

Prognostics of industrial plant components in the absence of fault data: comparison between multivariate control charts and one-class SVM.

Navicelli A.*, De Carlo F.*, Tucci M.*

* Dipartimento di Ingegneria Industriale, University of Firenze, Via di S. Marta, 3 50139 – Firenze – Italy
(andrea.navicelli@unifi.it, filippo.decarlo@unifi.it)

Abstract: Predictive maintenance for critical components’ monitoring in industrial plants has aroused the interest of many researchers in recent years. The typical phenomenology of industrial plants’ failures shows a degradation of performance before the occurrence of the failure event; therefore, predictive maintenance is the most suitable technique to intercept them. To implement prognostics is necessary to have a lot of data on system behaviour in both nominal and degraded conditions up to the failure event. With this information, it is possible to build a suitable prognostic model using a mathematical-statistical or machine learning technique. The advent of the fourth industrial revolution favoured the collection of real-time assets’ data. The low failure rate that characterizes most critical assets of industrial plants, result in a lot of nominal conditions’ and an absence of degraded conditions’ data, hampering the implementation of prognostic. In this article, have been developed, validated and compared on a case study two prognostic techniques using only nominal condition data. The first one is based on the multivariate control charts (Hotelling); the second one uses the one-class SVM model. Both techniques, combined with an ARIMA time series analysis model, allow the real-time prediction of anomalous operating condition of the monitored asset. Since we don't have any fault data acquired on field, both the prognostic models developed can predict significant deviations from nominal operating conditions due to the degradation phenomena, but they can't characterize the failure mode that will arise until the failure occurs for the first time. The two models were applied to a case study to verify their robustness in predicting deviations from the nominal operating conditions of a multistage compressor caused by surge phenomenon.

Keywords: Predictive maintenance, prognostic, multivariate control charts, Support Vector Machine

1.Introduzione

La manutenzione predittiva è una filosofia manutentiva che fa parte delle tecniche di manutenzione preventiva insieme a quella periodica e su condizione. (Ahmad and Kamaruddin, 2012) Si basa sul monitoraggio in real-time delle condizioni operative dei componenti di impianti industriali e dei sistemi di processo, al fine di stimare la Remaining Useful Life (RUL). Sulla base di questo parametro l'ingegneria di manutenzione può facilmente ottimizzare e schedare gli interventi manutentivi necessari al ripristino delle prestazioni del treno macchine, con una conseguente riduzione degli interventi correttivi, responsabili della maggior parte dei costi totali di manutenzione. (Löfsten, 1999) I costi di manutenzione sono complessivamente definiti dal costo per gli interventi ordinari, previsti ed organizzati con anticipo e gli interventi straordinari, dovuti a guasti imprevisi che hanno un maggiore impatto economico per l'azienda, dovuto generalmente ad un maggior periodo di fuori servizio del componente di impianto. (Komonen, 2002) Un altro evidente beneficio diretto di un sistema di manutenzione predittivo è dovuto all'incremento della disponibilità dei macchinari. (Si et al., 2011) L'implementazione di un sistema manutentivo di tipo predittivo, risulta inoltre essere

il più appropriato per l'89% dei modi di guasto, in confronto alla manutenzione periodica, sufficiente nel restante 11%. (Hashemian, 2010) Per questi motivi l'interesse dei ricercatori verso la manutenzione predittiva è incrementato esponenzialmente da quando è stato pubblicato il primo articolo nel 1952 come mostrato in Figure 1.

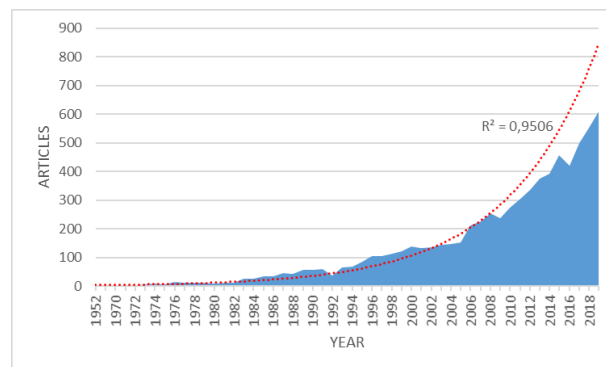


Figure 1: Numero di articoli pubblicati per anno (Keywords: Predictive Maintenance, Limits: Articles, Source: Scopus.com). In rosso la funzione esponenziale di miglior approssimazione con il relativo parametro R^2 .

Lo sviluppo di un sistema di manutenzione predittiva necessita di dati relativi al comportamento dell'asset analizzato durante la sua evoluzione verso il fenomeno di guasto. L'elevata affidabilità della maggior parte dei componenti critici di impianto, combinata ad una cultura di storicizzazione dei dati relativi ai guasti molto giovane, rende apparentemente impossibile la stima della RUL dei componenti di interesse se non in pochissimi casi. La letteratura propone in tal caso, tecniche diagnostiche basate sullo Statistical Process Control, che si limita a segnalare comportamenti anomali dell'asset (Alwan and Roberts, 1988; Bersimis et al., 2007; MacGregor and Kourti, 1995; Mason and Young, 2002; Oakland, 2007). In questo contesto, lo scopo dello studio è quello di costruire una metodologia prognostica utilizzabile in assenza di dati di guasto del componente di interesse. L'idea è quella di combinare un modello diagnostico ad una analisi delle serie temporali così da rendere possibile la stima della RUL. A tal scopo sono stati costruiti due differenti metodi diagnostici tipici dell'SPC: il primo basato sulle carte di controllo multivariate o carte di Hotelling ed il secondo che sfrutta il modello Machine Learning One-class Support Vector Machine (SVM). Entrambe le tecniche combinano le variabili di controllo selezionate come spia del modo di guasto da prevedere e ne distillano un parametro unico proxy delle performance del componente. Applicando all'output del modello diagnostico la tecnica di analisi delle serie temporali ARIMA (AutoRegressive Integrated Moving Average) è possibile prevedere un eccessivo abbassamento delle prestazioni del componente assimilabile ad un evento di guasto. I due modelli prognostici costruiti sono stati quindi confrontati applicandoli ad un caso studio.

Nel successivo capitolo, verrà presentata la metodologia sviluppata e le tecniche matematiche principali utili al suo sviluppo. Nel capitolo “Caso studio” si presenta la sua applicazione al modo di guasto di pompaggio del compressore centrifugo multistadio di un impianto geotermico di produzione di energia, ed infine sarà discussa la metodologia e gli sviluppi futuri nell'ultimo capitolo.

2 Metodologia

Come descritto nel capitolo precedente è stato pensato lo sviluppo di un modello prognostico a partire dai soli dati di buon funzionamento del componente sotto analisi. L'applicazione ad un caso studio dei due modelli sviluppati consente un confronto delle loro prestazioni nella previsione della RUL. Lo schema seguito per lo sviluppo e confronto dei modelli prognostici è rappresentato schematicamente in Figure 2.

La logica sviluppata per stimare la vita utile residua di componenti di impianti industriali prevede l'utilizzo di due diverse classi di modelli matematico statistici:

- Modello di controllo statistico di processo (modelli 1 e 2 in Figure 2);
- Modello di analisi serie temporali (modello 3 in Figure 2).

Il primo è un modello deputato all'individuazione di funzionamenti anomali del componente sotto controllo

mentre il secondo analizza l'evoluzione nel tempo delle sue performance e ne prevede l'andamento futuro. Impostando una soglia minima di accettabilità delle performance del componente, il modello è in grado di fornire una stima della RUL del componente in analisi.

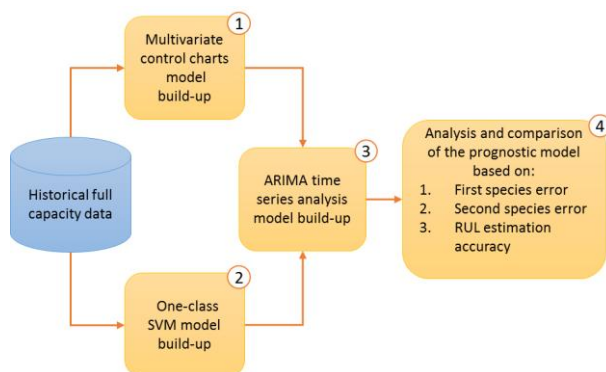


Figure 2: Schema di sviluppo e confronto dei modelli prognostici

Le performance dei macchinari di impianti industriali possono essere ricavate da un set di variabili di processo che ne descrivono il funzionamento. Per poter costruire un modello prognostico è necessario acquisire in real-time il valore di tali variabili ed elaborarle attraverso il modello di controllo statistico di processo per ottenere una unica variabile di controllo correlata a sua volta alle performance del componente analizzato. Le variabili di controllo acquisibili in real-time per la stima delle performance di componenti di impianti industriali sono tipicamente: vibrazioni, temperature, pressioni e portate. La combinazione di una o più di queste variabili di controllo permette di ottenere una variabile proxy delle performance dell'asset. A tale scopo sono stati utilizzati, applicati ad un singolo caso studio e confrontati, due diversi modelli di controllo statistico di processo multivariati:

- Carte di controllo multivariate o carte di Hotelling;
- One-class SVM;

Il primo è un classico modello statistico di controllo di processo mentre il secondo è un modello machine learning. La variabile di controllo proxy delle prestazioni del componente viene poi analizzata da un modello per l'analisi di serie temporali che prevede il suo andamento nel tempo. Il modello selezionato e sviluppato (attività 3 in Figure 2) è un modello Autoregressivo integrato a media mobile (ARIMA). Una volta costruiti i due modelli prognostici sui dati storici di buon funzionamento, l'applicazione al caso studio ci permette di confrontarne le performance. I parametri di confronto utilizzati sono 3:

- Errore di prima specie del modello diagnostico;
- Errore di seconda specie del modello diagnostico;
- Accuratezza nella stima della RUL.

2.1 Carte di controllo multivariate

Le carte di controllo multivariate, introdotte nel 1947 da Harold Hotelling, consentono di aggregare informazioni relative ad alcune variabili di processo su un diagramma utilizzando la cosiddetta statistica T^2 . La carta di controllo

T^2 è basata sulla distribuzione normale multivariata e sulla distanza di Mahalanobis, ovvero la distanza del set di variabili acquisite dalla media della distribuzione gaussiana multivariata. I valori del vettore medio e della matrice di covarianza della popolazione sono stimati a partire dal set di dati storici delle variabili di controllo selezionate che è il campione della popolazione della quale voglio stimare la funzione di distribuzione di probabilità multivariata. La statistica T^2 di Hotelling è definita come segue:

siano:

$$x_1, \dots, x_n \quad (1)$$

vettori colonna di numeri reali dove ogni colonna rappresenta lo storico di una variabile di controllo selezionata e:

$$\bar{x} = (x_1 + \dots + x_n)/n \quad (2)$$

Le loro medie. Sia \bar{x}

$$W = \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})' / (n - 1) \quad (3)$$

La matrice delle loro varianze e sia μ il vettore colonna delle medie stimate della popolazione, si ha:

$$T^2 = (\bar{x} - \mu)' W^{-1} (\bar{x} - \mu) \quad (4)$$

La statistica T^2 risulta essere la singola variabile di controllo proxy delle performance del componente analizzato da utilizzare per diagnosticare significative variazioni delle prestazioni e da analizzare tramite il modello di analisi di serie temporali per prevedere quando le prestazioni scenderanno sotto la soglia di accettabilità. (Bersimis et al., 2007; Ilin and Raiko, 2010)

Essendo T^2 la distanza del set di variabili dalla media della gaussiana multivariata costruita sui dati storici di buon funzionamento, ad un incremento del parametro T^2 corrisponde un abbassamento delle prestazioni del componente stesso.

2.2 One-class SVM

Lo one-class learning o SVM non supervisionata, mira a separare i dati dall'origine nello spazio n dimensionale, con n pari al numero di variabili di controllo. È un algoritmo utilizzato per il rilevamento di valori anomali. L'algoritmo viene addestrato sul set di dati storici delle variabili di controllo selezionate relative al buon funzionamento del componente analizzato con l'obiettivo di minimizzare la doppia espressione:

$$0,5 \sum_{jk} \alpha_j \alpha_k G(x_j, x_k) \quad (5)$$

Rispetto ad $\alpha_1, \dots, \alpha_n$ soggetto a:

$$\sum \alpha_j = nv \quad (6)$$

Con $0 \leq \alpha_j \leq 1$ per ogni $j = 1, \dots, n$. Il valore di $G(x_j, x_k)$ è un elemento (j, k) della matrice di Gram.

Un piccolo valore di v porta a un minor numero di vettori di supporto e, quindi, a un confine decisionale liscio e grezzo. Un grande valore di v porta a un maggior numero di vettori di supporto e, quindi, a un contorno decisionale sinuoso e flessibile. Il valore ottimale di v dovrebbe essere abbastanza grande da catturare la complessità dei dati e

abbastanza piccolo da evitare il sovrallenamento. Inoltre, $0 < v \leq 1$. (Schölkopf et al., 2001)

Una volta addestrato sui dati storici delle variabili di controllo, il modello attribuisce ad ogni set di nuovi dati il valore di verosimiglianza di appartenenza alla popolazione di addestramento. Il valore di verosimiglianza è quindi la singola variabile di controllo proxy delle performance del componente controllato da utilizzare per diagnosticare significative variazioni delle prestazioni e da analizzare tramite il modello di analisi di serie temporali per prevedere quando le prestazioni scenderanno sotto la soglia di accettabilità.

Il valore di verosimiglianza sarà tanto più elevato quanto più il nuovo set di dati elaborato dal modello addestrato è vicino allo storico dati utilizzati per l'addestramento. Siccome l'addestramento è fatto sui dati relativi alle condizioni di buon funzionamento del macchinario, maggiore è il valore di verosimiglianza e migliori saranno le performance del componente.

2.3 ARIMA

Il metodo ARIMA (AutoRegressive Integrated Moving Average) è un modello che dati in ingresso 3 parametri p, d, q è in grado di modellare l'andamento di una serie temporale e di prevedere ciò che accadrà i successivi istanti temporali. I 3 parametri stanno ad indicare:

- p : ordine del modello autoregressivo;
- d : grado del primo differenziale;
- q : ordine del modello a media mobile.

Con $p, d, q \in \mathbb{N}$.

La determinazione dei migliori parametri da utilizzare per la costruzione del modello ARIMA è fondamentale per ottenere una buona previsione.

Il modello utilizzato nel caso studio effettua una ottimizzazione automatica dei tre parametri tramite minimizzazione dell'Aikaike Information Criterion. Tale parametro assume valore pari a:

$$aic = -2(\log L) + 2(\text{numParam}) \quad (7)$$

Con:

$$\text{numParam} = p + d + q \quad (8)$$

Tale criterio predilige i modelli ARIMA con elevata loglikelihood rispetto ai dati utilizzati per la costruzione del modello e penalizza quelli con elevato numero di parametri (Box et al., 2015; Pai and Lin, 2005; Zhang, 2003). Impostando il numero desiderato di punti da prevedere il modello restituisce per ogni punto di previsione il suo valore atteso e la sua deviazione standard ipotizzando un rumore gaussiano intorno all'andamento della variabile.

Nel prossimo capitolo saranno applicati ad un singolo caso studio i due modelli prognostici sviluppati e verranno confrontate le performance di previsione della RUL.

3 Case study

La metodologia elaborata è stata applicata ad un impianto di produzione di energia geotermica. L'impianto segue un

ciclo a condensazione a singolo flash, ed è in grado di sviluppare una potenza a regime di circa 20 MW. Il macchinario studiato è il compressore per l'estrazione dei gas incondensabili. È una turbomacchina centrifuga a 3 stadi che riporta i gas incondensabili, elaborati assieme al vapore dalla turbina e dal condensatore, in condizioni di pressione atmosferica e ad una temperatura di circa 170 °C; condizioni che gli permettono di essere efficacemente trattati prima dell'immissione in atmosfera.

L'analisi dei dati storici del compressore assieme al personale specializzato dell'impianto caso studio, ci ha permesso di selezionare i dati relativi al funzionamento a regime del macchinario per l'addestramento dei modelli diagnostici e di utilizzare i dati relativi al fenomeno di pompaggio per verificare che questi diano un valore di performance del compressore significativamente al di sotto di quelle utilizzate per l'addestramento del modello. L'applicazione del modello di analisi di serie temporali, ci permette inoltre di valutare l'accuratezza della stima della RUL dei due modelli.

Data la natura del macchinario sono stati selezionati tutti i sensori relativi alle condizioni termofluidodinamiche (pressione, temperatura e portata) del gas lavorato in corrispondenza di ingresso ed uscita di tutti gli stadi di compressione per un totale di 16 variabili di controllo correlate alle performance del compressore.

Il periodo storico analizzato, corrispondente a circa 8 giorni di acquisizione dati con campionamento ad 1 secondo è stato suddiviso insieme al personale di impianto in 3 differenti periodi:

- Regime;
- Anomalia;
- Pompaggio.

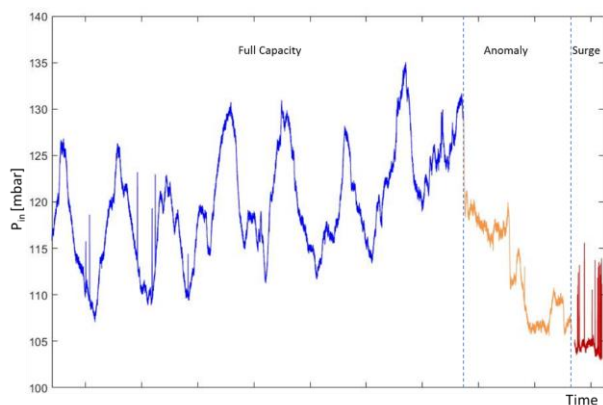


Figure 3: Example of operating conditions periods division

I dati relativi al periodo di regime contano 635782 elementi per ognuna delle 16 variabili di controllo, quelli relativi al periodo di funzionamento anomalo 166473 e quelli del periodo di pompaggio 42103. La grande quantità di dati nel periodo di regime, utilizzato per addestrare i due modelli di controllo statistico di processo ne garantisce la robustezza.

3.1 Applicazione delle carte di controllo multivariate

Sui dati di regime è stata costruita la carta di controllo multivariata utilizzando la funzione pca del software Matlab

tra i cui output compare la statistica T^2 , variabile di controllo correlata alle prestazioni del compressore. Il modello così addestrato è stato poi applicato anche ai dati del periodo di anomalia e di pompaggio. Data l'elevato rumore della statistica T^2 , si è deciso di effettuare una media mobile con bucket temporale pari a 300 secondi in modo da rendere più stabile la previsione ARIMA e più robusto il controllo statistico del processo. In Figure 4 è riportato l'andamento della media mobile della statistica T^2 nel tempo con il relativo limite superiore di controllo (UCL) pari a 25,06 con confidenza del 95% e limite inferiore di controllo relativo al periodo di pompaggio (SCL) pari a 77,36 con confidenza del 5%:

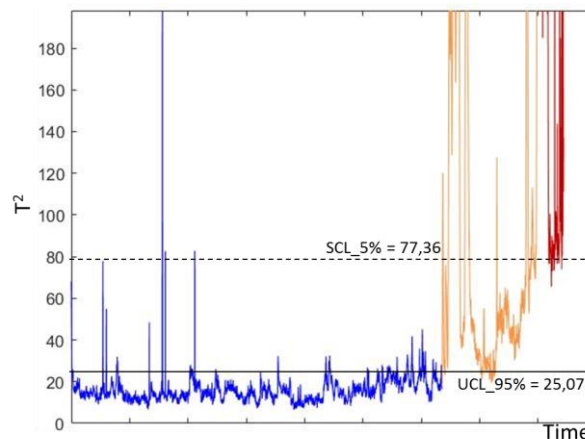


Figure 4: Evoluzione nel tempo della media mobile della variabile di controllo T^2

Dalla Figure 4 risulta evidente che i valori di T^2 relativi al periodo contrassegnato come pompaggio (in rosso) sono nettamente superiori al limite superiore di controllo, in particolare il valore minimo di T^2 nel periodo di pompaggio risulta pari a 65,61 ed il valor medio pari a 190,66. L'errore di prima specie del modello, ovvero la mancata segnalazione di pompaggio è quindi nullo per il periodo selezionato. Avendo scelto un limite superiore di controllo pari al 95%, l'errore di seconda specie, ovvero l'errata segnalazione di funzionamenti anomali risulta pari al 5%. Per ottenere anche un errore di seconda specie nullo è necessario utilizzare come logica di segnalazione del funzionamento anomalo del compressore la permanenza al di sopra della soglia UCL per un tempo maggiore di 2587 secondi. Utilizzando questa logica il modello è in grado di segnalare una anomalia di funzionamento 45,34 ore prima dell'insorgenza del modo di guasto in esame.

3.2 Applicazione del One-class SVM

Come nel precedente caso il modello è stato addestrato sui dati storici relativi al periodo di funzionamento di regime, rappresentato in Figure 3 in blu. Per l'addestramento è stata utilizzata la funzione fitsvm di Matlab con kernel di tipo gaussiano e ottimizzazione automatica del parametro di scale.

Dato l'elevato rumore della variabile di controllo output del modello addestrato, si è deciso di effettuare una media mobile con bucket temporale di 300 secondi come nel caso precedente.

In Figure 5 è riportato l’andamento della media mobile della variabile di controllo Likelihood nel tempo con il relativo limite inferiore di controllo (LCL) pari a 5883,6 con confidenza del 95% e limite superiore di controllo relativo al periodo di pompaggio (SCL) pari a -1,148 con confidenza del 5%:

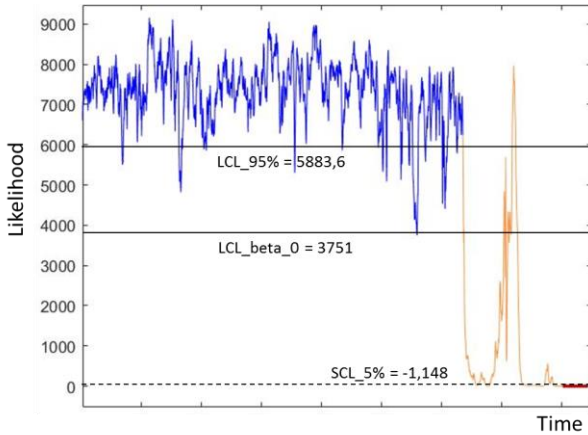


Figure 5: Evoluzione nel tempo della media mobile della variabile di controllo Likelihood

Anche in questo caso è evidente come i dati di pompaggio (in rosso in figura) siano estremamente al di sotto del limite di controllo LCL; in particolare il massimo valore della variabile di controllo nel periodo di pompaggio è di -1,06 ed ha un valor medio di -1,19. Il modello ha quindi una elevata capacità diagnostica di pompaggio del compressore analizzato. L’errore di prima specie sul periodo di analisi è nullo ed avendo scelto un limite inferiore del 95% l’errore di seconda specie è pari al 5%. Per ottenere un errore di seconda specie nullo, è necessario utilizzare come logica di segnalazione del funzionamento anomalo del compressore la permanenza al di sotto della soglia LCL per un tempo maggiore di 146732 secondi, vanificando il potere diagnostico del modello. In alternativa è possibile ottenere lo stesso risultato abbassando il limite di soglia LCL al valore di 3751 (LCL_beta_0). Utilizzando questo valore di soglia il modello è in grado di segnalare un’anomalia di funzionamento del componente 45,96 ore prima dell’insorgere del guasto.

3.3 Applicazione del modello ARIMA

Una volta verificato che entrambi i modelli hanno un elevato potere diagnostico per il caso studio analizzato, abbiamo costruito un metodo di valutazione della qualità della stima della RUL del compressore immaginando un suo utilizzo in realtime sull’impianto. In Figure 6 è riportato lo schema di funzionamento dell’algoritmo ARIMA messo a punto e la valutazione del modello prognostico:

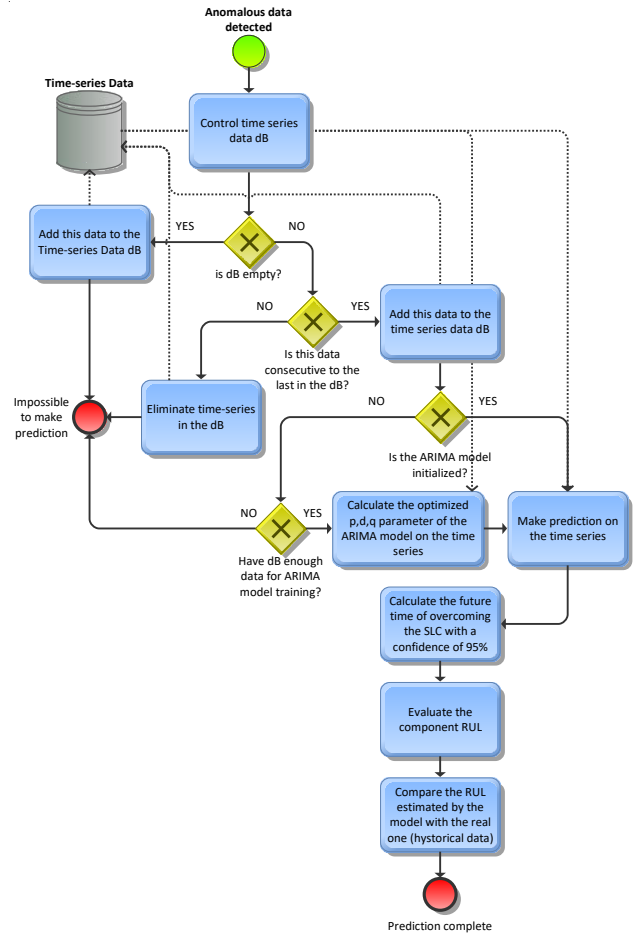


Figure 6: Logica di utilizzo del modello ARIMA nel caso studio e valutazione delle sue performance

Il modello ARIMA è stato testato sui dati di output dei due modelli diagnostici costruiti, ovvero sui valori storici di T^2 per le carte di controllo multivariate e sui valori di Likelihood output del modello One-class SVM, per confrontarne le performance predittive. Nel caso del modello che sfrutta le carte di Hotelling si iniziano a marcare i dati di T^2 come anomali dopo che questi risultano oltre la soglia $UCL_{95\%}$ per un tempo superiore a 2857s; nel caso del modello diagnostico one-class si considerano anomali tutti i dati al di sotto della soglia LCL_{beta_0} . In entrambi i casi il modello ARIMA è stato inizializzato sui primi 1000 valori anomali. L’inizializzazione del modello calcola i parametri p , d e q che meglio si adattano ai dati tramite ottimizzazione automatica basata sulla minimizzazione dell’Aikake Information Criterion. L’output del modello è il valore atteso e la relativa banda di confidenza del 95% dell’istante futuro in cui il valore della variabile di controllo proxy delle performance del compressore oltrepassa la soglia limite $SCL_{5\%}$. Il modello ARIMA inizializzato è per prima cosa adattato alla serie temporale di cui vogliamo la previsione (Time-series Data in Figure 6) utilizzando la funzione estimate e successivamente si effettua la previsione tramite la funzione forecast su software MATLAB.

La forte instabilità del parametro di controllo T^2 nel periodo anomalo, fa sì che la previsione ARIMA sia estremamente instabile e altalenante per tutti i dati anomali fino

all’insorgere dell’evento di pompaggio, rendendo assolutamente inaffidabile la previsione.

Per quanto riguarda il parametro Loglikelihood output del modello One-class SVM, questo presenta inizialmente un andamento discendente molto stabile. La previsione con il modello ARIMA infatti risulta anch’essa stabile e robusta fin dalle prime previsioni come mostrato in Figure 7.

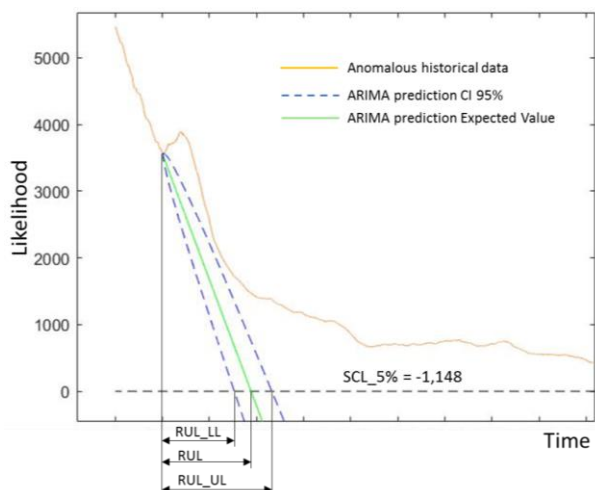


Figure 7: Prima previsione ARIMA della RUL

Il modello ARIMA, stima una RUL attesa del componente di 31,4 minuti, e considerando un intervallo di confidenza del 95% sulla previsione, un limite inferiore di 25,5 minuti ed uno superiore di 38,75. Confrontando questo valore con l’arrivo effettivo del periodo di pompaggio che volevamo prevedere, la RUL risulta sottostimata di circa 45,68 ore.

Successivamente, come mostrato in Figure 5, il valore di Loglikelihood diviene molto instabile ed addirittura ritorna all’interno del limite di controllo LCL_95%, rendendo impossibile continuare la previsione con il modello ARIMA.

4 Discussione e sviluppi futuri

Entrambi i modelli di controllo di processo (carte di controllo multivariate e One-class SVM) applicati al caso studio presentano elevate prestazioni in termini di potere diagnostico di funzionamenti anomali del compressore studiato.

Non possiamo dire lo stesso per quanto riguarda il modello ARIMA applicato alla variabile di controllo proxy delle prestazioni del compressore output dei due modelli diagnostici. In un caso infatti, l’andamento estremamente instabile delle prestazioni del compressore nel transitorio tra funzionamento a regime e pompaggio, non permette di effettuare alcuna previsione della RUL tramite il modello ARIMA. Nel secondo caso invece, la previsione è estremamente stabile nel periodo iniziale, ma la stima della RUL è fortemente sottostimata.

La tabella di seguito riporta una sintesi delle prestazioni dei due modelli prognostici studiati.

Table 1: your table’s caption

	Hotelling + ARIMA	One-Class SVM +ARIMA
Errore di prima specie	0%	0%
Errore di seconda specie	0%	0%
Diagnosi Precoce	45,34 ore	45,96 ore
Stima della RUL	n.d	31,4 minuti
RUL reale – RUL stima	n.d	45,68 ore

Il comportamento estremamente instabile del parametro proxy delle performance del compressore in entrambi i modelli diagnostici potrebbe essere dovuto alla particolare scelta del fenomeno di guasto in analisi. Il fenomeno del pompaggio infatti, non è un vero e proprio evento di guasto, quanto un fenomeno che se protratto nel tempo porta alla rottura del compressore per le elevate vibrazioni che ne comporta. Il pompaggio infatti non è causato da una cattiva condizione meccanica del compressore, ma dalle particolari condizioni termofluidodinamiche del gas lavorato e dalle condizioni del circuito in cui esso è inserito. In definitiva, dipendendo non solo dalle prestazioni del compressore ma complessivamente da tutto il circuito, questo fenomeno risulta molto difficile da prognosticare.

Nel momento in cui si applica questa metodologia prognostica sul campo, l’assenza di dati relativi alle condizioni di guasto, fa sì che il limite di controllo SCL_5% non possa essere calcolato a priori ma ipotizzato ad una certa distanza dalla popolazione di regime, utilizzata per addestrare i modelli diagnostici. Dal singolo caso studio analizzato risulta che per il fenomeno del pompaggio, la popolazione di regime e di guasto sono estremamente distanti, soprattutto utilizzando il modello One-Class SVM.

Inoltre, stesse condizioni di bassa performance del macchinario monitorato, potrebbero essere causate da diversi eventi di guasto. La scelta delle corrette variabili di controllo di addestramento dei modelli diagnostici può mitigare in parte questa incertezza. Nel momento in cui si hanno a disposizione dati relativi alle variabili spia del fenomeno di guasto anche nelle condizioni di cattivo funzionamento, è possibile implementare un sistema prognostico più robusto sfruttando un modello SVM a due classi. Questo sarà così in grado di riconoscere la classe di regime da quella dello specifico guasto che stiamo analizzando. (Navicelli et al., 2019)

In ottica di proseguimento del lavoro presentato risulta interessante applicare la stessa metodologia prognostica a differenti casi studio e modi di guasto per ottenere un risultato generalizzabile.

References

Ahmad, R., Kamaruddin, S., 2012. An overview of time-based and condition-based maintenance in industrial application. *Comput. Ind. Eng.* 63, 135–149.

- Alwan, L.C., Roberts, H.V., 1988. Time-series modeling for statistical process control. *J. Bus. Econ. Stat.* 6, 87–95.
- Bersimis, S., Psarakis, S., Panaretos, J., 2007. Multivariate statistical process control charts: an overview. *Qual. Reliab. Eng. Int.* 23, 517–543.
- Box, G.E., Jenkins, G.M., Reinsel, G.C., Ljung, G.M., 2015. *Time series analysis: forecasting and control*. John Wiley & Sons.
- Hashemian, H.M., 2010. State-of-the-art predictive maintenance techniques. *IEEE Trans. Instrum. Meas.* 60, 226–236.
- Ilin, A., Raiko, T., 2010. Practical approaches to principal component analysis in the presence of missing values. *J. Mach. Learn. Res.* 11, 1957–2000.
- Komonen, K., 2002. A cost model of industrial maintenance for profitability analysis and benchmarking. *Int. J. Prod. Econ.* 79, 15–31.
- Löfsten, H., 1999. Management of industrial maintenance—economic evaluation of maintenance policies. *Int. J. Oper. Prod. Manag.*
- MacGregor, J.F., Kourti, T., 1995. Statistical process control of multivariate processes. *Control Eng. Pract.* 3, 403–414.
- Mason, R.L., Young, J.C., 2002. *Multivariate statistical process control with industrial applications*. Siam.
- Navicelli, A., Vincitorio, M., De Carlo, F., Tucci, M., 2019. Predictive maintenance in industrial plants: real application of Machine Learning models for prognostics. XXIV Summer Sch. “Francesco Turco” – *Ind. Syst. Eng.* 165–171.
- Oakland, J.S., 2007. *Statistical process control*. Routledge.
- Pai, P.-F., Lin, C.-S., 2005. A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega* 33, 497–505.
- Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J., Williamson, R.C., 2001. Estimating the support of a high-dimensional distribution. *Neural Comput.* 13, 1443–1471.
- Si, X.-S., Wang, W., Hu, C.-H., Zhou, D.-H., 2011. Remaining useful life estimation—a review on the statistical data driven approaches. *Eur. J. Oper. Res.* 213, 1–14.
- Zhang, G.P., 2003. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* 50, 159–175.