviluppo di modellizzazione e considerazioni sull'impianto**

Si giunge ora alla summa della preparazione fatta, con cui si va a generare un modello di previsione della potenza dell'impianto PV descritto nel capitolo 3.

Ciò si è tradotto nel fare una serie di scelte di progetto sulla base delle informazioni presenti nella letteratura e sulle specifiche caratteristiche dei dati disponibili e del software utilizzato.

### **4.1 Selezione ed estrazione dei dati, scelta del tipo di modello**

Per motivare al meglio la scelta dei dati selezionati ed impiegati nella modellizzazione d'impianto, occorre approfondire oltre la logica con cui il lavoro di preparazione del modello è stato svolto.

Il focus resta predire andamenti in finestre temporali future a quella di generazione del modello, produrre i KPI misurati, stimati, d'errore e trarre (dove possibile) le dovute conclusioni sulla bontà del modello e sullo status dell'impianto. Terminando in tal senso, però, l'analisi svolta avrebbe validità limitata allo specifico caso proposto, correndo il rischio di ribaltamento del problema dei modelli analitici.

Se infatti i modelli analitici più tradizionali si allineano alle caratteristiche dei sistemi PV su cui vengono impiegati, fanno ciò individuando i valori più adatti di parametri già fissati come forma e scrittura nella formulazione del modello analitico. Si è potuto osservare questo concetto all'opera con il Single Diode Model: la formulazione generale è una, derivata da correlazioni fisiche (elettriche nello specifico) e soltanto caso per caso essa è "forzata" a riprodurre le caratteristiche di uno specifico pannello/impianto ecc.

Proprio in contrasto a questa meccanica di funzionamento delle modellizzazioni tradizionali, che generalizza per semplificazione, non si deve andare in totale antitesi, generando un modello funzionante solo per l'impianto in esame e non riproponibile in altri contesti, o il lavoro di tesi risulterebbe fine a sé stesso.

In tal senso, si è speso il dovuto impegno, più che nel singolo modello finale, nello sviluppare una tecnica di produzione di modelli applicabile in varie circostanze, usando diversi dati di partenza da cui trarre differenti conclusioni e senza cristallizzarsi in una forma troppo specifica.

Come tale intento abbia permeato il lavoro verrà mostrato passo dopo passo, nelle sezioni a seguire, ma il tutto comincia con la selezione dei parametri di input (e delle grandezze in output), nonché delle fonti da cui reperirne i valori.

I dati direttamente misurati dagli strumenti d'impianto in esame sono stati estratti dal sito VCOM, contenente le misure dell'impianto, in formato xls ora per ora di ogni settimana dal 2014 al 2021. Tutti i dati sono poi stati organizzati con Microsoft Excel, prima di caricarli in Matlab.

L'impianto ha iniziato la sua operatività a termine del 2012, mentre le misure sono state estratte dalla strumentazione e collegate al sito nel corso del 2013, portando il 2014 ad essere il primo anno con una buona disponibilità dei dati, tale da considerarlo nella sua interezza; sempre ricercando una certa completezza di informazioni a livello annuale, l'anno corrente (2022) è stato escluso.

Per gli output da stimare si sono selezionate la potenza complessiva d'impianto DC, ottenuta come sommatoria ora per ora dei valori di potenza in ingresso agli 8 inverter della struttura, ed ovviamente la potenza AC immessa in rete e misurata dal contatore d'immissione (come descritto nel capitolo 3).

La potenza AC è sempre il dato più importante da andare a stimare, poiché rappresentativa delle prestazioni d'impianto valutate dal punto di vista della rete a cui è connesso, ragione per cui compare nella maggior parte dei KPI di performance, risultando essenziale.

Si è deciso però di inserire anche la potenza DC in virtù della sua presenza in altri KPI, e della somiglianza senza diretta sovrapposizione con gli andamenti della potenza AC, offrendo ulteriori informazioni su cui basarsi per la modellizzazione.

Il numero di output da stimare è quindi uguale a 2.

Nello scegliere gli input, invece, le considerazioni si sono diramate maggiormente.

La grandezza con la massima correlazione rispetto alla produzione energetica è sempre l'irradianza in piano d'array, input fondamentale. Essa è misurata nell'impianto da una cella con la stessa tecnologia dei pannelli PV impiegati. Nell'ottica di mantenere alta la disponibilità dei dati, a parziale discapito della qualità degli stessi, la gestione dell'impianto ha optato per sostituire le ore carenti di informazioni dalla cella/sensore con stime fatte tramite dati satellitari ed un software privato.

Il secondo input imprescindibile è la temperatura dei moduli che, come visto in precedenza, influenza fortemente le prestazioni di tutto l'impianto [22],[29], al punto tale da sviluppare diversi metodi di correzione del performance ratio [31], proprio per tenere conto dell'effetto della variabilità della temperatura.

Non sempre, purtroppo, la misura diretta della temperatura di pannello è disponibile, ed in tal caso si può sopperire partendo dalla temperatura dell'aria e dalla velocità del vento [32 - 34] per modellarla.

Si possono usare modelli empirici più classici basati su tali misure [35 - 38], prelevate nel sito d'impianto (o in prossimità da stazioni meteorologiche disponibili), oppure impiegando per i dati ambientali valori stimati a loro volta [39]. Anche in questo ambito di modellizzazione, comunque, l'impiego di machine learning e reti neurali è in rapida crescita e pone una valida alternativa alle tecniche empiriche, anche per le più precise [40 - 44]. Usare una qualsiasi di queste metodologie è concettualmente irreprensibile, ma si deve porre particolare accortezza agli elementi di incertezza introdotti ad ogni stima e modellizzazione intermedie.

Fortunatamente, l'impianto in questione mette a disposizione le misure dirette di temperatura dei pannelli, ma i paper di riferimento pongono un'ottima base di partenza nel caso in cui si voglia compiere un lavoro di modellizzazione dell'impianto in assenza delle stesse; viene inoltre sottolineata l'elevata sensibilità dei valori di temperatura di pannello alle velocità del vento (che nella maggior parte dei climi aiuta nel raffreddamento dei moduli).

L'ultima grandezza scelta in input tra quelle disponibili in impianto è la temperatura degli inverter. Nonostante non abbia correlazione con la potenza al pari di irradianza e temperatura dei moduli, è disponibile per ogni inverter e si assume ragionevolmente affidabile, in quanto le misure realizzate dagli inverter sono ad alta precisione, poiché impattano sull'operazione in MPPT delle stringhe e forniscono le misure di potenza, che necessitano alta accuratezza, per poter giudicare le prestazioni degli impianti.

Un secondo gruppo di dati è stato invece recuperato da database regionali ARPAE, combinando le misure proposte da due stazioni meteorologiche, una situata a 7 km dall'impianto e l'altra a 9 km.

La scelta di impiegare anche dati regionali ha lo scopo di esibire la qualità di modelli di previsione della potenza prodotta nel caso non si disponga di misurazioni precise direttamente nell'impianto (come può essere il caso per impianti di piccola taglia o in assenza di anche solo una tra temperatura di pannello e irradianza in POA), per arricchire lo studio di una componente di generalità a cui si faceva cenno prima.

L'integrazione dei valori dai due database è stata imposta dalla mancanza di misurazioni nel 2017 per una delle due fonti e nel 2018 per l'altra. Entrambe non proseguivano oltre il 2018 e ulteriori stazioni erano particolarmente distanti. Per questa ragione, i modelli prodotti da dati ARPAE sono stati testati fino al 2018 e non al 2021.

Le grandezze prelevate sono state l'irradianza globale orizzontale GHI, la temperatura dell'aria a 2 m dal suolo e la velocità del vento a 10 m dal suolo, tutti come medie orarie.

Le disponibilità dei dati, togliendo quindi ore/giorni in cui almeno un parametro risultava mancante, sono del 91% per le misure d'impianto e del 100% per le ARPAE (dopo aver unito i due database delle stazioni). La qualità è invece migliore per i valori d'impianto perché le misure sono prese con strumenti più adatti e nel luogo preciso, mentre dati ARPAE possono ignorare elementi specifici e locali a causa della distanza.

I valori misurati in impianto sono già utilizzabili in ingresso a qualsiasi tipo di modello, poiché l'irradianza è valutata in POA e la temperatura è già estratta a livello del modulo.

Per i dati ARPAE la situazione potenzialmente si complica: nel caso di modelli analitici tradizionali sarebbe necessario prima trasformare i dati disponibili, ovvero scomporre la GHI in componenti diretta, diffusa e riflessa per poi applicare trasposizione riportando in POA e utilizzare una delle tecniche accennate in precedenza per derivare la temperatura dei pannelli da quella dei moduli e dalla velocità del vento. Tutte queste trasformazioni e modellizzazioni preliminari introducono ulteriori incertezze ed elementi di variabilità, ma una soluzione più congeniale sorge spontaneamente dalla scelta del tipo di modello di previsione da implementare.

Si opta infatti per un modello basato sul Machine Learning e sulle reti neurali ANN nello specifico.

Le ragioni sono la generale miglior precisione dei modelli ANN rispetto ad analitici tradizionali ed il supporto da parte del software Matlab allo sviluppo di reti neurali tramite il toolbox dedicato. La realizzabilità ed efficacia di un modello ANN di previsione della potenza PV sviluppato in Matlab sono confermate in letteratura [9].

Alcuni vantaggi offerti dai modelli ANN riguardano la flessibilità:

- In merito allo sviluppo del modello: come esposto nel capitolo 1, la rete realizzata non è mai univoca (per variabilità dei parametri strutturali e di quelli che cambiano durante l'addestramento), univoca può essere invece la strategia di ottimizzazione e sviluppo del modello ANN, per cui lavorare con le reti neurali permette di proporre una procedura di sviluppo e test delle reti applicabile anche ad impianti differenti.
- Nella natura degli input: che le grandezze specifiche misurate in situ abbiano maggiore correlazione con la potenza in output rispetto ad altre da trattare è evidente, ma a differenza di modelli empirici che richiedono valori già espressi nella forma corretta, una rete può trattare di tutto senza concetti pregressi sui legami fisici che legano input e output. Dato che, ad esempio, temperatura e vento influenzano in un determinato modo la potenza (come sappiamo accada), la rete imparerà il collegamento traducendolo in pesi e bias adeguati all'interno dei layer.  
Ciò significa che, anche con l'intento di evitare elementi aggiuntivi di incertezza e variabilità, non è necessario pretrattare i parametri di input in forme particolari, e se ne occuperà invece la rete stessa modificando la propria struttura interna durante l'addestramento: questo permette di usare i dati ARPAE esattamente come sono e lasciare alla rete l'onere di intuirne le correlazioni.

Ai dati ARPAE sono state affiancate anche le misure di temperatura d'inverter poiché, anche in assenza di sensori locali per le altre misure, gli inverter costituiscono un elemento imprescindibile dell'impianto e si può quindi assumere in generale l'accesso a tale informazione.

In aggiunta alle grandezze elencate finora, si è deciso di includere negli input l'ora e la stagione meteorologica corrispondenti ad ogni misura, dove con stagioni meteorologiche si intendono suddivisioni trimestrali di un anno basate soltanto su mutamenti climatici ed ambientali nelle zone temperate. Con questa definizione la primavera include marzo, aprile e maggio per intero senza considerare equinozi o solstizi e così per ogni gruppo di tre mesi a seguire.

Inserire questi elementi temporali permette di creare ulteriori collegamenti logici interni (ad esempio tra le ore e l'andamento d'irradianza, influenzata dall'altezza del sole e possibili ombreggiamenti in certe fasce orarie ecc.) e di aumentare il numero di input complessivo che nel caso di stima dell'output PV risulta una strategia migliore di qualsiasi tecnica di selezione delle feature [9].

Saranno comunque proposte varie reti, ognuna con una combinazione di input differente, per valutarne l'impatto sulla precisione di stima e l'applicabilità per chi avesse accesso solo a un determinato sotto-insieme dei parametri di input (variabile tra le installazioni).

## 4.2 Processo di sviluppo ed ottimizzazione della rete feed-forward

La rete sviluppata in Matlab è del tipo feed-forward FFNN, comunemente impiegato, particolarmente flessibile e semplice da personalizzare. Ha esattamente la struttura standard delle ANN, come descritta in Figura 1.5.1., con numero e taglia degli hidden layer variabili.

Questo capitolo è suddiviso negli step di preparazione dei dati, eventuale filtro degli stessi ed ottimizzazione delle reti.

### 4.2.1 Pre-processing dei dati d'impianto e ARPAE

Per entrambi i dataset si svolgono operazioni simili di pre-processing (o sarà indicato quando differiscono).

Un'operazione di filtro dei dati a monte è stata fatta nel prelevarli (per le misure d'impianto) dal sito VCOM: giorni e ore con totale discrepanza tra potenza e irradianza misurati (potenza nulla con elevato irraggiamento o viceversa) sono stati esclusi imputandoli a momentanei problemi alla strumentazione. Alcune ore si sono poi rivelate del tutto mancanti poiché manutenzioni o operazioni sugli strumenti avevano interrotto il prelievo di tutte le misure fino al ripristino.

I dati ARPAE hanno subito un taglio dei dati dal 2019 in poi per mancanza, mentre prima i dataset di due stazioni sono stati combinati.

Un primo script pone a 0 l'irradianza quando erroneamente segnata negativa (raramente e solo nei dati ARPAE notturni) o quando minore di  $3 \text{ W/m}^2$  (offset per i dati d'impianto durante la notte).

Sono poi state escluse le ore notturne sulla base dell'irradianza appena corretta.

Le ore con dati mancanti per anche solo una grandezza sono state eliminate.

In vista della preparazione degli input e output, la potenza DC è stata calcolata come la sommatoria di tutte le potenze in ingresso agli 8 inverter, la divisione ora/giorno/mese/anno dei dati è stata accresciuta con una colonna di valori contenenti le stagioni meteorologiche ed infine la temperatura degli inverter è stata resa in un'unica colonna di temperatura media d'inverter (realizzata tra i 6 valori intermedi, eliminando minimo e massimo ora per ora).

Tutti i dati vengono normalizzati rispetto alla propria grandezza (2), così da ottenere ogni parametro compreso tra 0 e +1, per semplificare l'apprendimento delle reti. Sono stati estratti i valori adibiti all'addestramento, ovvero tutti quelli appartenenti al 2014, poiché è il primo anno ad alta disponibilità di misure e per prassi si allena la rete sull'andamento del primo anno (che offre una situazione più sana e ideale dell'impianto) con la possibilità (ad addestramento terminato) di confrontarne le previsioni con il funzionamento reale negli anni successivi ed evidenziare eventuali anomalie.

In una matrice sono allocati gli output normalizzati (potenze DC e AC normalizzate), mentre la matrice contenente gli input normalizzati offre 4 possibili combinazioni per ognuno dei due dataset:

1. Dati d'impianto, 5 input, [ore, stagione, irradianza, temperatura pannelli, temperatura inverter]
2. Dati d'impianto, 4 input, [ore, stagione, irradianza, temperatura pannelli]
3. Dati d'impianto, 4 input, [ore, stagione, irradianza, temperatura inverter]
4. Dati d'impianto, 4 input, [ore, irradianza, temperatura pannelli, temperatura inverter]
5. Dati ARPAE, 6 input, [ore, stagione, irradianza, temperatura aria, velocità vento, temperatura inverter]
6. Dati ARPAE, 5 input, [ore, stagione, irradianza, temperatura aria, velocità vento]
7. Dati ARPAE, 5 input, [ore, stagione, irradianza, temperatura aria, temperatura inverter]
8. Dati ARPAE, 5 input, [ore, irradianza, temperatura aria, velocità vento, temperatura inverter]

\*Occorre sottolineare che si sottintende irradianza in POA per l'impianto e GHI per i dati ARPAE.

Si tratta quindi di allenare le reti con tutti i parametri a disposizione oppure eliminando rispettivamente la temperatura degli inverter, la stagione o la temperatura dei pannelli nel caso dei dati d'impianto, o temperatura degli inverter, stagione e velocità del vento nel caso ARPAE.

Si è deciso di non operare alcuna selezione delle feature di input perché in letteratura nell'ambito PV si è dimostrato più volte inutile o dannoso ai fini delle prestazioni del modello; si faranno le dovute considerazioni sul ruolo degli specifici parametri in modo indiretto, ovvero confrontando reti prodotte senza averli considerati in fase di addestramento.

La scelta dei parametri da escludere ha motivazioni ben precise legate proprio al confronto delle prestazioni dei modelli: per entrambi i dataset si escludono alternativamente il valore della stagione o della temperatura di inverter, per potersi poi esprimere sul reale impatto che hanno sulle reti che li contengono. Allenando una rete ARPAE senza la velocità del vento si intende verificare come la sensibilità della temperatura dei pannelli nei confronti di tale grandezza si traduca nel training del modello. Il motivo per il quale è stata allenata anche una rete senza i dati di temperatura di pannello, ripetutamente indicata come grandezza fondamentale nella previsione della potenza, è esposto retroattivamente nelle sezioni successive del capitolo corrente.

Successivamente si pone un'ulteriore distinzione tra le reti: se i valori del 2014 siano stati filtrati dagli outliers o meno, producendo quindi un totale di 16 reti, divise in 4 principali categorie sulla base del dataset di addestramento e dell'applicazione del filtro dei valori anomali. Per farlo si usa direttamente il comando specifico messo a disposizione da Matlab, con cui si cercano valori anomali per ogni grandezza indipendentemente, per poi cancellare quell'ora di dati per intero. I dati d'impianto, più precisi, hanno una percentuale di outliers rimossi dell'8%, mentre i dati ARPAE del 14%.

Sia che gli input e output di addestramento siano filtrati dagli outliers o meno, al momento del calcolo di scarti e coefficiente di determinazione, si considerano i valori nella loro interezza.

#### **4.2.2 Addestramento della rete con ottimizzazione Bayesiana**

Lo script Matlab principale si occupa di produrre le reti di interesse discriminando tra numero di input e dataset. Viene caricato l'anno di dati adibito all'addestramento, dopodiché inizia l'ottimizzazione della rete neurale. Per ragionare sulla struttura ed ottimizzazione di una rete neurale, occorre prima richiamare i concetti di parametri ed iperparametri:

- Parametri: sono gli elementi risultanti dall'addestramento, che la rete modifica internamente ed autonomamente per meglio predire gli output. Sono i bias ed i pesi applicati ai neuroni (nodi presenti nei vari layer) funzionali alla minimizzazione dell'errore di modello
- Iperparametri: sono fissati manualmente e preventivamente e non variano durante l'addestramento. Racchiudono il numero di hidden layer, il numero di neuroni per layer, le funzioni di trasferimento e quella di attivazione.

Gli iperparametri vanno ottimizzati prima di allenare la rete, in modo da minimizzarne poi l'errore in fase di convalida: sono solitamente selezionati in range standard iterando fino ad ottenere il comportamento sperato, poiché non c'è un'unica combinazione corretta di questi valori, né metodi diretti per ottenerli e sperimentare ogni combinazione non è pratico.

Si può allora ricorrere a metodi di ottimizzazione degli iperparametri che cercano di minimizzare una funzione obiettivo errore (spesso l'MSE in fase di convalida) tramite ricerca a griglia o casuale, che però ignorano ad ogni nuova valutazione i risultati di quelle precedenti.

Si impiega allora l'ottimizzazione Bayesiana, già disponibile all'interno del toolbox di deep learning di Matlab, che, a differenza della ricerca casuale o a griglia, tiene traccia dei risultati delle valutazioni passate, e le utilizza per formare un modello probabilistico che mappa gli iperparametri con un processo gaussiano: crea così un surrogato della funzione obiettivo che tenta di ottimizzare allo stesso modo.

Ad ogni step trova nuovi iperparametri per ottimizzare il surrogato, li applica alla funzione obiettivo, produce un nuovo surrogato e, continuando in questo modo, la conoscenza di step precedenti migliora l'approssimazione del surrogato alla funzione obiettivo reale finché gli iperparametri ottimali per il surrogato lo sono anche per la funzione obiettivo e l'errore di convalida viene minimizzato.

Nell'ottimizzazione Bayesiana, nell'addestramento e nella struttura della rete entrano in gioco diversi iperparametri e valori chiave; si vanno ora ad elencare mirando alla massima chiarezza espositiva.

Il dataset è suddiviso nelle 3 componenti:

- rapporto di training = 0.65
- rapporto di convalida = 0.2
- rapporto di test/verifica = 0.15

Tali valori sono stati scelti basandosi sulle indicazioni standard e sono presenti all'interno della funzione obiettivo da minimizzare, dove si ricorda che la funzione obiettivo fa un addestramento e produce l'MSE di convalida che l'ottimizzazione Bayesiana tenta di minimizzare per trovare i migliori iperparametri.

Dato che la suddivisione del dataset nelle sue componenti è realizzata in modo randomico (per non introdurre bias in fase di addestramento), non sempre si dimostra efficace. Una regola pratica per controllare la buona suddivisione dei dati è che l'epoch di minimo MSE di convalida sia prossimo all'epoch di minimo MSE di test. Gli epoch altro non sono che il numero di volte in cui la rete è stata esposta ai dati di addestramento durante lo stesso. L'ottimizzazione Bayesiana ripete l'addestramento varie volte, ma ogni addestramento prevede che la rete apprenda nuove informazioni (aggiornando bias e pesi) ad ogni epoch, perché l'apprendimento per ripetizione resta il punto focale del ML.

Il numero di volte in cui la Bayesiana ripete l'addestramento per minimizzare la funzione obiettivo ha un massimo fissato in partenza: per esperienza con questi specifici dati tale termine può essere tenuto tra 60 e 80, sopra ai 100 raramente l'MSE di convalida scende ulteriormente.

Si elencano ora gli iperparametri:

- numero di hidden layers: proposto tra 1 e 3, ottimizzato da Bayesiana
- numero di neuroni in ogni hidden layer: proposto tra 1 e 40, ottimizzato da Bayesiana
- funzione di trasferimento: fissata come da default ad una sigmoide tangente iperbolica
- funzione di attivazione: fissata come da default a lineare pura

Inoltre, l'algoritmo di training TA impiegato è la Levenberg-Marquardt backpropagation, il più usato e performante in letteratura, poiché rapido e preciso.

Le funzioni di trasferimento (imposte uguali in ogni layer) e quella di attivazione non sono state modificate rispetto allo standard perché sono iperparametri di tipo categorico e non numerico, più complessi da ottimizzare; l'esperienza con lo script ha dimostrato che i termini di default fossero la miglior combinazione disponibile. Le due funzioni sono riportate in Figura 4.2.1 e Figura 4.2.2.

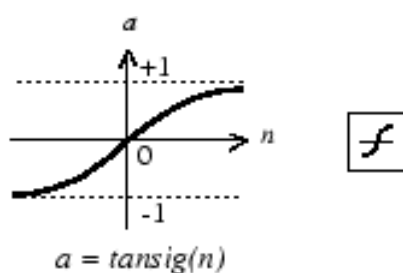


Figura 4.2.1 Funzione di trasferimento sigmoide tangenziale iperbolica

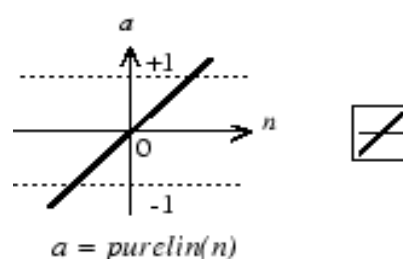


Figura 4.2.2 Funzione di trasferimento lineare

Si è posto un problema di natura fisica durante l'addestramento delle reti: avendo in tutti i layer funzioni con possibili output compresi tra -1 e +1, anche con tutti gli input iniziali tra 0 e +1 grazie alla normalizzazione, piccoli errori di stima possono accumularsi e causare una stima negativa di potenza in certi punti. Per sopperire a ciò si sono subito impiegate funzioni di attivazione con solo valori positivi che operassero da sostanziale filtro in output, ma non ha funzionato ed anzi tutte le funzioni di attivazione di quel tipo hanno impattato negativamente sulla precisione del modello. Per gestire anche le potenze basse, i *tansig* vanno alle volte a sovrastimare leggermente o sottostimare leggermente, e questo fa andare i valori anche sotto 0, ma impedirlo in fase di training limita la capacità generale della rete di seguire gli andamenti reali di produzione energetica e per quello le prestazioni peggiorano.

Si è allora optato per una semplice soluzione: annullare tutte le stime negative solo una volta in uscita dall'addestramento della rete, in modo tale da non turbarlo, raggiungere alte prestazioni ed ottenere solo valori fisicamente ragionevoli. Tale tecnica si è dimostrata ripetutamente efficace, senza abbassare i coefficienti di determinazione, poiché i valori negativi erano realmente di entità contenuta.

Alcuni iperparametri sono fissati, gli altri sono ottimizzati dalla Bayesiana minimizzando l'MSE di convalida, ma i parametri interni (bias e pesi), risultanti dall'iterazione con minimo MSE, non sono per forza i migliori. L'ottimizzazione ha infatti dato forma alla rete esternamente (con numero di layer e neuroni), ma parametri di apprendimento migliori si potrebbero comunque ottenere per cui, fissati gli iperparametri, si lancia un ciclo for (tra le 80 e 150 iterazioni solitamente) continuando ad allenare la rete per migliorare i parametri interni. Il criterio che si cerca continua ad essere quello standard nella creazione di una ANN, ovvero il minimo errore durante la fase di convalida.

Internamente al ciclo, le reti vengono applicate anche agli anni successivi, calcolando e salvando NMSE e coefficienti di determinazione mensili ed annuali.

Considerati gli step e script impiegati, il tutto risulta abbastanza rapido dal punto di vista computazionale (considerando anche che all'atto pratico potrebbe bastare il lancio degli script soltanto una volta al mese o all'anno), variando tra i 2 e gli 8 minuti di durata.

### 4.3 Applicazione delle reti e conseguenti considerazioni sull'impianto

Si elencano in tabella 4.1 sigle e corrispettivi significati per identificare le reti nei grafici a venire.

Tabella 4.1 Significato delle sigle per i nomi delle reti

A	Misure da stazione meteorologica ARPAE
P	Misure in situ d'impianto (PLANT)
4, 5, 6	Numero di parametri di input impiegati nel training
F/c.f.o.	Filtro degli outliers del 2014 durante il training
nS	Stagione non usata come input
nV	Velocità del vento non usata come input
nTinv	Temperatura inverter non usata come input
nTmod	Temperatura pannelli non usata come input

In questa prima fase di realizzazione dei modelli ANN e valutazione dei KPI ci si concentrerà sugli elementi ritenuti più rappresentativi, sia delle prestazioni delle reti, che dell'impianto.

Per quanto concerne le reti neurali, si prendono al momento in considerazione come KPI di previsione i valori NMSE ed  $R^2$ , per valutare l'accuratezza con cui il modello segue, ora per ora, l'andamento della potenza (viene svolto il calcolo rispetto alla componente AC immessa in rete) e l'MBE annuale, alla ricerca di eventuali bias di stima e per generare in anticipo un'opinione sulla qualità che i KPI legati alle prestazioni d'impianto (ma attesi, prodotti da modello) potrebbero avere.

Si è scelto l'NMSE (29) in particolare perché è un valore già relativo e adimensionale, per cui non servono frame di riferimento particolari per interpretarne il significato. La finestra temporale minima considerata è stata di tipo mensile poiché l'operazione di prelievo e pre-processing dei dati d'impianto ha causato la perdita di ore e giorni, vanificando ipotetiche suddivisioni giornaliere o settimanali, mentre si possono prendere in esame i valori complessivi mensili senza particolari remore.

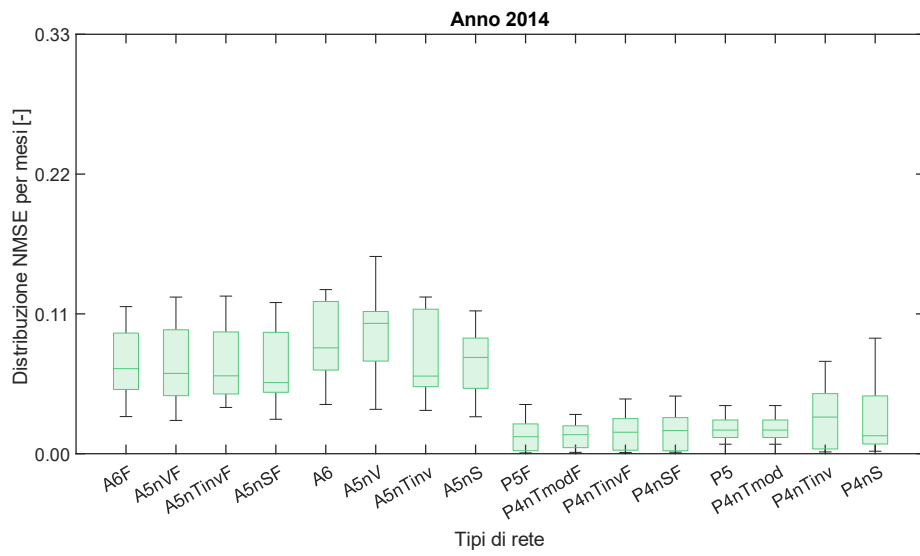
Tra i KPI legati alle performance reali d'impianto, invece, si calcola e rappresenta il più importante, ovvero il performance ratio, eventualmente accompagnato dalle versioni corrette in temperatura.

Questa sezione non prevede un'organizzazione troppo rigida perché KPI di previsione, di prestazione attesa e di prestazione reale possono dare informazioni reciproche e su particolari condizioni dell'impianto e delle misure ad esso associate, che preventivamente erano meno impattanti o non immediatamente riconoscibili. Starà poi all'accortezza dell'interpretazione distinguere dove termini l'influenza di un elemento (ad esempio un guasto all'impianto) ed inizi quella di un altro (come una scarsa accuratezza locale di una certa ANN).

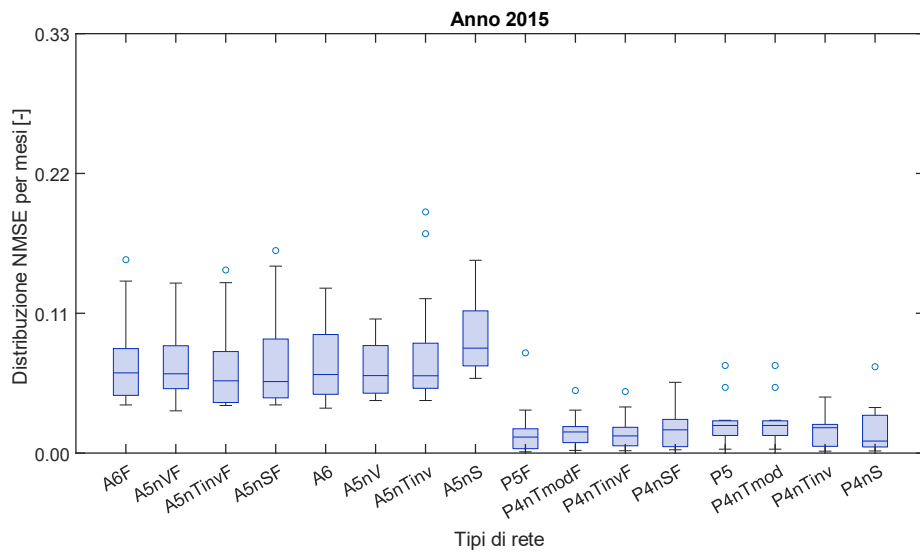
Si inizia col rappresentare, anno per anno, l'errore quadratico medio normalizzato NMSE mensile sotto forma di boxplot. Un boxplot è una rappresentazione di tipo statistico che racchiude nei box i valori dal primo al terzo quartile, ovvero esclude il 25% dei dati dal valore inferiore ed il 25% dei valori superiori. Le linee sopra e sotto i box sono definite "baffi" e dovrebbero includere tutti i valori restanti nel caso di distribuzioni relativamente uniformi, mentre alcuni punti al di fuori rappresenterebbero eventuali outliers. I tratti orizzontali contenuti in ogni box rappresentano la mediana di tale distribuzione di dati.

I boxplot contenenti gli NMSE mensili sono presentati dalla figura 4.3.1 alla figura 4.3.8.

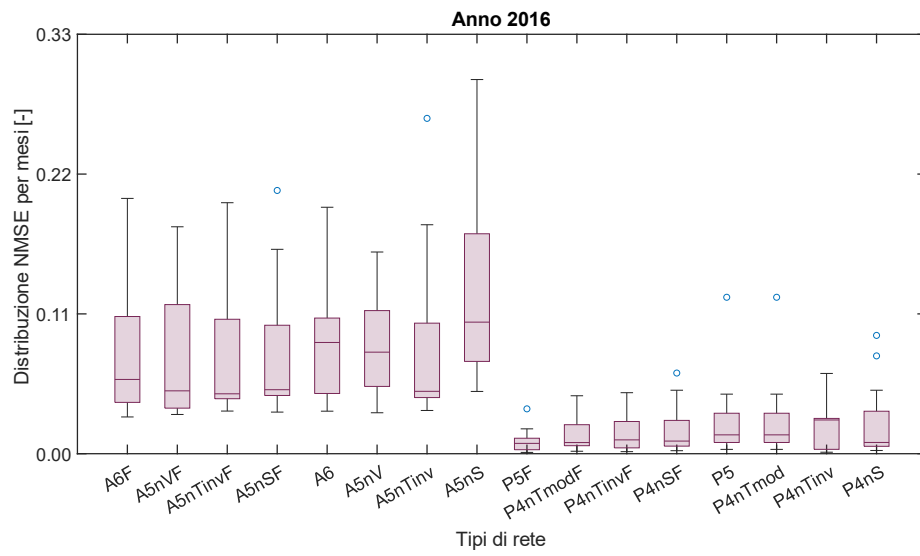




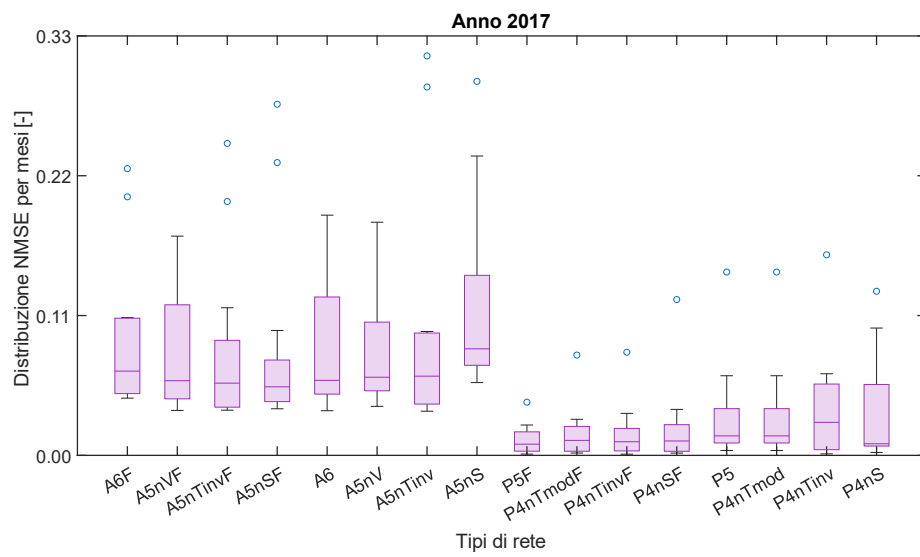
**Figura 4.3.1** boxplot dell'NMSE mensile 2014 per reti ARPAE e PLANT



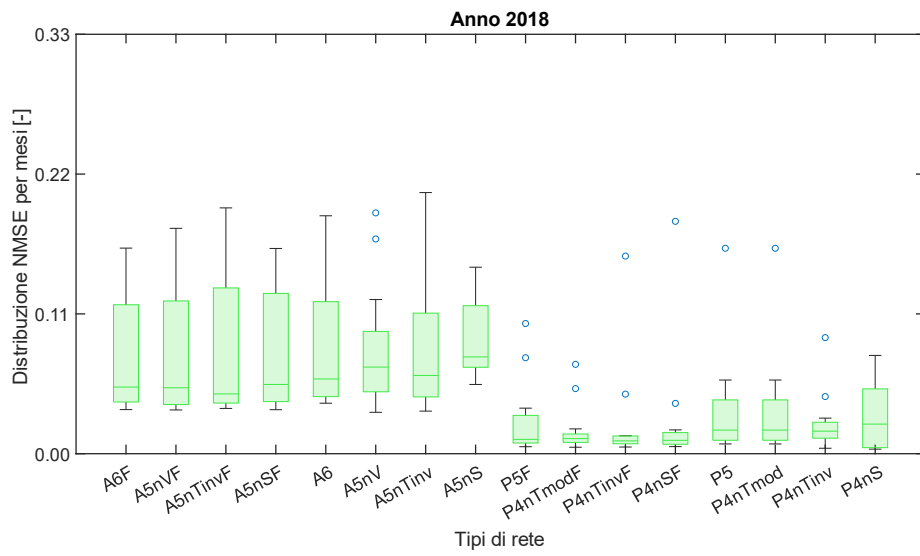
**Figura 4.3.2** boxplot dell'NMSE mensile 2015 per reti ARPAE e PLANT



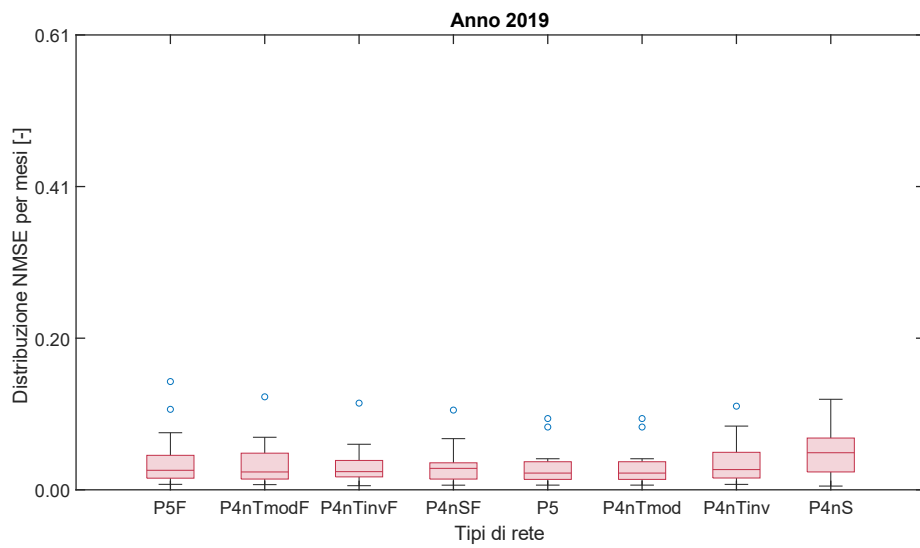
**Figura 4.3.3** boxplot dell'NMSE mensile 2016 per reti ARPAE e PLANT



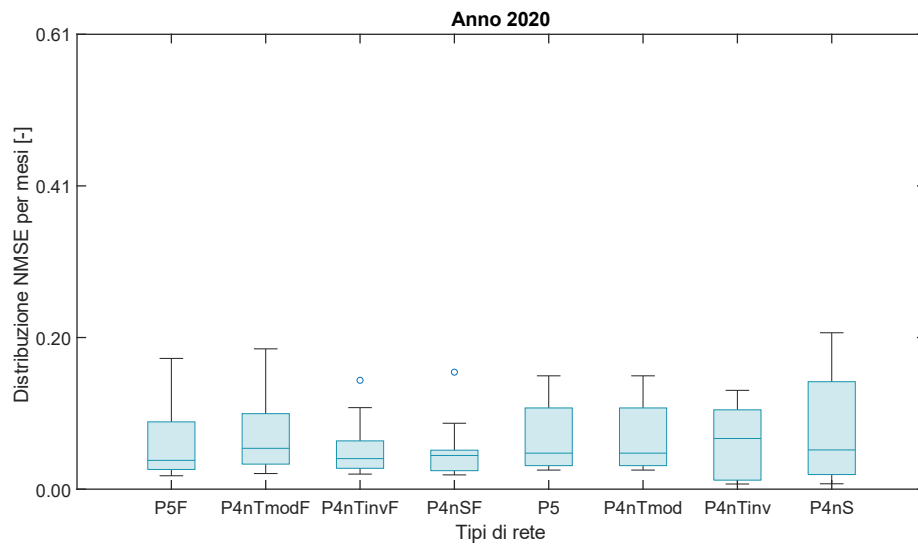
**Figura 4.3.4** boxplot dell'NMSE mensile 2017 per reti ARPAE e PLANT



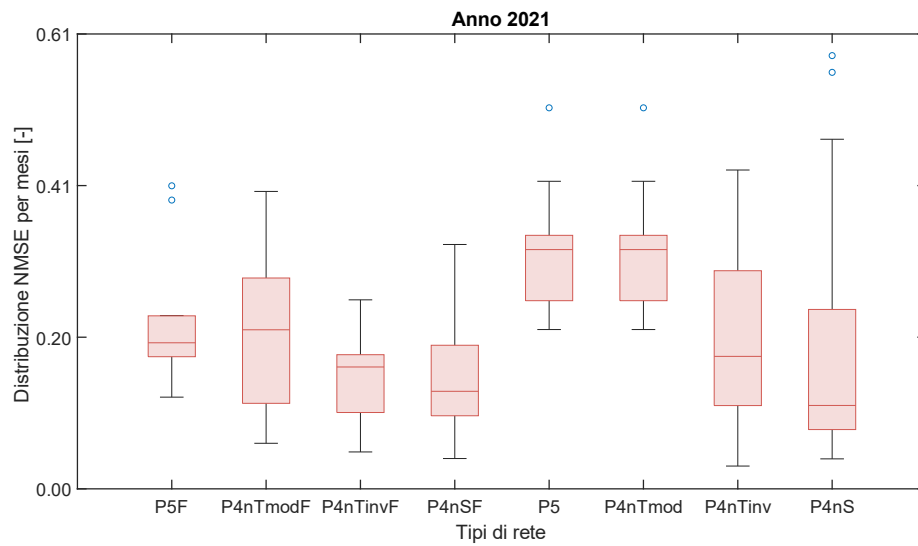
**Figura 4.3.5** boxplot dell'NMSE mensile 2018 per reti ARPAE e PLANT



**Figura 4.3.6** boxplot dell'NMSE mensile 2019 per reti ARPAE e PLANT



**Figura 4.3.7** boxplot dell'NMSE mensile 2020 per reti ARPAE e PLANT



**Figura 4.3.8** boxplot dell'NMSE mensile 2021 per reti ARPAE e PLANT

Già dai valori di NMSE si possono cominciare a fare alcune considerazioni:

- Le reti addestrate con dati ARPAE presentano costantemente scarti maggiori rispetto a quelle che usano dati d'impianto.
- Il filtro degli outliers in fase di addestramento offre un ottimo supporto alle reti ARPAE, che hanno variabilità maggiore dei dati (motivo per cui il comando di filtro automatico ne elimina una percentuale maggiore), mentre ha impatto più limitato ma comunque positivo sulle reti PLANT, in particolare a 5 input.
- I range di NMSE dipendono molto di più dalla famiglia di reti di appartenenza che dalla selezione specifica dei parametri di input. Stranamente ciò vale anche per le reti addestrate senza temperatura dei pannelli, che si appoggiano probabilmente agli andamenti dell'irradianza e della temperatura d'inverter.
- In genere le reti con più input hanno prestazioni migliori e quella che presenta complessivamente risultati più promettenti è la rete a 5 input di impianto con l'uso dei filtri.
- C'è una naturale e diffusa tendenza media all'aumento dell'errore con il passare degli anni, poiché per quanto precisa possa essere la modellizzazione di un impianto PV, impiegando il modus operandi standard dell'addestrare sul primo anno di dati, le ANN vanno a predire un funzionamento sempre simile a quello del primo anno e non possiedono strumenti per sapere di possibili interventi, manutenzioni o problemi, o percepire il lento invecchiamento delle componenti d'impianto. Questa ragione è plausibilmente dietro al continuo e lieve incremento degli scarti, ma non spiega perché tutti gli NMSE crescano in modo intenso negli ultimi 3 anni valutati (ad esempio dal 2018 al 2019 la mediana d'errore per P5F raddoppia da 0.011 a 0.026), come si vedrà più chiaramente con la rappresentazione dei coefficienti di determinazione (30).

Per dare maggiore chiarezza a quanto detto, si sfruttano gli andamenti dei coefficienti di determinazione mensili per proporre uno spunto grafico più intuitivo. Tali rappresentazioni si trovano dalla figura 4.3.9 a 4.3.12.

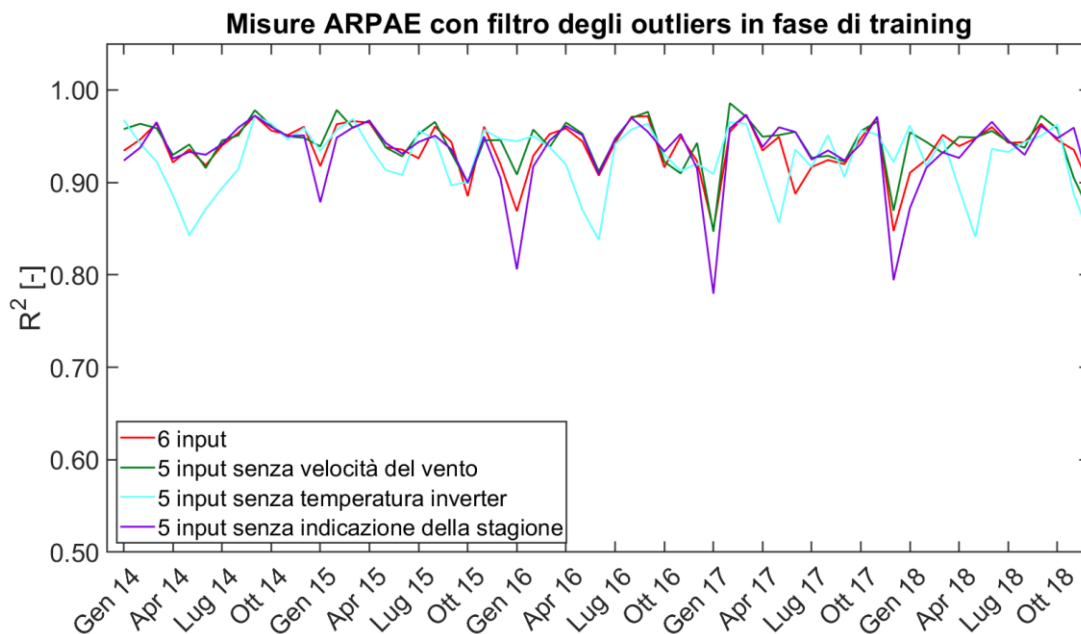


Figura 4.3.9 coefficiente di determinazione per reti ARPAE con filtro outliers

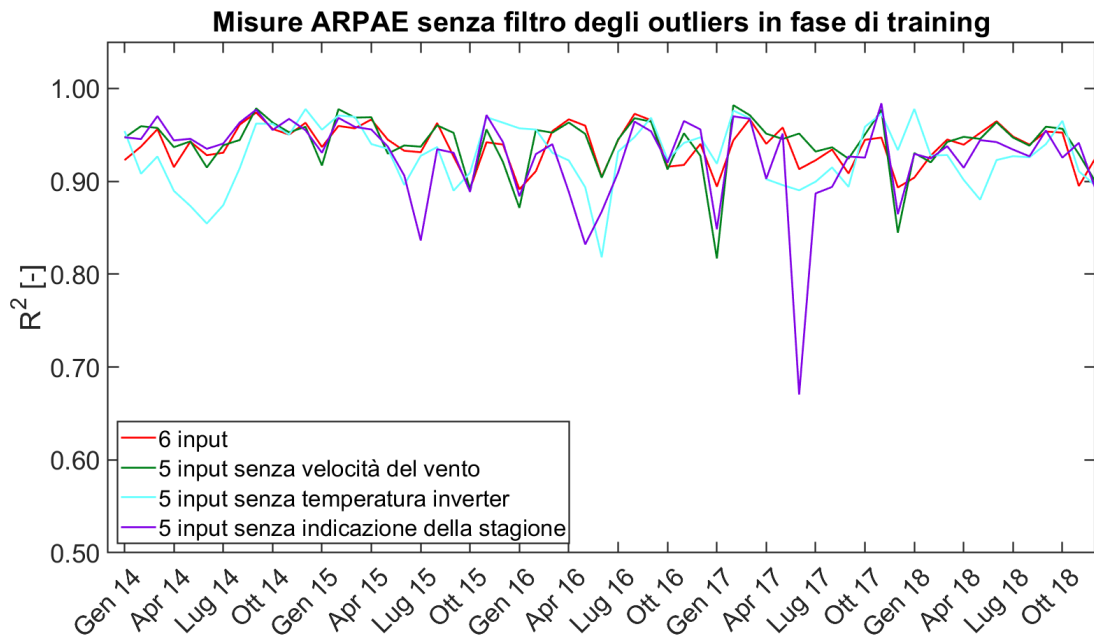


Figura 4.3.10 coefficiente di determinazione per reti ARPAE con filtro outliers

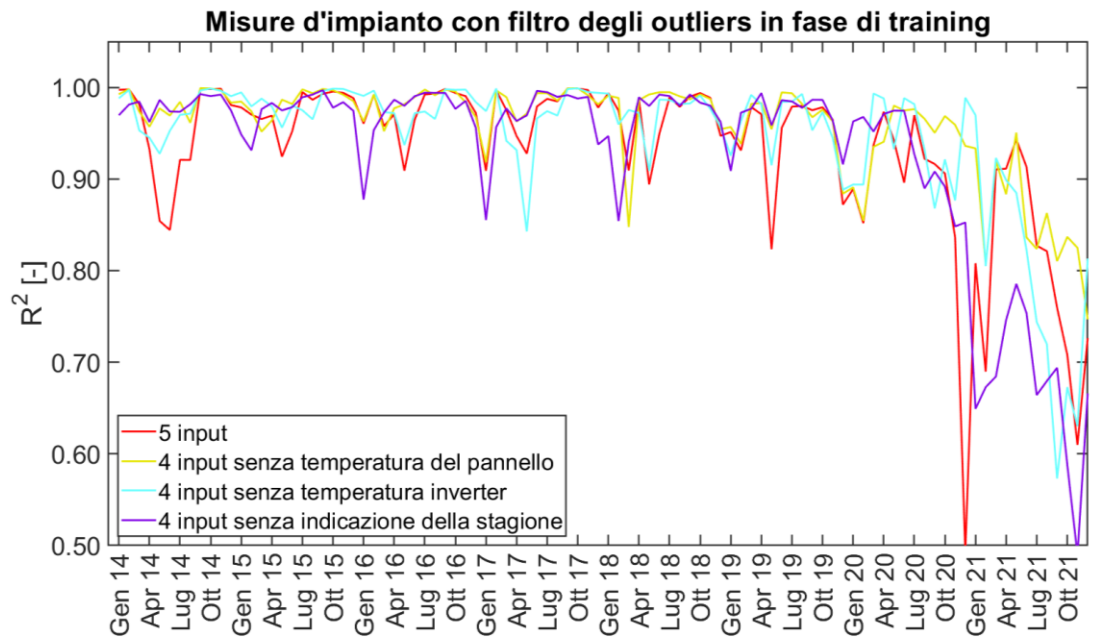


Figura 4.3.11 coefficiente di correlazione per reti PLANT con filtro outliers

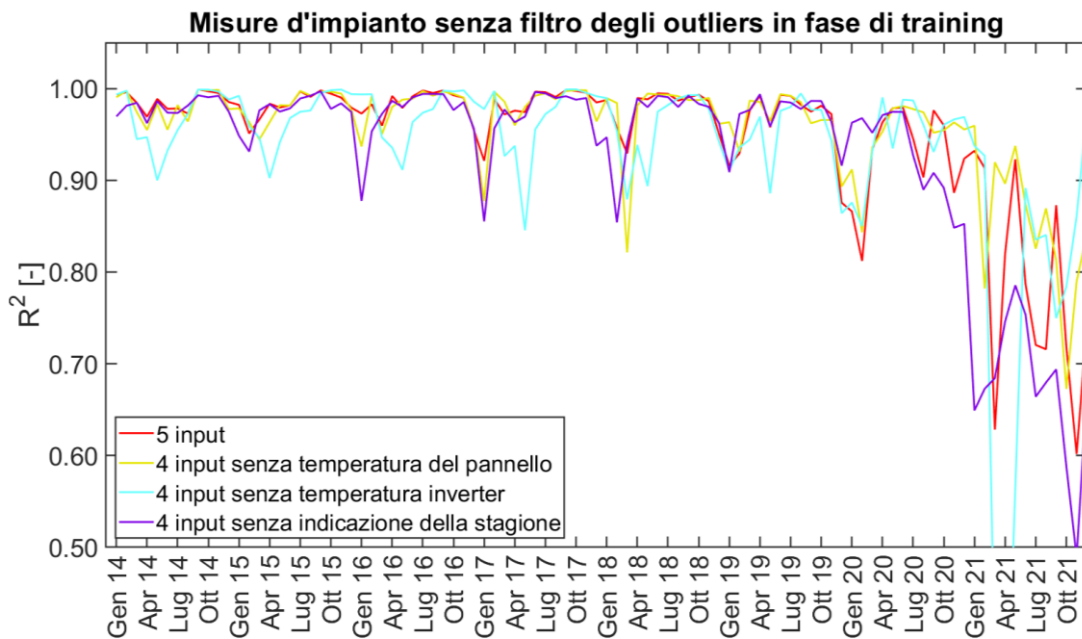


Figura 4.3.12 coefficiente di correlazione per reti PLANT senza filtro outliers

Questo KPI, che indica l'accuratezza (mensile in questo caso) con cui i modelli predicono la potenza PV, dà alcune informazioni aggiuntive rispetto alla distribuzione di errori vista nelle pagine precedenti:

- I minimi di precisione sono raggiunti sempre in corrispondenza dei mesi invernali, indipendentemente dalla rete considerata; possibili cause possono essere ombreggiamenti più frequenti, irradianza poco intensa misurata in modo errato ecc. Rimane da investigare al riguardo tra le grandezze misurate in impianto, perché pur essendo una problematica comune a entrambi i dataset, il fenomeno mostra un pattern più uniforme ed accentuato nel caso delle reti con dati di impianto.
- Nei mesi non invernali le reti PLANT hanno un ottimo comportamento, con coefficienti di determinazione prossimi a 1, mentre il massimo per le ARPAE sfiora di rado valori oltre 0.95.
- Nelle reti PLANT, le uniche a giungere oltre il 2018 come dati considerati, si inizia a notare un crollo dell' $R^2$  a cavallo tra 2019 e 2020, con le reti addestrate senza temperatura che mostrano una performance migliore, seppur peggiorata rispetto agli altri anni. L'addestramento senza questo parametro di input è stato eseguito proprio a causa del rilevamento di temperature anomale dei pannelli nel corso degli ultimi anni. Il ragionamento sulla temperatura sarà espanso a breve ma è bene notare che anche le reti addestrate senza questa grandezza in input subiscono il tracollo della capacità di previsione, per cui la ragione trainante deve essere un'altra, anch'essa da indagare.
- Non c'era un effetto evidente nelle distribuzioni di NMSE, ma si può osservare come i picchi negativi di coefficiente di determinazione siano raggiunti per le reti allenate senza il valore delle stagioni, con effetto particolarmente accentuato per i casi ARPAE. Questo significa che l'informazione offerta dalla stagione subentra come parametro di supporto durante i mesi freddi, quando altri parametri (da identificare) divengono più inaffidabili.

Come preventivato, l'analisi dei KPI di previsione del modello (ovvero i KPI espressi come scarti o complementari degli stessi) ha permesso non solo di evidenziare forze e debolezze dei modelli in sé, sulla base degli input e dataset impiegati, ma di riscontrare anomalie logicamente imputabili a funzionamenti subottimali di qualche tipo all'interno dell'impianto PV, il tutto favorito dalla suddivisione delle reti in tanti sottogruppi, che ha permesso di poter assegnare in modo più preciso certi effetti d'errore discriminando sulla base delle specifiche di creazione delle diverse reti.

D'ora in avanti i risultati esposti saranno stati raggiunti compiendo semplici analisi statistiche delle varie grandezze, ma motivati dallo spiegare i comportamenti dei KPI analizzati. Il performance ratio e le sue variazioni saranno valutati successivamente per dedurre informazioni aggiuntive sullo stato di funzionamento dell'impianto.

### 4.3.1 Problemi di misura della temperatura di pannello

In allarme a causa degli errori elevati riscontrati nella modellizzazione degli anni più recenti, si sono indagate le grandezze di input e output in cerca di possibili anomalie: la più lampante per la variazione repentina è stata la temperatura dei pannelli misurata nell'impianto.

Tale dato esibisce un'evidente e ripida diminuzione a tutto tondo, scendono i quartili e la mediana, abbastanza stabili negli anni precedenti, il che risulta difficilmente imputabile a variazioni climatiche di tale scala (Figura 4.3.13).

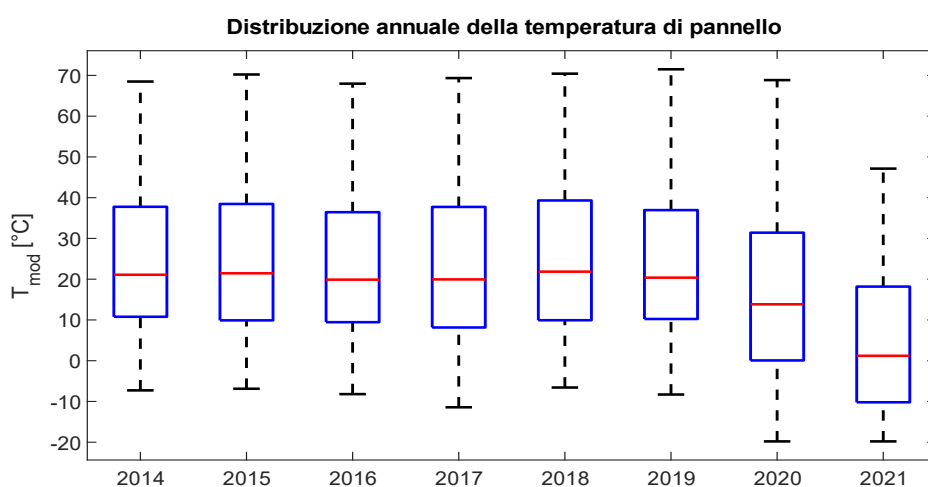


Figura 4.3.13 Boxplot della temperatura dei pannelli

Più nello specifico, riportando l'intero andamento delle misure di temperatura dei moduli in Figura 4.3.14, il cambiamento è evidente anche considerando la variabilità termica stagionale.

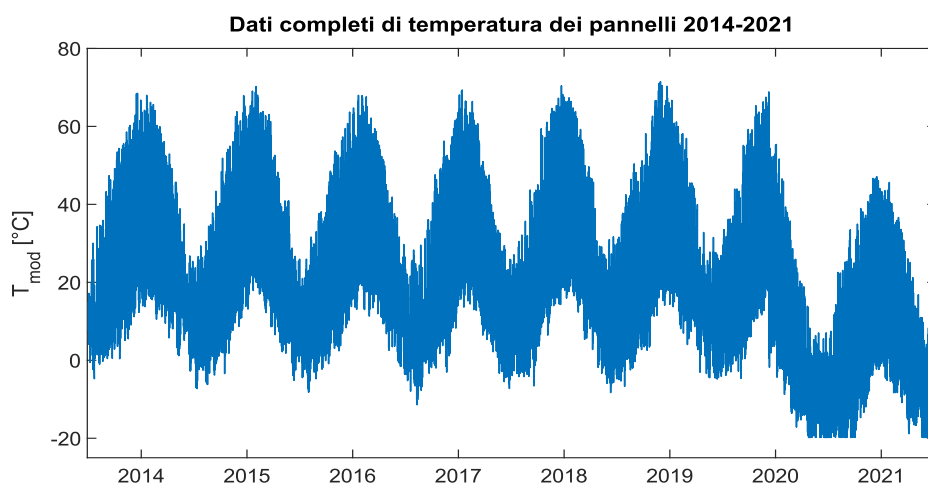
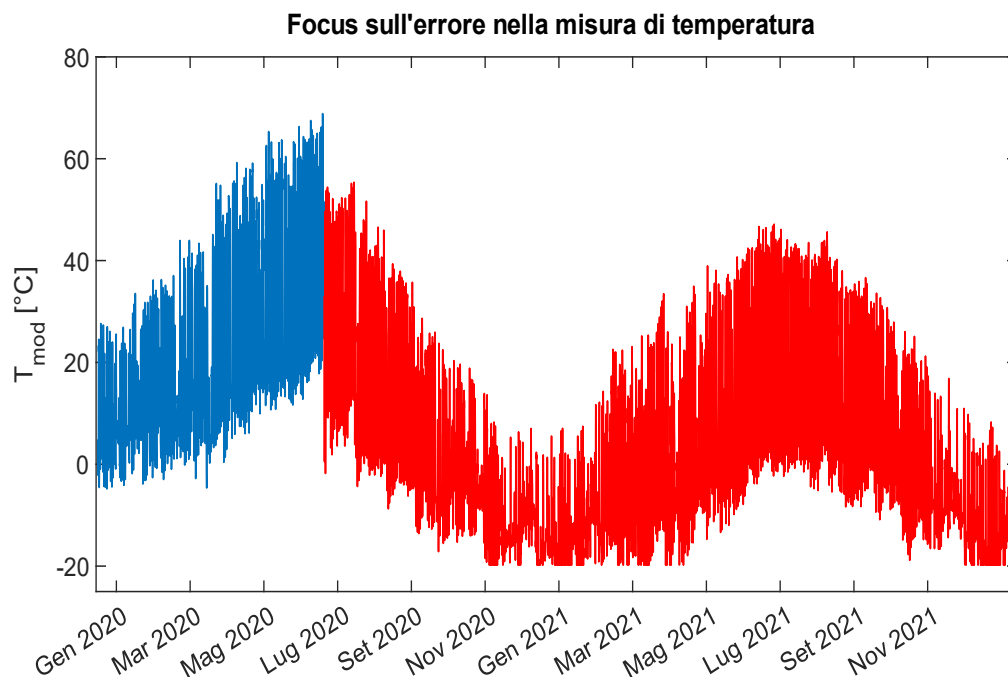


Figura 4.3.14 Andamento completo della misura temperatura di pannello



Guardando infine al dettaglio, si individua in Figura 4.3.15 un drastico shift verso il basso dei valori in data 7/7/20. Le misure del sensore di temperatura non solo scendono di molto in ampiezza, ma i valori più bassi sono come ‘tagliati’ in corrispondenza dei  $-20^{\circ}\text{C}$ , probabilmente a causa di un offset mal impostato o altri errori nella gestione dei dati del sensore, se non una vera e propria anomalia dello stesso. Questo ha implicato che, anche scalando rispetto alle temperature medie degli anni precedenti, il corretto andamento dei valori restasse irrecuperabile (alcune prove sono state eseguite, con scarsi risultati).



**Figura 4.3.15 Dettaglio del problema di misura della temperatura al 07/07/20**

Proprio in virtù di quanto scoperto, si è deciso di addestrare anche reti senza temperatura di pannello in input, nonostante normalmente la temperatura sia uno dei principali parametri ad influenzare corrente e tensione di funzionamento dei pannelli, e di conseguenza la potenza prodotta, diventando un input chiave nel training. Come visto con i coefficienti di determinazione, si è ottenuto un lieve miglioramento per alcuni punti degli anni finali, ma il problema del collasso di precisione persiste anche in tali reti, rendendo l'errore di misura di temperatura una concausa.

Per quanto la presenza di problematiche contemporanee sia un elemento di disagio per il funzionamento d'impianto, esalta invece l'utilità del lavoro fin qui svolto: il gruppo di reti nella sua interezza, ha evidenziato un disturbo in una certa finestra temporale, ed una volta riconosciuto uno dei valori responsabili, implementare reti che non lo utilizzassero e confrontarne le prestazioni con le altre ha permesso sia di ottenere conferma sul ruolo della temperatura, che portare alla luce la preziosa informazione che altri dati sono fuori posto nella stessa finestra temporale.

Non trovando grandi anomalie negli altri input, ci si è rivolti alle potenze in output.

Seppure anche nella potenza complessiva fosse difficile individuare problemi evidenti, si è giunti ad una conclusione analizzando gli andamenti di potenza dei singoli ingressi degli inverter (contenenti alternativamente una o due stringhe l'uno).

### 4.3.2 Guasto delle stringhe ed impatto sulle stime di energia e potenza

In Figura 4.3.16 sono raffigurati gli andamenti di tutte le potenze DC annuali medie in ingresso agli inverter, scalate rispetto al proprio valore medio di potenza DC prodotta nel 2014 (primo anno di misure).

Le potenze in questione sono state rese in forma relativa per equilibrare gli ingressi a doppia stringa e quelli a stringa singola. La crescita iniziale di potenza non è indice d'allarme, poiché l'irraggiamento del 2014 è stato di entità limitata.

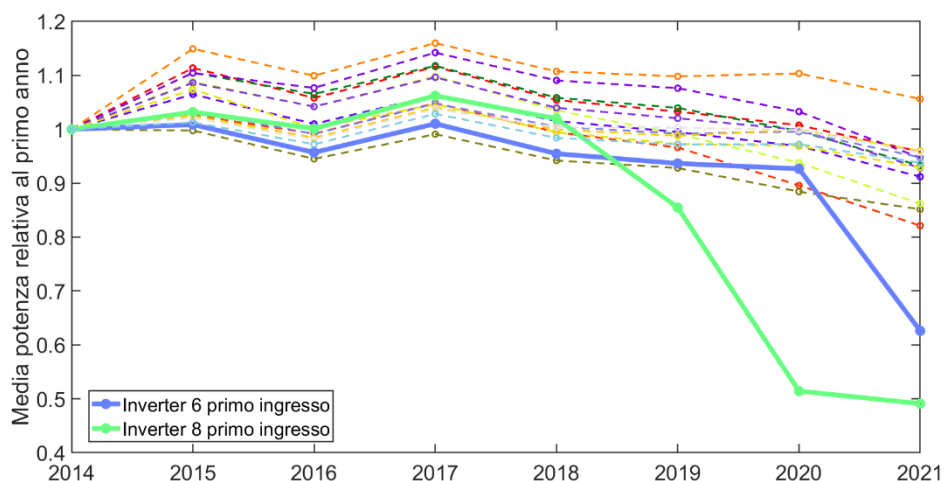


Figura 4.3.16 Potenze DC medie annuali relative

Subito si fanno evidenti gli andamenti anomali per i primi ingressi (a doppia stringa) degli inverter 6 e 8, che si spiegano con il distacco/guasto di una delle stringhe per ognuno, portando all'incirca ad un dimezzamento della potenza prodotta.

Il primo guasto a verificarsi è imputabile all'inverter 8, in data 06/08/19, Figura 4.3.17

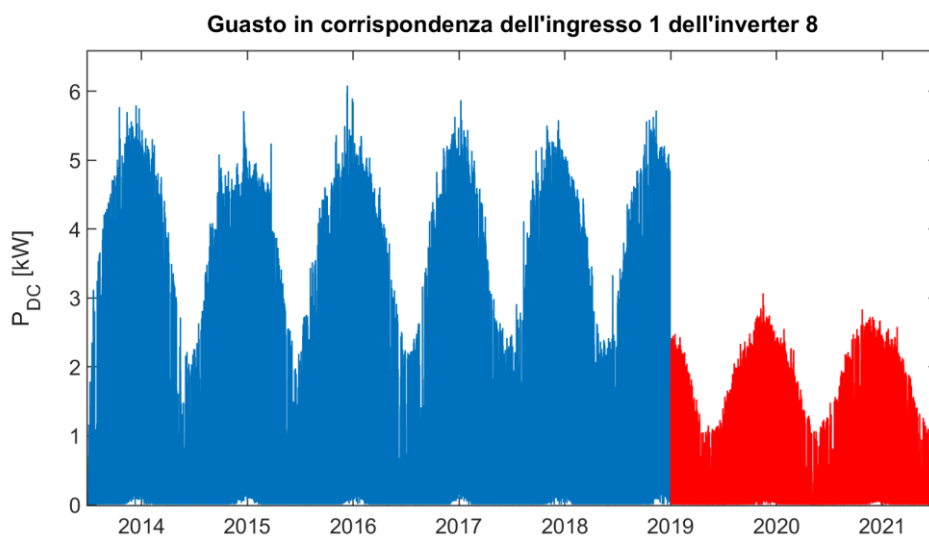


Figura 4.3.17 Guasto di una stringa da 16 moduli collegata all'inverter 8

La curva dei valori di potenza dopo il guasto mantiene la stessa forma annuale, soltanto dimezzata.

Lo stesso si verifica con un'altra stringa all'ingresso dell'inverter 6, in data 07/06/21, Figura 4.3.18.

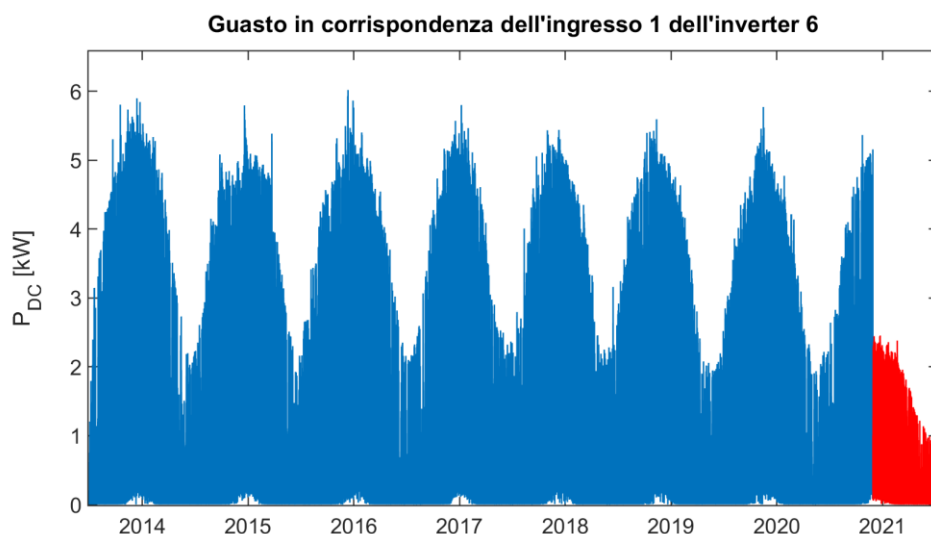


Figura 4.3.18 Guasto di una stringa da 16 moduli collegata all'inverter 6

La perdita della prima stringa provoca una diminuzione della potenza nominale dell'impianto pari a:

$$Loss_{inv8} = 16 \cdot 0.25 = 4 \text{ kW} \xrightarrow{\% P_0} 4.08\%$$

Da agosto 2019 in poi, quindi, la potenza nominale effettiva dell'impianto si riduce al 95.92% di quella iniziale. La perdita della seconda stringa impone un'ulteriore riduzione del 4.08%, portando, da giugno 2021 in poi, ad avere una potenza nominale effettiva di 90 kW, ovvero il 91.84% di quella iniziale.

Le reti neurali, allenate con input ambientali e output di potenza, quando poi applicate ai periodi successivi all'addestramento non hanno a disposizione informazioni sui guasti a livello produttivo perché le potenze devono stimarle. Unendo i due guasti all'errore di misura della temperatura, si giustificano in larga parte gli sviluppi altamente imprecisi ed instabili delle reti negli ultimi 3 anni, poiché operano come se tutti i dati ricevuti in input fossero affidabili e non ci fosse nessun fenomeno esterno ad inficiare i valori di potenza.

Questo non è un difetto della realizzazione delle reti, bensì l'opposto: nel confrontare valori stimati in condizione sana dell'impianto con quelli reali, esplicite discrepanze svolgono come desiderato il ruolo di indice di funzionamento (o misura) subottimale.

### 4.3.3 Calcolo dell'MBE

Un KPI di modello di interesse prima del calcolo dei KPI di prestazione dell'impianto è l'errore medio di bias MBE (32): poiché i termini di energia sono fondamentali nel calcolo dei KPI di prestazione e ci si è finora focalizzati su valori mensili ed annuali (anche perché una stima accurata a livello giornaliero è un'impresa più complessa da realizzare), ad essere rilevante è l'errore di energia (mensile ed annuale). Tenuto conto che tutto lo studio è stato eseguito con dati di potenza media oraria, il calcolo dell'energia è numericamente uguale alla sommatoria delle potenze e perciò la differenza tra energia stimata e misurata altro non è che l'MBE senza divisione per il numero delle ore.

Questo KPI ha anche il vantaggio di estrapolare dal segno se la rete stia sovrastimando o sottostimando e nel caso se un pattern o offset sia presente.

Come valore medio o meno, resta un KPI con specifica unità di misura e variabile con il range di potenze definito dalla taglia dell'impianto. Per una maggiore generalizzazione dei risultati si impiega, quindi, la versione normalizzata con la potenza media misurata nell'intervallo temporale:

$$MBE_{P,rel} = \frac{1}{N} \sum_{i=1}^{i=N} \frac{(PP_i - PM_i)}{\overline{PM}} = \frac{\overline{PP} - \overline{PM}}{\overline{PM}} \quad (34)$$

Tale scrittura è equivalente al calcolare lo scarto di energia mensile/annuale e normalizzarlo con l'effettiva energia prodotta nel lasso di tempo considerato. Risulta allora:

$$MBE_{P,rel} = \frac{E_{AC,P} - E_{AC,M}}{E_{AC,M}} \quad (35)$$

In Figura 4.3.19 sono mostrati gli MBE relativi annuali per tutte e 16 le reti, dove, da sinistra, si trovano le reti ARPAE con filtro degli outliers (indicato c.f.o.) e senza, le reti PLANT con filtro e senza.

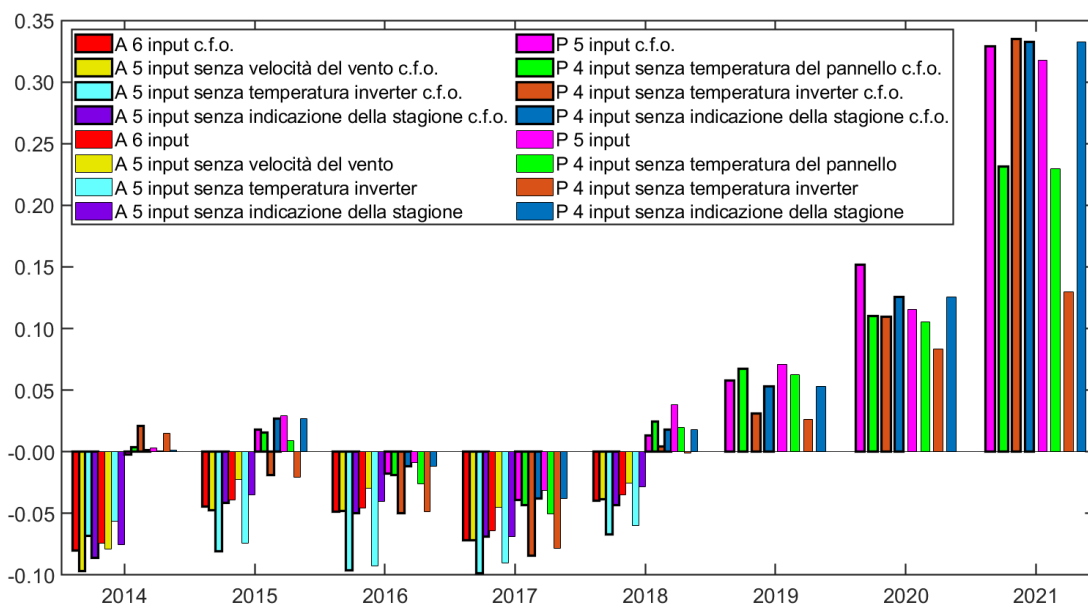


Figura 4.3.19 MBE relativi annuali per ogni rete

Il grafico è dominato dagli elevati errori di reti ARPAE e della tripletta di anni finali. Da questo punto del lavoro in poi, si porrà meno attenzione alle reti ARPAE in quanto gli errori sono considerevoli anche nell'anno di addestramento delle reti. La continua sottostima delle potenze sarà però un utile elemento di confronto.

In merito agli MBE di 2019 – 2021, essi rivelano una crescente sovrastima, causata dalla temperatura falsamente bassa (che la rete ha imparato ad associare ad incrementi della potenza prodotta) e dalla mancata conoscenza dei due guasti di stringa. Si potrebbero introdurre ipotesi riguardo un'ulteriore concausa, ma si lascia in sospeso il ragionamento per integrare le valutazioni dei KPI di prestazione della sezione seguente.

Tra le reti PLANT, le due a 5 input (con e senza filtro, in rosa) continuano a dominare dal punto di vista dell'accuratezza, con errori massimi nei primi 5 anni del 3% senza filtro, e in media ben più bassi.

Gli MBE relativi annuali per le due reti con dati d'impianto a 5 input sono riportati in Tabella 4.2.

**Tabella 4.2 MBE relativi [%] delle reti PLANT a 5 input**

Rete\anno	2014	2015	2016	2017	2018	2019	2020	2021
P5 c.f.o.	-0.25	+1.81	-1.78	-3.90	1.15	5.34	15.19	32.91
P5	-0.23	+2.90	-0.86	-3.15	2.75	6.39	11.53	31.73

Si denota un pattern simile negli MBE relativi, visibile anche per tutte le altre reti PLANT: l'errore è giustamente al minimo nell'anno di addestramento, poi diventa positivo nel 2015, negativo tra 2016 e 2017 e positivo di nuovo nel 2018.

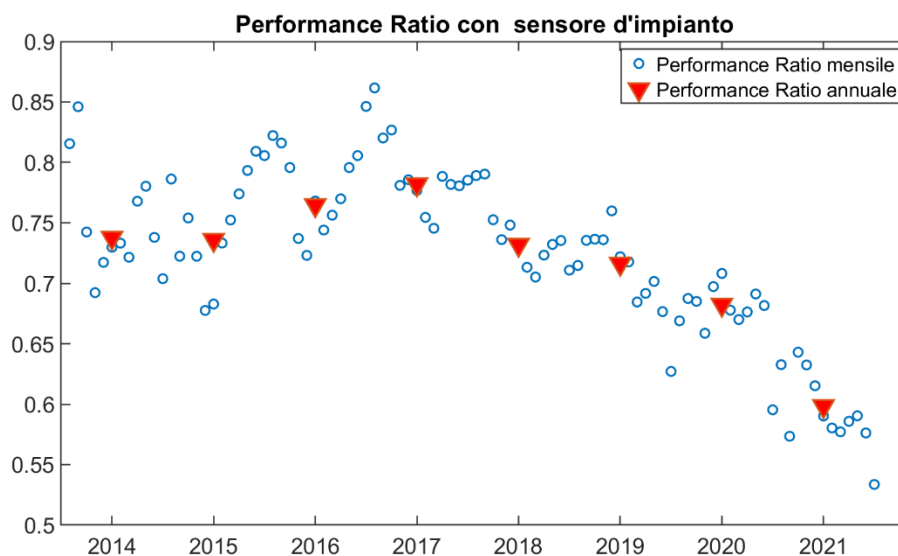
Giunge ora utile il confronto con le reti basate sui dati regionali, in quanto questo pattern è totalmente diverso dalla costante sovrastima ARPAE, e si può dedurre che a provocare tale tipo di andamento sia una differenza tra i due dataset, un altro parametro d'impianto che introduce possibili anomalie variabili.

Si andrà nel dettaglio valutando i KPI di prestazione dell'impianto.

#### 4.3.4 KPI di prestazione: PR, CPR

Ci si concentra ora sugli indicatori chiave di prestazione, da confrontare con gli stessi ottenuti da stime di potenza. Si svolge questo passaggio alla fine di una serie di valutazioni che hanno permesso di portare alla luce molteplici criticità nel funzionamento e nelle misure d'impianto, così da non giungere impreparati a valori peculiari ed alle cause che li generano.

Il performance ratio (10) calcolato con i dati di impianto è mostrato in Figura 4.3.20.



**Figura 4.3.20 PR calcolato dalle misure d'impianto**

Un rapido sguardo ai valori di PR annuali è sufficiente ad evidenziare numerose anomalie; per categorizzarle e comprendere al meglio la situazione, si considerano gli anni in tre gruppi distinti:

- A. [2014-2017]: il PR cresce costantemente, e questa è l'anomalia principale poiché la tendenza del PR dovrebbe generalmente essere monotona decrescente a causa dell'invecchiamento dei pannelli. Mensilmente (non considerando la crescita annuale atipica) si delinea la corretta forma di parabola inversa, dovuta all'innalzamento di prestazioni con il freddo invernale ed al peggioramento con il caldo dei mesi estivi. Fanno eccezione novembre e dicembre del 2014, esibenti valori particolarmente bassi.

- B. [metà 2017 – metà 2019]: termina l'andamento crescente ed inizia una più naturale decrescita, anche se il salto annuale dal PR del 2017 a quello del 2018 è particolarmente ripido a causa dei PR ancora molto elevati della prima metà del 2017. Mensilmente, i valori sono più compatti e non esibiscono una distinta distribuzione a parabola; si verifica nuovamente la caduta del PR mensile durante i mesi freddi.
- C. [metà 2019 – 2021]: il trend mensile non cambia nella forma rispetto alla sezione precedente, ma annualmente si verifica un crollo dei PR.

La motivazione di alcuni di questi fenomeni è intuibile, mentre altri sono più celati. Si indagherà sulla questione per step, ma prima si fa un breve richiamo al segno degli MBE della sezione precedente. Le reti, addestrate sul primo anno, apprendono un certo legame tra input e output, e quindi anche tra irradianza e potenza: ciò significa che negli anni successivi si aspetteranno un PR grossomodo stabile e legato a tale correlazione; per cui, dove il PR cresce, la rete sottostima e l'errore è negativo (2016, 2017), mentre dove il PR cala, la rete sovrastima e l'errore è positivo (dal 2018 in poi).

In C. la rapida diminuzione del PR è dovuta a ciò che si verifica nella potenza AC prodotta (che si trova al numeratore del PR), che abbiamo visto decrescere a causa della perdita di due stringhe. L'effetto è cumulativo con l'invecchiamento dei pannelli.

La ragione dietro alla differenza negli andamenti mensili ed annuali tra A. e B. non è rapidamente identificabile, e questo è dovuto alla composizione del vettore di valori di irradianza misurata nell'impianto. Reputando di massima affidabilità la misura di potenza immessa in rete (anche per le implicazioni economiche di avere una misura valida di energia prodotta), si è ricercata l'anomalia nei dati d'irraggiamento.

Come già accennato in precedenza, l'irradianza non corrisponde sempre ai valori di sensore, ma su VCOM (dove i dati d'impianto sono raccolti) può essere compiuta la sostituzione con dati trasposti in POA a partire dai valori satellitari.

Si sono quindi ispezionati a fondo i dati di irradianza, con un certo grado di difficoltà (poiché, quando i dati da satellite non necessitavano di essere calcolati, il relativo vettore di valori su VCOM veniva riempito con le misure di sensore invece che essere lasciato vuoto). Per i primi anni di operazione i dati satellitari erano quindi calcolati di rado e sovente indistinguibili dai valori misurati in impianto, ma a partire dalla seconda metà del 2017 in poi i dati satellitari hanno iniziato ad essere aggiornati coi valori coerenti e sono andati progressivamente a sostituire sempre più spesso quelli di sensore nel vettore d'irradianza complessivo.

Questa realizzazione permette di motivare la distinzione tra gli andamenti nei due periodi, ma serve ancora precisare perché le due circostanze impongano al PR proprio le tendenze annuali e mensili che manifesta. La curva mensile dei PR in A. è ben definita perché comunque i valori sono misurati all'inclinazione, orientamento e ubicazione corretti, mentre con i dati satellitari in B. l'accuratezza dei valori d'irradianza stimati ha dei limiti, con il vantaggio però di valori complessivi mensili ed annuali più attendibili, in questo specifico contesto. Ciò che si intende è che il PR in decremento dà conferma dell'attendibilità media dei valori satellitari.

Per quanto concerne la crescita annuale inattesa con i valori di sensore, questa implica una sottostima dell'irradianza. Assieme al calo di PR in alcuni mesi invernali, le evidenze possono puntare verso fenomeni di ombreggiamento non considerati.

Si è allora andati ad estrarre immagini satellitari ed al livello del suolo dell'impianto, rivelando la presenza di alberi lasciati a crescere di fronte alla falda Est.

Nella figura 4.3.21, risalente al 2019, si nota chiaramente che l'altezza degli alberi è sufficiente a creare irraggiamento frequentemente, soprattutto nei mesi invernali e alla mattina, con radiazione molto inclinata ed il sole basso nel cielo. Le foto in Figura 4.3.22 e 4.3.23 mostrano la crescita degli alberi tra gli anni.



**Figura 4.3.21 Vista da Nord della falda est e degli alberi**



**Figura 4.3.22 Vista dall'alto dell'impianto a luglio 2015**

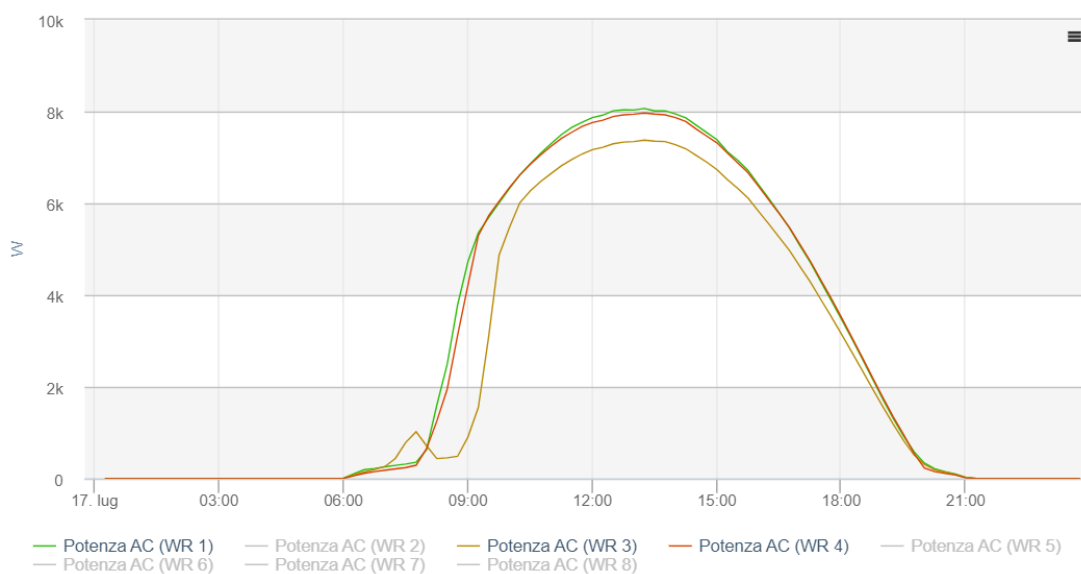


**Figura 4.3.22 Vista dall'alto dell'impianto a luglio 2017**

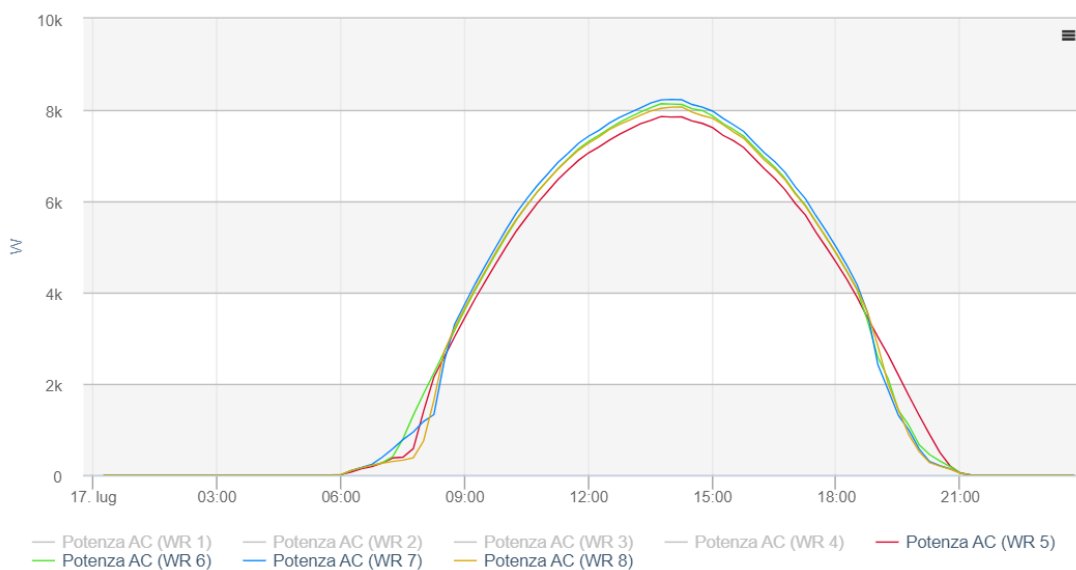
Essendo gli alberi posizionati ad Est della falda Est stessa, sono prони a ripetere questo fenomeno d'ombreggiamento, danneggiando le performance della falda Est e di conseguenza dell'intero impianto.

Come se ciò non bastasse, il sensore è particolarmente esposto a tale ombreggiamento, più del resto dell'impianto, per cui, con la crescita degli alberi ed il passare degli anni, sottostima sempre più l'irradianza (anche in momenti in cui il resto della falda non è particolarmente ombreggiato) e sovrastima quindi il PR, mentre i valori satellitari in parte sovrastimano l'irradianza perché non hanno informazioni sull'ombreggiamento locale, e questo spiega ulteriormente la pendenza (seppur nella direzione corretta) dei PR calcolati dal 2018 in poi.

La Figura 4.3.23 sottostante mostra chiaramente il fenomeno in atto durante un giorno di cielo sereno di luglio 2016 sulle stringhe appartenenti alla falda est. L'immagine è rappresentativa, ma molti giorni si presentano in questo modo. Di confronto per verificare la presenza dell'ombreggiamento soltanto sulla falda est, in Figura 4.3.24 sono mostrate le potenze di falda ovest prodotte nello stesso giorno.



**Figura 4.3.23 Dimostrazione dell'impatto dell'irraggiamento sulla falda est**



**Figura 4.3.24 Falda ovest non soggetta all'ombreggiamento degli alberi**



Una conferma più tecnica e meno legata alla mera rappresentazione visiva si può ottenere proprio applicando il PR in modo specifico alle due falde. Ovviamente non sarà del tutto rappresentativo del collegamento fisico tra le due grandezze di potenza ed irradianza poiché, sebbene si possa ottenere la potenza complessiva di falda sommando quella di tutti gli inverter presenti sulla stessa, la misura di irradianza è una soltanto.

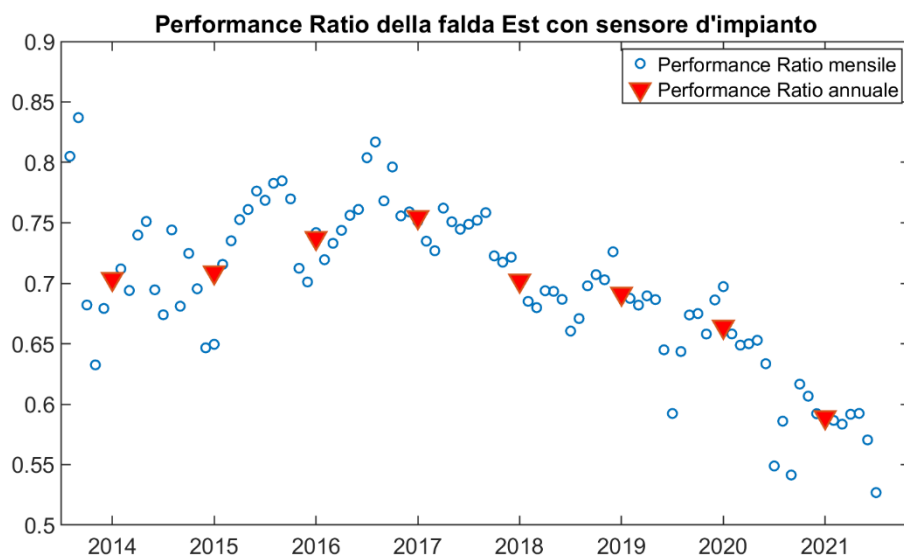


Figura 4.3.25 Performance ratio calcolato da potenza della falda est ed irradianza comune

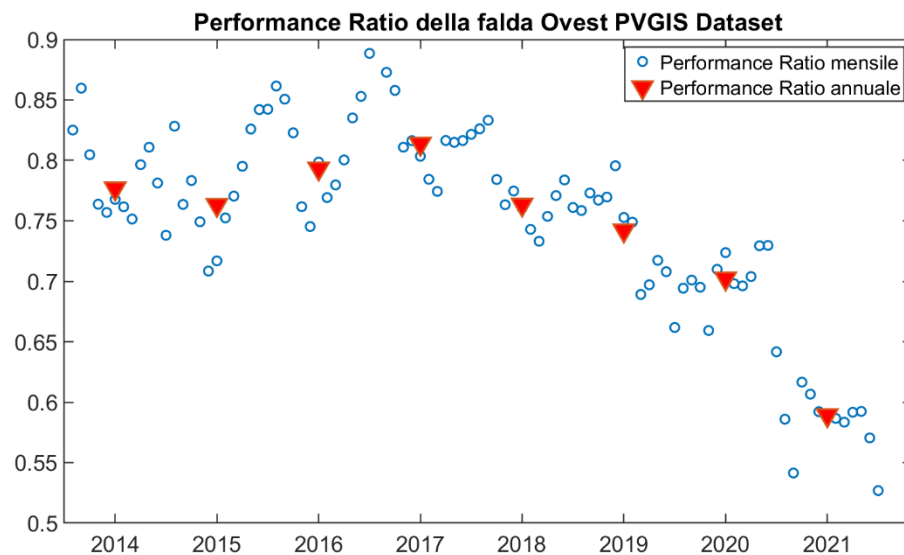
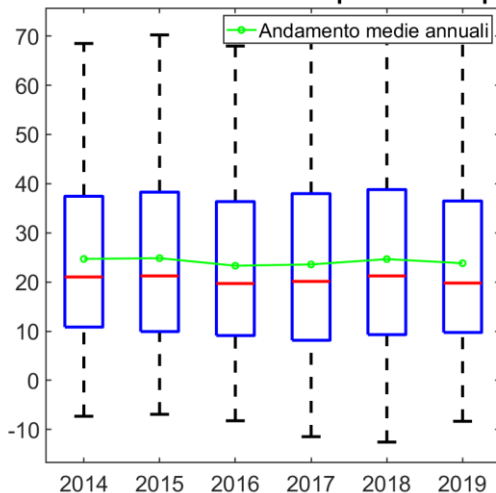


Figura 4.3.25 Performance ratio calcolato da potenza della falda ovest ed irradianza comune

Confrontando le due figure dei PR di falda, l'andamento rimane pressoché invariato come forma (fortemente influenzato dall'irradianza in comune), mentre le ampiezze sono maggior per la falda ovest che, libera dall'ombreggiamento, produce sempre più della falda est. L'unica eccezione riguarda l'ultimo anno, in cui ormai entrambe le stringhe sono andate perse (e si ricorda che appartengono entrambe alla falda ovest).

Un altro elemento ha lievemente aggravato la situazione del PR, in particolare nel 2016 e 2017, ovvero l'abbassamento sia di media che mediana delle temperature dei pannelli (Figura 4.3.26):

**Distribuzione annuale della temperatura dei pannelli**

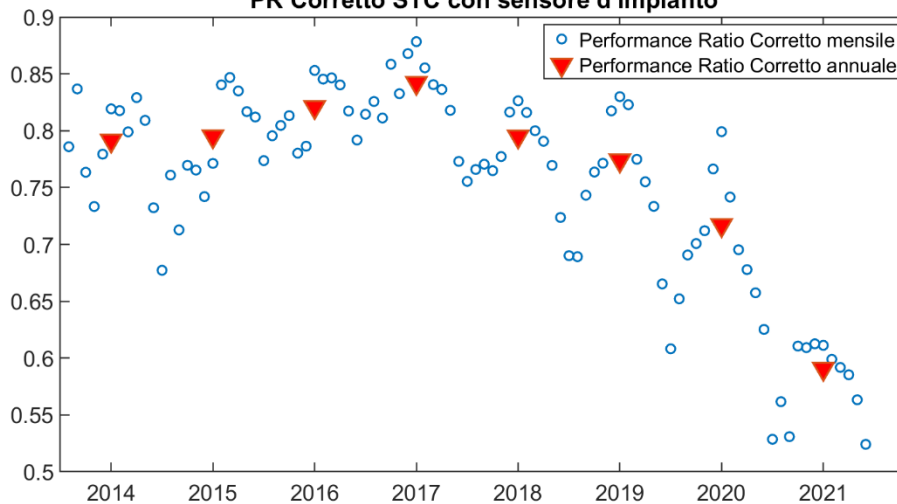


**Figura 4.3.26 Temperature medie e mediane dei pannelli, minime durante 2016, 2017**

Tale informazione va inclusa nel calcolo dei PR corretti, per poterne apprezzare l'eventuale rilevanza.

Occorre ricordare che il CPR calcolato rispetto alla temperatura media pesata tende ad addensare i dati di PR mensili, senza però variare quelli annuali, a differenza del metodo STC, che varia tutto.

**PR Corretto STC con sensore d'impianto**



**Figura 4.3.27 Performance ratio corretto con la temperatura STC di 25°C**

Correggendo in temperatura con il metodo STC, l'incremento relativo dei primi anni è lievemente diminuito, a prova del fatto che le temperature minori durante il 2016 ed il 2017 avessero un'influenza, ma tale influenza si mostra come minima. A livello mensile, per gli anni successivi che impiegano l'irradianza di satellite, essa è sovrastimata durante i mesi invernali a causa dell'ombreggiamento non considerato, e quindi il PR in tali mesi è sottostimato. La correzione in temperatura, che abbassa i valori di PR dei mesi freddi (che normalmente sono più alti) ed innalza quelli dei mesi caldi, in realtà spinge

ulteriormente verso il basso PR già contenuti in partenza, ecco il perché della particolare redistribuzione dei dati nel caso del CPR in STC.

In figura 4.3.28, è invece riportato il CPR calcolato rispetto alla temperatura media pesata annuale, con i valori mensili affiancati a quelli del PR normale. I valori annuali sono invece coincidenti.

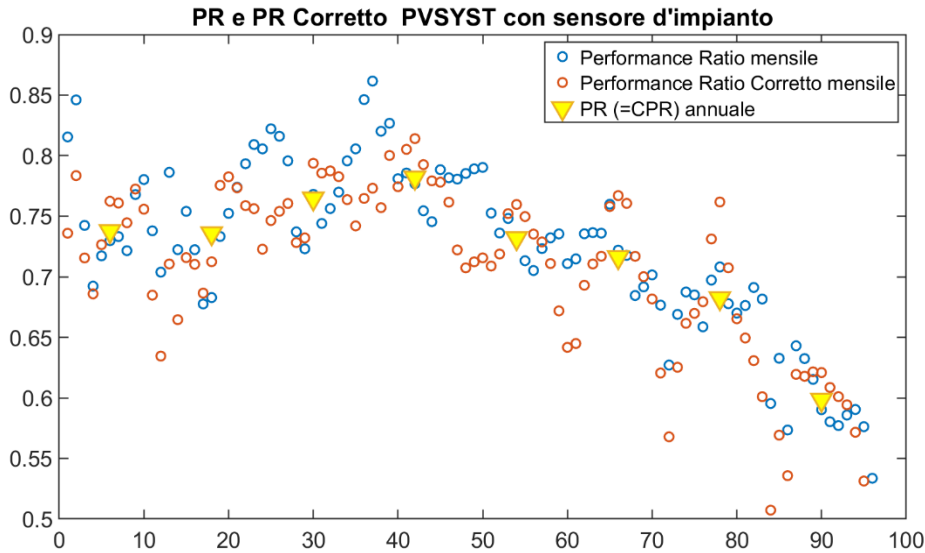


Figura 4.3.28 CPR corretto con la temperatura pesata annuale medi

Essendo scalato in relazione più precisa con i dati dell'impianto, il CPR calcolato con questa tecnica mostra maggiore compattezza mensile, ma i fenomeni spiegati nel caso precedente si ripetono senza particolari variazioni.

### 4.3.5 Confronto con il Performance Ratio dei modelli

L'ultimo elemento da valutare è il confronto tra KPI di prestazione basati sulle misure e KPI di prestazione stimati. Visto che in questo capitolo ci si è focalizzati sul PR nel valutare le prestazioni reali, si userà questo parametro anche per eseguire il paragone. La figura è la seguente:

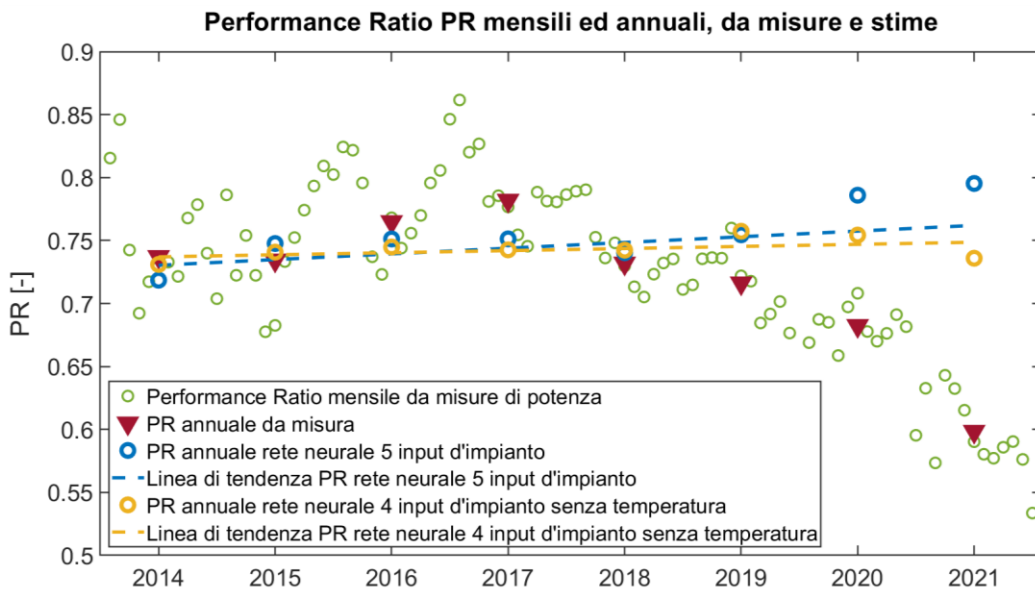


Figura 4.3.29 PR di misura confrontato al PR atteso da due reti, con linee di tendenza

In genere, si può tirare una retta di approssimazione dei PR annuali ed ottenere una linea di tendenza che descriva in modo lineare l'andamento medio delle prestazioni d'impianto PV. Richiamando tutti i ragionamenti svolti sull'irradianza ed il PR dell'impianto in analisi, risulta chiaro che cercare di approssimare il comportamento nel tempo con una semplice dipendenza lineare non abbia validità. In tal senso è impossibile identificare l'effetto dell'invecchiamento dei pannelli.

Lo stesso ragionamento non vale, però, per l'evoluzione del PR stimato con le potenze previste dalle reti neurali. Nella Figura 4.3.29, sono stati riportati i valori per due reti esemplari e gli annessi comportamenti peculiari.

Le reti selezionate sono state la più performante nella previsione della potenza, ovvero la 5 input con filtro, e la 4 input con filtro e senza la temperatura dei pannelli, al fine di mostrare l'effetto di errore nella temperatura misurata.

I PR delle reti sono abbastanza simili a quelli d'impianto nei primi 3 anni, che, essendo temporalmente i più prossimi all'addestramento, sono anche quelli ottimizzati meglio.

Si verifica la situazione anticipata con il calcolo degli MBE, ovvero che i PR stimati dalle reti tendono a mantenere una certa uniformità nel tempo, la stessa instauratasi in fase di training tra input e output.

Sarebbe quindi difficile che una ANN sia in grado di seguire l'andamento dei PR futuri e riconoscere, all'interno della linea di tendenza, un invecchiamento dei pannelli inferiore o uguale allo 0.7%, ma in questa precisa circostanza sarebbe ugualmente mascherato dall'effetto di crescita del PR presente nei primi 4 anni, e che provoca un lieve innalzamento dei PR stimati, invece di mantenerli pressoché costanti, com'è logico aspettarsi da modelli allenati su un anno solo.

Le linee di tendenza per le due reti sono ovviamente state estratte per gli anni fino al 2019, per escludere gli effetti della temperatura errata. Dopo il 2019, infatti, la rete a 5 input addestrata anche con la temperatura di pannello tende a salire innaturalmente, perché le vengono proposte temperature in rapida decrescita in quegli anni. Lo stesso, com'era corretto aspettarsi, non si verifica per la rete allenata senza la temperatura, i cui PR continuano all'incirca nella direzione orizzontale.

### **4.3.6 Impiego dati PVGIS nei KPI d'impianto**

Per fare ulteriori considerazioni sui KPI d'impianto (PR e PR corretti in particolare) occorre trovare fonti alternative di irradianza, che abbiano un comportamento omogeneo nel corso degli anni (a differenza del dataset ibrido messo a disposizione dai dati di impianto).

Sono stati inizialmente presi in esame i valori ARPAE, ma si sarebbero potuti valutare soltanto 5 anni di vita dell'impianto e sarebbe occorso praticare un'opera di suddivisione e trasposizione per tradurre l'irradianza globale orizzontale GHI in irradianza nel piano di array.

Si è optato infine per l'impiegare i valori di irradianza offerti dal programma di calcolo della radiazione solare PVGIS (Photovoltaic Geographical Information System), che impiega immagini satellitari ad alta risoluzione ed integra informazioni su velocità del vento e temperatura per produrre le stime di irradianza. Il software permette di estrarre dati con diversa precisione temporale ed adattandoli ad inclinazione ed orientamento della superficie di interesse (nel caso di impianti PV si tratta del POA, che sono in realtà uno per falda nel caso in analisi). Come spiegato in letteratura, i valori satellitari non sono caratterizzati da alta precisione locale (sia dal punto di vista temporale che spaziale), a causa dei processi di elaborazione dei dati e delle informazioni a terra non accessibili (come, per l'appunto, ombreggiamenti di vario genere), ma danno ottime rappresentazioni dei valori mensili ed annuali di irraggiamento che interessano una determinata zona. In virtù di questo, sembra l'opzione migliore per un calcolo di KPI più rappresentativo del funzionamento d'impianto.

I dataset prelevati sono stati scelti come medie orarie dal 2014 al 2020 (ultimo anno disponibile), in formato SARAH-2 (il più aggiornato e preciso per l'ambito europeo), con POA diversi per le due falde dell'impianto, per poi operarne la media oraria e confrontarla con la totalità della potenza AC.

I risultati vengono proposti nella pagina seguente.

Innanzitutto, viene ricalcolato il Performance Ratio dell'impianto PV, sostituendo agli irraggiamenti problematici di misura quelli satellitari PVGIS, dopodiché si applica la correzione con la temperatura media pesata annuale.

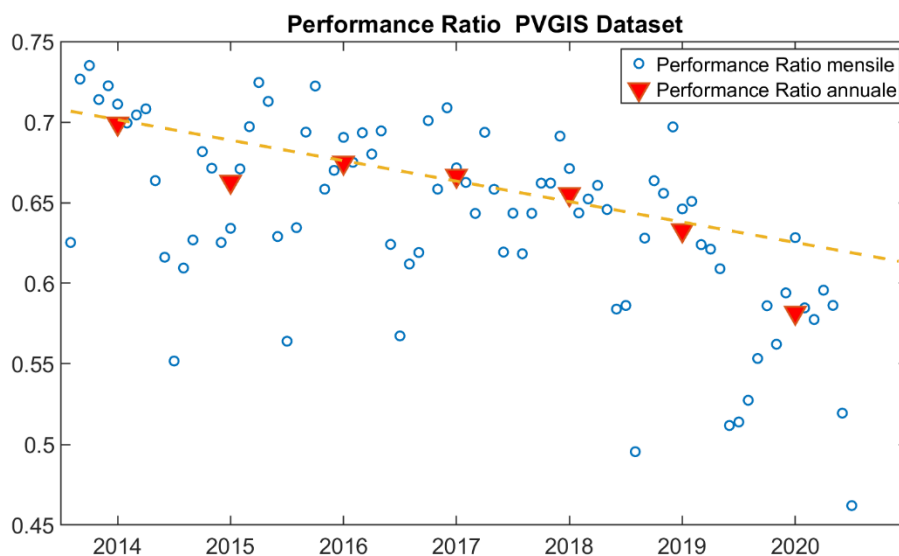


Figura 4.3.30 Performance ratio calcolato rispetto agli irraggiamenti PVGIS

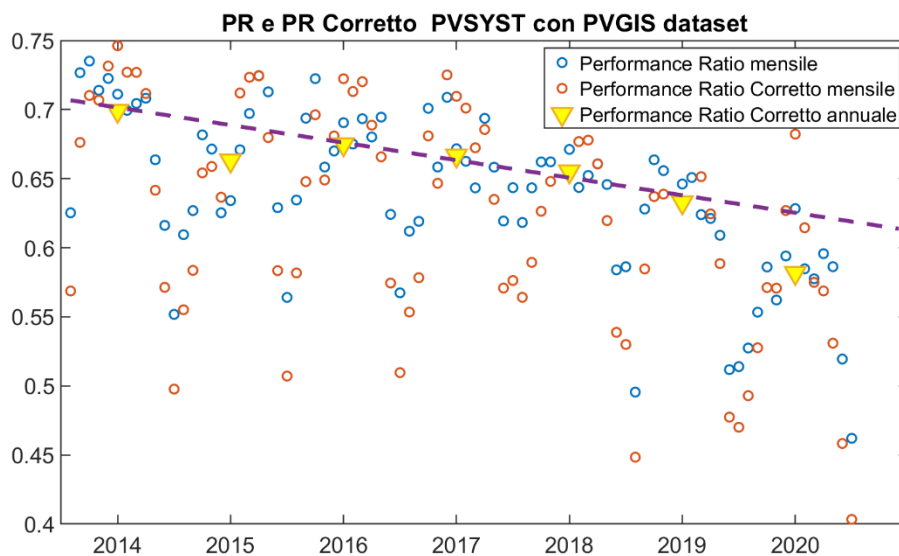


Figura 4.3.31 PR e CPR corretto con la temperatura media, dati PVGIS

La tendenza del PR è ora ben più chiara a livello annuale, anche se chiaramente la distribuzione mensile soffre della correlazione più limitata della potenza con i dati satellitari.

L'invecchiamento dei pannelli è più facilmente interpretabile adesso, anche se la pendenza della linea di tendenza (calcolata con i valori precedenti ai guasti di stringa) rappresenta una diminuzione annua del 1.1%, ben diverso dallo 0.7% garantito per i pannelli. Va anche conteggiato l'effetto di ombreggiamento crescente, che potrebbe giustificare in parte il termine aggiuntivo, Di certo le basse prestazioni invernali sono evidenti dai valori mensili di PR e CPR (che di nuovo riduce ulteriormente valori già limitati di PR invernali).

L'andamento ben più chiaro appena trovato con l'ausilio del nuovo set di dati di radiazione solare potrebbe invitare a preparare modelli basandosi sullo stesso, ma molteplici tentativi hanno dimostrato che la precisione oraria non è sufficiente ad allenare correttamente una ANN in questo caso.

#### **4.3.7 Altre opzioni di sviluppo delle reti prese in considerazione**

In più punti, e valutando differenti KPI di previsione, si è dimostrata la generale accuratezza delle reti neurali sviluppate nel corso di questo capitolo 4 sia nel prevedere i valori di potenza (e KPI associati) che di energia mensile ed annuale, al netto delle problematiche riscontrate direttamente nei dataset (ed in particolare nelle misure d'impianto).

Il fatto che le reti modellizzino al meglio i dati sani di training (corrispondenti al primo anno) è sia un punto di forza che di debolezza, in quanto permette di sfruttarle per contrasto con i valori osservati ed individuare anomalie e indici di allarme (come svolto ripetutamente nel corso del capitolo 4), ma questa "rigidità" implica anche l'impossibilità di integrare nel modello effetti a lungo termine, che la rete è troppo statica per poter apprezzare. Nello specifico, includere l'invecchiamento naturale dei pannelli non è considerabile negli input in quanto è un valore costante (o quantomeno assunto tale), e non può essere collegato all'ottimizzazione degli iperparametri della rete.

Si è deciso quindi di creare un nuovo gruppo di reti, cambiando approccio ed inserendo tale fattore in seconda fase di training, ovvero dopo l'ottimizzazione degli iperparametri ma prima di scegliere bias e pesi ottimali.

Per compiere ciò, gli output di potenza stimati ad ogni iterazione vengono scalati al ribasso di un termine dello 0.7% annuale diviso per le ore di training ed applicato ad ogni valore di potenza. Nel cercare la rete migliore, quindi, lo script non valuta più la rete originariamente prodotta, ma la sua versione contenente il fattore di invecchiamento. Per ovvie ragioni, una volta selezionata la rete considerata migliore, i suoi valori dovranno subire l'applicazione a posteriori dell'invecchiamento, prima che sia impiegata come modello.

Un'altra variazione introdotta in questo secondo metodo di sviluppo delle ANN è l'aver sostituito la tradizionale minimizzazione dell'errore di convalida con la minimizzazione degli MBE mensili (sempre considerati entro il dataset di training, le reti non possono essere messe a conoscenza di futuri valori da predire quando sono ancora nel processo di addestramento), nell'ottica di comparare poi i KPI (mensili ed annuali) di prestazione, invece di focalizzarsi sulla capacità del modello di seguire le osservazioni ora per ora (collegata ai valori di NMSE ed  $R^2$  precedentemente calcolati).

Pur essendo il concetto interessante in principio, l'attuazione si è rivelata inconcludente, per cui si è deciso di non riportare i risultati numerici, dal significato ridotto, e sono state invece esposte le criticità.

Gli MBE mensili erano più alti di quelli delle reti sviluppate con la tecnica di modellizzazione principale e, sebbene gli expected yield fossero ragionevoli, PR ed EPI non mostravano valori promettenti, con EPI tendenzialmente maggiori di 1, quando l'opposto dovrebbe essere la norma per una corretta modellizzazione.

Le possibili ragioni dietro questo tipo di risultato sono molteplici, ma fondamentalmente apportare modifiche considerevoli al processo di ottimizzazione delle reti ne vanifica le performance.

Per ottenere risultati migliori, che includano anche un'incorporazione dell'invecchiamento dei pannelli, si fa una proposta nell'ultima sezione.

#### **4.3.8 Modifiche implementabili per potenziare l'efficacia ANN-KPI**

Lo scopo del lavoro di tesi era produrre un modello affidabile per la previsione della potenza, stimarla in una determinata finestra temporale, calcolare i principali KPI prodotti dalla stima e confrontarli con quelli risultanti da valori osservati, così da sfruttare eventuali discrepanze come indici di funzionamento subottimale dell'impianto PV, e successivamente applicare la routine desiderata di monitoraggio.

La serie di script per la creazione, ottimizzazione ed applicazione delle ANN al problema, si sono dimostrati in grado di generare modelli flessibili, ma affidabili, anche in contesti di diffusa incertezza sulla qualità dei parametri di input. Le reti sono state addestrate con il primo anno disponibile e poi applicate ai successivi, eccellendo ovviamente nella previsione dell'anno immediatamente successivo, a livello annuale, mensile, orario. Per questo motivo applicazioni di questa metodologia di modellizzazione possono riproporre l'addestramento annualmente o mensilmente con un'ottima garanzia della precisione di previsione, ed aggirare così l'insufficiente sensibilità della rete a termini di perdita graduale e contenuta: l'informazione sul deterioramento sarebbe inclusa di volta in volta nel dataset in aggiornamento, producendo una nuova rete, maggiormente a conoscenza dello stato attuale dell'impianto.

Ad ogni nuovo addestramento (di per sé non superiore al quarto d'ora di durata grazie alla semplicità strutturale del FFNN) si potrebbe mantenere la prima rete, addestrata sulla condizione ideale d'impianto e automatizzare la produzione ed il confronto di triplette dei KPI ritenuti più rilevanti, con un valore da osservazione, uno dalla rete originale ed uno dalla rete più giovane. A seconda degli scarti e dei KPI nei quali si verificassero, lanciare semplici script secondari di controllo delle grandezze ambientali ed elettriche.

Questa dovrebbe rivelarsi una strategia dall'impegno computazionale richiesto contenuto, con elevato potenziale nel campo del monitoraggio e analisi delle prestazioni.

Versioni più complesse potrebbero affiancare al DFNN una rete neurale di classificazione, che scaldi gli script secondari e si attivi per identificare guasti ed anomalie sulla base degli scarti dei KPI prodotti dalle due reti principali. D'altronde, vari esempi in letteratura (nominati a più riprese nel capitolo 1) di impiego di reti neurali di classificazione in questo ambito dimostrano la possibilità di integrazione in un approccio misto del tipo proposto.

## CONCLUSIONI

Nel panorama odierno di diffusione delle tecnologie che impiegano risorse rinnovabili, il fotovoltaico è divenuto uno standard diffuso e ricercato, grazie alla sua modularità, applicabilità anche in piccole zone e sostanziale semplicità della tecnologia alla base. Il fotovoltaico è un ambito sdoganato nel suo senso più ampio: esistono pannelli più performanti e tecnologie differenti, ma ciò che fa davvero la differenza è come un impianto viene gestito, quanto rapidamente e con che precisione si identificano cause di peggioramento delle prestazioni e con quale livello di tempestività si riesce ad intervenire.

Che si tratti di monitoraggio o valutazione delle performance, i KPI svolgono, come il nome suggerisce, un ruolo chiave. Ce ne sono di più peculiari, soprattutto per tecnologie più inusuali, ma in genere il loro ruolo è esibire la correlazione tra grandezze fondamentali dell'impianto, con formulazioni standardizzate che permettano di comparare tali correlazioni rispetto ad altri impianti simili (e non), o rispetto a valori di riferimento. I Key Performance Indicators di prestazioni riconosciuti ed applicati sono ben noti e sono stati trattati nel dettaglio, ma la vera opportunità di sviluppo di KPI nasce con la modellizzazione degli impianti.

Produrre un modello il più preciso possibile dell'impianto da gestire introduce automaticamente due nuovi set di KPI, il primo uguale nella forma (si ripetono le formulazioni) ma differente nell'origine (stime del modello invece dei valori osservati), ed il secondo di natura totalmente diversa, ovvero KPI della previsione/modellizzazione espressi come scarti di vario tipo.

In questo modo si possono paragonare le prestazioni dell'impianto non soltanto con i valori esterni provenienti da altre strutture, ma con valori strettamente legati alle specifiche dell'impianto, offrendo una panoramica ben più personalizzata per sostenere una gestione ottimizzata e contemporaneamente monitorare rispetto ai valori previsti. L'esistenza stessa del modello apporta informazioni utili: come si è potuto apprezzare, valori atipici di KPI relativi al confronto tra misure e modello hanno permesso di individuare numerose anomalie, non facilmente rintracciabili con la semplice applicazione dei KPI standard alle misure.

L'approccio tramite Machine Learning, e Reti Neurali nello specifico, si è rivelato flessibile e potente, producendo numerosi modelli diversi ad alta accuratezza e permettendo di individuare moltissime informazioni celate e funzionamenti subottimali, riuscendo così a lavorare con dataset in condizioni particolarmente ostiche. Il tutto è risultato alla fine in un'analisi approfondita dell'impianto trattato, offrendo spunti interessanti in merito al riutilizzo ed ampliamento degli approcci e modelli sviluppati, con ampie potenzialità unite ad una generale semplicità di impiego degli strumenti realizzati per l'analisi.



## BIBLIOGRAFIA

- [1] Y. Chouay and M. Ouassaid, "An intelligent method for fault diagnosis in photovoltaic systems," 2017 International Conference on Electrical and Information Technologies, Singapore, Nov. 2017, pp. 1-5
- [2] W. Chine and A. Mellit, "ANN-based fault diagnosis technique for photovoltaic strings", 2017 5th International Conference on Electrical Engineering, Boumerdes, Oct. 2017, pp. 1-4
- [3] Y. Tian, C. Chen, K. Su, J. Yuan and J. Zhang, "Design of Photovoltaic Array Fault Online Evaluation System", 2020 5th International Conference on Computer and Communication Systems, Shanghai, May 2020, pp. 912-916
- [4] M. Hojabri, S. Kellerhals, G. Upadhyay and B. Bowler, "IoT-Based PV Array Fault Detection and Classification Using Embedded Supervised Learning Methods", *Energies*, Vol. 15 (6), pp. 1-18, Mar. 2022
- [5] T. Berghout, M. Benbouzid, X. Ma, S. Djurović and L. H. Mouss, "Machine Learning for Photovoltaic Systems Condition Monitoring: A Review", IECON 2021 - 47th Annual Conference of the IEEE Industrial Electronics Society, Toronto, Oct. 2021, pp. 1-5
- [6] F. Harrou, Y. Sun and A. Saidi, "Online model-based fault detection for grid connected PV systems monitoring", 2017 5th International Conference on Electrical Engineering, Boumerdes, Oct. 2017, pp. 1-5
- [7] S. Samara and E. Natsheh, "Intelligent real-time photovoltaic panel monitoring system using artificial neural networks", *IEEE Access*, Vol. 7 (1), pp. 50287-50299, Apr. 2019
- [8] U. K. Das, K. S. Tey, M. Seyedmahmoudian, S. Mekhilef, M. Y. I. B. Idris, W. Van Deventer, B. Horan and A. Stojcevski, "Forecasting of photovoltaic power generation and model optimization: A review", *Elsevier*, vol. 81, pp. 912-928, Jan. 2018
- [9] F. Touati, A. Khandakar, M. E. Chowdhury, A. J. S. Gonzales, C. K. Sorino, and K. Benhmed, *Renewable Energy - Technologies and Applications*, London, IntechOpen, 2020, Ch. Photo-Voltaic (PV) Monitoring System, Performance Analysis and Power Prediction Models in Doha, Qatar
- [10] M. Omar, A. Dolara, G. Magistrati, M. Mussetta, E. Ogliari and F. Viola, "Day-ahead forecasting for photovoltaic power using artificial neural networks ensembles", 2016 IEEE International Conference on Renewable Energy Research and Applications, Birmingham, Nov. 2016, pp. 1152-1157
- [11] L. Liu, D. Liu, Q. Sun, H. Li and R. Wennersten, "Forecasting Power Output of Photovoltaic System Using A BP Network Method", *Energy Procedia*, Vol. 142 (12), pp. 780-786, Dic. 2017
- [12] G. Almonacid-Olleros, G. Almonacid, J. I. Fernandez-Carrasco and J. M. Quero, "Deep Learning Modelling for Photovoltaic System Monitoring" *Multidisciplinary Digital Publishing Institute Proceedings*, Vol. 31 (1), pp. 1-10, Nov. 2019
- [13] E. Natsheh and S. Samara, "Toward Better PV Panel's Output Power Prediction; a Module Based on Nonlinear Autoregressive Neural Network with Exogenous Inputs", *Applied Sciences*, Vol. 9 (18), pp. 1-20, Sep. 2019
- [14] A. Raghoebarsing and A. Kalpoe, "Performance and economic analysis of a 27KW grid-connected photovoltaic system in Suriname", *IET Renewable Power Generation*, Vol. 11 (12), pp. 1545-1554, Oct. 2017
- [15] A. Boretti and S. Castelletto, "Trends in performance factors of large photovoltaic solar plants", *Journal of Energy Storage*, Vol. 30 (1), pp. 1-14, Aug. 2020
- [16] S. S. Inamdar, S. K. Singh, A. Sharma and S. Chakrabarti, "Performance Evaluation of 5kWp grid connected polycrystalline Si-based Photovoltaic Plant", 2018 20th National Power Systems Conference, Tiruchirappalli, Dec. 2018, pp. 1-6

- [17] R. Dahlioui, B. Laarabi, S. M. Alaoui, G. Dambrine, E. Menard, J. Boardman and A. Barhdadi, "Evaluation of Losses in PV Performances Due to Soiling Effect in Rabat" 2018 6th International Renewable and Sustainable Energy Conference, Rabat, Dec. 2018, pp. 1-3
- [18] S. A. M. Said, G. Hassan, H. M. Walwil and N. Al-Aqeeli, "The effect of environmental factors and dust accumulation on photovoltaic modules and dust-accumulation mitigation strategies" *Renewable and Sustainable Energy Reviews*, Vol.82, (1), pp. 743-760, Feb 2018
- [19] M. Gostein, J. R. Caron and B. Littmann, "Measuring soiling losses at utility-scale PV power plants", 2014 IEEE 40th Photovoltaic Specialist Conference, Denver, Jun. 2014, pp. 885-890
- [20] E. Urrejola, J. Antonanzas, P. Ayala, M. Salgado, G. Ramírez-Sagner, C. Cortés, A. Pino and R. Escobar, "Effect of soiling and sunlight exposure on the performance ratio of photovoltaic technologies in Santiago, Chile", *Energy Conversion and Management*, Vol. 114 (1), pp. 338-347, Apr. 2016
- [21] M. Ali, M. H. Iqbal, N. A. Sheikh, H. M. Ali, M. S. Manzoor, M. M. Khan and K. F. Tamrin, "Performance Investigation of Air Velocity Effects on PV Modules under Controlled Conditions", *International Journal of Photoenergy*, Vol. 2017 (1), pp. 1-10, Dec. 2017
- [22] J. K. Kaldellis, M. Kapsali and K.A. Kavadias, "Temperature and wind speed impact on the efficiency of PV installations. Experience obtained from outdoor measurements in Greece", *Renewable Energy*, Vol. 66 (1), pp. 612-624, Jun 2014
- [23] M. Dhimish, "Thermal impact on the performance ratio of photovoltaic systems: A case study of 8000 photovoltaic installations", *Case Studies in Thermal Engineering*, Vol. 21 (1), pp. 1-5, October 2020
- [24] K. Hasan, S. B. Yousuf, M. S. H. K. Tushar, B. K. Das, P. Das and M. S. Islam, "Effects of different environmental and operational factors on the PV performance: A comprehensive review", *Energy Science & Engineering*, Vol. 10 (2), pp. 656-675, Dec. 2021
- [25] M. Malvoni, M. G. De Giorgi and P. M. Congedo, "Study of degradation of a grid connected photovoltaic system", 72nd Conference of the Italian Thermal Machines Engineering Association, Lecce, September 2017, pp. 644-650
- [26] Y. Lyu, A. Fairbrother, M. Gong, J. H. Kim, X. Gu, M. Kempe, S. Julien, K.T. Wan, S. Napoli, A. Hauser, G. O'Brien, Y. Wang, R. French, L. Bruckman, L. Ji and K. Boyce, "Impact of environmental variables on the degradation of photovoltaic components and perspectives for the reliability assessment methodology", *Solar Energy*, Vol. 199 (1), pp. 425-436, Feb. 2020
- [27] A. Mohd Khalid, I. Mitra, W. Warmuth, V. Schacht, "Performance ratio – Crucial parameter for grid connected PV plants", *Renewable and Sustainable Energy Reviews*. 65, pp. 1139-1158, 2016
- [28] J. Leloux, L. Narvarte, M. Collares-Perreira, A. Desportes, *Analysis of the state of the art of PV systems in Europe*, 2015
- [29] R.K. Prakhya, S.R. Kotha, "Temperature effect on the performance metrics of gridtied SPV plant", *International Journal of Engineering Research and Technology*. 11. 909-923, (2018)
- [30] A. Gopi, K. Sudhakar, N. W. Keng, A. R. Krishnan, "Comparison of normal and weather corrected performance ratio of photovoltaic solar plants in hot and cold climates", *Energy for Sustainable Development*, Volume 65, pp. 53-62, 2021
- [31] R. A. Fajarin and E. A. Setiawan "Analysis Corrected Performance Ratio on Photovoltaic Through Four Temperature Cell Model", The 3<sup>rd</sup> International Renewable Energy Conference "Sustainable Development of Tropical Renewable Energy", Bali , Sep. 2018, pp.1-4
- [32] C. Coskun, U. Toygar, O. Sarpdag and Z. Oktay, "Sensitivity analysis of implicit correlations for photovoltaic module temperature: A review", *Journal of Cleaner Production*, Volume 164 (1), pp. 1474-1485, Ott. 2017
- [33] J. S. Griffith, M. S. Rathod and J. Paslaski, "Some tests of flat plate photovoltaic module cell temperatures in simulated field conditions", 15<sup>th</sup> Photovoltaic Specialists Conference, Kissimmee, May 1981, pp. 822-830

- [34] E. Kaplani and S. Kaplanis, "Thermal modelling and experimental assessment of the dependence of PV module temperature on wind velocity and direction, module orientation and inclination", *Solar Energy*, Vol. 107 (1), pp. 443-460, Sep. 2014
- [35] T. A. Olukan, M. Emziane, "A Comparative Analysis of PV Module Temperature Models", *Energy Procedia*, vol. 62, pp. 694-703, Dic. 2014
- [36] K. Y. Lai and B. H. Lim, "Comparative Study for Time-specific Ross Coefficient and Overall Ross Coefficient for Estimation of Photovoltaic Module Temperature", 2019 IEEE Conference on Sustainable Utilization and Development in Engineering and Technologies, Penang, Nov. 2019, pp. 251-256
- [37] A. Obiwulu, M. Chendo, N. Erusiafe, and S. Nwokolo, "Implicit meteorological parameter-based empirical models for estimating back temperature solar modules under varying tilt-angles in Lagos, Nigeria", *Renewable Energy*, Vol. 145 (1), pp. 1-38, Jun. 2019
- [38] P. Mora Segado, J. Carretero and M. Sidrach de Cardona, "Models to predict the operating temperature of different photovoltaic modules in outdoor conditions", *Progress in Photovoltaics: Research and Applications*, Vol. 17 (6), pp. 372-393, Oct. 2014
- [39] N. Aoun, "Methodology for predicting the PV module temperature based on actual and estimated weather data", *Elsevier Energy Conversion and Management*, Vol. 14 (1), pp.1-14, May 2022
- [40] C. Li, S. V. Spataru, K. Zhang, Y. Yang and H. Wei, "A Multi-State Dynamic Thermal Model for Accurate Photovoltaic Cell Temperature Estimation," in *IEEE Journal of Photovoltaics*, vol. 10, no. 5, pp. 1465-1473, Sep. 2020
- [41] O. May Tzuc, A. Bassam, P. E. Mendez-Monroy and I. S. Dominguez, "Estimation of the operating temperature of photovoltaic modules using artificial intelligence techniques and global sensitivity analysis: A comparative approach", *Journal of Renewable and Sustainable Energy*, Vol. 10 (3), pp. 1-19, May 2018
- [42] G. Ciulla, V. Lo Brano and E. Moreci, "Forecasting the Cell Temperature of PV Modules with an Adaptive System", *Hindawi Publishing Corporation International Journal of Photoenergy*, Vol. 2013 (1), pp. 1-10, Sep. 2013
- [43] A. Bassam, O. May Tzuc, M. Escalante Soberanis, L. J. Ricalde and B. Cruz, "Temperature estimation for photovoltaic array using an adaptive neuro fuzzy inference system", *Sustainability*, Vol. 9 (8), pp. 1-16, Aug. 2017
- [44] İ. Ceylan, O. ErKaymaz, E. Gedik and A. E. Gürel, "The prediction of photovoltaic module temperature with artificial neural networks", *Case Studies in Thermal Engineering*, Vol. 2014 (3), pp. 11-20, Mar. 2014
- [45] G. I. Diaz, A. Fokoue-Nkoutche, G. Nannicini and H. Samulowitz, "An effective algorithm for hyperparameter optimization of neural networks," in *IBM Journal of Research and Development*, vol. 61, no. 4/5, pp. 9:1-9:11, Jul.-Sep. 2017
- [46] E. Arena, A. Corsini, R. Ferulano, D. A. Iuvara, E. S. Miele, L. Ricciardi Celsi and M. Villari, "Anomaly detection in photovoltaic production factories via Monte Carlo pre-processed principal component analysis" *Energies*, Vol. 14 (15), p. 3951, Jul. 2021
- [47] S. Roy, J. Byrne and C. Pickering, "A systematic quantitative review of urban tree benefits, costs, and assessment methods across cities in different climatic zones", 37<sup>th</sup> European Photovoltaic Solar Energy Conference and Exhibition, Lisbona, Sep. 2020, pp. 351-363,
- [48] Y. Avenas, L. Dupont, N. Baker, H. Zara and F. Barruel, "Condition Monitoring: A Decade of Proposed Techniques" *Industrial Electronics Magazine*, Vol. 9 (4), pp. 22-36, Dec. 2015
- [49] C. H. Huang, Y. H. Lian and H. L. Tsai, "Design for microcontroller-based photovoltaic monitoring system using MATLAB/Simulink", 2016 International Automatic Control Conference, Taichung, Nov. 2016, pp. 1-6
- [50] M. Campanelli and K. Emery, "Device-dependent light-level correction errors in photovoltaic I-V performance measurements", 2013 IEEE 39th Photovoltaic Specialists Conference, Tampa, Jun. 2013, pp. 67-72

- [51] A. Betti, M. Tucci, E. Crisostomi, A. Piazzzi, S. Barmada and D. Thomopoulos, "Fault prediction and early-detection in large pv power plants based on self-organizing maps", *Sensors*, Vol. 21 (5), pp. 1-20, Mar. 2021
- [52] S. Sarikh, M. Raoufi, A. Bennouna, A. Benlarabi and B. Ikken, "Fault diagnosis in a photovoltaic system through IV characteristics analysis", 2018 9th international renewable energy congress, Hammamet, Mar. 2018, pp. 1-6
- [53] M. Aghaei, F. Grimaccia, C. A. Gonano and S. Leva, "Innovative automated control system for PV fields inspection and remote control", *IEEE Transactions on Industrial Electronics*, Vol. 62 (11), pp. 7287-7296, Nov. 2015
- [54] M. Ejgar, B. Momin and T. Ganu, "Intelligent monitoring and maintenance of solar plants using real-time data analysis", 2017 International Conference on Consumer Electronics-Asia, Bengaluru, Oct. 2017, pp. 133-138
- [55] M. Lave, W. Hayes, A. Pohl and C. W. Hansen, "Evaluation of Global Horizontal Irradiance to Plane-of-Array Irradiance Models at Locations Across the United States," *IEEE Journal of Photovoltaics*, Vol. 5 (2), pp. 597-606, Mar. 2015
- [56] R. Colmenares-Quintero, E. Rojas-Martinez, F. Macho-Hernantes, K. Stansfield and J. C. Colmenares, "Methodology for automatic fault detection in photovoltaic arrays from artificial neural networks", *Cogent Engineering*, Vol. 8 (1), pp. 1-22, Jan. 2021
- [57] H. Rezk and E. S. Hasaneen, "A new MATLAB/Simulink model of triple-junction solar cell and MPPT based on artificial neural networks for photovoltaic energy systems", *Ain Shams Engineering Journal*, Vol. 6 (3), pp. 873-881, Sep. 2015
- [58] K. Prerna and G. Shankar, "Modelling and simulation of photovoltaic system for isolated and grid connected mode", 2016 3rd International Conference on Recent Advances in Information Technology, Dhanbad, Mar. 2016, pp. 316-320
- [59] S. Lindig, A. Louwen, D. Moser and M. Topic, "Outdoor PV System Monitoring—Input Data Quality, Data Imputation and Filtering Approaches", *Energies*, Vol. 13 (5099), pp. 1-18, Set. 2020
- [60] F. Rodríguez, A. Fleetwood, A. Galarza and L. Fontán, "Predicting solar energy generation through artificial neural networks using weather forecasts for microgrid control", *Renewable energy*, Vol. 126 (1), pp. 855-864, Oct. 2018
- [61] P. Guerriero, F. Di Napoli and S. Daliento, "Real time monitoring of solar fields with cost/revenue analysis of fault fixing", 2016 IEEE 16th International Conference on Environment and Electrical Engineering, Firenze, Jun. 2016, pp. 1-6,
- [62] B. Andò, S. Baglio, A. Pistorio, G. M. Tina and C. Ventura, "Sentinella: Smart Monitoring of Photovoltaic Systems at Panel Level", *IEEE Transactions on Instrumentation and Measurement*, vol. 64 (8), pp. 2188-2199, Aug. 2015
- [63] S. Armstrong and W. G. Hurley, "A thermal model for photovoltaic panels under varying atmospheric conditions", *Applied Thermal Engineering*, Vol. 30 (11-12), pp. 1488-1495, Aug. 2010
- [64] T. M. Klucher, "Evaluation of models to predict insolation on tilted surfaces" *Solar Energy*, Vol. 23 (1), pp. 111-114, Dec. 1978
- [65] M. Gulin, M. Vasak and M. Baotic, "Estimation of the global solar irradiance on tilted surfaces", 17th International Conference Electric Drives Power Electron, Dubrovnik, Jan. 2013 , pp. 334-339
- [66] S. V. Oprea and A. Bâra, "Ultra-short-term forecasting for photovoltaic power plants and real-time key performance indicators analysis with big data solutions. Two case studies - PV Agigea and PV Giurgiu located in Romania", *Computers in Industry*, Vol. 120 (1), pp. 1-22, Settembre 2020
- [67] M. Mansouria, A. Al-khazrajib, M. Hajjic, M. F. Harkatd, H. Nounoua and M. Nounoud, "Wavelet optimized EWMA for fault detection and application to photovoltaic systems", *Solar Energy*, Vol. 167, pp. 125–136, Jun. 2018

[68] C. Schwingshackl, M. Petitta, J. E. Wagner, G. Belluardo, D. Moser, M. Castelli, M. Zebisch and A. Tetzlaff, "Wind effect on PV Module Temperature: Analysis of Different Techniques for an Accurate Estimation, Amsterdam" *Energy Procedia*, Vol. 40, pp. 77-86, Aug. 2013

[69] C. B. Jones, B. H. King, J. S. Stein, J. S. Fada, A. J. Curran, R. H. French, E. Schnabel, M. Koehl and O. Lavrova, "Quantify Photovoltaic Module Degradation using the Loss Factor Model Parameters", 44th Photovoltaic Specialist Conference, Washington, Jun. 2017, pp. 3488-3493

*I ringraziamenti più sentiti sono rivolti al Prof. Bignucolo, che non solo mi ha introdotto alla proposta di tesi supportandomi in questo arco conclusivo del mio percorso accademico, ma mi ha contagiato con l'amore per gli impianti di produzione dell'energia elettrica e le tecnologie associate, già dai tempi della partecipazione al Suo corso.*

*Un secondo grazie è sicuramente da rivolgersi all'Ing. Riccardo Clementi, pronto a sostenermi passo dopo passo con dritte nate dall'esperienza, esperienza di cui sono stato compartecipe all'interno dello stimolante ambito lavorativo di STE Energy, dove mi sono sentito accolto e spronato.*

*Partendo da casa, sono davvero grato per i numerosi momenti in cui sono stato appoggiato ed incoraggiato, anche quando CMP e Pitrazia non sapevano come aiutare, ma sono semplicemente stati presenti, ed è più di quanto si possa desiderare alle volte.*

*Mille volte grazie Botti, per avermi sopportato fino alla fine con i miei dubbi strani e la ricerca continua di consigli, anche se tra elettrici e chimici ci si capisce poco.*

*Emma, hai continuato ad essere quella calda voce della ragione che permette di mettere le cose in prospettiva, e farne a meno avrebbe cambiato tutto.*

*Stefano, nell'augurarti di raggiungermi a breve in questo traguardo (e so che ce la farai), ti ringrazio per le chiacchierate sciocche dei nostri "podcast", perché le piccolezze di tutti i giorni aiutano a tenere assieme i pezzi della vita di una persona.*

*Pietro, le conversazioni dell'ultimo periodo di laurea per entrambi si sono rivelate esilaranti, anche a distanza eravamo sulla stessa barca, anche se hai attraccato prima, sto arrivando haha.*

*Pippo ed Oscar, so che ancora mi aspettate per i nostri "Friday" assieme, grazie per non aver mai mollato ed esserci stati anche da distanza, vi voglio bene.*

*Braggio, mi fa sempre piacere parlare assieme ed aggiornarci, ma le chiamate notturne ci hanno privato di ore di sonno già al minimo (sì lo rifarei).*

*Pocci, Leo, non ci siamo sentiti così spesso, ma avete sempre tenuto caldo un posto accogliente per me nel party in quelle lunghe serate faticose, e lo apprezzo.*

*Pier, Gianpi, grazie per i pensieri semplici, quelle chiamate ogni tanto per sentire come andava, che mi hanno tenuto compagnia.*

*Infinite volte grazie a chi ho già ringraziato e a chi ho sbadatamente dimenticato, grazie per essermi stati vicini a modo vostro in questo periodo importante della mia vita, nella speranza che rimarrete a lungo in futuro.*