



**Politecnico
di Torino**

Politecnico di Torino

Collegio di Ingegneria Gestionale – Classe LM-31
Corso di Laurea Magistrale in Ingegneria Gestionale

Tesi di Laurea di II livello

Analisi di un sistema CONWIP con job difettosi,
nastri trasportatori e loop: confronto di strategie per
migliorarne le prestazioni

Relatore:
Prof.ssa Arianna Alfieri

Candidato:
Alice Pecchio

Corelatore:
Prof. Claudio Castiglione

Anno accademico 2025-2026

Sommario

1. Introduzione: Contesto e Motivazione della Ricerca	3
2. Revisione della letteratura	8
2.1 Introduzione e Contesto di Ricerca	8
2.2 Introduzione al Sistema CONWIP	9
2.3 Scheduling e Meccanismi di Prioritizzazione in Sistemi a Macchine Parallele	10
2.4 Il Digital Twin (DT) per l'Industria 4.0.....	12
2.4.2 Architetture e Implementazione	12
2.5 Ottimizzazione della Schedulazione Basata su Digital Twin: Il Contesto del Job Shop Scheduling e la Necessità di Dinamicità	14
2.6 Applicazioni Settoriali e Contesto Specifico della Ricerca	15
2.6.1 Il Parallelo con i Sistemi Logistici a Flusso Rigido	15
2.7 Contributo Applicativo e Rilevanza del lavoro: Ottimizzazione della Logica in Ambienti CONWIP con Rework	16
3. Descrizione del Problema.....	19
3.1 Architettura e Layout del Sistema Logistico	19
3.2 Dinamiche del flusso e logiche operative	20
3.1.2 Flusso delle Entità e Vincolo di Congestione Logistica	21
3.2 Parametri e Stocasticità del Modello	21
3.2.1 Tempi di processo e distribuzioni.....	21
3.2.2 Dinamica Rework e variabilità	22
3.3 Identificazione e definizione delle regole di priorità	23
3.3.1 Variabili di controllo: regole di priorità del DT.....	23
3.4 Vincoli del sistema e Implicazioni modellistiche	24
4. Metodologia dello studio	26
4.1 Modellazione dei flussi post-ispezione e bilanciamento della capacità	26
4.2 Dimensionamento del limite CONWIP tramite la Legge di Little	28
4.3 Impostazioni sperimentali.....	30
4.4 Definizione del numero di repliche.....	31
4.5 Key Performance Indicators	32

4.6 Scenari simulativi.....	34
5. Descrizione degli esperimenti.....	35
5.1 Scelta del numero di pallet e del WIP del sistema	35
5.2 Confronto tra i modelli con e senza DT	38
5.3 Scelta della finestra di attivazione del DT	40
5.4 Determinazione della regola di priorità più efficace.....	43
5.5 Analogie e differenze con l'indicatore Remaining Passages	44
6. Discussione dei risultati e implicazioni manageriali.....	46
6.1 La gestione del WIP e il limite della resilienza operativa	46
6.2 Il paradosso del Digital Twin: stabilità contro reattività	47
6.3 Analisi della finestra decisionale e limiti statistici	48
6.4 Conflitto tra KPI e ottimizzazione dei flussi interni	48
6.5 Conclusioni e implicazioni per il management.....	49
7. Conclusioni.....	50
7.1 Conclusioni generali.....	50
7.2 Contributi della tesi	50
7.3 Limiti del lavoro svolto	51
7.4 Prospettive di estensione	52
8. Bibliografia	54

1. Introduzione: Contesto e Motivazione della Ricerca

L'attuale panorama industriale, profondamente influenzato dai principi dell'Industria 4.0, esige che i sistemi manifatturieri e logistici operino al massimo della loro efficienza, garantendo al contempo elevata flessibilità e una rapida capacità di adattamento alle fluttuazioni della domanda. In contesti produttivi caratterizzati da un'alta complessità strutturale e da processi stocastici, la gestione ottimale dei flussi interni è un fattore critico di successo.[1]

I sistemi di convogliamento (conveyors), costituiscono un'importante supporto ai sistemi di movimentazione dei materiali. Essi determinano la velocità e la sequenza con cui i prodotti vengono instradati alle diverse stazioni di lavorazione.

L'obiettivo di questa tesi è analizzare un sistema di convogliamento di materiali operante con un vincolo di Work-In-Process costante (CONWIP)[2], la cui complessità è data dalla presenza di diversi anelli di flusso e da punti decisionali multipli che gestiscono i prodotti in attesa di lavorazione.

A differenza dei sistemi tradizionali di tipo "push", guidati da previsioni della domanda che spesso generano eccessi di scorte e lunghi lead time, l'approccio adottato in questa sede si fonda sulla logica "pull". In questo modello, la produzione è trainata dalla domanda reale, riducendo drasticamente il rischio di sovrapproduzione. Nello specifico, il sistema analizzato implementa la metodologia Constant Work In Process (CONWIP), che agisce a livello globale mantenendo costante il numero di job circolanti attraverso un meccanismo di pallet e permette di redistribuire il flusso modificando il numero di entità che accedono contemporaneamente alla linea.

L'efficacia dei sistemi di movimentazione meccanica è messa a dura prova dalla presenza di layout non lineari, come quelli a configurazione chiusa con caroselli.

In tali contesti, la gestione dei nodi di "merge" e "split" (punti di convergenza e diramazione) diventa critica: una mancata sincronizzazione può infatti innescare fenomeni di "starving" nelle stazioni a valle per mancanza di materiale, o di "blocking" in quelle a monte per eccesso di saturazione nei buffer.[3] Tali dinamiche non solo riducono l'efficienza, ma richiedono un monitoraggio costante dei flussi fisici per evitare che le code si propaghino all'intero impianto.[4]

L'obiettivo fondamentale di questo lavoro di tesi è quello di migliorare il throughput (TH) del sistema. Il TH è la metrica per eccellenza dell'efficienza produttiva e la sua massimizzazione è direttamente correlata al miglioramento della capacità di risposta del sistema al mercato.[5], [6]

Per conseguire questo obiettivo, l'analisi si concentra sulla gestione della logica di lavorazione, piuttosto che sulla modifica della capacità fisica delle macchine. Nello specifico, si intende migliorare l'efficienza imponendo differenti regole di priorità al modello, focalizzando l'attenzione sul secondo anello critico del sistema. Questa è l'area identificata come il maggiore generatore di congestione, a causa anche della sua posizione a valle della stazione di rilavorazione.

L'approccio si basa sulla simulazione del sistema per valutare l'impatto di diverse logiche di priorità.

Le motivazioni della ricerca nascono da una chiara diagnosi di inefficienza. In merito a quanto presente in letteratura sulla gestione dei flussi produttivi, le parziali inefficienze del sistema sono molteplici.

L'indicatore più preoccupante è l'analisi del tempo di ciclo (flow time) dei job: non perturbando le regole di priorità, i job tendono a passare gran parte del loro flow time in coda o nei loop del sistema e a ridurre ad una percentuale molto piccola il tempo impiegato per le lavorazioni.

Questo squilibrio, dove il tempo di attesa supera spesso il tempo di servizio, è la prova di una difficoltà logistica nell'allocazione delle risorse e dimostra che c'è un'importante possibilità di miglioramento. L'eccessivo delay è attribuito a un sovraccarico del sistema, la cui origine è complessa e multifattoriale. Il problema non è la scarsa velocità delle macchine, ma il sistema inefficace di instradamento e di dispatching in presenza di variabilità, che viene creata in particolare dalle possibilità di rilavorazioni alle quali possono essere sottoposti i job.

Un'altra tematica importante per questa tesi è proprio il sovraccarico, che non è solo un effetto, ma è alimentato da caratteristiche strutturali del sistema. Nel presente documento l'analisi si concentra su due fonti primarie di congestione e instabilità: il Traffico Non Produttivo e la Dinamica di Rework.

Relativamente al primo, il nastro trasportatore svolge una doppia funzione: muovere i job da lavorare e supportare lo spostamento di pallet vuoti che occupano spazio prezioso sull'anello. Se non gestiti in buffer dedicati (come la stazione di accumulo pallet prevista nel modello), costringono a un'ulteriore

regola di priorità complessa che deve risolvere il trade-off tra job in lavorazione e il riposizionamento dei pallet. Questa componente logistica, se non separata e prioritizzata correttamente, contribuisce a limitare il sistema.

Per quanto riguarda invece la Dinamica Rework, la causa principale del sovraccarico è la possibilità di rilavorare i job e di rimandarli alla stazione di provenienza o a quella successiva dopo il controllo qualità (WSQC).[7]

Questo meccanismo, che garantisce la qualità del prodotto, ha tre impatti negativi sul flusso:

1. Ingresso di Job in Ritardo (Delayed Entries): la stazione WSQC è il punto in cui avviene la rilavorazione. Un job che non supera la fase di rettifica viene espulso dal sistema, mentre nel caso in cui il difetto venga risolto il job avrà la possibilità di essere reindirizzato verso la WS di origine per compiere nuovamente la lavorazione, oppure verso la successiva per procedere nel suo flusso produttivo. Entrambe le operazioni comportano che il job occupi il sistema per un tempo maggiore rispetto a quello previsto.[8]
2. Congestione Lunga e Stocastica: la modellazione prevede che un job rilavorato debba compiere un intero giro sul conveyor prima di essere reintrodotta in circolo. Questo vincolo amplifica l'effetto congestionante: un job non a buon fine occupa risorse di trasporto per lungo lasso di tempo, rallentando l'intero flusso. Inoltre, poiché non possiamo definire un numero esatto di job da rilavorare ma solo una probabilità, la natura non statica del processo introduce un livello di variabilità e incertezza che rende inefficace qualsiasi regola di instradamento fissa. [3]
3. Rischio di Deadlock e l'Attesa Circolare: nei sistemi di produzione automatizzati (AMS) dotati di loop chiusi e buffer limitati, il sovraccarico può evolvere nel fenomeno del deadlock. Questa condizione critica si verifica quando le risorse condivise (nastri, macchinari o buffer) vengono occupate da entità che si bloccano reciprocamente nell'attesa di accedere l'una alle risorse dell'altra. In un sistema altamente interconnesso come quello in esame, una congestione locale può propagarsi rapidamente, generando uno stato di attesa circolare che impedisce l'avanzamento di qualsiasi job. La prevenzione del deadlock diventa quindi un requisito fondamentale.[7], [9]

Per affrontare il problema del delay, la ricerca si è concentrata su un punto specifico del sistema: un anello (loop) che funge da imbuto logistico. Questa sezione della linea è sottoposta a confluenza di flussi di rilavorazione da più stazioni e presenta la massima complessità decisionale (più opzioni di instradamento parallele). È in questo punto critico che l'analisi si concentra per trovare la massima leva sull'ottimizzazione del throughput.

Gli obiettivi specifici della tesi sono molteplici, in primo luogo vi è quello di misurare e confrontare il miglioramento del TH nel loop critico (Figura 1), e di conseguenza sull'intero sistema, ottenuto attraverso l'unione di quattro regole di priorità implementate. Le regole, basate sullo status di lavorazione del job, devono essere valutate rispetto a un sistema baseline senza priorità.

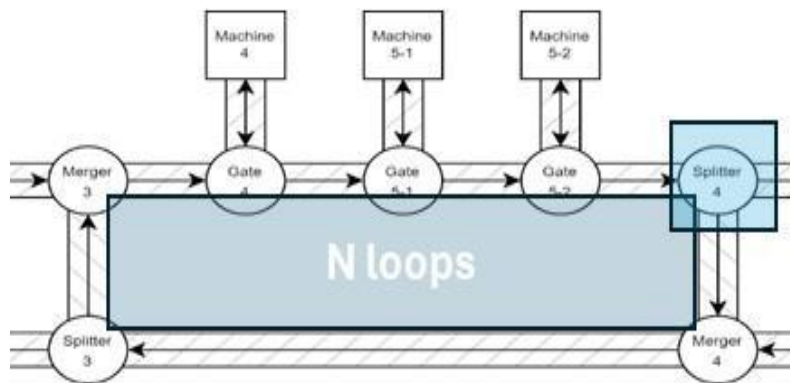


Figura 1: Loop WS4 e WS5

Un secondo obiettivo di questo documento è quello di valutare l'efficacia del controllo dinamico delle differenti regole di priorità implementate nel Digital Twin. Questo strumento rappresenta una copia digitale del sistema in fase di studio [9], esso a differenza della linea non è costituita da moduli fisici, ma è un algoritmo che ne valuta le potenzialità future simulandone il comportamento e, in base a quanto suggerito dalle regole di priorità, è in grado di dare un'indicazione su quale potrebbe essere lo scenario più favorevole. Ciò implica dimostrare che la strategia di "simulazione nella simulazione" sia in grado di selezionare la regola di controllo che massimizza il TH in tempo quasi reale, superando le performance della migliore regola di priorità implementata in modo statico.

In quest'ottica, lo studio si focalizza in modo particolare sull'integrazione del Digital Twin direttamente all'interno dell'ambiente di simulazione.

Oltre ai suddetti punti, si aggiunge anche la possibilità di analizzare l'impatto delle regole dinamiche sulla distribuzione del flow time dei job, verificando se il controllo dinamico riesca a invertire il trend osservato, ovvero ridurre in maniera significativa il tempo trascorso in coda a favore del tempo di lavorazione.

Un altro aspetto significativo di questo lavoro risiede nell'implementazione del Digital Twin (DT) non solo come strumento di analisi e visualizzazione, ma come agente di controllo attivo e decisionale all'interno di un ambiente di Simulazione a Eventi Discreti (DES). La tesi realizza un'architettura di simulazione dove un modello (il DT) simula il futuro prossimo del sistema più ampio (il "sistema fisico") per guidare le sue decisioni operative. Questa metodologia è all'avanguardia nell'ingegneria dei sistemi e fornisce un modello replicabile per l'integrazione di sistemi cyber-fisici in ambienti virtuali. Inoltre, il modello di tesi valida un algoritmo di controllo basato su un ciclo di simulazione condensata. Questa è una dimostrazione concreta di come i principi dell'ottimizzazione predittiva (dove le decisioni sono prese a valle delle simulazioni) possano essere applicati per gestire variabili logiche (regole di dispatching) in sistemi manifatturieri complessi.[10]

A seguire il focus dello studio si è poi rivolto verso l'impatto che i supporti fisici hanno per la definizione del TH, infatti stabilendo un numero preciso di pallet da utilizzare si va ad influenzare quanto la linea risulta più meno carica di job e si definiscono equilibri diversi per quanto riguarda l'utilizzo delle macchine, la gestione delle rilavorazioni e la congestione del sistema.

In conclusione, vi è l'obiettivo di identificare la regola ottimale, ovvero di determinare se, a seguito della sperimentazione con il DT, emerga una regola di priorità che si possa definire "assoluta" in quanto viene selezionata e utilizzata dal DT con un livello di confidenza pari al 95% dei casi. La sua identificazione avrebbe una profonda implicazione pratica, suggerendo che una regola fissa, se supportata da una conoscenza a priori della sua efficacia nella maggior parte degli scenari, può semplificare il sistema di controllo in applicazioni reali.

2. Revisione della letteratura

2.1 Introduzione e Contesto di Ricerca

Il secondo capitolo presenta una panoramica dei concetti teorici e degli studi precedenti necessari per inquadrare il problema analizzato. La trattazione parte dall'analisi del sistema CONWIP, spiegando come questo metodo di controllo della produzione permetta di gestire i flussi di lavoro in modo intermedio tra logica "push" e "pull". Si passa poi all'approfondimento della simulazione e della teoria delle code, che rappresentano gli strumenti tecnici usati per studiare il comportamento dei materiali e i tempi di attesa nel sistema. Una parte centrale è dedicata al Digital Twin, descritto come l'architettura digitale capace di collegare i dati reali con i modelli virtuali per supportare le decisioni. Infine, vengono esaminate le diverse regole di schedulazione e le strategie di gestione dei pezzi difettosi, con l'obiettivo di evidenziare come l'integrazione di questi elementi possa migliorare l'efficienza complessiva e ridurre i tempi di attraversamento nel processo produttivo.

La ricerca si colloca all'intersezione di tre domini accademici e applicativi: la modellazione di sistemi dinamici tramite Simulazione a Eventi Discreti (DES), le architetture e le funzionalità del Digital Twin (DT) nel contesto dell'Industria 4.0, e le strategie di ottimizzazione della schedulazione in tempo reale.

La Quarta Rivoluzione Industriale ha reso la capacità di prendere decisioni real time e di adattare i processi operativi a fronte di perturbazioni improvvise un imperativo strategico.

Le sfide legate alla gestione di sistemi di produzione e reti di trasporto, spesso caratterizzati da flussi rigidi e schedulati, richiedono metodologie in grado di mappare la complessità del mondo fisico nel dominio virtuale e di intervenire attivamente sulla realtà. In questo contesto, l'evoluzione della simulazione verso il concetto di Digital Twin rappresenta il cuore dell'analisi e l'ottimizzazione di sistemi altamente regolari.[1], [10]

2.2 Introduzione al Sistema CONWIP

Il sistema CONWIP (Constant Work-In-Process) rappresenta una logica di controllo della produzione ibrida che combina i vantaggi dei sistemi push e pull. Introdotto per la prima volta da Spearman et al. nel 1990[2], il CONWIP si differenzia dal sistema Kanban tradizionale poiché, invece di monitorare il numero di pezzi per ogni singola stazione, stabilisce un limite massimo di unità lavorative (WIP) per l'intera linea produttiva o per specifici segmenti di essa, come visibile in Figura 2.

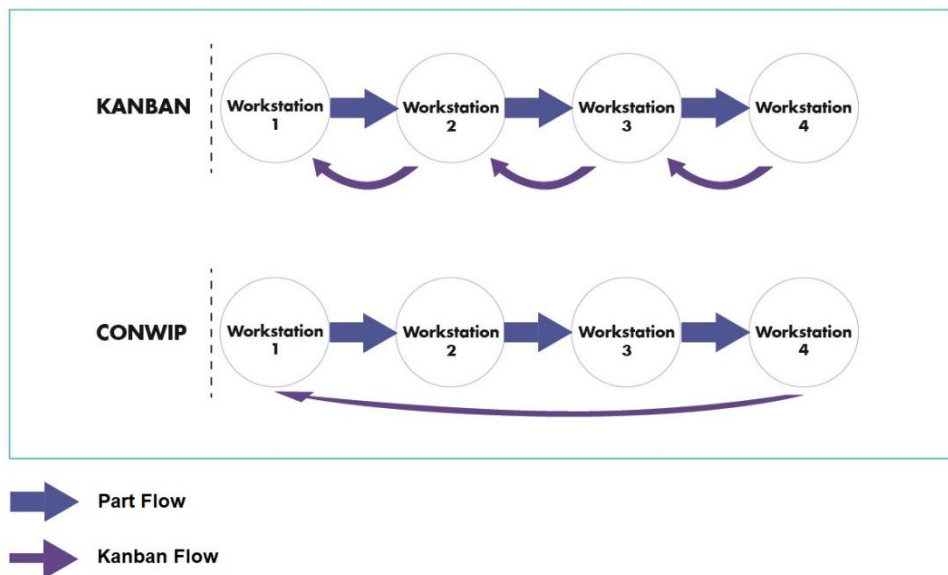


Figura 2: logica CONWIP a confronto con Logica KANBAN

Fonte: UTEK Lean Production Division

In questo contesto, il rilascio di un nuovo ordine nel sistema è subordinato al completamento di un pezzo in uscita: tale meccanismo di "backlog" garantisce che il carico di lavoro circolante rimanga costante, riducendo la congestione e stabilizzando i tempi di attraversamento (flow time). Lo studio condotto da Hopp e Spearman ha evidenziato che l'efficacia del CONWIP risiede nella sua capacità di mantenere un throughput elevato pur operando con livelli di inventario minimi, proteggendo il sistema dalle fluttuazioni causate dalla variabilità dei tempi di servizio.[2], [11]

Nelle implementazioni più complesse, l'efficacia del controllo CONWIP è strettamente legata alla gestione della capacità fisica del sistema.

Studi successivi da parte di Alfieri, Castiglione e Pastore si sono concentrati su due aspetti fondamentali e hanno dimostrato che, in presenza di sistemi di trasporto automatizzati come i nastri trasportatori, la definizione del livello ottimale di WIP deve tenere conto non solo della capacità delle macchine, ma anche dello spazio fisico di accumulo disponibile sui nastri, per evitare fenomeni di saturazione o stallo circolatorio (deadlock) e inoltre la scelta del WIP migliore è cruciale per un corretto bilanciamento del TH e della sua variabilità.[12], [13] A tal proposito, lo studio di Ashrafian et al. risulta di fondamentale importanza per la validazione della metodologia qui adottata: gli autori dimostrano come l'identificazione di questo livello ottimale non possa basarsi su calcoli statici, ma debba essere supportata da simulazioni dinamiche in grado di individuare il punto di equilibrio operativo tra la saturazione del sistema e l'incremento del flow time, prevenendo la congestione logistica che spesso inficia le prestazioni reali.[3]

Un elemento di ulteriore complessità, centrale nella presente trattazione, è rappresentato dai flussi di rilavorazione. Come evidenziato da Florescu et al., la presenza di rework trasforma il sistema in una rete a ciclo chiuso dove il carico generato dai pezzi non conformi può saturare precocemente il limite di WIP.[8] Questo studio è cruciale per giustificare la necessità di un monitoraggio costante: Florescu dimostra infatti che, in presenza di rilavorazioni, il throughput non solo si stabilizza, ma rischia di crollare drasticamente se il sistema viene inondato di pallet che riempiono i loop di ritorno, impedendo l'ingresso di nuovi job. Inoltre, lo studio sottolinea come il WIP sia differente e oscilli in maniera indipendente a seconda delle regole di priorità impostate. Ciò che ancora risulta inesplorato è l'impatto di regole di priorità variabili rispetto alla definizione di un livello di CONWIP. A supporto dello studio in corso vi è la letteratura recente e alcuni studi effettuati da Bortolini et al. suggeriscono infatti che l'integrazione di regole di priorità dinamiche all'interno di un'architettura CONWIP sia la chiave per ottimizzare le prestazioni in ambienti caratterizzati da alta variabilità e flussi di rilavorazione ed è proprio questo il focus del presente studio. [14]

2.3 Scheduling e Meccanismi di Prioritizzazione in Sistemi a Macchine Parallele

All'interno di workstation configurate con risorse parallele identiche, è importante definire un'efficace architettura di scheduling nell'ottica finale della massimizzazione del Throughput e dell'ottimizzazione del bilanciamento del

carico tra le unità produttive. In tali contesti, la letteratura scientifica distingue le logiche di gestione in due ambiti decisionali complementari: le regole di sequenziamento (o priorità) e le regole di allocazione (o assegnazione).[5], [6]

La letteratura suggerisce che non esista una configurazione universalmente ottima; l'efficacia di una combinazione di regole è strettamente correlata alla variabilità del sistema. In ambienti caratterizzati da loop di ricircolo e nastri trasportatori, la sfida risiede nel coordinare queste scelte logiche con i vincoli fisici del sistema, assicurando che la massimizzazione del throughput non si traduca in una saturazione critica dei buffer di trasporto, con conseguente rischio di stallo operativo.[6]

L'integrazione tra la gestione dei flussi e l'architettura fisica del sistema trova un riscontro cruciale negli studi recenti applicati al settore automotive, dove la complessità dei componenti per veicoli elettrici richiede soluzioni di bilanciamento dinamiche e altamente automatizzate. Il problema del bilanciamento delle linee non può prescindere dalla scelta del sistema di movimentazione dei materiali, poiché l'uso di caroselli e nastri trasportatori permette di trasformare le strutture di trasporto in veri e propri loop di buffer capaci di assorbire la variabilità operativa senza ricorrere a ingombranti buffer statici. Tale configurazione è particolarmente efficace in sistemi regolati dalla logica CONWIP, dove la determinazione del WIP critico risulta fondamentale per ottimizzare il throughput a fronte di macchine non perfettamente affidabili.[13] È stata poi ulteriormente approfondita la questione, sottolineando come la semplice regolazione del livello globale di WIP non sia sufficiente a garantire la stabilità in ambienti caratterizzati da un alto mix produttivo e mass customization. Gli esperti propongono l'adozione di politiche di sequenziamento dei job gestite tramite loop di convogliamento, le quali permettono di riconfigurare i flussi in tempo reale e prevenire fenomeni di congestione o deadlock critici vicino alle stazioni collo di bottiglia.[12] La sinergia tra queste metodologie conferma che l'efficienza di una linea di assemblaggio moderna dipende dalla capacità di far dialogare le politiche di controllo logico (CONWIP e sequenziamento) con le caratteristiche fisiche della movimentazione, utilizzando la simulazione ad eventi discreti come strumento indispensabile per validare la robustezza del sistema prima dell'effettiva implementazione fisica.

2.4 Il Digital Twin (DT) per l'Industria 4.0

La necessità di un'analisi e di un intervento in tempo reale ha spinto la DES verso la sua evoluzione più sofisticata, il Digital Twin, che rappresenta una delle tecnologie chiave dell'Industria 4.0.[10], [15]

Il Digital Twin (DT) è un modello virtuale dinamico e in tempo reale di un oggetto, processo o sistema fisico. La sua prima definizione risale all'anno 2003 ed è ad opera del professor Micheal Grieves e secondo lo studioso esso costituisce un componente cruciale per la gestione del life cycle dei prodotti manifatturieri, come strumento di replica della realtà e di ingegnerizzazione delle fasi di sviluppo.[9]

La caratteristica fondamentale che lo distingue da una semplice simulazione è la connessione bidirezionale con la sua controparte fisica (Physical Twin), che consente non solo il monitoraggio continuo, ma anche l'invio di comandi correttivi al sistema stesso.

Le principali caratteristiche del DT, necessarie per affrontare la schedulazione dinamica, includono:

- Connessione in Tempo Reale: scambio continuo e a bassa latenza di dati tra il mondo fisico e il modello virtuale.
- Capacità Predittiva e Prescrittiva: utilizzo di algoritmi di simulazione per prevedere il comportamento futuro (ad esempio, l'accumulo di ritardo) e suggerire o implementare azioni correttive (prescrizione).

La natura bidirezionale del DT lo rende l'unica configurazione in grado di implementare l'algoritmo di ottimizzazione in tempo reale richiesto per correggere dinamicamente un sistema a schedulazione rigida in presenza di ritardi.[2], [10]

2.4.2 Architetture e Implementazione

L'implementazione di un Digital Twin per sistemi operativi complessi non si esaurisce nella creazione di un modello statico, ma richiede un'architettura dinamica e modulare che integri i flussi di dati fisici nel motore di simulazione in modo trasparente. Tale struttura deve permettere non solo l'osservazione del sistema, ma soprattutto la chiusura del ciclo di controllo, dove l'output del modello virtuale torna a influenzare l'entità fisica.

Secondo la letteratura più recente, in particolare il modello a cinque dimensioni proposto da Tao et al. (2018), l'architettura tipica di un Digital Twin è strutturata su diversi livelli interconnessi (Figura 3), ciascuno con compiti specifici per garantire il funzionamento continuo e sincrono:[16], [17]

- Livello Fisico (Physical Layer): rappresenta il sistema reale nella sua interezza, comprendendo macchinari, sensori, attuatori e l'infrastruttura di movimentazione (nastri trasportatori). In questo livello avviene la generazione dei dati grezzi di stato; ogni movimento di un pallet o completamento di una lavorazione viene registrato per alimentare il gemello virtuale.[17]
- Livello di Dati (Data Layer): costituisce il sistema nervoso del Digital Twin. È la piattaforma incaricata della raccolta, pulizia e instradamento delle informazioni provenienti, ad esempio, dai PLC (Programmable Logic Controllers) o dai sistemi SCADA. Questo livello deve garantire una comunicazione bidirezionale in tempo reale e a bassa latenza, assicurando che la replica virtuale sia sempre allineata allo stato attuale della linea produttiva. La gestione dei dati qui include anche l'archiviazione di serie storiche, fondamentali per addestrare algoritmi predittivi.[18]
- Livello di Modellazione e Simulazione (Virtual Layer): è il nucleo computazionale dell'architettura, dove risiede il modello di Simulazione a Eventi Discreti (DES). Questo livello esegue la replica virtuale del sistema e, grazie alla sua capacità di accelerare il tempo simulato, svolge analisi predittive sulle prestazioni future. In questo ambiente vengono testate diverse regole di priorità e configurazioni logistiche, agendo come una "sandbox" sicura in cui gli algoritmi di ottimizzazione possono ricercare la soluzione migliore senza rischiare interruzioni nel mondo fisico.[15]
- Livello di Servizi (Service Layer): rappresenta il ponte tra l'intelligenza del sistema e l'operatività. Include l'interfaccia utente (dashboard) per il monitoraggio dei KPI e, soprattutto, il modulo di ottimizzazione. Una volta che il livello di simulazione ha identificato la strategia ottimale (ad esempio la regola di dispatching più efficace per sbloccare un loop di lavorazione), il Service Layer traduce questa decisione in comandi operativi. Questi vengono inviati nuovamente al livello fisico attraverso il Data Layer, ripristinando la schedulazione ottimale e garantendo il mantenimento del Throughput desiderato.[4]

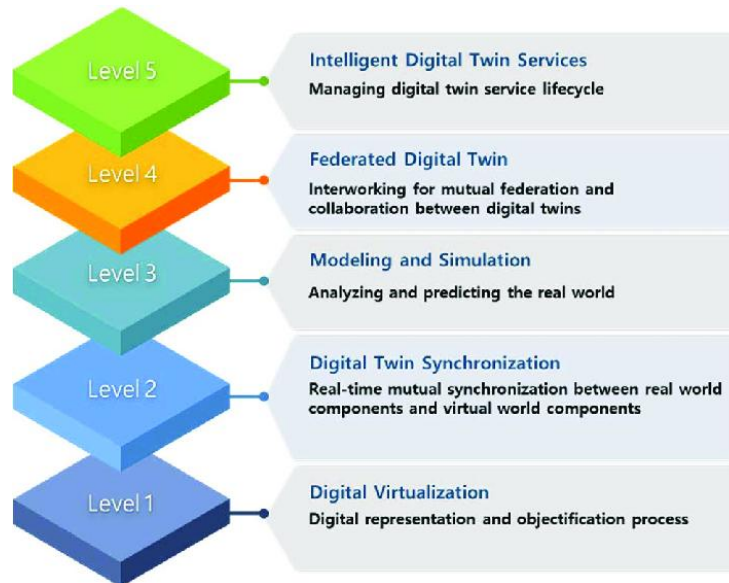


Figura 3: i 5 livelli del Digital Twin

Fonte: IEEE Access

Questa architettura stratificata assicura che il Digital Twin non rimanga un'entità isolata, ma diventi uno strumento di supporto alle decisioni dinamico, capace di rispondere tempestivamente alle derive stocastiche del processo produttivo, come variabilità nei tempi di lavorazione o picchi imprevisti di pezzi difettosi (rework). [2]

2.5 Ottimizzazione della Schedulazione Basata su Digital Twin: Il Contesto del Job Shop Scheduling e la Necessità di Dinamicità

L'integrazione della DES nel framework DT ha come obiettivo principale l'ottimizzazione delle performance operative attraverso la schedulazione dinamica.[19], [20]

La schedulazione (Scheduling) è il processo di assegnazione di compiti (jobs) a risorse limitate nel tempo, per ottimizzare metriche di performance definite. Il Job Shop Scheduling (JSS), in cui i job seguono percorsi variabili attraverso diverse macchine, è particolarmente rilevante e noto per essere un problema con elevata complessità computazionale.

In un contesto Smart Manufacturing, le incertezze operative (ritardi inattesi, deviazioni nei tempi di processo, guasti) non possono più essere gestite da schedulazioni statiche. Le strategie di schedulazione rigida, sebbene efficienti

teoricamente, sono inadeguate a mantenere la puntualità quando insorgono i disturbi.

L'analisi condotta da Fang, Peng, Lou et al. segna un punto di svolta metodologico in tale ambito, introducendo il Digital Twin non come una mera replica speculare e passiva, ma come uno strumento correttivo attivo capace di neutralizzare le inefficienze derivanti da eventi imprevisti e asimmetrie informative. [19] Attraverso una sincronizzazione bidirezionale tra l'entità fisica e la sua controparte virtuale, il Digital Twin trasforma la gestione del JSS in un sistema di controllo a ciclo chiuso (closed-loop), dove il feedback in tempo reale permette di rilevare derive stocastiche e ricalibrare costantemente i flussi operativi, garantendo una precisione altrimenti irraggiungibile in contesti ad alta variabilità. In questa prospettiva, il lavoro di Sitahong et al. [20] consolida la cornice sistematica necessaria alla manifattura intelligente, evidenziando come la reattività del sistema non dipenda esclusivamente dalla pervasività della raccolta dati, ma dalla profonda integrazione tra modellazione accurata, algoritmi di ottimizzazione evoluti e l'impiego della Discrete Event Simulation (DES) come motore decisionale. Tale approccio trasforma la simulazione in uno strumento di validazione preventiva: prima che ogni azione correttiva venga implementata sul campo, essa viene sottoposta a un'analisi predittiva rigorosa che permette di testare scenari "what-if" e assicurare che la risposta ai disturbi sia sempre allineata agli obiettivi di efficienza globale, conferendo alla logistica di officina una resilienza dinamica indispensabile per la sopravvivenza nei moderni mercati competitivi.[19], [20]

2.6 Applicazioni Settoriali e Contesto Specifico della Ricerca

2.6.1 Il Parallelo con i Sistemi Logistici a Flusso Rigido

Le metodologie basate su DES e Digital Twin trovano una naturale applicazione in quegli ambiti dove l'affidabilità e la sincronia temporale sono requisiti non negoziabili, come nel trasporto pubblico su ferro o nella logistica integrata. Un analogia particolarmente calzante per il Job Shop Scheduling è quella dei sistemi di convogliatori modulari utilizzati nei grandi centri di smistamento. Questi impianti sono progettati per movimentare carichi in modo costante e prevedibile, seguendo sequenze predefinite che richiedono una stretta aderenza alla pianificazione originale per evitare congestioni. Tuttavia, proprio come accade in una linea di smistamento, la rigidità del flusso diventa il punto debole del sistema

non appena si manifesta un'anomalia: un ritardo minimo su un modulo può propagarsi rapidamente, declassando l'efficienza dell'intera catena.

In questo contesto, il parallelo con la logistica a flusso rigido evidenzia la necessità di passare da una gestione puramente meccanica a una supervisione "intelligente". Se nei sistemi tradizionali il disturbo viene gestito con logiche di arresto o buffer fisici, l'integrazione di un gemello digitale permette di trattare la rigidità non come un limite, ma come una base di precisione su cui innestare una capacità di risposta dinamica. L'obiettivo non è dunque lo smantellamento del flusso programmato, quanto l'introduzione di uno strato di controllo capace di simulare in anticipo le conseguenze di un intoppo, ricalibrando le velocità o i percorsi prima che il blocco diventi critico. Questa visione permette di mutuare la regolarità tipica della logistica pesante traslandola nella complessità del Job Shop, ottenendo un sistema che conserva la puntualità del trasporto rigido pur possedendo la flessibilità necessaria per gestire job con priorità e cicli di lavoro variabili. [21]

2.7 Contributo Applicativo e Rilevanza del lavoro: Ottimizzazione della Logica in Ambienti CONWIP con Rework

Dal punto di vista industriale, la ricerca si concentra sulla risoluzione di un problema comune, ma altamente complesso, in ambito manifatturiero: la gestione del throughput in presenza di rilavorazioni stocastiche e vincoli di trasporto.

Le principali differenze con quanto già presente in letteratura sono le seguenti:

A differenza di studi che si concentrano sull'ottimizzazione dei parametri fisici (come la velocità del nastro o la dimensione dei buffer), questo lavoro dimostra il valore di un intervento puramente logico. Mentre la letteratura classica tende a considerare il layout come una variabile modificabile, la realtà industriale analizzata impone vincoli stringenti. L'analisi dettagliata di quattro regole di priorità nel punto di massima congestione fornisce un set di strategie operative immediatamente applicabili a linee CONWIP simili, tenendo come cardine le necessità aziendali reali, come macchinari già acquistati e processi che non possono variare nell'ordine e nella durata. Questo approccio riflette la visione di ElMaraghy (2019) sulla Smart Changeability, la quale suggerisce che la

flessibilità logica sia spesso più efficace e sostenibile di quella fisica in contesti produttivi consolidati.[4]

Un secondo elemento di confronto riguarda la mitigazione dell'overhead logistico. La tesi fornisce dati chiave sull'interazione tra la variabilità del rework e la gestione dello spazio sul nastro, con particolare attenzione al vincolo del "giro intero". Sebbene Fernandes et al. (2005) abbiano identificato il rework come un fattore di instabilità per il throughput CONWIP, questo studio va oltre la semplice diagnosi: i risultati indicano se una gestione dinamica delle priorità è sufficiente a compensare l'overhead logistico indotto dal sistema di ritorno, offrendo una soluzione operativa alla saturazione precoce dei pallet descritta in letteratura.[7]

Infine, il lavoro affronta una dimensione di complessità spesso semplificata nei modelli teorici: la gestione dei vettori fisici. In un contesto reale, è importante valutare che anche i pallet sono fondamentali per la movimentazione e, come tali, devono avere la possibilità di compiere lo stesso percorso dei job e di ritornare alla stazione di partenza. Questo rappresenta un limite oggettivo per le entità che possono viaggiare sul nastro e crea un'ulteriore dimensione di complessità che si allinea alle osservazioni di Alfieri, Castiglione e Pastore. Il confronto con i loro studi conferma che il pallet non è un'entità neutra, ma un vincolo di traffico che, se non correttamente gestito tramite le regole di priorità qui analizzate, può trasformare una linea bilanciata in un sistema soggetto a stalli critici.[12], [13]

La revisione della letteratura dunque ha fornito una solida base per la ricerca. È stato dimostrato che la Simulazione a Eventi Discreti è lo strumento analitico appropriato per la modellazione di sistemi dinamici. Questa metodologia è potenziata dall'architettura del Digital Twin, che permette di operare in tempo reale e supportare la schedulazione dinamica e interattiva. Algoritmi avanzati sono disponibili per eseguire l'ottimizzazione in questo contesto altamente reattivo, con l'obiettivo specifico di ridurre la deviazione di esecuzione (*scheduling execution deviation*).

Il quadro concettuale per questo lavoro è definito dall'intersezione di tre pilastri metodologici:

1. DES e Flusso Deterministico (D/G/1): la necessità di modellare con estrema fedeltà la schedulazione rigida impone l'uso della Simulazione a Eventi Discreti. Questo approccio è superiore all'analisi stocastica tradizionale, poiché la performance del sistema è misurata non sulla

gestione della casualità, ma sulla robustezza del determinismo a fronte di eventi esogeni.

2. Digital Twin e Prescrizione: il DT rappresenta il framework architetturale essenziale per chiudere il loop di feedback tra l'informazione fisica (dati di stato e deviazione) e l'azione virtuale/prescrittiva (nuova schedulazione). Il DT trasforma il modello DES da strumento di analisi what if a motore di decision making in tempo reale.
3. Ottimizzazione Dinamica e Interattiva: la funzione prescrittiva del DT richiede algoritmi in grado di eseguire la SBO in pochi minuti. La letteratura conferma che solo la schedulazione dinamica e interattiva che aggiorna continuamente i parametri di risorsa e ricalcola la soluzione ottimale può efficacemente prevenire la propagazione del ritardo, che è la principale minaccia per i sistemi schedulati.

3. Descrizione del Problema

3.1 Architettura e Layout del Sistema Logistico

Il sistema produttivo oggetto della presente analisi visibile in Figura 4, si configura come una linea a flusso controllato, governata dalla logica CONWIP (Constant Work-In-Process), la cui struttura è stata progettata per bilanciare l'efficienza produttiva con la flessibilità necessaria alla gestione dei rientri qualitativi.

Il layout si sviluppa attorno a un anello principale di movimentazione, denominato Conveyor Forward, che funge da canale principale per l'intero processo. Lungo questo percorso si incontrano sei workstation primarie, numerate da WS1 a WS6, a cui si aggiunge una stazione dedicata esclusivamente al controllo qualità finale, la WSQC.

Un elemento caratteristico di questa configurazione risiede nella differenziazione della capacità produttiva tra le stazioni, mentre la maggior parte delle unità opera in modalità monomacchina, le stazioni WS2 e WS5 sono state potenziate con una configurazione a due macchine identiche poste in parallelo.

Tale scelta non è casuale, ma risponde alla necessità di gestire carichi di lavoro superiori in punti critici del flusso: nella WS2 per assorbire la variabilità in ingresso e nella WS5 per gestire una specifica complessità operativa che richiede a ogni singolo job di subire due lavorazioni consecutive.

La movimentazione fisica dei componenti avviene tramite l'impiego di supporti standardizzati, i pallet, che circolano lungo il nastro trasportatore. Il sistema prevede una baia di carico iniziale dove il pezzo grezzo viene accoppiato al pallet e una baia di scarico finale dove avviene la separazione dell'entità prodotta dal suo supporto. Quest'ultimo, una volta svuotato, viene ricondotto automaticamente all'inizio della linea tramite un percorso di ritorno, garantendo così la chiusura del ciclo e il mantenimento costante del numero di unità in circolazione, come previsto dal paradigma CONWIP. L'integrazione tra flussi produttivi e flussi logistici dei pallet vuoti rappresenta una delle principali sfide di gestione, poiché la coesistenza di queste due entità sullo stesso convogliatore genera una fonte strutturale di congestione che deve essere regolata in maniera corretta secondo instradamenti ottimizzati.

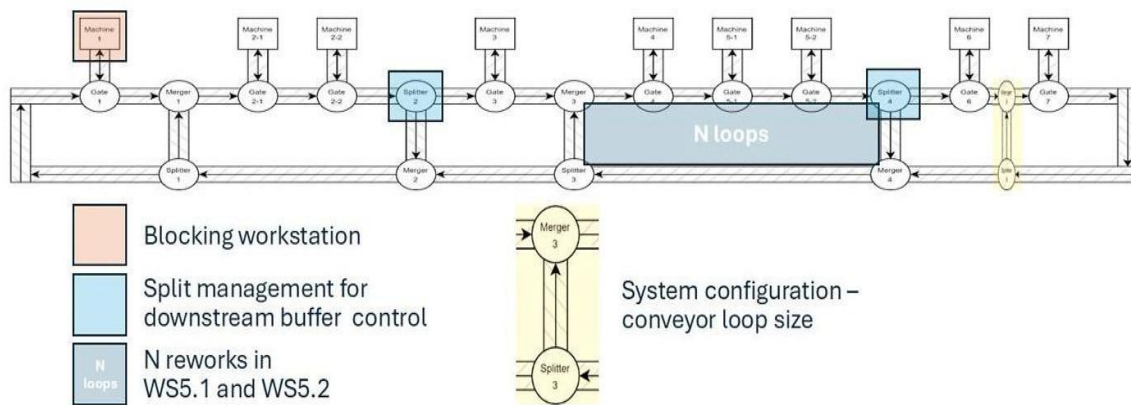


Figura 4: schema del Modello

3.2 Dinamiche del flusso e logiche operative

Il comportamento dinamico del sistema è regolato da una combinazione di vincoli fisici e regole di priorità che mirano a prevenire la saturazione delle risorse. Un meccanismo fondamentale in tal senso è rappresentato dal "blocking after service" applicato alla prima stazione. Secondo questa logica, un job che ha completato la lavorazione presso la WS1 non può liberare la risorsa finché la stazione successiva, la WS2, non presenta almeno una delle due macchine disponibile. Questo approccio, pur potendo generare ritardi localizzati, è essenziale per garantire la stabilità dell'intera linea, evitando che l'accumulo incontrollato di materiali tra le stazioni iniziali comprometta la fluidità dei passaggi successivi.

La gestione delle attese e dei colli di bottiglia è ulteriormente resa articolata dall'impiego di caroselli o loop di accumulo. Questi elementi, posizionati strategicamente in corrispondenza delle stazioni WS2 e WS5, permettono ai job di rimanere in circolazione dinamica qualora la risorsa di destinazione risulti occupata. Invece di una coda statica che saturerebbe lo spazio fisico del nastro principale e che porterebbe al blocco del nastro, i pezzi continuano a ruotare attorno alla workstation.

Il percorso dei job lungo il sistema è governato da nodi di smistamento (splitter) e di confluenza (merger). Gli splitter distribuiscono i carichi di lavoro tra le macchine parallele o indirizzano i pezzi verso i loop di attesa, mentre i merger riuniscono i flussi per garantire che la sequenza temporale e logica delle operazioni venga rispettata. In questo contesto, la WS5 emerge come il vero cuore critico del sistema: il requisito del doppio passaggio di lavorazione

trasforma intrinsecamente questa stazione in un collo di bottiglia potenziale, dove la gestione delle priorità tra chi deve effettuare il primo ciclo e chi il secondo diventa determinante per il throughput finale dell'intero impianto.

3.1.2 Flusso delle Entità e Vincolo di Congestione Logistica

Uno dei temi fondamentali del problema in esame è quello di utilizzare una logica CONWIP, genera due flussi di entità che competono per l'uso del nastro trasportatore:

1. Entità Produttive (Job in Lavorazione)

Rappresentano l'oggetto primario di studio (il prodotto sul pallet). Il loro obiettivo è completare l'intera sequenza di lavorazioni e uscire dal sistema. La metrica di successo è la massimizzazione del throughput (TH) di queste entità. Sono le entità a cui si applicano le regole di priorità dinamiche testate nel Loop WS5. Il Digital Twin è progettato esclusivamente per ottimizzare la gestione di questo traffico.

2. Entità Logistiche (Pallet Vuoti)

Rappresentano il supporto fisico senza il prodotto. Queste entità sono necessarie per mantenere il vincolo CONWIP: un nuovo job produttivo può entrare nel sistema solo quando un pallet vuoto ritorna al punto di carico. Queste entità costituiscono il traffico logistico non produttivo. Il loro transito è un overhead logistico obbligato che occupa spazio e tempo prezioso sul convogliatore, agendo come una fonte strutturale di congestione. Sebbene non siano ottimizzate direttamente dal DT (che si concentra sul loop WS5), la loro presenza costante deve essere considerata.

3.2 Parametri e Stocasticità del Modello

3.2.1 Tempi di processo e distribuzioni

Tutte le WS hanno un tempo di processo pari a 20 minuti, ad eccezione della WS2, che ha una durata pari a 40 minuti, infatti sono state inserite due macchine per la stessa lavorazione.

Workstation	Numero di Macchine	Durata (min)	Logica di Lavoro	Note
WS1	1	20	Singolo passaggio	Punto di carico/ingresso
WS2	2	40	Singolo passaggio	Loop dedicato per l'attesa
WS3	1	20	Singolo passaggio	-
WS4	1	20	Singolo passaggio	-
WS5	2	20	Doppio passaggio	2 cicli/job
WS6	1	20	Singolo passaggio	Ultima fase prima del QC
WSQC	1	20	Ispezione	Scelta se il job verrà reintrodotta nel sistema o scartato definitivamente

Tabella 1: Durata delle lavorazioni

La velocità dei conveyor è stata impostata a 3 m/s e di conseguenza ha un impatto seppur minimo sul determinismo e sulla conseguente variabilità del sistema, mentre è più rilevante per quanto riguarda il Flow Time dei job.

3.2.2 Dinamica Rework e variabilità

Ogni workstation, ad eccezione della prima, presenta una probabilità, fissata pari al 12% indicata come α , che il processo di lavorazione non risulti perfettamente conforme ai requisiti. In tali circostanze, il sistema attiva una procedura di re-routing verso la stazione di controllo qualità (WSQC). Una volta che il job raggiunge la WSQC, viene sottoposto a un'ispezione rigorosa che può dare origine a tre esiti differenti, ciascuno con un impatto diverso sulle performance del sistema.

Il primo scenario prevede che il job venga considerato rilavorabile; in questo caso, esso viene rimandato alla workstation di provenienza per una fase aggiuntiva di correzione. Questo genera un carico extra non solo per la macchina coinvolta, ma per l'intero sistema logistico, poiché il job deve compiere un ulteriore giro completo del nastro trasportatore, occupando spazio e risorse pallet. Il secondo scenario è quello della conformità, in cui il pezzo viene rilavorato e poi reinviato verso la stazione successiva rispetto a quella da cui è partito. Infine, esiste la possibilità di scarto definitivo qualora vengano riscontrati difetti critici non riparabili.

Questi tre scenari sono identificati da 3 priorità distinte:

- Con probabilità γ il job viene rimandato alla workstation di provenienza per una rilavorazione aggiuntiva;
- Con probabilità δ il job viene inviato alla workstation successiva, considerandolo conforme;
- Con probabilità $1 - \gamma - \delta$ il job viene scartato a causa di difetti critici.

3.3 Identificazione e definizione delle regole di priorità

La WS5, avendo due macchine dovrebbe avere la possibilità di lavorare un numero maggiore di job contemporaneamente, ma ricordiamo anche che è previsto che tutti i job facciano due passaggi in questa WS e questo rallenta di molto il flusso. Quando poi nessuna delle due stazioni è disponibile, il job rimane nel loop e continua a girare intorno fino a quando una delle due stazioni della WS5 non è disponibile per la lavorazione.

3.3.1 Variabili di controllo: regole di priorità del DT

Quando si parla di DT, si intendono 4 repliche virtuali semplificate della linea che vengono eseguite contemporaneamente ogni due ore per 6 minuti, che si distinguono tra loro per le regole di priorità implementate, e che recepiscono i dati del sistema principale.

Viene poi effettuata da parte del DT un'analisi interna al modello che definisce qual è stato in termini di TH il migliore dei modelli, che è poi quello che viene eseguito almeno fino alle due ore successive, ovvero fino a quando il modello non viene nuovamente verificato, il sistema confronta i risultati in termini di

throughput (TH) e di riduzione del carico residuo, applicando al modello reale la prima logica.

Andando a definire più nel dettaglio la struttura dei sotto modelli, fin da subito è importante ricordare che le 4 repliche del modello interne al DT lavorano in parallelo. Qui sotto sono state definite le regole di priorità per interpretarne meglio i risultati:

- DT1: priorità Macchina 5.1 a chi non ha ancora lavorato, priorità Macchina 5.2 a chi ha già lavorato.
- DT2: priorità di entrambe le macchine a chi ha già lavorato.
- DT3: priorità Macchina 5.1 a chi ha già lavorato, priorità Macchina 5.2 a chi non ha ancora lavorato.
- DT4: priorità di entrambe le macchine a chi non ha ancora lavorato.

Il DT è solamente una rappresentazione virtuale di quello che accadrebbe nel modello se si scegliesse quella regola di priorità, per questo motivo anche le durate delle workstation sono state ridotte di un fattore pari a 1/20 del tempo originale, così come la durata della simulazione del Digital Twin è di soli 6 minuti, che hanno lo scopo di replicare il sistema durante due ore di funzionamento.

Per poter misurare il reale effetto del DT sul modello, sono state definite differenti metriche e KPI che definissero in maniera univoca il DT più significativo.

Poiché la durata della simulazione dei modelli è stata impostata come identica, il criterio di valutazione si è basato sul numero di job che riuscivano a completare l'intero processo.

L'obiettivo finale resta la possibilità di ridurre al minimo il tempo che i job passano all'interno del sistema e che la lavorazione preveda meno tempi di attesa e di transito possibili.

3.4 Vincoli del sistema e Implicazioni modellistiche

Al fine di garantire la massima focalizzazione sulle dinamiche di flusso e sull'efficacia delle logiche di priorità, il modello è stato costruito sulla base di alcune assunzioni semplificative essenziali. In primo luogo, i tempi accessori di movimentazione, quali le fasi di discesa dal pallet, l'ingresso e l'uscita dalle macchine e il riposizionamento sul pallet al termine del ciclo, sono stati trascurati. Si è ipotizzato, infatti, che il loro impatto temporale sia marginale rispetto ai tempi

di lavorazione effettivi e che la loro esclusione non alteri la validità dei risultati complessivi.

Per quanto riguarda l'operatività delle workstation, il modello presuppone che ogni risorsa lavori ininterrottamente, senza considerare fermate per manutenzioni pianificate, guasti improvvisi o cambi di setup tra diversi job. Questa astrazione permette di rappresentare un flusso produttivo uniforme e continuo, isolando l'analisi dalle variabili accidentali. Allo stesso modo, la stazione di controllo qualità è stata modellata come un'unica unità funzionale regolata da logica FIFO e anch'essa prevede che i job entrino in loop se una volta arrivati in prossimità della risorsa questa risulta occupata. Non è stata introdotta alcuna distinzione tra le diverse tipologie di difetto o i livelli di gravità; il sistema si basa su una rappresentazione aggregata in cui le probabilità di rilavorazione (re-routing) sono considerate note e costanti nel tempo.

L'analisi è stata condotta considerando un turno produttivo di 24 ore. Tale scelta riflette fedelmente il contesto operativo delle grandi realtà industriali, dove il ciclo a ciclo continuo rappresenta la configurazione standard per ottimizzare la saturazione degli impianti e stabilizzare i flussi. Operare su un arco temporale completo permette di osservare il sistema in condizioni di regime, superando le oscillazioni iniziali e garantendo una valutazione solida delle performance.

Infine, è importante sottolineare che il modello non include variabilità esterne quali assenze del personale, ritardi nelle forniture o fluttuazioni della domanda. L'intero sistema viene analizzato in condizioni di regime controllato, permettendo uno studio estremamente dettagliato e rigoroso degli effetti prodotti dall'applicazione del Digital Twin ad un contesto che, per come viene posto, potrebbe definirsi altamente robotizzato ed autonomo.

4. Metodologia dello studio

Il modello di simulazione è stato progettato per validare l'efficacia di un Digital Twin (DT) integrato nell'ambiente Arena, concepito come un modulo decisionale autonomo rispetto alla simulazione principale. L'obiettivo primario del DT è identificare, tra diverse regole di priorità, quella più performante in termini di ottimizzazione del throughput.

Il coordinamento tra i due sistemi è gestito da un meccanismo di controllo che, con una frequenza di 120 minuti, genera un'entità deputata all'attivazione del Digital Twin; da questo momento, i job che transitano attraverso i checkpoint predefiniti vengono reindirizzati nel modello parallelo. Dopo un intervallo di 6 minuti, una seconda entità provvede alla disattivazione del DT, innescando l'analisi dei KPI per determinare la configurazione ottimale.

All'interno del Digital Twin, le entità sono sottoposte alla medesima sequenza di lavorazioni e controlli prevista nel modello principale, ma con una velocità di esecuzione accelerata: questa sfasatura temporale consente di anticipare gli scenari futuri e definire preventivamente la strategia operativa da implementare nelle due ore di funzionamento successive. Per garantire la massima efficienza computazionale e snellezza del DT, il flusso dei job in ingresso è stato opportunamente filtrato, escludendo i componenti soggetti a rilavorazioni derivanti da workstation precedenti o reintrodotti nei loop 4/5 per mancata disponibilità della Macchina 6. Tale semplificazione permette di evitare ritardi superflui su entità non determinanti per l'analisi. Infine, la gestione del sistema si avvale di una logica CONWIP basata su un numero fisso di pallet; questa scelta non costituisce un vincolo limitante, bensì rappresenta una strategia deliberata per stabilizzare la congestione del sistema, minimizzando il rischio di saturazione dei nastri trasportatori attraverso la limitazione del Work In Process attivo sulla linea.

4.1 Modellazione dei flussi post-ispezione e bilanciamento della capacità

Un elemento di fondamentale complessità del modello è rappresentato dalla gestione delle non conformità presso la stazione di controllo qualità (WSQC). Al termine dell'ispezione, il flusso produttivo si scinde in tre percorsi probabilistici

che riflettono la realtà dei processi industriali moderni, dove l'errore umano o meccanico deve essere previsto e gestito in modo sistematico:

- **Rilavorazione:** Con una probabilità $\gamma = 5\%$, il job viene identificato come difettoso ma recuperabile. In questo caso, il pezzo viene reindirizzato alla workstation di provenienza per una ripetizione parziale o totale delle operazioni. Questo feedback genera un carico di lavoro aggiuntivo che deve essere assorbito dalle risorse senza compromettere il throughput globale.
- **Idoneità:** Con una probabilità $\delta = 55\%$, il job supera con successo i test qualitativi e viene diretto alla stazione successiva o verso l'uscita del sistema, rappresentando il flusso ideale del valore.
- **Scarto:** Con la probabilità residua $1 - \gamma - \delta = 10\%$, il job presenta difetti critici non emendabili e viene rimosso definitivamente dal sistema. Sebbene elevata, questa quota è funzionale all'analisi della resilienza del sistema in condizioni di stress.

L'intero dimensionamento del sistema ruota attorno all'identificazione della risorsa critica, ovvero il collo di bottiglia, individuato con certezza nella Workstation 5. La WS5 rappresenta il punto di massima pressione poiché ogni unità di prodotto richiede due lavorazioni complete presso questa stazione. Nonostante la presenza di due macchine identiche in serie (WS5.1 e WS5.2), ciascuna lavorazione può essere eseguita indifferentemente su una delle due con probabilità del 50%. Il tempo medio di lavorazione associato all'intera fase WS5 risulta quindi pari a:

$$t_{WS5} = 40 \text{ minuti.}$$

Considerando la presenza di due risorse, il tasso di servizio complessivo della fase WS5 è dato da:

$$\mu_{WS5} = \frac{n \text{ macchine}}{t_{WS5}} = \frac{2}{40 \text{ min}} = 3 \text{ job/h} = TH_{target}$$

Questo valore rappresenta il limite teorico del throughput dell'intero impianto. Al di sopra di tale soglia, un aumento forzato del WIP non si tradurrebbe in un incremento della produttività, ma in un progressivo aumento dei tempi medi di attraversamento e del rischio di saturazione dei buffer.

4.2 Dimensionamento del limite CONWIP tramite la Legge di Little

Per definire correttamente il numero di pallet necessari si è seguito un ragionamento rigoroso basato sulla quantificazione del tempo medio di attraversamento minimo del job e sulla successiva applicazione della legge di Little. In primo luogo è stato valutato che il job attraversa workstation lineari come WS1, WS3, WS4 e WS6 con un tempo di processing pari a 20 minuti ciascuna:

$$T_{base} = T_{WS1} + T_{WS3} + T_{WS4} + T_{WS6} = 4 \times 20 = 80 \text{ min.}$$

La stazione WS2 è composta da due macchine identiche in parallelo entrambe con un tempo di lavorazione di 20 minuti. Poiché le due risorse operano contemporaneamente, il tempo medio equivalente per ogni singolo job si dimezza:

$$T_{WS2} = 20/2 = 10 \text{ min}$$

Aggiungendo il contributo del collo di bottiglia WS5

$$T_{WS5} = 40 \text{ min.}$$

si ottiene il tempo di processamento netto atteso:

$$T_{proc} = 80 + 10 + 40 = 130 \text{ min.}$$

A questo valore deve essere sommato il contributo statistico della stazione di controllo qualità. L'ispezione avviene con probabilità $\alpha = 12\%$ e richiede esattamente 20 minuti. Se l'esito è una rilavorazione $\gamma = 55\%$, si assume che essa aggiunga un'operazione supplementare da 20 minuti, altrimenti il job può

essere scartato e di conseguenza risulterà inutile ai fini statistici per definire il TH del sistema. Il valore medio del Flow Time è quindi:

$$\begin{aligned} T_{QC_atteso} &= \gamma \times (20 + \gamma \times 20) = 0,12 \times (20 + 0,55 \times 20) = \\ &= 0,12 \times (20 + 11) = \\ &= 3,72 \text{ min.} \end{aligned}$$

Infine, si stima un tempo medio totale dedicato ai movimenti logistici (nastri trasportatori e caroselli) pari a

$$T_{mov} = 28,33 \text{ min.}$$

Sommando tutti i contributi, si ricava il Flow Time minimo:

$$FT_{min} = T_{proc} + T_{QC_atteso} + T_{mov} = 130 + 3,72 + 28,33 = 162,05 \text{ min} \approx 2,7 \text{ h}$$

Questo valore rappresenta il limite inferiore del tempo di attraversamento, escludendo i tempi di attesa in coda. Per includere un margine di sicurezza che consenta code fisiologiche e assorba la variabilità stocastica, si assume un Flow Time desiderato

$$FT_{desired} = 4,2 \text{ h} = 252 \text{ min.}$$

Usando la legge di Little

$$WIP = TH \times FT$$

si ottiene il limite CONWIP ottimale per l'avvio della simulazione:

$$W^* = TH_{target} \times FT_{desired} = 3 \text{ job/h} \times 4,2 \text{ h} \approx 13 \text{ job.}$$

Questo limite di 13 unità garantisce teoricamente che il sistema lavori alla massima efficienza possibile senza incorrere in fenomeni di congestione che degraderebbero le prestazioni complessive, tra gli obiettivi della nostra analisi vi è proprio l'indagine di quanto questo valore si discosti dalla condizione ottimale del sistema.

4.3 Impostazioni sperimentali

Il sistema produttivo esaminato è stato studiato adottando la logica della simulazione all'equilibrio, una configurazione tipica dei processi che operano a ciclo continuo e senza interruzioni programmate su archi temporali molto estesi.

Questo approccio metodologico permette di valutare il comportamento del modello in regime stazionario, ovvero una condizione di stabilità dinamica in cui le metriche fondamentali, come la capacità produttiva, il volume dei materiali in lavorazione e i tempi medi di permanenza, risultano statisticamente costanti nel tempo.

L'obiettivo primario di una simulazione in stato stazionario è infatti l'analisi delle prestazioni ordinarie e permanenti del sistema, neutralizzando le anomalie generate dalle condizioni di avvio che caratterizzano la fase iniziale di start-up. Per assicurare che le misurazioni riflettano fedelmente la realtà operativa, è indispensabile scartare i valori relativi al periodo transitorio, durante il quale il sistema si evolve dallo stato iniziale fino al raggiungimento del punto di equilibrio.

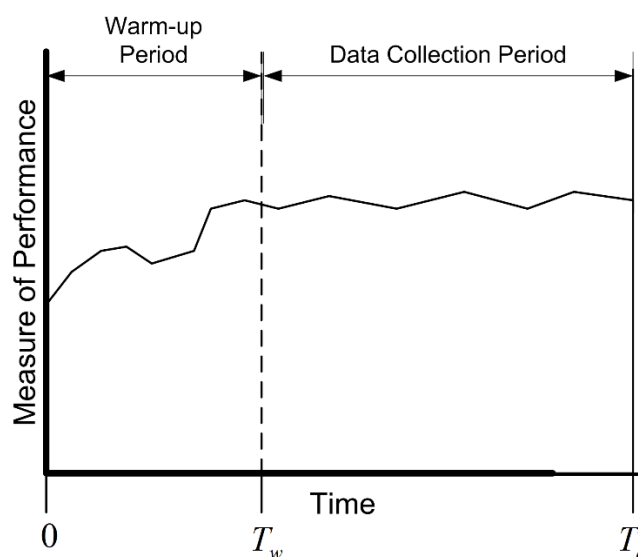


Figura 5: warm Up e Data Collection Period

Questo intervallo critico è definito tempo di warm-up e la sua evoluzione è chiaramente distinguibile nella Figura 5, che illustra il passaggio cruciale tra l'instabilità della fase di riscaldamento e la regolarità del regime stazionario. La

determinazione di un tempo di warm up appropriato è essenziale per isolare i dati rappresentativi dell'equilibrio, evitando che le fluttuazioni iniziali non significative vadano a inquinare la validità statistica dei risultati finali. Per quanto riguarda la durata complessiva della simulazione, essa non deriva da calcoli precostituiti ma rappresenta una decisione progettuale legata alla natura specifica del caso studio; nella fattispecie, è stato stabilito un orizzonte di 182,5 giorni totali con un periodo di warm up fissato a 182,5 giorni. Tale impostazione, che destina egual misura del tempo simulato alla stabilizzazione del sistema, costituisce una scelta prudentiale volta a garantire il completo superamento di ogni fenomeno di transizione e a permettere la raccolta di campioni statistici estremamente solidi in condizioni di stabilità accertata. In ultima analisi, per conferire piena significatività scientifica allo studio, alla gestione del warm-up deve affiancarsi una stima rigorosa del numero di repliche da eseguire, parametro fondamentale per attestare l'affidabilità e la precisione dei risultati ottenuti attraverso il modello stocastico.

4.4 Definizione del numero di repliche

Al fine di garantire la robustezza statistica dei risultati e minimizzare l'impatto della componente stocastica del modello, si è proceduto alla determinazione del numero ottimale di repliche basandosi sulla variabile del throughput totale annuo. La teoria della simulazione prevede che, per stimare un parametro prestazionale con una precisione prefissata, sia necessario eseguire un numero di prove indipendenti tale da contenere l'errore di stima entro un intervallo di confidenza noto.

Inizialmente, è stato condotto un set di $n_0 = 5$ repliche pilota, i cui risultati in termini di pezzi totali prodotti sono stati: 10252, 10222, 10273, 10250 e 10223. Da questi valori è stata calcolata una media campionaria $\bar{x} = 10244$ pezzi e uno scarto quadratico medio campionario $s^2 = 466,5$ e la deviazione standard pari a $s = 21,60$.

L'obiettivo prefissato è contenere l'errore di stima entro lo 0,5% della media campionaria, corrispondente a circa 51,22 pezzi. Utilizzando la distribuzione T di Student con $n_0 - 1 = 4$ gradi di libertà per un livello di confidenza del 95% ($\alpha = 5\%$), il valore critico individuato è $t(4;0,975) = 2,776$. Il numero di repliche necessario n è stato quindi ricavato mediante la seguente relazione:

$$n \geq \left(\frac{t_{n_0-1, 1-\alpha/2} \cdot s}{E} \right)^2$$

Sostituendo i valori ottenuti:

$$n^* = \left(\frac{2,776 \cdot 21,60}{51,22} \right)^2 = (1,1706)^2 \approx \mathbf{1,37}$$

L'esito del calcolo indica che il numero di repliche iniziali ($n_0 = 5$) è ampiamente sufficiente per garantire la precisione richiesta, poiché la variabilità riscontrata tra le diverse corse del modello è estremamente ridotta rispetto alla soglia di errore del 0,5%. Pertanto, le 5 repliche eseguite permettono già di ottenere stime statisticamente significative e affidabili per l'analisi delle prestazioni in regime stazionario.

Quanto è stato definito però, è stato di condurre comunque 15 repliche, in modo tale da assicurarsi che la variabilità non si amplificasse e per poter raccogliere informazioni statistiche rilevanti come varianza e deviazione standard, che il software di Rockwell non riuscirebbe a produrre con un numero minimo di repliche.

4.5 Key Performance Indicators

Per garantire un esame accurato e profondo dell'impatto derivante dalle strategie di gestione del materiale in lavorazione e dalle logiche di sequenziamento adottate, è stato selezionato un pacchetto di indicatori di prestazione chiave (KPI).

Tali parametri sono stati individuati per restituire una prospettiva equilibrata e multi-prospettica del rendimento totale della struttura, agendo come criteri di confronto oggettivi per i diversi scenari analizzati. La scelta di queste metriche permette di coprire simultaneamente la dimensione produttiva, le tempistiche di risposta e la solidità dei processi. Il quadro valutativo si focalizza quindi su variabili in grado di restituire una fotografia fedele dell'efficacia operativa e della dinamicità del flusso.

Il throughput (TH) è stato identificato come la variabile cardine per quantificare la produttività, poiché esprime la cadenza con cui i lavori ultimati abbandonano il processo nel tempo. Tale metrica rivela l'attitudine del sistema nel trasformare i materiali in lavorazione in prodotti finiti, rappresentando il traguardo principale della logica di gestione CONWIP, mirata proprio all'ottimizzazione del rendimento globale.

A completamento dell'analisi è stato introdotto l'Average Flow Time, il quale traccia la durata media della permanenza di un'entità all'interno del sistema, dal rilascio iniziale fino al completamento. Questo indice, operando in sinergia con il throughput secondo i dettami della Legge di Little permette di misurare accuratamente la tempestività dei cicli e la reattività della struttura. L'impiego combinato di tali indicatori garantisce un'indagine approfondita che intreccia i volumi produttivi con la fluidità del transito, assicurando un equilibrio tra la massima capacità di output e la stabilità dei processi interni.

Per completare l'analisi dei criteri di valutazione, è necessario introdurre due indicatori statistici e operativi fondamentali per la validazione delle politiche di sequenziamento:

Il primo è il Most Chosen Model, uno strumento statistico utilizzato per determinare con quale frequenza una specifica regola di priorità sovraperforma le altre nelle diverse repliche. Nel presente studio, è stata fissata una soglia di confidenza del 95%: il raggiungimento di tale valore indicherebbe che una determinata regola non è solo marginalmente migliore, ma garantisce una superiorità sistematica. Tale evidenza aprirebbe la strada a una revisione delle strategie aziendali, suggerendo il passaggio dal tradizionale approccio FIFO a una scelta statica differente, basata su criteri di efficienza più rigorosi.

Il secondo indicatore è il Remaining Passages, il quale offre una prospettiva complementare al Most Chosen Model focalizzandosi sul carico di lavoro residuo nel sistema. Questo parametro è particolarmente critico per l'analisi della Workstation 5 (WS5), che nel modello rappresenta il punto focale a causa della doppia lavorazione richiesta. L'indicatore permette di distinguere se un eventuale miglioramento delle performance sia dovuto a un reale incremento dei pezzi lavorati o, piuttosto, a una distribuzione più fluida e razionale dei job tra le stazioni. Monitorare i passaggi rimanenti è essenziale per identificare inefficienze nascoste: una riduzione di questo valore conferma che la regola di priorità

adottata sta effettivamente snellendo il processo, evitando che i pezzi ristagnino nei loop di rilavorazione della WS5.

4.6 Scenari simulativi

Scenario 1: Analisi di Sensibilità del Parco Pallet (baseline con DT)

In questa prima fase, il sistema opera sotto la supervisione del Digital Twin attivo. L'obiettivo è analizzare la risposta del throughput (TH) al variare del numero di pallet circolanti, partendo da un valore teorico iniziale. Questo scenario funge da analisi di sensibilità: si osservano le oscillazioni del TH in risposta alle modifiche del WIP (Work-In-Process) e si verifica la convergenza statistica del modello al termine delle repliche prefissate, validando la stabilità del sistema in modalità dinamica.

Scenario 2: Configurazione Statica e Benchmark (DT Off)

Il secondo scenario è volto alla definizione di un benchmark prestazionale. Viene utilizzato il numero di pallet identificato come "ottimale" dal Digital Twin nello scenario precedente, ma operando con il DT disattivato. In questa configurazione, il sistema non riceve input correttivi e adotta una logica di sequenziamento standard FIFO (First-In-First-Out). I risultati ottenuti in termini di throughput e tempo medio nel sistema (Flow Time) rappresentano il termine di paragone per quantificare il reale valore aggiunto dalle logiche di ottimizzazione.

Scenario 3: Ottimizzazione Dinamica e Performance del Digital Twin (DT On)

L'ultimo scenario rappresenta la configurazione a regime: il sistema utilizza il numero ottimale di pallet con il Digital Twin pienamente operativo. L'analisi è focalizzata sulla verifica del miglioramento dei KPI rispetto al benchmark dello Scenario 2, monitorando in particolare l'incremento del throughput e la contrazione dei tempi di attraversamento. Inoltre, viene analizzato il comportamento intrinseco del Digital Twin, misurandone l'occorrenza di ogni ramo e il numero medio di job processati e rilasciati con successo dalla logica prescrittiva.

5. Descrizione degli esperimenti

L'analisi si sviluppa attraverso il confronto di tre diverse configurazioni d'impianto, distinte in base alle regole di priorità implementate nella workstation critica soggetta a rilavorazioni. L'obiettivo primario è valutare l'impatto delle politiche operative mediante l'analisi di indicatori chiave (KPI) quali il throughput (TH), il tempo medio di attraversamento (Flow Time), la probabilità di deadlock e il volume produttivo totale, un focus specifico sui passaggi eseguiti tra le due unità della stazione WS5.

5.1 Scelta del numero di pallet e del WIP del sistema

Per calibrare correttamente il modello, è stata condotta una campagna sperimentale variando la dotazione di pallet tra 13 e 50 unità, in particolare, questa scelta coincide con una limitazione in termini di WIP del sistema ed ha un importante impatto sull'esecuzione degli esperimenti, per questo è stata considerata prioritaria.

L'instabilità più grave emersa durante i test con un numero eccessivo di pallet è stata l'insorgenza di deadlock (stalli logici). In queste configurazioni, il sistema risultava talmente congestionato da impedire il completamento delle quindici repliche statistiche previste, bloccandosi sistematicamente (vedi Figura 6). Tale fenomeno è dovuto alla saturazione completa dei buffer: in presenza di loop di rilavorazione, i pallet finiscono per occupare ogni posizione disponibile, impedendo reciprocamente il movimento dei pezzi e paralizzando l'intero flusso. La scelta di fissare la dotazione a 29 pallet è stata quindi dettata dalla necessità di bilanciare l'output con la resilienza operativa. Questa configurazione è l'unica che ha garantito la continuità del processo, permettendo alle macchine di lavorare a ritmi sostenibili ed evitando che il sistema operasse costantemente sull'orlo del collasso funzionale e con un utilizzo ottimale considerando la configurazione del sistema, seppur discretamente alto.

È stata dunque condotta un'analisi che ha correlato il numero di pallet inseriti all'insorgenza di deadlock e dunque al numero di repliche complete che il modello riusciva ad eseguire prima di andare in blocco, che ha portato ai seguenti risultati:

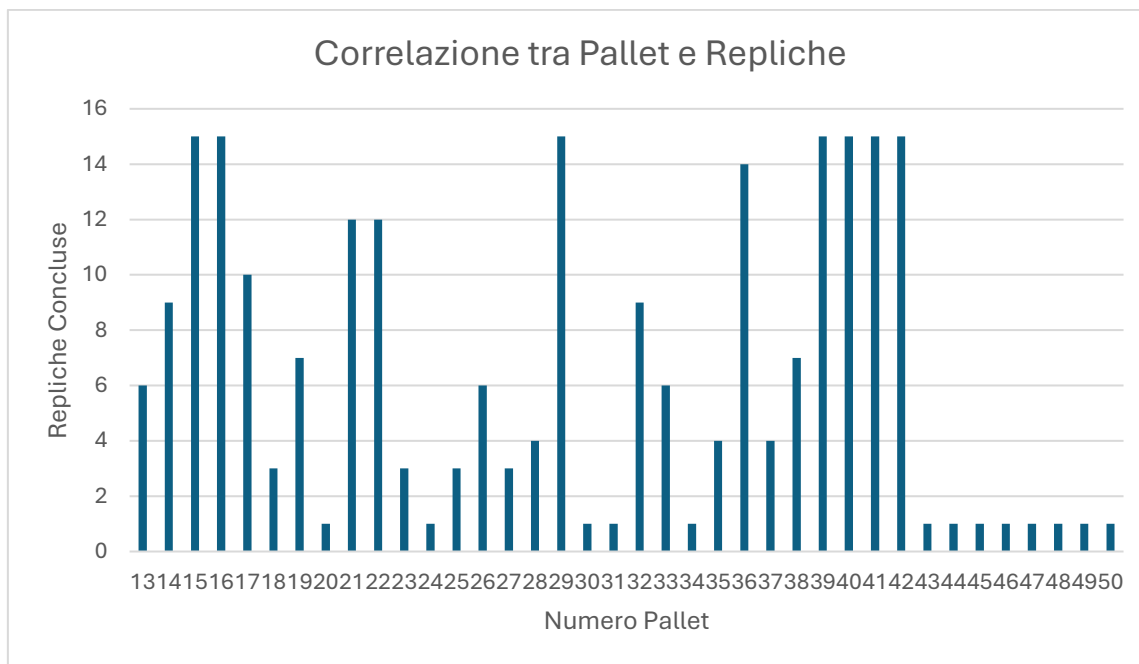


Figura 6: Correlazione Numero di Pallet e Numero di Repliche Concluse

In particolare, un aspetto molto importante è stato notare come l'utilizzo di alcune risorse, man mano che il numero di pallet aumentava, raggiungevano valori così alti da sfiorare l'unità e dunque, siccome quasi tutte le WS prevedono una singola macchina, creando una forte situazione di instabilità e rischio di Bottleneck.

Si può osservare come non si configuri un vero e proprio trend tra il numero di pallet e le repliche concluse e il pattern risulta piuttosto causale. Un segnale più netto deriva invece dai WIP superiori a 43, infatti è visibile come il sistema non sia più in grado di gestire tutti questi job contemporaneamente in lavorazione.

Nella definizione del WIP ottimale per il sistema, un'ulteriore analisi ha riguardato l'influenza che il numero di pallet ha sul FT medio dei job, sottolineando come un numero di supporti eccessivo non porti migliorie significative in termini di rapida fuoriuscita di pezzi, bensì anzi, crei ingorgo e renda difficile la finalizzazione del processo e aumenti di più del 130% il FT che i job avevano nel caso di un numero di pallet più contenuto.

Nella Figura 7 e 8 sono state riportate soltanto le casistiche che prevedevano il completamento delle 15 repliche e per rendere il campione più omogeneo per cui il report di SIMAN veniva correttamente emesso.

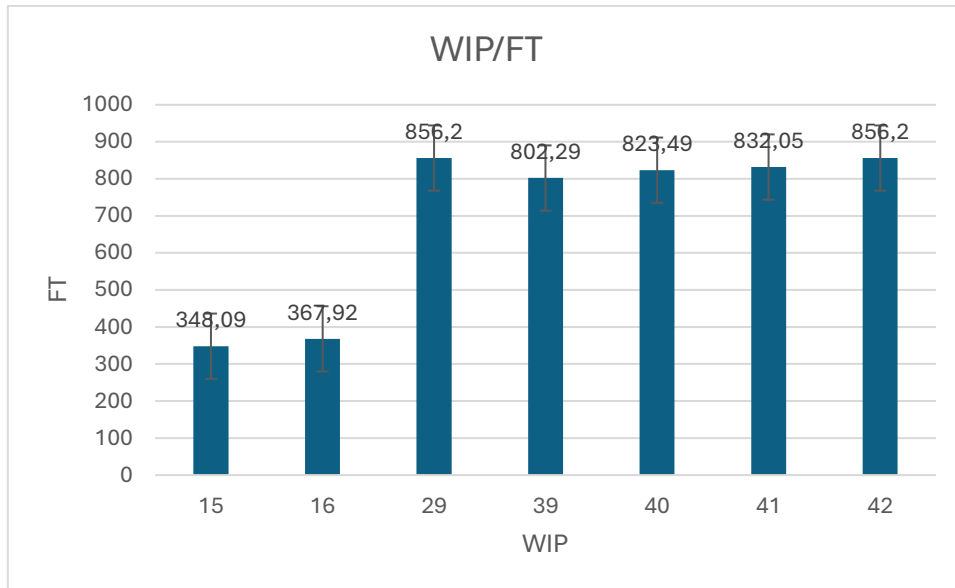


Figura 7: Rapporto WIP e FT medio

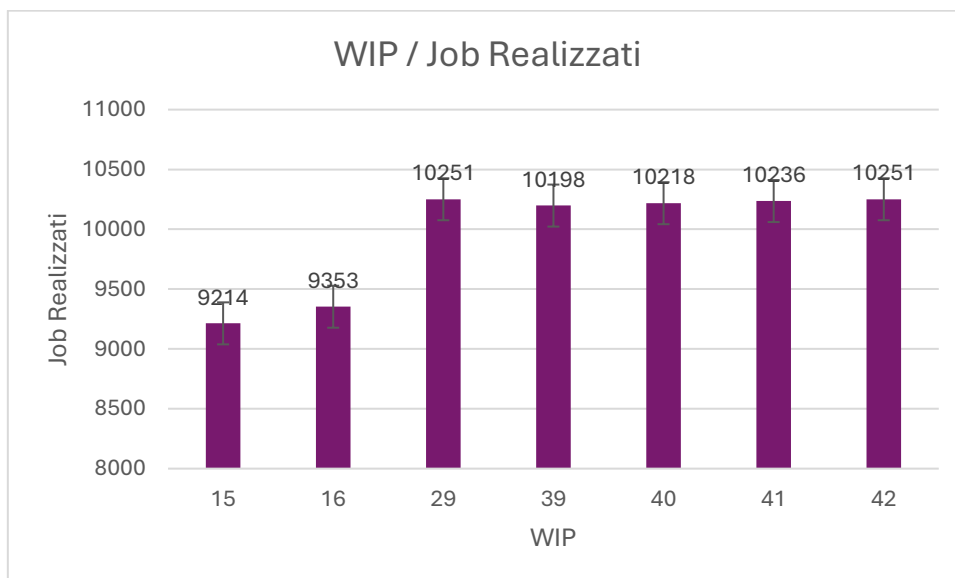


Figura 8: Rapporto WIP e Numero di Job Realizzati

E' importante sottolineare che in sistemi così complessi, la enorme differenza tra il FT atteso e quello reale è ciò che rende fondamentale la parte simulativa, infatti tramite i calcoli teorici sarebbe stato impossibile rilevare gli effetti del sovraccollamento della linea.

L'andamento del FT e del numero di WIP realizzati risulta essere il medesimo e questo risulta in linea con quanto atteso, infatti un numero di pallet molto piccolo

lascia che le macchine trascorrono molto tempo in stato Idle e non massimizza il loro utilizzo, prediligendo però una fuoriuscita più rapida dei job, e contestualmente, poiché le macchine non sono al massimo della loro produttività si lascia molto margine ad un aumento del throughput, che infatti si verifica aumentando il WIP.

In termini di TH, la scelta dei 29 pallet ha portato ad una produzione di +898 pezzi nell'arco della simulazione e quindi si è passati da un TH pari a 2,14 job/h a quello definitivo di 2,34 job/h, con un incremento di 9,34 punti percentuali.

Da questo momento in poi, in tutte le analisi successive è stato assunto come valore target del WIP quello di 29 pallet, definito come migliore per il bilanciamento tra gli utilizzi, le entità prodotte e il flow time medio delle entità.

5.2 Confronto tra i modelli con e senza DT

Passando ora all'analisi del punto cruciale del documento, si è deciso di valutare le prestazioni del modello con DT, confrontandolo con un sistema parallelo privato di questo controllo dinamico

La Figura 9 riporta il numero di entità correttamente lavorate che il sistema è riuscito a processare avvalendosi delle decisioni intelligenti del DT nel sistema.

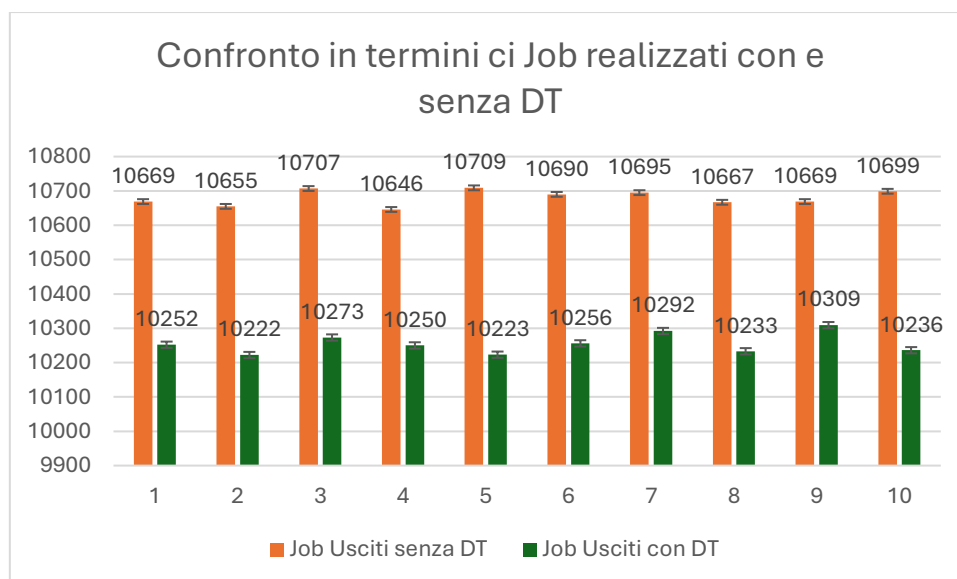


Figura 9: Confronto numero Job realizzati con e senza DT per replica

Come si può osservare dal grafico, in tutte le repliche il modello senza DT ha performato in modo più efficiente e ha prodotto circa 400 unità in più sull'intero orizzonte simulativo, pari a 6 mesi.

Di conseguenza il TH passa da circa 2,45 job/h nel modello con solo FIFO a 2,34 job/h nel modello ibrido con il DT, questo è dovuto in parte alla possibilità che le regole non siano tanto efficienti quanto il FIFO, ma anche al fatto che risulta estremamente complicato gestire i due processi in contemporanea, infatti quando il DT lavora, tutto il resto del sistema dovrebbe "congelarsi" e riprendere l'esecuzione dopo che la decisione della miglior regola di priorità da implementare è stata presa. Questo con i tools a nostra disposizione risulta essere estremamente complesso e non realizzabile, di conseguenza si accetta di avere un'approssimazione del risultato.

Dal punto di vista teorico, l'adozione di diverse regole di priorità influisce direttamente sulla gestione delle code e sulle prestazioni complessive, poiché determina l'ordine con cui i job vengono processati, modificando il tempo medio di attesa.

Per quanto riguarda invece il FT, nel caso del sistema statico regolato dal FIFO, risulta essere in media di 526 minuti, contro i 857 minuti del DT, il dettaglio dei confronti nei vari casi è riportato nella Figura 10.

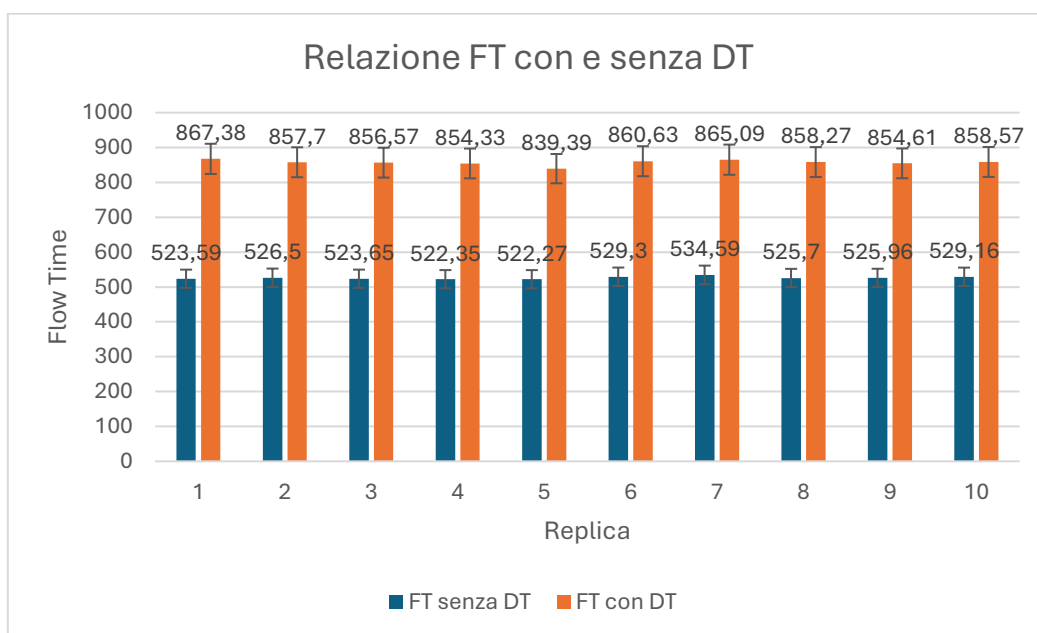


Figura 10 Relazione tra il Flow Time nella simulazione con DT e senza DT

Lo scenario di riferimento è basato sulla regola FIFO, ovvero il primo elemento che entra nella coda è anche il primo a essere servito. Questa politica rappresenta la configurazione più semplice e intuitiva, spesso utilizzata come default nei sistemi automatizzati. Tale approccio non richiede informazioni aggiuntive sui job né una stima dei tempi residui, risultando adatto a contesti in cui l'obiettivo principale è mantenere un flusso regolare e prevedibile. La FIFO si configura come una strategia neutrale e stabile, garantendo un comportamento ordinato e privo di favoritismi, ma non sempre conduce alla massima efficienza. In presenza di variabilità nei tempi di lavorazione può infatti manifestarsi l'effetto convoy, in cui i job con tempi più lunghi rallentano l'avanzamento di quelli più rapidi, comportando un aumento del tempo medio di attraversamento. Questa possibilità non si verifica nel sistema in questione, che ha tutte quante le lavorazioni di pari durata.

5.3 Scelta della finestra di attivazione del DT

Nella scelta del periodo di replica del DT, sono state effettuate differenti prove, in particolare in merito alla durata delle repliche e nella loro periodicità. La scelta è poi ricaduta su un intervallo di 2 ore, che quindi prevedere che possano essere prodotti dalla WS5 al più 6 pezzi. Questa scelta è dovuta al fatto che nel momento in cui il DT conclude la sua fase sperimentale, le statistiche vengono salvate e tutto ciò che esce dal modello ridotto da quel momento in poi va perso, in termini di aggiornamento delle statistiche. L'orizzonte temporale che era subito stato individuato in maniera teorica era di un'ora, che si è rivelata troppo breve per riuscire ad osservare un numero discreto di uscite dal modello che fosse in grado di rendere lo studio degno di attenzione.

Qui di seguito le regole di funzionamento che regolano la WS5 nelle diverse varianti del DT:

DT	Priorità Macchina 5.1	Priorità Macchina 5.2
1	Chi non ha ancora lavorato	Chi ha già lavorato
2	Chi ha già lavorato	Chi ha già lavorato
3	Chi ha già lavorato	Chi non ha ancora lavorato
4	Chi non ha ancora lavorato	Chi non ha ancora lavorato

Tabella 2: Riassunto delle regole di priorità

La figura 11 dà un'indicazione riguardo quale mix di regole di priorità risulta più utilizzato sui due orizzonti temporali di simulazione.

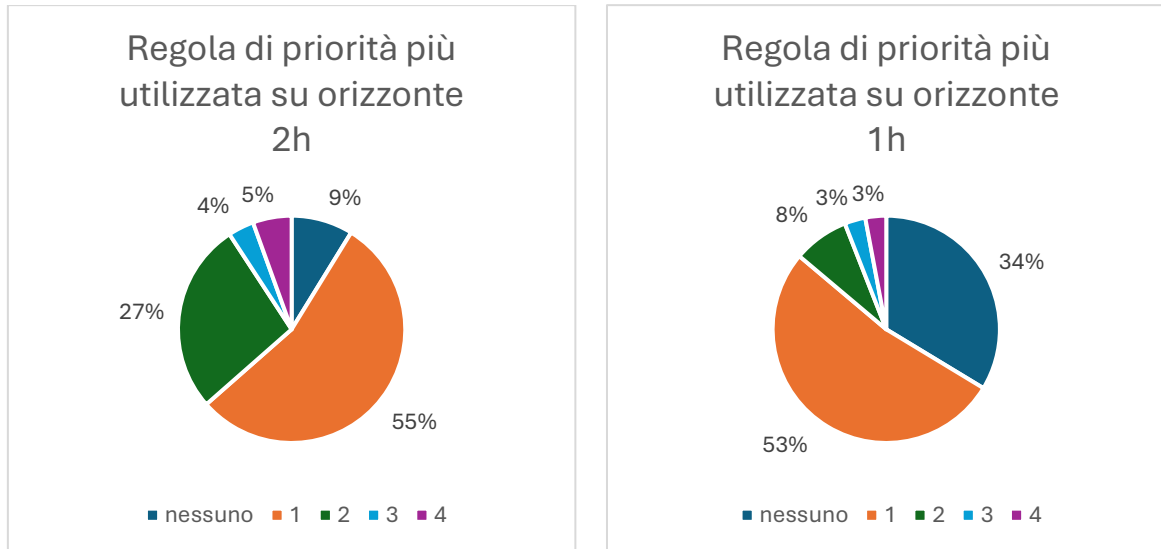


Figura 11: Regola di priorità più utilizzata su un orizzonte temporale di 2 ore e di 1 ora

Come si può osservare dai grafici le distribuzioni non sono identiche in termini di preferenza delle regole di priorità, come ci si potrebbe aspettare da un'indagine iniziale.

Si osserva che impostare come priorità della prima macchina i job non ancora lavorati e come priorità della seconda quelli che invece hanno già effettuato un passaggio all'interno della Workstation 5 risulta essere l'opzione vincente in entrambi i casi, ma, come previsto, nella simulazione di un'ora, l'impatto della impossibilità di scegliere in tempo un DT, risulta di gran lunga superiore.

Questo è un dato intermedio che è stato fondamentale per la costruzione di un sistema che avesse un fondamento industriale, infatti ampliando l'intervallo di simulazione è aumentata la consistenza dell'indagine e ha portato a risultati industrialmente più significativi e apprezzabili, non inquinati da un limite troppo serrato.

Nello specifico, per osservare questo dettaglio, si tiene in considerazione quanto è impattante la voce nessuno sulla percentuale totale; questa dicitura sta ad indicare che il sistema non è riuscito a concludere la lavorazione di nessuna unità

e quindi qualsiasi DT fosse stato eseguito non avrebbe condotto ad un miglioramento delle performance, per questo si è deciso di distinguere questa casistica nell'analisi, al fine di individuare davvero quale regola, in maniera distinta e pulita, risultasse più efficace.

Un dettaglio più preciso sull'efficienza è dato infatti dai passaggi totali che le macchine riescono a realizzare e infatti anche questo indicatore è stato oggetto della nostra indagine, ma questo aspetto sarà oggetto di indagine nei paragrafi successivi.

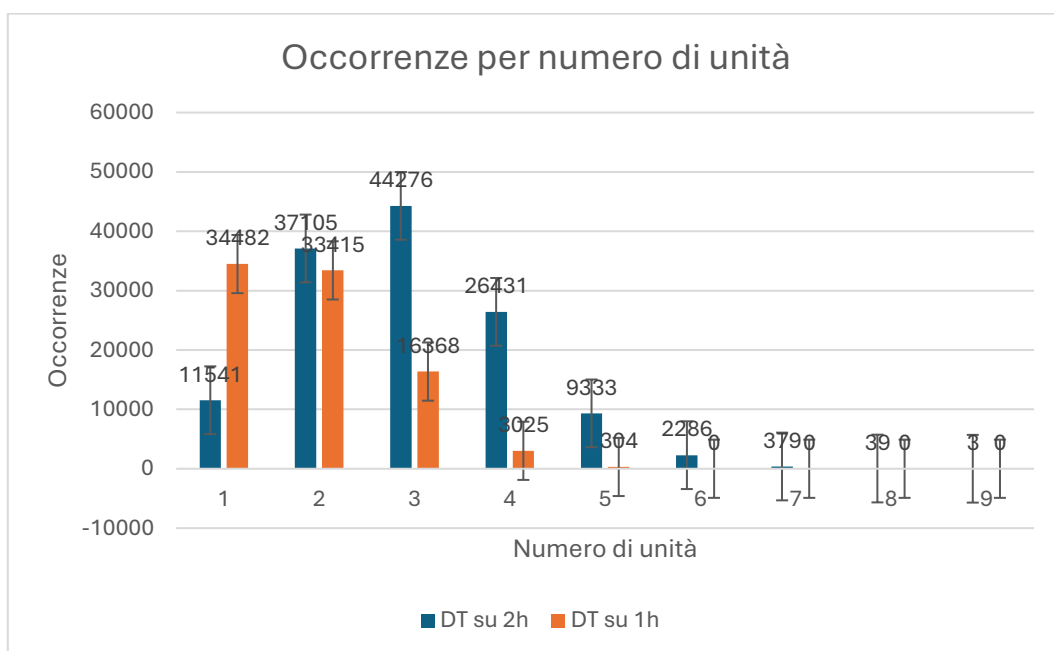


Figura 12: Occorrenza per numero di unità

Ampliando l'intervallo a due ore, invece, si può osservare la bontà del modello e in proporzione il numero di entità che si vedono uscire dal modello è maggiore. Il limite del DT, è legato ad una sua interruzione in un intervallo definito in maniera statica, e non in base alle performance, quindi per esempio imponendo un target ai modelli e definendo il migliore in base a quale di questi lo raggiunge per primo; questo sarebbe meno discriminatorio verso determinate regole, che potrebbero reagire meglio su un arco temporale più prolungato.

L'importanza della Figura 12 non è data dal numero in sé di volte in cui effettivamente è uscito un preciso numero di entità, ma dall'andamento, che nella

simulazione di due ore è approssimabile ad una gaussiana, mentre nella simulazione di una sola ora, poiché il lasso temporale è estremamente breve, tutta la coda sinistra della distribuzione risulta inosservabile e si sottolinea l'inefficienza del modello dal fatto che in circa 35000 casi nessuna entità è riuscita a terminare il suo passaggio nella WS5, annullando in questi casi di fatto l'utilità del DT.

Da questa analisi in poi, la finestra di simulazione è stata fissata pari a 2 ore, in quanto, come evidente dalle analisi qui sopra, più rappresentativa per il sistema in questione.

5.4 Determinazione della regola di priorità più efficace

L'analisi approfondita dello scenario di simulazione con orizzonte temporale di 2 ore, identificato come il parametro ottimale per la stabilità statistica, ha permesso di far emergere dinamiche non immediatamente prevedibili a livello teorico. Sebbene la capacità nominale della workstation indichi un output teorico di circa 6 entità nell'arco delle due ore, l'osservazione dei dati rivela una realtà più complessa e dinamica.

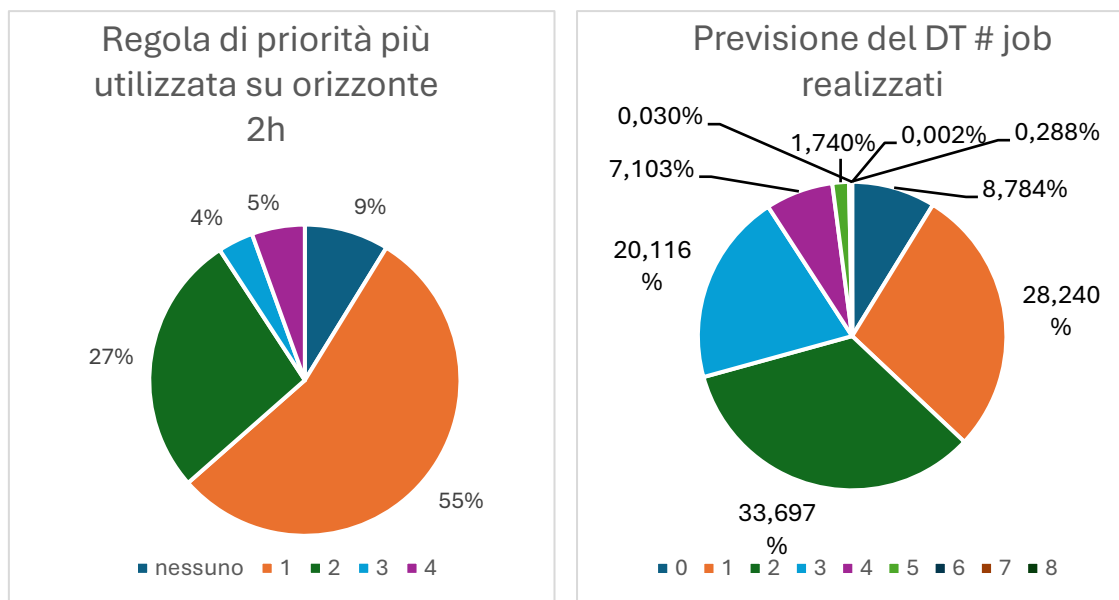


Figura 13: Regola di priorità più utilizzata su orizzonte temporale di 2 ore e aspettativa del numero di job realizzati

È stato infatti riscontrato che il sistema riesce, in alcuni casi specifici, a concludere il processamento di 7 o 8 entità, come rappresentato in Figura 13. Questo fenomeno è riconducibile alla natura dei flussi in ingresso al modello: le entità non sono costituite esclusivamente da nuovi job ("nuovi ingressi"), ma includono una quota significativa di job già precedentemente processati dalla workstation e rientrati nel loop per operazioni successive. La presenza di questi "pezzi circolanti" riduce il tempo netto necessario per la finalizzazione del job all'interno della finestra osservata, portando a picchi di produttività locale che superano la media teorica.

Analizzando la distribuzione della produzione per singola finestra, emerge che l'opzione più frequente è il completamento di 2 job, seguita in ordine di ricorrenza da 1 e 3 job. La variabilità di questi risultati è un indicatore cruciale: essa non è strettamente vincolata a una specifica configurazione di priorità, bensì rappresenta la chiave del dinamismo intrinseco del sistema. Questa fluttuazione dimostra che la performance non dipende solo dalla regola decisionale impostata, ma dallo stato istantaneo di saturazione dei buffer e dalla mutua interazione tra i job nel loop.

In sintesi, l'efficacia di una regola di priorità non può essere valutata in modo isolato, poiché il sistema risente di una forte componente stocastica. La scelta della regola "ottimale" diventa quindi un tentativo di stabilizzare un flusso che, per sua natura, tende a creare cluster di produzione (come i 2-3 job frequenti) alternati a momenti di bassa produttività, condizionati dall'accumulo di pezzi che necessitano di più passaggi per essere completati.

5.5 Analogie e differenze con l'indicatore Remaining Passages

Per il modello con simulazione di durata pari a 2 ore, è stata condotta un'ulteriore analisi comparativa focalizzata sulle divergenze che emergono impostando criteri di priorità differenti. In particolare, si è voluto indagare se la scelta della regola decisionale (DT) rimanesse invariata al variare del KPI di riferimento.

La Figura 14 mostra come, nel 66% dei casi esaminati, la scelta basata sul throughput (TH) della workstation WS5 non coincida con quella che sarebbe stata effettuata basandosi sul numero di passaggi totali che i job hanno compiuto sulle due macchine nell'arco di tempo dedicato.

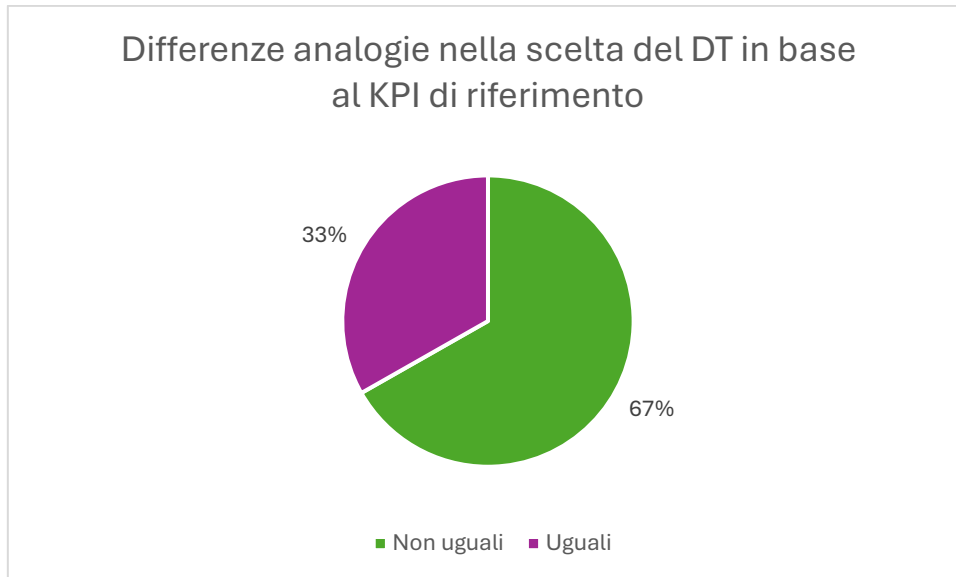


Figura 14: Differenze analogie nella scelta del DT in base al KPI di riferimento

Questa discrepanza suggerisce che ottimizzare la produzione locale (quanti pezzi escono dalla stazione) non garantisce necessariamente una gestione ottimale dei flussi interni (quante operazioni totali vengono eseguite). Un DT focalizzato solo sul TH potrebbe infatti privilegiare i job prossimi all'uscita, trascurando quelli che necessitano di più passaggi e che, se accumulati, rischiano di saturare il loop nel lungo periodo.

6. Discussione dei risultati e implicazioni manageriali

L'obiettivo di questo capitolo è analizzare criticamente le evidenze emerse dalle simulazioni, contestualizzandole all'interno delle dinamiche operative dell'impianto. La discussione non si limita alla mera esposizione dei dati, ma cerca di interpretare i nessi di causalità tra le scelte di configurazione (WIP, orizzonti temporali, KPI) e le performance globali osservate, fornendo al contempo una guida strategica per le decisioni manageriali.

6.1 La gestione del WIP e il limite della resilienza operativa

La determinazione del numero ottimale di pallet ha rappresentato una prima verifica per la stabilità del sistema. I risultati hanno evidenziato un fenomeno chiaro di rendimenti marginali decrescenti: superata la soglia dei 29 pallet, ogni unità aggiuntiva immessa nel sistema ha prodotto incrementi produttivi trascurabili, spesso limitati a poche decine di pezzi su base annua.

L'aspetto più critico emerso dai test con una dotazione superiore (fino a 50 pallet) è stata l'insorgenza sistematica di deadlock, ovvero stalli logici che hanno impedito il completamento delle repliche statistiche. È fondamentale sottolineare che l'incapacità di concludere queste simulazioni non deve essere interpretata come un malfunzionamento del software, bensì come un segnale di inaffidabilità strutturale del layout fisico.

In presenza di loop di rilavorazione, i pallet finiscono per occupare ogni posizione disponibile nei buffer; quando ogni slot di attesa è saturo, il movimento reciproco dei pezzi diventa impossibile, paralizzando l'intero flusso. La scelta dei 29 pallet è dunque il risultato di una ricerca del punto di massima pressione ammissibile: una quota che garantisce un utilizzo delle macchine discretamente alto (sfiorando l'unità in alcune stazioni critiche come la WS1) senza però condurre il sistema al collasso funzionale.

I dati mostrano che un numero eccessivo di supporti causa un aumento del Flow Time medio superiore al 130% rispetto alle configurazioni più snelle. In termini manageriali, ciò significa che "riempire" la linea oltre il necessario non protegge dalla variabilità, ma la amplifica. Oltre i 29 pallet, lo spazio fisico necessario per assorbire le fluttuazioni dei tempi di lavorazione e delle rilavorazioni scompare, trasformando ogni minimo ritardo in un potenziale blocco permanente.

Gli aumenti non sono soltanto legati al FT, ma anche alla produttività, infatti nell'inserimento di 29 pallet, rispetto all'opzione con 16, il TH aumenta di 0,2 job/h, per un totale di quasi 1000 pezzi in 6 mesi di simulazione.

6.2 Il paradosso del Digital Twin: stabilità contro reattività

Un punto centrale della ricerca riguarda il confronto tra il modello regolato dal Digital Twin (DT) e quello statico basato sulla logica FIFO. I risultati hanno mostrato una realtà controintuitiva: il sistema privo di controllo dinamico ha performato in modo più efficiente, producendo circa 2000 unità in più nell'orizzonte simulativo. Il throughput è passato da 2,45 job/h del modello FIFO a 2,34 job/h del modello ibrido.

Questa perdita di efficienza è riconducibile a due fattori principali. In primo luogo, la cosiddetta "latenza decisionale" o "congelamento" del sistema. In un contesto industriale reale, mentre il Digital Twin esegue le simulazioni per determinare la miglior regola di priorità, l'impianto fisico continua a muoversi. In secondo luogo, la natura stessa del sistema analizzato influisce sul risultato: poiché tutte le lavorazioni hanno durata uguale, l'effetto convoy (tipico dei sistemi con alta variabilità) non si manifesta, rendendo la semplicità della FIFO intrinsecamente robusta e difficile da superare in termini di pura velocità produttiva.

Anche l'analisi del Flow Time conferma questa tendenza: il sistema statico ha registrato una media di 526 minuti contro gli 857 minuti del sistema con DT. È probabile che il Digital Twin, nel tentativo di ottimizzare localmente la workstation critica, finisca per generare code più lunghe per i job complessi, favorendo l'uscita rapida solo di una parte del campione e aumentando così la permanenza media totale dei pezzi nel sistema.

Un limite strutturale emerso nella sperimentazione è la sensibilità del modello alla qualità e alla tempistica dei dati. In un contesto di controllo dinamico, anche una minima latenza informativa trasforma una decisione ottimale in una scelta obsoleta. Questo studio conferma che l'aggiornamento del dato è più importante della complessità dell'algoritmo di calcolo: nel nostro caso, la robustezza della logica FIFO ha superato la reattività del Digital Twin proprio perché non soggetta ai rischi di disallineamento temporale tra stato fisico e stato simulato.

6.3 Analisi della finestra decisionale e limiti statistici

La scelta dell'orizzonte temporale di 2 ore per l'attivazione del DT è stata dettata dalla necessità di superare il "rumore statistico" che caratterizzava la finestra di un'ora. Con un output massimo teorico di soli 6-8 pezzi per finestra, il Digital Twin si trova a prendere decisioni strategiche basandosi su un campione estremamente ridotto.

Statisticamente, 6 osservazioni non sono sufficienti per garantire che una regola di priorità sia effettivamente superiore a un'altra in modo deterministico. Il DT rischia di essere vittima della componente stocastica della simulazione: una regola potrebbe apparire vincente solo perché, in quella specifica finestra, sono casualmente arrivati job meno complessi o già parzialmente lavorati (i cosiddetti "pezzi circolanti"). Questa variabilità è la chiave del dinamismo del sistema, ma rappresenta anche un ostacolo alla precisione del controllo dinamico, rendendo difficile distinguere tra un reale miglioramento delle performance e un evento fortuito dovuto alla saturazione istantanea dei buffer.

6.4 Conflitto tra KPI e ottimizzazione dei flussi interni

L'analisi condotta sulle analogie e differenze tra gli indicatori di performance ha rivelato un dato di estrema rilevanza manageriale: nel 66% dei casi, la scelta del miglior DT basata sul throughput della WS5 non coincide con quella basata sul numero di passaggi totali effettuati dai job. Questa discrepanza suggerisce che ottimizzare la produzione locale (quanti pezzi escono dalla stazione) non garantisce affatto una gestione ottimale dei flussi interni.

Un Digital Twin focalizzato esclusivamente sul throughput immediato tende a privilegiare i job prossimi alla finalizzazione, trascurando quelli che necessitano di più cicli di lavorazione. Se questi pezzi "complessi" si accumulano, rischiano di saturare il loop nel lungo periodo, creando colli di bottiglia che il monitoraggio locale non è in grado di prevedere. Inoltre, l'uscita non omogenea dalle workstation può creare ondate di job che confluiscono contemporaneamente verso le macchine successive (come la WS6), le quali, non potendo accoglierli tutti nei propri buffer, costringono i pezzi in eccesso a percorrere un intero giro aggiuntivo del Loop 5, con un impatto devastante sul Flow Time complessivo.

6.5 Conclusioni e implicazioni per il management

In conclusione, lo studio dimostra che l'implementazione di logiche di controllo dinamico come quelle utilizzate nel Digital Twin non garantisce automaticamente un miglioramento delle performance, specialmente in sistemi già operanti vicino al limite di saturazione e caratterizzati da bassa variabilità nei tempi di servizio.

Per il management, le implicazioni sono chiare:

1. **Priorità alla stabilità del WIP:** prima di investire in sistemi di sequenziamento avanzati, è fondamentale identificare il "punto di rottura" del WIP (nel nostro caso 29 pallet). Operare oltre questa soglia annulla ogni beneficio derivante dalle regole di priorità.
2. **Visione Sistemica vs Locale:** l'ottimizzazione deve spostarsi dalla singola workstation all'intero loop. La discrepanza del 66% tra i KPI indica che un approccio locale può essere miope e controproducente per la fluidità globale.
3. **Adattabilità della strategia:** in situazioni di "rush" produttivo, potrebbe avere senso accettare un Flow Time inferiore a scapito del volume totale, mentre in condizioni di regime la stabilità del FIFO rimane la scelta più affidabile per massimizzare l'output annuo.

7. Conclusioni

7.1 Conclusioni generali

Il presente lavoro di tesi ha permesso di analizzare nel dettaglio le dinamiche di un sistema produttivo complesso, caratterizzato da loop di rilavorazione e stazioni critiche, attraverso l'integrazione di modelli di simulazione ad eventi discreti e logiche di controllo basate sul paradigma del Digital Twin. Il percorso intrapreso ha portato alla luce evidenze che sfidano alcune assunzioni classiche della gestione della produzione, offrendo una visione pragmatica e basata sui dati della stabilità industriale.

La conclusione principale riguarda la gestione del Work In Process (WIP). La ricerca ha dimostrato che esiste un limite strutturale di "capacità protettiva", identificato in 29 pallet, oltre il quale il sistema degrada inevitabilmente. L'insorgenza di deadlock e l'esplosione del Flow Time riscontrata con dotazioni superiori (fino a 50 pallet) confermano che la massimizzazione del carico non coincide con la massimizzazione dell'efficienza. Al contrario, in un sistema pull con alta interdipendenza, la saturazione delle risorse diventa il principale catalizzatore di inaffidabilità.

Un secondo tema rilevante riguarda l'efficacia del Digital Twin. Sebbene la letteratura specialistica identifichi nei sistemi di controllo dinamico la chiave per la reattività industriale, la sperimentazione ha evidenziato come, in contesti a bassa variabilità di durata delle operazioni, la stabilità della logica FIFO (statico) risulti superiore. Il minor throughput della simulazione che comprendeva il supporto del Digital Twin (circa 2000 unità in meno nell'orizzonte simulativo) e l'incremento del Flow Time suggeriscono che l'introduzione di priorità dinamiche debba essere attentamente pesata rispetto alla latenza informativa che essa stessa genera. La decisione del DT, basata su campioni statistici ridotti e su una "fotografia" del sistema potenzialmente obsoleta, ha dimostrato che un controllo eccessivamente frequente può introdurre instabilità anziché risolverla.

7.2 Contributi della tesi

L'analisi condotta permette di approfondire il dibattito sui sistemi di produzione snella (Lean Manufacturing) attraverso evidenze metodologiche e applicative. In primo luogo, i dati offrono una conferma empirica della Legge di Little e della

Teoria delle Code in contesti non lineari. L'individuazione del punto di deadlock operativo oltre la soglia dei 29 pallet suggerisce che il WIP non debba essere considerato esclusivamente un parametro di calcolo, quanto piuttosto un limite fisico necessario per l'assorbimento della variabilità stocastica del processo. Inoltre, lo studio dei sistemi Multi-KPI evidenzia un disallineamento nel 66% dei casi tra l'ottimizzazione del throughput locale e i passaggi totali del job. Tale fenomeno indica un rischio di 'miopia decisionale', in cui l'efficientamento di una singola stazione (WS5) può determinare il sovrappeso dei loop secondari, degradando le prestazioni globali dell'impianto

7.3 Limiti del lavoro svolto

Il limite principale risiede nell'ottica ottimistica del modello. Per isolare l'impatto delle regole di priorità e del WIP, sono stati trascurati fattori critici della realtà industriale come i guasti improvvisi alle macchine, i tempi di setup per il cambio di produzione e le operazioni di manutenzione programmata. In un contesto reale, tali eventi ridurrebbero ulteriormente la finestra di stabilità, rendendo probabilmente il sistema ancora più fragile e suscettibile di stallo anche con dotazioni di pallet inferiori a 29.

Un altro limite riguarda la latenza informativa e il tempo di calcolo. Nel modello simulato, il "congelamento" del sistema durante la scelta del Digital Twin è stato modellato in modo ideale. Nella pratica industriale, la trasmissione dei dati dai sensori al modello di calcolo e il ritorno della decisione alle macchine introducono ritardi che potrebbero degradare ulteriormente le performance del controllo dinamico, un aspetto che la simulazione attuale riesce solo parzialmente a catturare.

Infine, la dimensione del campione decisionale rappresenta un limite statistico intrinseco alla finestra di 2 ore. Con un output di soli 6-8 pezzi per intervallo, il Digital Twin è stato spesso costretto a prendere decisioni basate su eventi stocastici di breve durata (rumore statistico), non avendo una visibilità sufficientemente ampia per intercettare trend di produzione di lungo periodo.

7.4 Prospettive di estensione

Il lavoro svolto apre la strada a diverse direttrici di ricerca e sviluppi futuri, mirati a superare i limiti sopra descritti e ad approfondire le potenzialità del controllo dinamico.

1. Integrazione della variabilità reale: una naturale estensione del lavoro consiste nell'introdurre nel modello di simulazione profili di guasto (MTBF/MTTR) e tempi di setup variabili. Questo permetterebbe di testare la resilienza del Digital Twin in condizioni di "stress test" più vicine alla realtà, verificando se, in presenza di forti perturbazioni, il controllo dinamico riesca finalmente a superare le performance della logica FIFO.
2. Espansione dell'orizzonte di visibilità: per risolvere il problema della "miopia locale", la ricerca futura dovrebbe orientarsi verso un Digital Twin a visibilità globale. Invece di ottimizzare solo l'uscita dalla WS5, l'algoritmo dovrebbe considerare lo stato di saturazione di tutte le stazioni del loop (inclusa la WS6) e il numero di passaggi residui totali di ogni pezzo circolante. Questo permetterebbe di bilanciare meglio il carico, evitando cluster di produzione che intasano le fasi a valle.
3. Algoritmi di Machine Learning per la predizione dei KPI: un'estensione interessante riguarda l'uso di tecniche di apprendimento automatico per "prevedere" la migliore regola di priorità senza dover eseguire simulazioni complete in tempo reale. Questo ridurrebbe drasticamente la latenza decisionale, permettendo al sistema di reagire quasi istantaneamente ai cambiamenti di stato dei buffer.
4. Analisi economica del WIP: sarebbe opportuno integrare nel modello una dimensione di costo legata al mantenimento del WIP e alle penali per i ritardi di consegna (Flow Time). Questo permetterebbe di identificare non solo il punto di ottimo tecnico (29 pallet), ma anche il punto di ottimo economico, bilanciando il costo del capitale immobilizzato nei pallet con la massimizzazione del profitto derivante dal throughput.

In conclusione, questa tesi dimostra che il passaggio verso una fabbrica intelligente e data-driven non può prescindere da una solida conoscenza delle leggi fisiche che regolano i flussi produttivi. L'innovazione tecnologica deve essere supportata da un'analisi critica della complessità operativa, ponendo la

stabilità e la resilienza del sistema come pre-requisiti essenziali per ogni strategia di ottimizzazione dinamica.

8. Bibliografia

- [1] Lasi, H., Fettke, P., Kemper, H. G., Feld, T., e Hoffmann, M. (2014). «Industry 4.0». *Business & Information Systems Engineering*, vol. 6, fasc. 4, pp. 239–242. doi: 10.1007/s12599-014-0334-4.
- [2] Hopp, W. J., & Spearman, M. L. (2008). CONWIP: A Pull-Alternative to Kanban. In *Factory Physics: Foundations of Manufacturing Management* (3rd ed., pp. 477-505). McGraw-Hill/Irwin.
- [3] A. Ashrafian et al., «Full-Scale Discrete Event Simulation of an Automated Modular Conveyor System for Warehouse Logistics», in *Advances in Production Management Systems. Towards Smart Production Management Systems*, vol. 567, F. Ameri, K. E. Stecke, G. Von Cieminski, e D. Kiritsis, A c. di, in *IFIP Advances in Information and Communication Technology*, vol. 567. , Cham: Springer International Publishing, 2019, pp. 35–42. doi: 10.1007/978-3-030-29996-5_4.
- [4] H. ElMaraghy, «Smart changeable manufacturing systems», *Procedia Manufacturing*, vol. 28, pp. 3–9, 2019, doi: 10.1016/j.promfg.2018.12.002.
- [5] Pinedo, M. L. (2008). *Scheduling: Theory, Algorithms, and Systems*. 3^a Edizione, Springer Science & Business Media, New York, NY, USA. doi: 10.1007/978-0-387-78935-4.
- [6] M. Gong e C. Williamson, «Quantifying the properties of SRPT scheduling», in *11th IEEE/ACM International Symposium on Modeling, Analysis and Simulation of Computer Telecommunications Systems*, 2003. MASCOTS 2003., Orlando, FL, USA: IEEE Comput. Soc, 2003, pp. 126–135. doi: 10.1109/MASCOT.2003.1240650.

[7] Fernandes, F. C. F., Filho, M. G., e Bonney, M. (2005). «Integrating Materials Flow, Production Control and Quality Control: a Proposal and Case Study». *Journal of Manufacturing Technology Management*, vol. 16, fasc. 1, pp. 81–104. doi: 10.1108/17410380510574100.

[8] A. Florescu e S. A. Barabas, «Modeling and Simulation of a Flexible Manufacturing System—A Basic Component of Industry 4.0», *Applied Sciences*, vol. 10, fasc. 22, p. 8300, nov. 2020, doi: 10.3390/app10228300.

[9] Grieves, M. (2003). «Digital Twin: Concept Model and Life Cycle Management». White Paper, Product Lifecycle Management Center, University of Michigan, Ann Arbor, MI, USA. [URL: <https://www.researchgate.net/publication/307509727>]

[10] Attaran, M., e Celik, B. G. (2023). «Digital Twin: Benefits, use cases, challenges, and opportunities». *Decision Analytics Journal*, vol. 6, Art. 100165. doi: 10.1016/j.dajour.2023.100165.

[11] Alfieri Arianna, Castiglione Claudio e Pastore Erica, «Implementing Job Sequencing in a CONWIP fully automated assembly line», presentato al Italian Manufacturing Association Conference, 2025.

[12] Alfieri Arianna, Castiglione Claudio, e Pastore Erica, «Integrating buffer “stock” and buffer loops in an assembly line with conveyors: An automotive case study», presentato al Italian Manufacturing Association Conference, 2025.

[13] Bortolini, M., Faccio, M., Galizia, F. G., Gamberi, M., e Pilati, F. (2020). «Design, engineering and testing of an innovative adaptive automation assembly system». *Assembly Automation*, vol. 40, fasc. 3, pp. 531–540. doi: 10.1108/AA-12-2019-0215.

[14] Kritzinger, W., Karner, M., Traar, G., Henjes, J., e Sihn, W. (2018). «Digital Twin in manufacturing: A categorical literature review». IFAC-PapersOnLine, vol. 51, fasc. 11, pp. 1016–1022. doi: 10.1016/j.ifacol.2018.08.474.

[15] Tao, F., Zhang, H., Liu, A., e Nee, A. Y. C. (2019). «Digital Twin in Industry: State-of-the-Art». IEEE Transactions on Industrial Informatics, vol. 15, fasc. 4, pp. 2405–2415. doi: 10.1109/TII.2018.2873186.

[16] ISO 23247-2:2021. Automation systems and integration — Digital twin framework for manufacturing — Part 2: Reference architecture. International Organization for Standardization, Ginevra, Svizzera. [URL: <https://www.iso.org/standard/75067.html>]

[17] E. Negri, L. Fumagalli, e M. Macchi, «A Review of the Roles of Digital Twin in CPS-based Production Systems», Procedia Manufacturing, vol. 11, pp. 939–948, 2017, doi: 10.1016/j.promfg.2017.07.198.

[18] Fang, Y., Peng, C., Lou, P., Zhou, Z., Hu, J., e Yan, J. (2019). «Digital-Twin-Based Job Shop Scheduling Toward Smart Manufacturing». IEEE Transactions on Industrial Informatics, vol. 15, fasc. 4, pp. 2392–2402. doi: 10.1109/TII.2019.2894858.

[19] A. Sitahong, Y. Chen, Y. Yuan, A. Wubuli, J. Ma, e P. Mo, «Research Review on Workshop Scheduling for Intelligent Manufacturing: Digital Twin Modeling, Optimization Algorithm, and System Architecture», Machines, vol. 13, fasc. 11, p. 1021, nov. 2025, doi: 10.3390/machines13111021.

[20] M. Javaid, A. Haleem, e R. Suman, «Digital Twin applications toward Industry 4.0: A Review», Cognitive Robotics, vol. 3, pp. 71–92, 2023, doi: 10.1016/j.cogr.2023.04.003.