

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:
Aurora Pujia

Corelatore:
Prof. Claudio Castiglione

Anno accademico 2024-2025

Sommario

1. INTRODUZIONE	4
1.1 Contesto dei sistemi produttivi pull: introduzione al CONWIP	4
1.2 Sfide nella gestione dei caroselli e nastri trasportatori	5
1.3 Congestione, deadlock e impatto sul throughput.....	7
1.4 Obiettivo della tesi	7
2. REVIEW DELLA LETTERATURA – SISTEMI PRODUTTIVI, CONTROLLO DEL WIP E SIMULAZIONE	9
2.1 Sistemi di controllo della produzione: push, pull, CONWIP e Kanban.....	9
2.2 Strategie di controllo dinamico del WIP nei sistemi CONWIP.....	13
2.3 Deadlock e congestione in sistemi con loop e buffer limitati	16
2.4 Scheduling e prioritizzazione in workstation parallele	19
2.5 Sequenziamento e gestione delle rilavorazioni nella stazione WSQC.....	21
2.6 Arena Simulation	23
3. DESCRIZIONE DEL PROBLEMA	26
3.1 Descrizione del layout del sistema CONWIP modellato	26
3.2 Logiche operative	27
3.3 Obiettivo tecnico: controllo dinamico del WIP per massimizzare il throughput e ridurre i deadlock	28
3.4 Ipotesi modellistiche e semplificazioni.....	29
4. CONFIGURAZIONE E SPERIMENTAZIONE DEL MODELLO DI SIMULAZIONE	32
4.1 Descrizione concettuale del modello di simulazione	32
4.2 Parametri e condizioni di simulazione	34
4.3 Impostazioni sperimentali.....	38
4.4 Scenari ipotizzati per l’analisi del sistema CONWIP	40
4.5 Indicatori di performance	42
4.6 Semplificazione del modello e validità dei risultati.....	43
5. ANALISI QUALITATIVA E DISCUSSIONE DEI RISULTATI.....	45
5.1 Obiettivi e impostazione dell’analisi	45
5.2 Analisi teorico-simulativa del comportamento del sistema	45
5.2.1 Effetto del controllo del WIP	46
5.2.2 Effetto delle regole di priorità	52
6. DISCUSSIONE MANAGERIALE DEI RISULTATI	57
6.1 Implicazioni per la definizione di metodologie di controllo di sistemi produttivi pull.....	57
6.2 Benefici e limiti del controllo dinamico del WIP	58
6.3 Riflessioni sulla gestione della flessibilità e congestione	59
6.4 Suggerimenti per implementazione industriale	61
7. CONCLUSIONI	63

7.1 Conclusioni generali	63
7.2 Contributi della tesi.....	63
7.3 Limiti del lavoro svolto	64
7.4 Prospettive di estensione.....	66
APPENDICI.....	68
Appendice A – Modello con nastri su ARENA Simulation	68
Appendice B – Modello senza nastri CONWIP dinamico su ARENA Simulation.....	72
Appendice C – Modello senza nastri CONWIP statico su ARENA Simulation	74
BIBLIOGRAFIA	76

1. INTRODUZIONE

1.1 Contesto dei sistemi produttivi pull: introduzione al CONWIP

Un sistema produttivo è l'insieme coordinato di risorse, attività e procedure che consentono di trasformare le materie prime in prodotti finiti, con l'obiettivo di soddisfare le richieste del mercato. Al suo interno troviamo non solo elementi tangibili come materiali, macchinari, forza lavoro e impianti, ma anche flussi immateriali, ad esempio informazioni, documenti e dati, che ne garantiscono il corretto funzionamento. Un ruolo fondamentale è svolto inoltre dalle attività gestionali, che comprendono la pianificazione, la progettazione delle linee, il controllo qualità e il coordinamento delle operazioni quotidiane.

La gestione di un sistema produttivo, però, non è priva di complessità. Le difficoltà possono derivare da più fattori come l'interdipendenza tra le varie stazioni di lavoro, la variabilità qualitativa e quantitativa dei materiali in ingresso, eventuali guasti ai macchinari e l'assenteismo degli operatori. A ciò si aggiungono elementi esterni, come l'instabilità della domanda o cambiamenti di natura economica, normativa o tecnologica. Tutto questo comporta inevitabilmente scostamenti rispetto a quanto pianificato, con possibili conseguenze su costi, efficienza e affidabilità complessiva [1] [2] [3].

Per affrontare tali criticità sono stati sviluppati diversi approcci e uno dei più diffusi è la logica di produzione *pull*. In questo modello il processo produttivo si attiva sulla base della domanda effettiva e non su stime o previsioni. In un sistema pull, la produzione, quindi, viene "trainata" dalla domanda reale, con il vantaggio di ridurre il rischio di sovrapproduzione e di mantenere più contenuti i livelli di scorte. Questo approccio favorisce una maggiore flessibilità e consente di adattarsi meglio alle esigenze del mercato, anche se talvolta può comportare tempi di consegna più lunghi e costi unitari superiori, perché non sfrutta al massimo le economie di scala tipiche della produzione di massa [4] [5].

La logica pull è inoltre spesso associata alla filosofia *Lean* e al modello *just-in-time* (JIT), che mirano all'eliminazione di attività che non aggiungono valore e al miglioramento continuo dei processi produttivi. Questi sistemi, migliorando il controllo del WIP e del TH, permettono una gestione più sostenibile dal punto di vista ambientale, grazie a una riduzione dell'impatto derivante da magazzini sovradimensionati e da produzione eccessiva.

In questo contesto si inserisce il *Constant Work In Process* (CONWIP), un metodo di controllo dei flussi produttivi che combina i vantaggi del sistema pull con un controllo centralizzato e globalizzato del WIP, cioè l'insieme dei job che si trovano in lavorazione o

in attesa lungo la linea produttiva. Mentre nei tradizionali sistemi pull come il *Kanban* il controllo del flusso avviene stazione per stazione con cartellini legati ai singoli prodotti, nel CONWIP si opera a livello globale, mantenendo un numero costante di cartellini a livello del sistema intero. Questo significa che un nuovo ordine di produzione può entrare solo quando un ordine precedente è stato completato ed è uscito dal sistema, mantenendo così invariato il livello globale di WIP [6] [7].

Questa caratteristica rende il CONWIP particolarmente indicato per sistemi produttivi con alta varietà di prodotti e complessità, dove il controllo localizzato può essere inefficace o troppo rigido. Il modello, infatti, contribuisce a velocizzare il flusso produttivo, ridurre i tempi di attraversamento e limitare gli accumuli eccessivi di materiali, garantendo un equilibrio tra flessibilità e controllo della produzione. Grazie a questa metodologia è possibile migliorare la sincronizzazione tra le diverse fasi produttive e ridurre i rischi di colli di bottiglia e ritardi, mantenendo sempre un allineamento stretto con la domanda reale.

In sintesi, il CONWIP può essere visto come un'alternativa dei sistemi pull: non si limita a guidare la produzione sulla base delle esigenze reali del mercato, ma fornisce anche un meccanismo di coordinamento che tiene sotto controllo l'intero processo. Grazie a questo approccio, la produzione risulta più equilibrata, sincronizzata e meno soggetta a blocchi dovuti a congestioni interne.

Questo studio si inserisce proprio in tale contesto, analizzando un sistema produttivo di tipo pull regolato da una politica CONWIP, nella quale il numero massimo di job presenti nel sistema viene gestito centralmente. Sebbene il livello globale di WIP rimanga costante, possono comunque verificarsi situazioni di congestione o blocco dovute alla configurazione fisica dell'impianto, ad esempio in presenza di loop, merge o limiti di capacità nei buffer. Tali problematiche mettono in evidenza la necessità di approfondire l'analisi delle interazioni tra i flussi fisici e le logiche di controllo al fine di ottimizzare la gestione complessiva del sistema.

1.2 Sfide nella gestione dei caroselli e nastri trasportatori

Nei moderni impianti produttivi una componente fondamentale è rappresentata dai sistemi di trasporto meccanico, come nastri trasportatori, rulliere e caroselli. Questi strumenti permettono il trasferimento efficiente di materiali tra le diverse stazioni operative, contribuendo a ridurre il carico di lavoro manuale e i tempi complessivi, aumentare la produttività e migliorare la sicurezza. Nel corso degli anni, questi sistemi hanno raggiunto un elevato livello di automazione, grazie all'integrazione con tecnologie avanzate e

intelligenti in grado di gestire flussi complessi e volumi estremamente elevati di prodotti dalle caratteristiche diversificate [8].

Tuttavia, vi sono comunque sfide significative legate alla progettazione e alla gestione dei sistemi di movimentazione interna. Le sfide affrontate derivano dalla variabilità della domanda, dalla diversificazione dei prodotti e, in particolare, dalla crescente pressione competitiva a livello globale [6]. In particolare, la presenza di layout non lineari, come quelli a configurazione chiusa o con caroselli, presenta criticità maggiori rispetto alle linee di produzione tradizionali. Il movimento ciclico dei materiali, combinato con la necessità di sincronizzare ingressi e uscite dai caroselli, richiede un controllo preciso per evitare problemi di congestione, blocco reciproco (*deadlock*) o accumulo di ritardi.

Un aspetto particolarmente problematico è rappresentato dai nodi di *merge* e *split*, ovvero i punti di convergenza o diramazione dei flussi. La mancata sincronizzazione in questi nodi può provocare *starving* (mancanza di materiale nelle stazioni a valle) o *blocking* (accumulo eccessivo di WIP nelle stazioni a monte), con evidenti ripercussioni negative sull'efficienza e sul rendimento dell'impianto. Questi problemi si accentuano in presenza di buffer limitati, dove lo spazio per gestire variazioni temporanee nei flussi risulta insufficiente. Inoltre, nei sistemi dotati di workstation parallele risulta essenziale ottimizzare la distribuzione dei carichi e stabilire quale macchina attribuire a ciascun job. Questo consente di assicurare un funzionamento armonioso ed equilibrato dell'intero impianto.

Un ulteriore elemento di sfida emerge dalla variabilità nei tempi di servizio tra le diverse stazioni e sottotask. Nei modelli che seguono logiche split-merge, le operazioni devono essere coordinate per garantire un corretto ricombinamento dei job a valle. Una strategia efficace in questo scenario potrebbe essere il ritardo controllato di alcune lavorazioni al fine di allineare i tempi di completamento e ridurre l'affollamento nei buffer in uscita. [9].

Infine, in ambienti caratterizzati da elevata variabilità di prodotto e processi organizzati a batch, la gestione dinamica del WIP e l'analisi preventiva attraverso modelli di simulazione ad eventi discreti rappresentano strumenti indispensabili per prevenire colli di bottiglia e valutare l'effetto di modifiche impiantistiche o logistiche [6].

Il sistema analizzato in questo lavoro evidenzia numerose criticità simili: flussi chiusi all'interno di caroselli, nodi di merge/split nelle stazioni e buffer limitati. Questi fattori rendono la regolazione del WIP una componente determinante per evitare deadlock e ottimizzare il throughput complessivo dell'impianto.

1.3 Congestione, deadlock e impatto sul throughput

Nei sistemi di produzione automatizzati (*Automated Manufacturing Systems – AMS*), il fenomeno del deadlock rappresenta una condizione critica e particolarmente indesiderata. Questa situazione si verifica quando un insieme di risorse condivise, come buffer, nastri trasportatori, macchinari o dispositivi di movimentazione, viene occupato da entità che si bloccano reciprocamente, nell'attesa di accedere alle risorse dell'altra. Questo genera uno stato di attesa circolare, impedendo l'avanzamento di qualsiasi processo e causando una stagnazione, parziale o completa, dell'intero sistema.

Il rischio di deadlock diventa più elevato in sistemi caratterizzati da loop chiusi, buffer limitati e processi concorrenti che competono per le stesse risorse. In tali contesti, anche un semplice guasto a una risorsa o un accumulo eccessivo di WIP in un punto specifico può innescare una catena di blocchi che compromette l'avanzamento dei materiali attraverso il sistema. La congestione locale, ad esempio in un carosello o in una stazione di lavoro con capacità di output limitata, tende a propagarsi al resto del sistema attraverso i collegamenti logistici, con effetti amplificati in sistemi altamente interconnessi. Questo fenomeno comporta un aumento complessivo del WIP e una significativa riduzione del *throughput*, ossia la quantità di output generata nel tempo. Paradossalmente, superato un certo limite critico di WIP, il sistema non migliora la produttività bensì si satura, entrando in una fase di rallentamento che può sfociare in un blocco totale.

Per evitare queste problematiche è fondamentale implementare un controllo intelligente e dinamico del WIP. Questo controllo deve essere in grado di regolare dinamicamente il numero di pallet o processi attivi nel sistema, monitorando costantemente lo stato delle aree critiche e prevenendo situazioni potenziali di deadlock. I sistemi di controllo devono essere progettati per aumentare il WIP quando necessario, ottimizzando l'utilizzo delle risorse, ma anche per ridurlo in modo proattivo, evitando condizioni di blocco.

Lo studio si focalizza sulla progettazione e sull'analisi di una soluzione finalizzata al controllo dinamico del WIP, tenendo conto della presenza di loop chiusi, stazioni parallele e buffer con capacità limitata. L'applicazione pratica di questa soluzione verrà approfondita nella relativa sezione.

1.4 Obiettivo della tesi

L'obiettivo principale di questa tesi è la costruzione di un modello di simulazione rappresentativo di un sistema manifatturiero complesso, sviluppato nel software *Arena Simulation*, attraverso un'accurata progettazione logica e metodologica che tenga conto delle

caratteristiche strutturali e operative del sistema. Il lavoro si inserisce nell'ambito dei sistemi produttivi a controllo pull e, in particolare, nella modellazione di una linea di assemblaggio regolata secondo la logica CONWIP, caratterizzata dalla presenza di loop chiusi, nodi di merge e split, workstation parallele e buffer con capacità limitata.

In questa fase del lavoro non essendo ancora disponibile un modello pienamente funzionante con il sistema di movimentazione completo, si è affiancato un secondo modello semplificato, sviluppato senza nastri trasportatori, con l'obiettivo di testare le logiche implementate e condurre una prima analisi del comportamento del sistema. L'obiettivo è comprendere come differenti strategie di regolazione del Work In Process e diverse regole di priorità applicate a una workstation composta da due macchine identiche, sulla quale i job richiedono una seconda lavorazione, possano influenzare la stabilità del flusso produttivo, il throughput complessivo e il rischio di congestione o deadlock.

A tal fine, vengono individuati e proposti quattro scenari di simulazione di riferimento, concepiti per evidenziare le principali differenze operative e prestazionali del sistema.

La tesi si articola in sette capitoli. Dopo un inquadramento teorico sui sistemi produttivi pull e sulle logiche di controllo CONWIP, viene presentata la descrizione del sistema analizzato e delle scelte modellistiche adottate. Successivamente, si illustrano le fasi di costruzione del modello di simulazione, seguite da una discussione sulle implicazioni prestazionali nei diversi scenari proposti. Il lavoro si conclude con una riflessione critica sui risultati concettuali ottenuti e con alcune considerazioni sulle possibili estensioni future, orientate alla validazione quantitativa e all'ottimizzazione del controllo dinamico del WIP.

2. REVIEW DELLA LETTERATURA – SISTEMI PRODUTTIVI, CONTROLLO DEL WIP E SIMULAZIONE

2.1 Sistemi di controllo della produzione: push, pull, CONWIP e Kanban

Nel corso degli anni sono stati sviluppati diversi sistemi di controllo della produzione, che si classificano principalmente in *push*, *pull* e modelli ibridi.

Nei sistemi *push*, la produzione non risponde direttamente alla domanda reale, ma è guidata da stime e piani predisposti in anticipo. Un esempio di logica *push* è mostrato in Figura 1. Questo approccio consente una gestione preventiva di approvvigionamenti, assicurando una certa continuità operativa anche di fronte a cambiamenti improvvisi nella domanda e permette anche di massimizzare le economie di scala e abbattere i costi unitari. Tuttavia, il modello presenta anche dei rischi perché previsioni errate possono portare a sovrapproduzione, con accumulo di scorte, aumento dei costi di magazzino e rischio di obsolescenza; viceversa, stime troppo conservative possono causare rotture di stock e perdite di opportunità di vendita. In generale, i sistemi *push* mostrano notevoli limiti in ambienti caratterizzati da forte variabilità della domanda, generando frequentemente alti livelli di WIP, lunghi lead time e amplificazione delle oscillazioni lungo la supply chain (*bullwhip effect*).

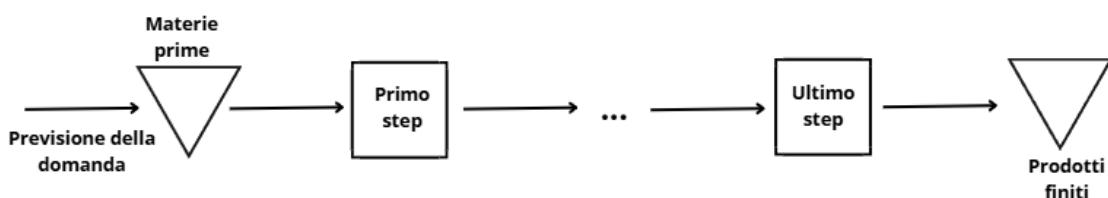


Figura 1: esempio di sistema produttivo con logica *push*: i job vengono immessi nel flusso secondo un piano predefinito, senza un controllo diretto sul livello di WIP nelle diverse workstation.

I sistemi *pull*, al contrario, attivano la produzione solo in risposta a un ordine reale. Questo modello, coerente con i principi della Lean Production e del Just-In-Time, consente di evitare la sovrapproduzione, ridurre i livelli di scorte e migliorare la flessibilità nella risposta alla domanda. Inoltre, tale logica si fonda su un flusso informativo che parte dal cliente e risale a monte del processo, regolando in modo sincronizzato le fasi produttive. Un esempio schematizzato del funzionamento di un sistema *pull* è riportato in Figura 2. Garantisce quindi maggiore flessibilità e rapidità nel gestire le richieste, oltre a migliorare l'efficienza dei magazzini. Tuttavia, presenta alcune criticità in quanto produrre esclusivamente su richiesta

può allungare i tempi di consegna e incrementare i costi unitari, poiché le economie di scala non vengono sfruttate al massimo. Inoltre, richiede un'organizzazione altamente efficiente e una gestione degli ordini impeccabile, risultando meno indicato in settori con domanda fortemente instabile o processi produttivi complessi e prolungati.

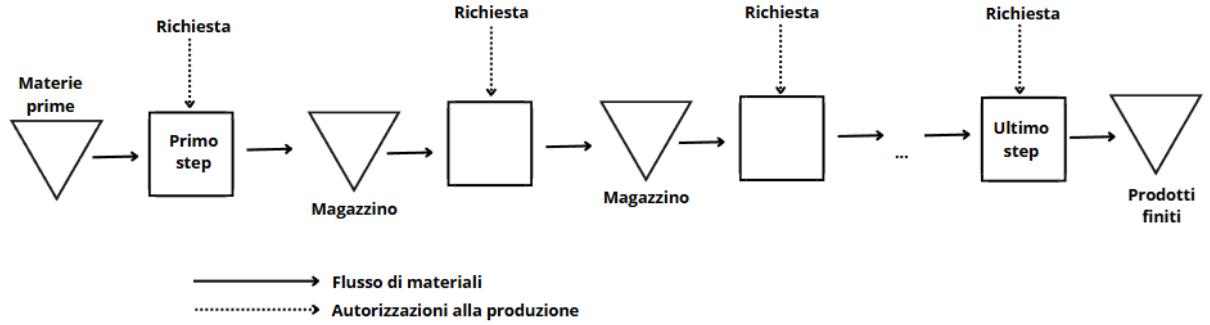


Figura 2: esempio di sistema produttivo con logica pull: l'ingresso di nuovi job avviene solo quando le stazioni a valle liberano capacità, mantenendo sotto controllo il livello complessivo di WIP.

Tra i sistemi pull, il Kanban è uno dei più noti ed è stato originariamente sviluppato nel contesto della produzione snella. Questo modello utilizza schede di autorizzazione, chiamate Kanban, per gestire il rilascio dei materiali e mantenere sotto controllo il WIP in ogni stazione produttiva. Ogni scheda simboleggia l'autorizzazione a produrre o trasferire un'unità, mentre la quantità di schede presenti in ogni punto della linea determina il limite massimo di unità ammesse in lavorazione o in transito. La logica di funzionamento del sistema si fonda su due regole fondamentali: in primo luogo, non si può produrre senza possedere un cartellino Kanban, il che implica che i reparti a monte debbano realizzare soltanto le parti effettivamente consumate da quelli a valle; in secondo luogo, i reparti a valle possono prelevare da quelli a monte solo i pezzi necessari, nella quantità e nel momento del consumo. Questo metodo facilita una gestione visiva dei livelli di inventario intermedio, contribuisce a prevenire la sovrapproduzione e assicura un flusso di lavoro ben orchestrato per rispondere alle esigenze del cliente [10]. Un esempio schematico di sistema Kanban è riportato in Figura 3.

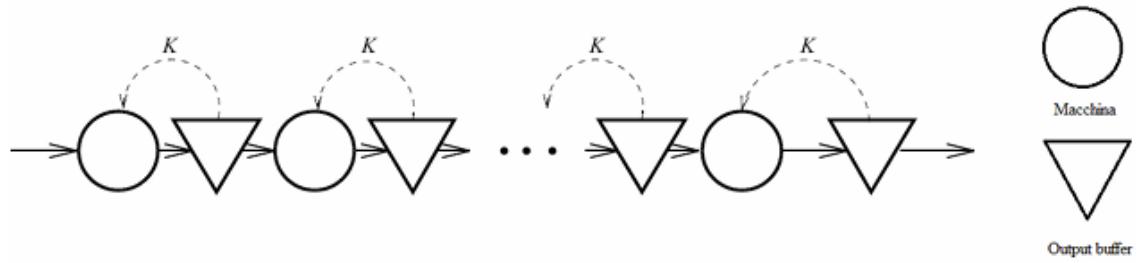


Figura 3: rappresentazione schematica di un sistema Kanban in logica pull, in cui il flusso dei materiali è regolato da cartellini di autorizzazione K.

Fonte: Gallo M., Guizzi G., Naviglio G., *Pull Production Policies: Comparative Study Through Simulative Approach*, Univ. di Napoli “Federico II”, 2011.

Una variante del Kanban è rappresentata dal sistema Constant Work-In-Process. Questo metodo utilizza un unico formato di scheda per gestire il WIP complessivo lungo tutta la linea, anziché suddividerlo per ciascuna stazione. Diversamente dal Kanban, che agisce su singole operazioni, il CONWIP regola il WIP globalmente, come illustrato in Figura 4, consentendo l’ingresso di un nuovo job solo quando un altro esce dal sistema e questo comporta che il livello complessivo di Work In Process all’interno del sistema rimanga costante nel tempo, da cui deriva l’acronimo CONWIP.

Nel modello CONWIP, il rilascio di nuove entità (materiali o ordini) nel sistema è attivato principalmente da un segnale legato al completamento e all’uscita di un lotto di prodotto finito dal sistema produttivo. Più precisamente, i cartellini CONWIP, che rappresentano il numero massimo di WIP consentito nel sistema, vengono "rilasciati" solo quando un lotto completo raggiunge l’uscita e termina il processo produttivo.

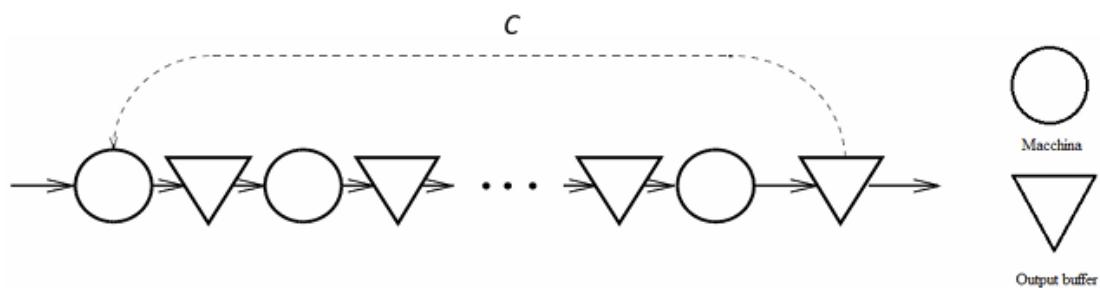


Figura 4: rappresentazione schematica di un sistema produttivo CONWIP con controllo globale del WIP

Fonte: Gallo M., Guizzi G., Naviglio G., *Pull Production Policies: Comparative Study Through Simulative Approach*, Univ. di Napoli “Federico II”, 2011.

Quindi, i principali eventi che attivano il rilascio di nuove entità sono:

- L’uscita di un lotto o entità completata dal sistema produttivo, che libera un cartellino CONWIP.

- Il rilascio del cartellino torna all'inizio del ciclo per autorizzare l'ingresso di un nuovo lotto dal magazzino.
- L'autorizzazione da parte del magazzino, che determina quale prodotto o ordine deve entrare in produzione in base alla priorità e alla disponibilità di cartellini.

Questo significa che nel CONWIP il rilascio è controllato in modo da mantenere costante il livello complessivo di WIP nell'intero sistema, garantendo che una nuova entità possa entrare solo se un'entità precedente è uscita completata, evitando così sovraccarichi e accumuli inutili nel processo produttivo. Questa filosofia è particolarmente adatta a flussi produttivi flessibili e complessi, riducendo i rischi di accumuli localizzati e semplificando la gestione del WIP [15].

Una differenza concettuale tra un sistema CONWIP e un sistema Kanban è rappresentata in Figura 5, dove si evidenzia come nel primo il controllo del WIP avvenga a livello globale sull'intera linea produttiva, mentre nel secondo sia localizzato tra le singole workstation, tramite schede di segnalazione che regolano i flussi tra stazioni adiacenti.

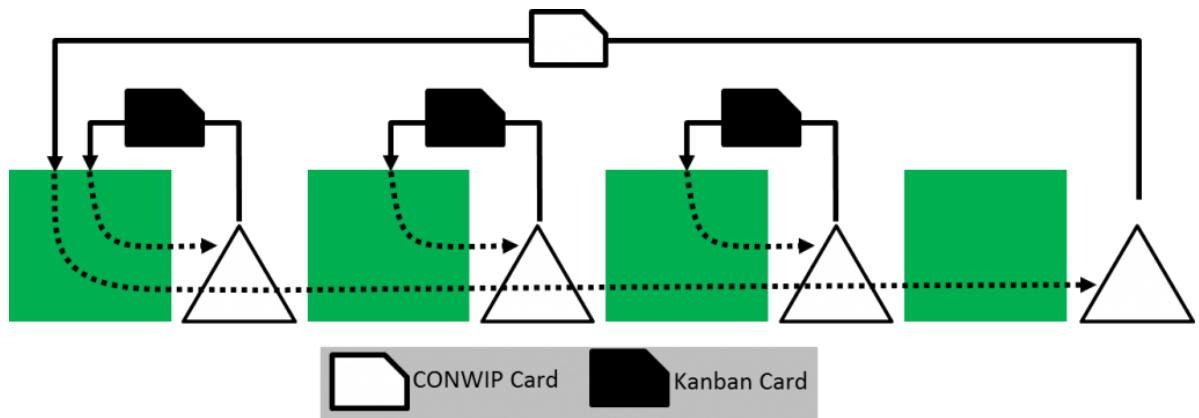


Figura 5: confronto tra i sistemi di controllo della produzione Kanban e CONWIP
Fonte: Qualitiamo, "Sistemi Pull – Kanban & CONWIP"

L'applicazione dei diversi sistemi varia in base al contesto operativo. Il Kanban è utilizzato maggiormente nelle linee di assemblaggio caratterizzate da processi ripetitivi, come nell'industria automobilistica o elettronica, dove risulta fondamentale bilanciare e limitare il WIP. Il sistema CONWIP, invece, si distingue per la gestione del Work In Process a livello globale, rivelandosi particolarmente vantaggioso in situazioni produttive con percorsi flessibili, cicli variabili e necessità di rilavorazione, tipiche di ambienti come i job-shop o le produzioni multiprodotto.

Sul piano delle prestazioni, studi comparativi evidenziano come il sistema CONWIP tenda a mantenere un livello medio di WIP più basso rispetto al Kanban, a parità di domanda.

Questo aspetto comporta diversi vantaggi, tra cui la riduzione dei costi di magazzino e un flusso operativo più regolare. Inoltre, la maggiore concentrazione di schede nelle singole postazioni del Kanban può portare a un aumento delle unità complessive per soddisfare le esigenze del cliente, mentre il CONWIP garantisce una distribuzione più equilibrata lungo la linea grazie al controllo sul rilascio iniziale [16].

In sintesi, i sistemi push mostrano una maggiore sensibilità alla variabilità e tendono a generare alti livelli di WIP in situazioni incerte, mentre i modelli pull, con particolare riferimento al CONWIP, offrono un controllo più efficace del flusso produttivo e una superiore capacità di adattamento. Queste qualità hanno spinto un crescente interesse da parte di accademici e imprese verso soluzioni pull e CONWIP, soprattutto in ambienti caratterizzati da cicli chiusi, rilavorazioni e domanda fluttuante [11] [12] [13] [14].

2.2 Strategie di controllo dinamico del WIP nei sistemi CONWIP

Il modello CONWIP standard presenta alcune limitazioni dovute alla natura rigida del numero di cartellini, che costituisce il principale parametro di controllo del sistema. Sebbene tale vincolo sia definito proprio per bilanciare throughput, scorte e backlog, una logica fissa può rivelarsi poco efficace in presenza di elevata variabilità o di condizioni operative dinamiche, dove il sistema rischia di trovarsi in situazioni di congestione o sottoutilizzo difficili da compensare. Per affrontare questa criticità, sono state sviluppate strategie di CONWIP adattivo, che consentono una regolazione dinamica del WIP in funzione dello stato reale del sistema e delle sue prestazioni operative.

Le modalità di controllo del WIP possono essere classificate principalmente in approcci statici e dinamici. Il controllo statico prevede limiti di WIP predeterminati e invariabili, definiti sulla base di analisi storiche, capacità produttiva o obiettivi di produzione. Questo metodo si distingue per la semplicità operativa, la prevedibilità e la stabilità, risultando particolarmente efficace in ambienti produttivi caratterizzati da condizioni relativamente costanti. Tuttavia, la sua rigidità può rivelarsi problematica in caso di variazioni repentine, portando a inefficienze, congestioni o tempi di attraversamento eccessivi. Nei sistemi complessi o soggetti a forte variabilità, infatti, un limite fisso può risultare talvolta sovradimensionato o altre volte insufficiente, compromettendo la fluidità complessiva del flusso produttivo.

Diversamente, il controllo dinamico permette di adattare i limiti del WIP in tempo reale sulla base di variabili come tempi di attraversamento, tassi di produzione o livelli di congestione. Questo approccio si avvale di sistemi di monitoraggio continuo e algoritmi per la regolazione

automatica, spesso basati su feedback, regole adattative o modelli predittivi. In questo modo, il sistema acquisisce maggiore flessibilità e resilienza, riuscendo a minimizzare colli di bottiglia, ottimizzare l'uso delle risorse e mantenere un flusso produttivo bilanciato anche in presenza di elevate variazioni. Tuttavia, l'implementazione richiede dati sempre aggiornati e affidabili, oltre a strumenti analitici e software avanzati per la simulazione, imprescindibili per una gestione efficace.

Oltre agli approcci statici e dinamici, in letteratura emerge la proposta di strategie ibride che combinano la stabilità dei limiti statici con la capacità adattiva dei meccanismi dinamici.

Le strategie di controllo dinamico del WIP nei sistemi CONWIP possono essere classificate in tre principali approcci: *Dynamic CONWIP*, *Generalized Kanban* e *Reinforcement Learning* (RL).

1. Il primo approccio, il Dynamic CONWIP, introduce una gestione flessibile del numero massimo di prodotti finiti nel buffer. Questo non è fisso, ma varia entro un intervallo predeterminato in base a un obiettivo di throughput. Ogni volta che viene completato un prodotto, si confronta il throughput attuale (TH) con il valore target (T). Se il TH è al di sotto dell'obiettivo, il numero massimo di prodotti finiti nel buffer può aumentare; viceversa, può ridursi o rimanere invariato. Questa logica implementa un controllo soglia, permettendo di rispondere a congestioni o inefficienze. Rispetto al CONWIP tradizionale, garantisce una maggiore reattività senza perdere la semplicità tipica dei sistemi pull.
2. Il Generalized Kanban rappresenta un'evoluzione del sistema CONWIP che mantiene l'essenza del controllo del WIP attraverso autorizzazioni alla produzione, ma introduce una maggiore flessibilità adattativa. A differenza del Kanban tradizionale, che utilizza cartellini prestabiliti per sincronizzare e limitare rigidamente i flussi tra le fasi produttive in base a quantità fisse di scorte, il Generalized Kanban modula le autorizzazioni di produzione in funzione di condizioni operative specifiche, quali la presenza di backorder o l'assenza di scorte disponibili al momento della domanda. Questo significa che, a livello pratico, il sistema non si limita più a un numero statico di cartellini, ma adatta dinamicamente il controllo delle scorte e del WIP in base allo stato reale del sistema produttivo. Una rappresentazione del Generalized Kanban System è mostrata in Figura 6. Tale approccio consente di mantenere un equilibrio dinamico tra domanda e disponibilità, gestendo in modo più efficace situazioni complesse e variabili, tipiche di ambienti

produttivi moderni che includono loop, nodi di merge e split, e buffer limitati. Il Generalized Kanban si configura quindi come una soluzione intermedia che combina i vantaggi di sistemi statici, come il Kanban classico, con un controllo adattivo che si avvicina a sistemi avanzati basati su logiche predittive o intelligenza artificiale, pur mantenendo una maggiore semplicità operativa e robustezza di implementazione.

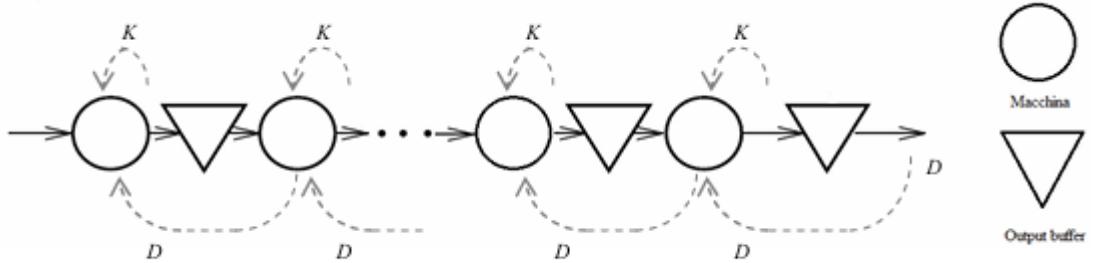


Figura 6: schema concettuale di un sistema Generalized Kanban

Fonte: Gallo M., Guizzi G., Naviglio G., *Pull Production Policies: Comparative Study Through Simulative Approach*, Univ. di Napoli "Federico II", 2011.

3. Parallelamente, la letteratura recente ha approfondito l'impiego di metodi avanzati di intelligenza artificiale per il controllo dinamico dei sistemi produttivi, con particolare attenzione al Reinforcement Learning. Questo approccio prevede che un agente software interagisca in modo iterativo e continuo con l'ambiente produttivo, apprendendo progressivamente una strategia decisionale ottimale basata sul feedback ricevuto sotto forma di ricompense o penalità. In letteratura, il controllo CONWIP è stato formalizzato come un problema decisionale sequenziale in cui l'agente di apprendimento per rinforzo RL osserva lo stato corrente del sistema, rappresentato da variabili quali i livelli di WIP, i backorder e le autorizzazioni temporanee aggiuntive. In base a queste informazioni, l'agente decide se mantenere la politica di rilascio standard ("do nothing"), concedere autonomamente un'autorizzazione supplementare alla produzione oppure ritirarne una nel momento in cui un prodotto lascia il buffer finale. Attraverso numerose simulazioni ad eventi discreti, l'agente apprende a ottimizzare la sequenza di decisioni al fine di minimizzare i costi complessivi nel lungo termine, incluse le spese legate al mantenimento di scorte elevate e i costi derivanti da backorder o ritardi. Questo paradigma si distingue per la capacità di adattarsi in maniera dinamica e proattiva alle condizioni variabili del sistema produttivo, rappresentando un'evoluzione significativa rispetto ai tradizionali metodi statici di controllo basati su regole fisse o soglie predeterminate [17].

I risultati sperimentali indicano che il Reinforcement Learning offre prestazioni superiori in tutti gli scenari considerati, adattandosi efficacemente a condizioni caratterizzate da elevata variabilità e carichi intensi. Tuttavia, richiede una significativa calibrazione dei parametri e implica costi computazionali rilevanti. Il Generalized Kanban si posiziona al secondo posto, superando il Dynamic CONWIP e il modello CONWIP tradizionale grazie alla sua flessibilità operativa. Il Dynamic CONWIP presenta miglioramenti modesti rispetto alla versione classica, ma rimane meno efficace rispetto agli altri approcci adattivi.

Da quanto emerso si può affermare che il Reinforcement Learning è la soluzione più potente per sistemi complessi con elevate variabilità e carichi pesanti. Il Generalized Kanban rappresenta invece un'alternativa pratica e più facilmente implementabile nei contesti caratterizzati da una domanda moderata. Da qui emerge una regola empirica: adottare l'RL in ambienti ad alta complessità e utilizzare il Generalized Kanban come soluzione robusta per scenari meno critici.

Infine, gli studi indicano che il futuro delle strategie CONWIP dinamiche si orienterà verso una maggiore integrazione tra simulazione e intelligenza artificiale. Questi strumenti serviranno come supporto decisivo per definire politiche stazionarie ottimali da applicare direttamente nei sistemi reali, riducendo i rischi e i costi associati alla sperimentazione diretta sugli impianti produttivi [17].

2.3 Deadlock e congestione in sistemi con loop e buffer limitati

Nei sistemi produttivi e logistici modellati come processi a eventi discreti, il fenomeno del deadlock rappresenta una delle principali criticità operative. Questo si verifica quando un gruppo di entità, come job o pallet, resta bloccato in una condizione di attesa reciproca circolare, impedendo al sistema di continuare le operazioni. Questa problematica risulta particolarmente significativa nei sistemi con loop chiusi, buffer limitati e instradamenti circolari, dove l'uso sequenziale delle risorse condivise può portare a stalli irreversibili, spesso risolvibili solo tramite interventi esterni. I deadlock rappresentano un elemento altamente indesiderato poiché riducono in maniera consistente l'efficienza nell'utilizzo delle risorse e, in contesti critici, possono mettere a rischio la sicurezza operativa.

Dal punto di vista teorico, il deadlock si manifesta quando si verificano simultaneamente le quattro condizioni definite da Coffman:

- Mutua esclusione: una risorsa può essere utilizzata esclusivamente da un solo processo alla volta. Questo fenomeno è comune nei sistemi produttivi, dove

macchine, robot o attrezzature non possono essere condivise contemporaneamente tra più operazioni.

- Hold-and-wait: un processo che possiede già alcune risorse può richiederne altre senza rilasciare quelle già possedute.
- Assenza di preemption: le risorse detenute da un processo non possono essergli sottratte forzatamente; esse possono essere liberate soltanto al termine dell'uso o attraverso un rilascio volontario da parte del processo stesso.
- Attesa circolare: si forma una catena chiusa di processi in cui ciascun processo è in attesa di una risorsa posseduta dal processo successivo, generando un ciclo di dipendenze che porta al deadlock.

Nei sistemi manifatturieri automatizzati, le prime tre condizioni che favoriscono il verificarsi di deadlock sono strettamente connesse alle proprietà fisiche delle risorse e, pertanto, risultano difficilmente eliminabili. Per prevenire i deadlock, l'intervento deve concentrarsi sull'attesa circolare, avvalendosi di politiche specifiche di schedulazione e supervisione.

La letteratura identifica quattro strategie principali per affrontare questo problema:

1. *Deadlock ignoring*: utilizzata quando la probabilità di blocco è estremamente bassa. In tali casi, il problema viene deliberatamente ignorato per ridurre la complessità del sistema. È una scelta applicabile in situazioni in cui il rischio è accettabile e i costi di prevenzione sarebbero troppo elevati.
2. *Deadlock prevention*: si basa su modifiche strutturali al sistema o sulla ridefinizione delle regole di gestione delle risorse, con l'obiettivo di evitare a priori qualsiasi condizione di attesa circolare. Questo approccio garantisce l'assenza di deadlock, ma comporta una diminuzione della flessibilità operativa e del livello di concorrenza.
3. *Deadlock avoidance*: permette l'allocazione delle risorse solo se il nuovo stato risultante è considerato “safe”, ossia tale da assicurare almeno una sequenza completabile di operazioni. Tuttavia, questa strategia richiede un monitoraggio continuo dello stato globale, comportando notevoli costi computazionali.
4. *Deadlock detection and recovery*: consente di rilevare il deadlock una volta che si verifica e di adottare misure correttive, come la terminazione dei processi coinvolti o il rilascio forzato delle risorse bloccate. È una soluzione indicata quando i deadlock sono eventi rari e i costi associati alla loro gestione risultano sostenibili.

[21].

Le reti di Petri rappresentano uno strumento matematico e di modellazione particolarmente efficace per analizzare e gestire i deadlock all'interno dei sistemi produttivi a eventi discreti, caratterizzati dalla condivisione di risorse, buffer limitati e strutture cicliche. Questo approccio consente di visualizzare in modo sia grafico che analitico lo stato delle risorse, dei processi e delle transizioni, fornendo una prospettiva chiara sui concetti di concorrenza, sulle limitazioni dei buffer e sulla sincronizzazione operativa. Un esempio di rete di Petri è illustrato in Figura 7: il modello si configura come un grafo bipartito composto da posti (che rappresentano risorse o condizioni di stato) e transizioni (che rappresentano eventi o attività). I *token*, distribuiti nei posti, descrivono la marcatura del sistema, ossia la configurazione corrente delle risorse.

A partire da tale rappresentazione è possibile costruire la matrice di incidenza, che descrive le relazioni tra posti e transizioni. L'analisi di questa matrice consente di individuare invarianti di posto e invarianti di transizione, strumenti utili per verificare proprietà di conservazione, raggiungibilità e assenza di deadlock nel sistema produttivo modellato.

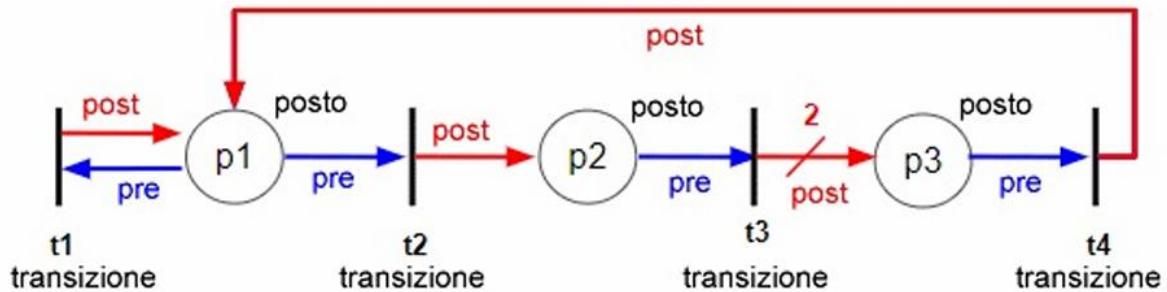


Figura 7: esempio di rete di Petri per la rappresentazione dei flussi produttivi
Fonte: Andrea Minini, “Le reti di Petri”

Grazie alle loro caratteristiche, le reti di Petri permettono non solo di identificare formalmente le condizioni che determinano i deadlock, ma anche di progettare sistemi di controllo capaci di prevenire tali problematiche. Si rivelano quindi uno strumento versatile per la modellazione e il controllo, applicabile in tutte e quattro le principali strategie di gestione dei deadlock: ignorare il problema, prevenirne l'insorgenza, evitare il blocco tramite scelte operative mirate o rilevare e recuperare il sistema una volta che il problema si è manifestato. Questa flessibilità consente di affrontare i deadlock con un grado di complessità computazionale che varia in funzione delle dimensioni e della struttura del sistema considerato [18] [19] [20].

Per gestire la complessità dei sistemi con flussi produttivi paralleli e una significativa condivisione di risorse, le tradizionali reti di Petri possono essere perfezionate attraverso l'utilizzo delle *Parallel Process Nets with Resources* (PPNR). Questo metodo avanzato permette di modellare e analizzare in profondità le dinamiche tra processi paralleli e risorse limitate, facilitando sia l'identificazione che la gestione efficace dei deadlock su scala estesa. Le PPNR separano il flusso dei processi dalla gestione delle risorse, rappresentandoli rispettivamente attraverso posti e transizioni, che indicano le operazioni eseguite e l'assegnazione delle risorse. Tale struttura permette di simulare il comportamento dei job lungo percorsi paralleli, identificando situazioni in cui più processi competono per le stesse risorse, con il rischio di blocchi operativi.

Per rilevare i deadlock senza analizzare ogni possibile *siphon*, le PPNR sfruttano i *transition vectors* (TVs). Questi vettori rappresentano in modo sistematico il flusso di token all'interno della rete, traducendo lo stato dei processi e delle risorse in una mappa comprensibile. Grazie ai TV, è possibile identificare con precisione situazioni di attesa circolare o blocchi imminenti. Una volta individuato un potenziale deadlock, il sistema può attivare procedure di recupero, come il rilascio controllato di risorse o la riprogrammazione dei job, prevenendo interruzioni totali nel sistema.

L'integrazione delle PPNR con i transition vectors offre un approccio efficace per gestire flussi produttivi complessi, ottimizzando l'utilizzo delle risorse, riducendo i tempi di fermo e assicurando la continuità operativa, anche in ambienti su larga scala [22].

Nel caso di sistemi caratterizzati da loop chiusi, caroselli e buffer limitati, i rischi di congestione e di deadlock tendono a intersecarsi in modo significativo. La saturazione dei buffer locali, infatti, può propagarsi fino a determinare uno stallo globale del sistema. Per affrontare tali criticità, è possibile ricorrere a diversi approcci di modellazione e controllo, come le reti di Petri, gli *Event Relationship Graphs* (ERG) o modelli di simulazione a eventi discreti, che consentono di analizzare la propagazione dei blocchi e di progettare politiche di supervisione capaci di garantire la continuità operativa e ridurre l'impatto delle congestioni sul throughput complessivo.

2.4 Scheduling e prioritizzazione in workstation parallele

Nelle workstation con più macchine parallele identiche la definizione di regole di scheduling adeguate è fondamentale per massimizzare il throughput e garantire un uso bilanciato delle risorse.

Le regole di scheduling sono generalmente suddivise in due principali categorie: regole di priorità e regole di assegnazione.

Le prime vengono applicate quando una singola macchina deve determinare l'ordine di lavorazione dei job presenti nella propria coda. Tra le principali troviamo:

- FIFO (*First In First Out*): i job vengono processati seguendo rigorosamente l'ordine di arrivo.
- SPT (*Shortest Processing Time*): privilegia i job con tempi di lavorazione più brevi, ideale per minimizzare il tempo medio di completamento.
- EDD (*Earliest Due Date*): favorisce i job con scadenze più ravvicinate.
- MST (*Minimum Slack Time*): dà priorità ai job il cui slack, ossia la differenza tra la scadenza residua e il tempo di lavorazione necessario, è minimo, concentrandosi sui più urgenti.
- CR (*Critical Ratio*): utilizza un indice calcolato considerando il rapporto tra il tempo residuo alla scadenza e il tempo di lavorazione richiesto, individuando i job maggiormente critici.
- LRPT (*Longest Processing Time*): assegna prioritariamente i job con tempi di lavorazione più lunghi, contribuendo a ridurre il makespan globale.
- SRPT (*Shortest Processing Time*): indirizza i job con tempi più brevi verso le macchine disponibili, facilitando la diminuzione dei tempi medi di completamento.

Le regole di assegnazione, invece, intervengono quando un job può essere eseguito su diverse macchine parallele identiche e si deve stabilire quale macchina specifica debba occuparsi del lavoro. Le più comuni includono:

- *Least Loaded*: assegna il job alla macchina con minor carico residuo, promuovendo un bilanciamento dinamico delle risorse.
- *Round Robin*: distribuisce i job in maniera ciclica tra le macchine, privilegiando semplicità d'implementazione ma senza ottimizzazione mirata.
- *Random*: smista i job casualmente, spesso utilizzata come baseline per analisi comparative.

La scelta della regola effettuata influenza direttamente i parametri chiave come throughput, bilanciamento del carico e tempi di completamento, variando in base agli obiettivi specifici di ottimizzazione perseguiti. Studi recenti evidenziano l'efficienza delle regole LRPT e SRPT nella riduzione rispettiva del tempo di completamento totale. In contesti dove

l'equilibrio dinamico tra le risorse è prioritario, invece, le regole come "Least Loaded" trovano una loro applicazione ideale.

2.5 Sequenziamento e gestione delle rilavorazioni nella stazione WSQC

I controlli e le rilavorazioni rivestono un ruolo fondamentale all'interno delle linee produttive poiché garantiscono il mantenimento di alti standard di qualità e contribuiscono a minimizzare sprechi e difetti. In questo contesto si inserisce la stazione di controllo qualità (WSQC), che ha il compito di verificare la conformità dei prodotti e, in caso di difetto, indirizzarli verso le opportune operazioni di rilavorazione. In tal modo si assicura che ogni unità soddisfi i requisiti richiesti dal processo produttivo.

Lo scopo principale è intercettare e correggere eventuali non conformità prima che il prodotto arrivi al cliente finale, evitando problemi di utilizzo o insoddisfazione. Per conseguire questo obiettivo, le aziende implementano sistemi di monitoraggio e controlli distribuiti lungo il processo produttivo, permettendo di rilevare i difetti in fase precoce e agire rapidamente per risolverli [24].

Accanto a queste soluzioni tecnologiche impiegate nei sistemi produttivi, un metodo particolarmente diffuso è il *routing probabilistico* applicato alla WSQC. Questo meccanismo, utilizzato nel controllo qualità condizionato, assegna a ogni prodotto una probabilità specifica di essere accettato, scartato oppure sottoposto a rilavorazione dopo l'ispezione. Basandosi su modelli statistici o empirici, questo approccio gestisce i flussi di prodotti difettosi in maniera più dinamica e realistica, favorendo l'ottimizzazione di costi e tempi legati al processo di rilavorazione.

Rispetto all'approccio deterministico, il routing probabilistico presenta numerosi benefici. Anzitutto consente di descrivere la realtà in maniera più accurata, simulando itinerari alternativi per i prodotti difettosi in base alla probabilità di rilavorazione o scarto. Ciò permette di valutare in modo più preciso il carico di lavoro delle stazioni coinvolte, ottimizzando così la distribuzione delle risorse disponibili. La flessibilità intrinseca del modello ne consente inoltre l'adattamento rapido a variazioni nella qualità del processo o nelle prestazioni dei macchinari, offrendo una maggiore reattività del sistema ai cambiamenti.

Gestire in chiave probabilistica i percorsi dei prodotti non conformi risulta particolarmente vantaggioso anche per bilanciare meglio i carichi di lavoro, minimizzare i colli di bottiglia e favorire la fluidità del flusso dei materiali. Questo approccio comporta un doppio beneficio:

una riduzione dell'accumulo di WIP e dei tempi di attraversamento più contenuti, contribuendo al funzionamento più armonioso ed efficiente dell'impianto.

È importante sottolineare che questi benefici non si manifestano automaticamente. La loro utilità dipende dalla definizione accurata delle probabilità e dalla stretta integrazione del modello con la logica di controllo e lo scheduling della produzione. Di conseguenza, questo approccio non garantisce automaticamente una riduzione dei tempi o dei costi, ma rappresenta uno strumento più realistico e informativo rispetto al routing deterministico [25].

All'interno della WSQC, tuttavia, la gestione del flusso dei pezzi e il loro sequenziamento assumono un ruolo determinante. L'ordine con cui i job vengono ispezionati o rilavorati influisce direttamente sui tempi di attesa, sulla saturazione della stazione e sulla stabilità del flusso produttivo complessivo. Il problema del sequenziamento nella WSQC può quindi essere trattato come un problema di scheduling operativo, in cui l'obiettivo è bilanciare i tempi di ispezione e rilavorazione ottimizzando l'utilizzo delle risorse.

L'ottimizzazione del sequenziamento presso la stazione WSQC rappresenta un aspetto cruciale per garantire la regolarità del flusso produttivo. Poiché l'ordine con cui i job vengono ispezionati e rilavorati influisce direttamente sui tempi di attesa e sulla saturazione della stazione, la gestione del sequenziamento può essere affrontata come un problema di scheduling operativo.

In letteratura, questo tipo di problematica è stato analizzato attraverso diversi approcci di ottimizzazione. Nei modelli più strutturati si fa ricorso a metodi esatti, come la programmazione dinamica o la programmazione intera, che consentono di individuare la sequenza ottimale delle operazioni in base a vincoli di capacità, priorità o tempi massimi di ciclo. Tuttavia, tali metodi risultano rapidamente complessi e difficilmente applicabili a sistemi produttivi di grandi dimensioni.

Per gestire scenari più realistici e variabili, vengono spesso impiegate tecniche euristiche o metaeuristiche (ad esempio algoritmi genetici o simulated annealing), integrate con strumenti di simulazione a eventi discreti per valutare gli effetti delle decisioni di sequenziamento su indicatori come il WIP o i tempi di attraversamento.

L'obiettivo comune a tutti questi approcci è trovare un equilibrio tra il numero di ispezioni necessarie per assicurare la qualità del prodotto e la minimizzazione dei periodi di inattività o del sovraccarico delle stazioni WSQC. Un sequenziamento ottimale dei pezzi permette

infatti di ridurre i colli di bottiglia, migliorare l'utilizzo delle risorse e assicurare la continuità operativa del sistema [26].

Il ciclo continuo ispezione → rilavorazione → uscita introduce una dinamica complessa:

- Incrementa il WIP perché i prodotti difettosi ritornano indietro nella linea per essere corretti.
- Può causare congestioni nella WSQC se il flusso di rilavorazioni è elevato.
- Aumenta i tempi di attraversamento e può ridurre il throughput complessivo se non gestito correttamente.
- Richiede un bilanciamento fra capacità della WSQC, risorse per la rilavorazione e limiti di inventario per mantenere l'efficienza produttiva.

Alla luce di quanto descritto, il controllo qualità e le regole di sequenziamento in WSQC permettono di gestire efficacemente i difetti, minimizzando l'impatto negativo sul flusso produttivo. Tuttavia, è fondamentale monitorare il loop ispezione-rilavorazione per evitare sovraccarichi e garantire un utilizzo ottimale delle risorse produttive.

2.6 Arena Simulation

La simulazione è uno strumento utile per comprendere e gestire la complessità dei sistemi produttivi perché attraverso un modello che riproduce il funzionamento del sistema reale, è possibile osservare come le diverse variabili evolvono nel tempo e come i vari elementi interagiscono tra loro. In questo modo diventa più semplice interpretare i meccanismi che regolano il comportamento complessivo del sistema e avere una visione più chiara di come potrebbe svilupparsi in futuro, facilitando così le decisioni operative e strategiche.

Uno dei vantaggi più rilevanti della simulazione è la possibilità di sperimentare scenari diversi senza dover intervenire direttamente sul sistema reale. Questo riduce i rischi economici, accelera i tempi di analisi e limita gli errori che potrebbero derivare da scelte organizzative non ottimali. A differenza della realtà produttiva, dove modifiche ed esperimenti possono risultare molto costosi o difficili da replicare, il modello simulato consente di riprodurre le stesse condizioni più volte e confrontarne facilmente i risultati. Inoltre, offre uno spazio sicuro per testare soluzioni innovative, senza compromettere la continuità della produzione, creando così le basi per introdurre con maggiore fiducia nuove strategie di miglioramento.

Si tratta di una simulazione a eventi discreti (*Discrete Event Simulation*, DES), uno dei tipi di simulazione più diffusi per la modellazione di sistemi dinamici. Questo approccio si basa sul meccanismo di avanzamento del tempo al prossimo evento: il tempo della simulazione non scorre in modo continuo, ma “salta” da un evento significativo al successivo. Durante l’esecuzione, viene mantenuta una lista di eventi futuri, che raccoglie tutti gli eventi conosciuti in ordine di occorrenza. Ogni volta che è necessario un avanzamento, l’evento con il tempo minimo viene estratto dalla lista, e l’orologio della simulazione viene aggiornato a tale istante. La lista deve quindi essere costantemente aggiornata e ordinata, affinché l’evento più imminente sia sempre elaborato per primo.

Il processo tipico di una simulazione a eventi discreti può essere descritto in una serie di passaggi fondamentali:

1. Si imposta l’orologio della simulazione a zero.
2. Si inizializzano tutte le variabili e lo stato del sistema.
3. Si determinano i tempi di arrivo iniziali per ciascun flusso di entità e si inseriscono tali entità nella lista degli eventi futuri.
4. Si rimuove l’evento in cima alla lista e si aggiorna l’orologio simulato al tempo corrispondente.
5. Se l’evento è un arrivo, si genera la prossima entità e la si inserisce nella lista; se si tratta dell’evento di stop, la simulazione termina.
6. Si aggiornano lo stato del sistema e le statistiche necessarie per la descrizione desiderata.
7. Si ritorna al passo 4 fino a quando la lista degli eventi futuri non risulta vuota.
8. Una volta conclusa la simulazione, si calcolano le statistiche finali e si producono i risultati.

La simulazione a eventi discreti, dunque, permette di rappresentare con precisione il comportamento temporale di un sistema complesso, seguendo l’evoluzione degli eventi nel tempo e aggiornando lo stato del sistema in funzione delle transizioni che si verificano [27].

Con Arena, questa metodologia consente di modellare non solo le workstation, ma anche le attività degli operatori, la logistica interna e le interazioni tra le stazioni. Questo approccio genera indicatori quantitativi come WIP medio, throughput, costi per unità e reattività del

sistema. Grazie a una visione globale e dinamica, è possibile valutare l'efficacia delle strategie adottate e individuare soluzioni capaci di integrare efficienza, qualità e flessibilità nel panorama produttivo.

3. DESCRIZIONE DEL PROBLEMA

3.1 Descrizione del layout del sistema CONWIP modellato

Il sistema produttivo analizzato è costituito da sei workstation principali (WS1–WS6), da una stazione finale di controllo qualità (WSQC) e da due baie, rispettivamente di carico e di scarico, da cui i job entrano ed escono dal processo. Tutte le workstation sono configurate come mono macchina, ad eccezione di WS2 e WS5, che dispongono invece di due macchine identiche in serie per aumentare la capacità di lavorazione.

Il flusso inizia con i pallet vuoti, il cui numero è determinato dal parametro di livello di WIP imposto dalla logica CONWIP. I pallet vengono caricati con il pezzo grezzo nella baia di ingresso e poi instradati verso la WS1, dove inizia la sequenza delle lavorazioni. Tale sequenza prevede il passaggio progressivo attraverso tutte le workstation, fino alla WS5, nella quale i pezzi devono essere lavorati per due volte consecutive da una qualsiasi delle due macchine identiche disponibili. Successivamente, i pallet proseguono verso la WS6 per l'ultima fase di lavorazione. In coda al processo si trova la WSQC, stazione dedicata al controllo qualità finale.

La movimentazione è garantita da un nastro trasportatore che collega in sequenza tutte le workstation, dalla baia di carico fino all'uscita.

Un aspetto molto importante riguarda la gestione delle attese tramite caroselli (*loop*). Qualora la stazione successiva non sia disponibile, infatti, i job vengono temporaneamente trattenuti all'interno del loop corrispondente, rimanendo in circolazione fino a quando la risorsa non si libera. In questo modo, il sistema implementa un meccanismo di sincronizzazione che evita congestioni a monte e consente di mantenere il flusso continuo.

Il percorso dei job è guidato da nodi di smistamento (*splitter*) e di confluenza (*merger*). Gli splitter hanno il compito di suddividere i flussi, distribuendo i job verso le diverse macchine disponibili o verso percorsi alternativi previsti dal layout. I merger, invece, riuniscono più rami del *conveyor* in un unico flusso, consentendo ai job di proseguire secondo la sequenza prestabilita. Questi elementi, combinati con le regole di instradamento e con le probabilità di invio alla WSQC, determinano l'evoluzione del percorso di ciascun job lungo l'intero ciclo produttivo.

Infine, una volta che il job ha completato tutte le lavorazioni richieste (inclusi eventuali controlli e rilavorazioni), esso esce dal sistema attraverso la baia di scarico. In parallelo, il pallet vuoto che lo ha trasportato viene automaticamente ricondotto all'area iniziale, dove si

incolonna nuovamente insieme agli altri pallet disponibili. Questo meccanismo consente di mantenere costante il livello di Work In Process, come imposto dalla logica CONWIP, e garantisce la disponibilità di pallet per i job successivi in ingresso.

La struttura complessiva del modello di simulazione, comprendente le workstation, i moduli di controllo dei flussi e i loop di rilavorazione, è illustrata in Figura 8.

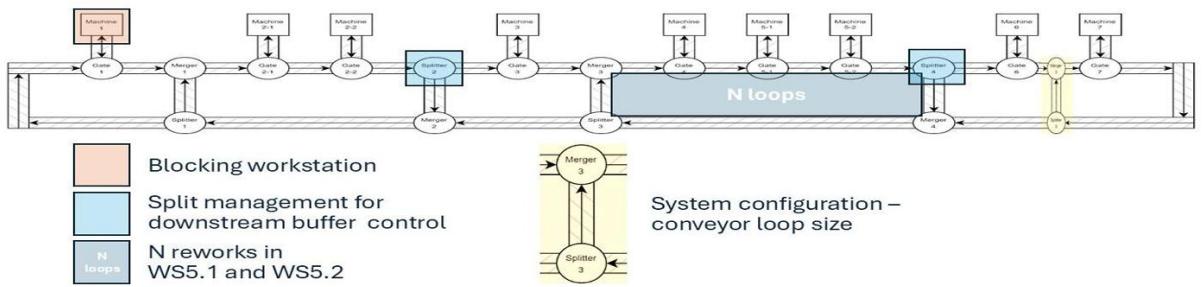


Figura 8: struttura del modello di simulazione.
Fonte: documentazione del caso di studio fornita dal relatore (2025).

3.2 Logiche operative

Le logiche operative definiscono il comportamento dinamico del sistema e regolano il flusso dei job attraverso le workstation, il conveyor e la WSCC. Esse combinano vincoli fisici, regole di priorità e probabilità di rilavorazione, assicurando che il modello riproduca fedelmente le dinamiche reali della linea produttiva.

Nel sistema, la WS1 opera secondo la regola del *blocking after service*, secondo la quale un job completato non può lasciare la workstation finché la workstation successiva non è disponibile. Di conseguenza, il pezzo rimane nella WS1, bloccando momentaneamente il flusso, fino a quando almeno una delle due macchine della WS2 risulta libera. Questo sistema aiuta a prevenire la saturazione delle aree successive e assicura una sincronizzazione efficace del flusso produttivo. L'approccio permette di mantenere stabile il livello di WIP lungo la linea e di ridurre il rischio di congestioni immediate, anche se può generare ritardi nei tempi di attraversamento complessivi.

In modo analogo, i job lavorati dalla WS2 rimangono nel carosello (loop) che circonda la stazione finché la WS3 non risulta disponibile. Solo quando quest'ultima si libera, il job successivo può lasciare il loop e procedere alla lavorazione successiva. Questa logica di accumulo locale consente di gestire in modo ordinato le code intermedie, evitando che il materiale si accumuli eccessivamente nei tratti a valle del sistema.

Un comportamento simile si osserva nella parte finale della linea: i job che hanno completato le due lavorazioni richieste presso la WS5 procedono verso la WS6 solo quando quest'ultima è libera; in caso contrario, restano temporaneamente nel carosello della WS5. Tale meccanismo di attesa controllata garantisce una migliore sincronizzazione tra le due stazioni, prevenendo situazioni di blocco dovute all'indisponibilità di buffer intermedi e contribuendo al mantenimento della stabilità del flusso complessivo.

Ogni workstation, dalla WS2 alla WS6, ha una probabilità α di inviare il job appena lavorato alla stazione di controllo qualità WSQC. Una volta arrivati alla WSQC, i job vengono ispezionati e possono seguire tre percorsi distinti:

- 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.

Questi valori probabilistici consentono di rappresentare realisticamente la variabilità della qualità dei prodotti e le decisioni legate alla rilavorazione. In questo modo si influenzano direttamente il carico delle workstation, il flusso dei job tra le stazioni e il livello di WIP lungo l'intera linea produttiva.

Nel complesso, le logiche operative combinano regole fisiche, probabilità di rilavorazione e requisiti di lavorazione multipla per garantire un funzionamento coerente e controllato del sistema, consentendo di valutare in maniera quantitativa l'effetto delle strategie di instradamento e di controllo qualità sulle performance complessive della linea produttiva.

3.3 Obiettivo tecnico: controllo dinamico del WIP per massimizzare il throughput e ridurre i deadlock

L'analisi si propone di esaminare il comportamento del sistema produttivo gestito secondo la logica CONWIP, valutandone le prestazioni operative in termini di throughput, livelli medi di WIP e utilizzo delle workstation. L'obiettivo è osservare come la configurazione del sistema, caratterizzata da loop chiusi, buffer limitati e macchine parallele, influenzi la fluidità del flusso e il rischio di congestione. L'approccio adottato consente di analizzare in modo dettagliato le dinamiche interne del processo e di individuare eventuali condizioni critiche che possono compromettere l'efficienza complessiva del sistema.

Uno degli aspetti chiave riguarda il controllo dinamico del Work in Progress, cioè il numero totale di pallet che circolano contemporaneamente all'interno del sistema. Nel modello CONWIP, il WIP è soggetto a un limite massimo stabilito dal numero di pallet disponibili, noto come livello critico di WIP. Il valore del WIP critico è una variabile di controllo fondamentale: se troppo basso, porta a una sottoutilizzazione delle risorse produttive e una riduzione del throughput complessivo; se troppo alto, invece, può saturare caroselli e buffer intermedi, incrementando il rischio di congestione e potenzialmente causando deadlock.

Il controllo dinamico del WIP consiste nell'adattare il numero di pallet disponibili in base allo stato corrente del sistema. Quando i dati indicano che la linea si sta comportando fluidamente, senza colli di bottiglia evidenti, è possibile aumentare il numero di pallet per massimizzare la capacità produttiva. Al contrario, nelle situazioni in cui un'eccessiva concentrazione di job si verifica in uno dei caroselli, il sistema deve essere capace di ridurre temporaneamente il WIP, favorendo lo smaltimento dei job già presenti e diminuendo il rischio di blocchi.

La scelta della frequenza con cui aggiornare i limiti di WIP rappresenta un aspetto delicato per l'equilibrio del sistema. L'intervallo deve essere sufficientemente rapido per cogliere le variazioni più significative, come oscillazioni nei flussi o cambiamenti della capacità, ma non così breve da generare instabilità dovuta a continue modifiche.

Un aggiornamento troppo frequente rischia infatti di rendere il sistema instabile, con oscillazioni e variazioni difficili da gestire; al contrario, un aggiornamento troppo lento non consente di reagire in tempo a situazioni di congestione o rallentamento. Per questo motivo, una soluzione pratica consiste nell'adottare un intervallo di aggiornamento regolare, ad esempio orario, oppure legato a eventi particolari osservati nel sistema.

In fase applicativa si può partire da un intervallo di riferimento e, attraverso l'osservazione del comportamento del sistema, adattarlo progressivamente fino a individuare un compromesso soddisfacente tra reattività e stabilità. Nella maggior parte dei casi, valori compresi tra alcune ore e una giornata garantiscono un controllo sufficientemente adattivo senza introdurre eccessiva variabilità.

3.4 Ipotesi modellistiche e semplificazioni

Per rendere il modello implementato in Arena più chiaro e focalizzato sugli aspetti chiave dello studio, sono state introdotte alcune ipotesi semplificative e approcci modellistici essenziali. In particolare, i tempi di processo delle workstation sono stati rappresentati

mediante distribuzioni triangolari, allo scopo di simulare in modo realistico la variabilità operativa, senza però perdere la facilità di parametrizzazione. Per le workstation principali è stata adottata una distribuzione di tempi di processo con valori medi e una variabilità contenuta, in modo da rappresentare un ciclo di lavorazione regolare e uniforme. La workstation dedicata al controllo qualità, invece, è stata modellata con tempi di processo più lunghi e caratterizzata da una variabilità maggiore rispetto alle altre macchine, a riflettere la natura più complessa e articolata delle attività di ispezione e verifica rispetto alle lavorazioni standard.

Inoltre, i tempi accessori relativi alla movimentazione dei job (discesa dal pallet, ingresso e uscita dalle macchine, risalita sul pallet) sono stati trascurati, ipotizzando che il loro impatto sia marginale rispetto ai tempi effettivi di lavorazione.

Relativamente alle workstation, è stato supposto che ogni risorsa operi continuamente senza interruzioni dovute ad attrezzaggi, manutenzioni pianificate o cambi di setup tra differenti job. Questo consente al modello di rappresentare un flusso produttivo uniforme e continuo.

Il controllo qualità è stato inoltre modellato come un'unica workstation caratterizzata da regole di sequenziamento. Non è stata fatta alcuna distinzione tra tipologie di difetti o livelli di gravità; invece, si è utilizzata una rappresentazione aggregata basata su probabilità di re-routing, considerate note e costanti nel tempo.

La lunghezza dei nastri trasportatori è stata stimata in modo da mantenere una disposizione equilibrata e pressoché equidistante tra le diverse workstation. Questa scelta consente di rappresentare un flusso di movimentazione regolare e realistico, evitando che le distanze influenzino in modo significativo i tempi complessivi di attraversamento o generino accumuli non rappresentativi nelle sezioni intermedie del sistema.

È stato considerato un turno operativo di otto ore, ipotizzando un ciclo produttivo continuo all'interno di questa finestra temporale.

Tale scelta consente di rappresentare una giornata lavorativa tipica, sufficiente per osservare l'evoluzione del sistema e il raggiungimento di un comportamento stabile, mantenendo al tempo stesso una durata di simulazione realistica e coerente con i contesti industriali di riferimento.

È importante sottolineare che il modello non tiene conto di variabilità esterne. Non sono stati inclusi guasti alle macchine, assenze di personale, ritardi nelle forniture o fluttuazioni della domanda dei clienti. L'intero sistema viene quindi analizzato in condizioni di regime

controllato, permettendo uno studio più dettagliato degli effetti delle regole di priorità e del meccanismo di controllo del WIP sul throughput e sulla probabilità di deadlock.

4. CONFIGURAZIONE E SPERIMENTAZIONE DEL MODELLO DI SIMULAZIONE

4.1 Descrizione concettuale del modello di simulazione

Il modello di simulazione è stato sviluppato sulla base delle logiche operative illustrate nel capitolo precedente, focalizzandosi in particolare sull'integrazione di un vincolo globale di tipo CONWIP, che regola il numero massimo di pallet in circolazione all'interno del sistema. La presenza di loop produttivi con caroselli, soggetti a condizioni di blocco, unitamente ai requisiti di lavorazione multipla specifici di WS5 e al sistema di controllo qualità WSQC, che introduce probabilità di rilavorazione, contribuisce a rendere il sistema particolarmente vulnerabile a fenomeni di congestione e potenziali situazioni di deadlock.

Per contrastare efficacemente tali criticità, è stato progettato un meccanismo di controllo dinamico del WIP basato su un sistema di feedback continuo, articolato in due moduli principali:

1. Valutazione del rischio di deadlock: il sistema monitora l'occupazione relativa di ciascun anello produttivo tramite il calcolo dell'indicatore di *Ring Occupancy*, definito come il rapporto tra il numero di unità presenti nel singolo anello e la capacità massima, quest'ultima determinata considerando la lunghezza complessiva dei pallet e dei nastri trasportatori associati.
2. Monitoraggio del throughput rispetto a un target di riferimento: il sistema rileva costantemente il throughput corrente e lo confronta con un valore target prestabilito, in modo da attuare politiche adattive sull'ingresso di nuovi job.

Il modello applica quindi una regolazione dinamica del limite massimo di WIP al termine di ogni intervallo orario di simulazione, attraverso le seguenti logiche decisionale:

- Se almeno uno tra gli anelli critici supera il 90% di occupazione, il limite massimo di WIP disponibile per il sistema viene ridotto, al fine di evitare che la congestione peggiori e si raggiungano situazioni di blocco irreversibile.
- Se, viceversa, tutti e tre gli anelli mantengono un'occupazione inferiore al 60%, il limite di WIP viene incrementato di un'unità, permettendo un maggior flusso di pallet e sostenendo quindi il throughput.

- Se non si verifica nessuna delle due condizioni precedenti, il limite massimo di WIP rimane invariato, mantenendo stabile il livello di circolazione interna senza introdurre variazioni non necessarie.

La scelta di aggiornare il limite dinamico ogni ora deriva da un equilibrio tra reattività del sistema e stabilità operativa. Un intervallo temporale di aggiornamento troppo breve potrebbe generare variazioni frequenti e instabili del limite WIP, causando oscillazioni eccessive nel flusso produttivo e potenziale inefficienza. Al contrario, un intervallo troppo lungo rischierebbe di rendere il sistema poco reattivo ai mutamenti delle condizioni operative, ritardando interventi correttivi necessari per prevenire congestioni o deadlock. L'aggiornamento orario rappresenta quindi un compromesso efficace, sufficientemente frequente per permettere un monitoraggio tempestivo e l'applicazione di correttivi, ma al contempo in grado di mantenere un funzionamento stabile e coerente del sistema nel tempo.

Sul fronte del controllo del throughput, la logica ibrida adottata prevede:

- In presenza di un throughput inferiore al target, l'ingresso di nuovi job nel sistema è sempre consentito, indipendentemente dall'attuale livello di WIP. Questo favorisce il riempimento e il riavvio del flusso produttivo.
- Qualora il throughput sia pari o superiore al target, l'accesso di nuovi job è subordinato al rispetto del limite dinamico di WIP calcolato in precedenza. Se il limite è raggiunto, nuovi job rimangono in attesa, prevenendo un eccessivo accumulo e un potenziale deadlock.

Questa configurazione, schematizzata in Figura 9, implementa un controllo proattivo sulla saturazione dei caroselli produttivi, combinato con un meccanismo adattivo basato sulle performance operative del sistema. Tale strategia consente di bilanciare efficacemente due obiettivi fondamentali: massimizzare il throughput di sistema, evitando al contempo di incorrere in deadlock causati da congestionamenti eccessivi nei loop di produzione.

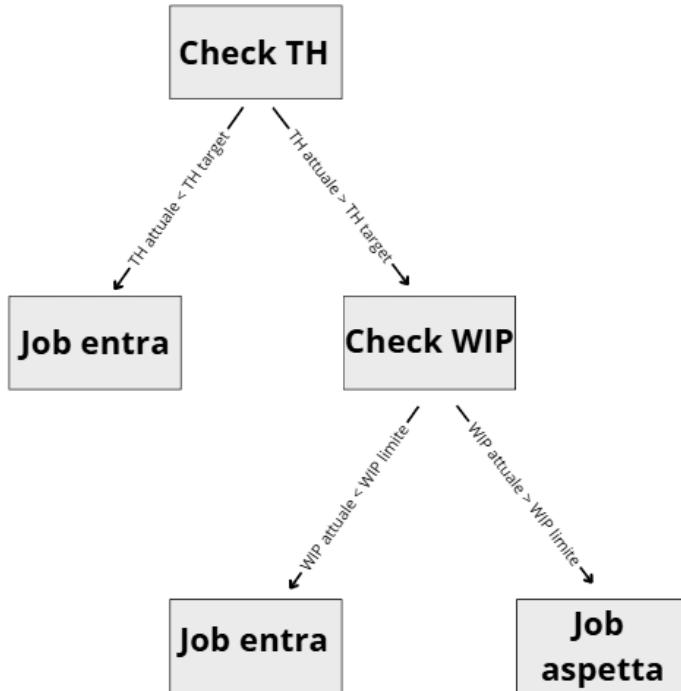


Figura 9: schema logico semplificato della logica di controllo dei job nel sistema

L’obiettivo principale dell’aggiornamento dinamico del livello di WIP è triplice: massimizzare il throughput complessivo del sistema, ridurre il rischio di deadlock dovuto alla congestione nei loop produttivi e ottimizzare l’utilizzo delle macchine.

Inoltre, si mira a contenere il *flow time*, cioè il tempo medio di attraversamento dei job nel sistema, con effetti positivi sia sulla velocità di lavorazione che sulla fluidità complessiva del processo produttivo.

Questi obiettivi sono fondamentali per garantire un equilibrio efficace tra efficienza produttiva e stabilità operativa in un sistema complesso come il CONWIP con rilavorazioni e loop soggetti a blocchi, migliorando così le performance quantitativamente rilevanti e la robustezza gestionale del sistema stesso.

4.2 Parametri e condizioni di simulazione

Ai fini della sperimentazione sono stati definiti i parametri numerici di input che caratterizzano i tempi di processo, le probabilità di rilavorazione e il livello iniziale del WIP, insieme alle condizioni di partenza per gli esperimenti.

In particolare, per la modellazione dei tempi di processo delle workstation principali si è adottata una distribuzione triangolare con parametri (15, 20, 25) minuti. Tale scelta consente di rappresentare un tempo medio approssimativo di 20 minuti, introducendo una variabilità

limitata che rispecchia l'incertezza operativa tipica senza generare una dispersione eccessiva nei valori simulati. La distribuzione triangolare è stata preferita per la sua efficacia nel contesto di dati limitati o stimati, in quanto consente di definire con facilità i valori minimo, modale e massimo, offrendo una rappresentazione flessibile e realistica dei tempi di ciclo. Per la workstation dedicata al controllo qualità, il modello prevede tempi di processo più lunghi e caratterizzati da maggiore variabilità, rappresentati tramite una distribuzione triangolare (25, 30, 40) minuti. Questo riflette la natura intrinsecamente più lenta e meno standardizzabile delle attività di ispezione e controllo qualità rispetto alle operazioni produttive standard, consentendo di simulare accuratamente le possibili fluttuazioni temporali di tali attività.

Relativamente alla gestione delle rilavorazioni, al termine di ogni lavorazione un job ha una probabilità α pari al 10% di essere indirizzato alla WSQC, ipotizzando che una percentuale minoritaria ma significativa di pezzi presenti difetti rilevanti che richiedono un'ispezione specifica. Questa probabilità consente di modellare un sistema realistico in cui il controllo qualità è frequente ma non sistematico su ogni job.

Dopo il processo di controllo qualità, si considerano tre potenziali esiti per ciascun job, con probabilità specifiche:

- Con probabilità $\gamma = 40\%$, il job viene rimandato alla workstation di provenienza per una rilavorazione, rappresentando quei difetti che richiedono la ripetizione parziale o totale di alcune operazioni.
- Con probabilità $\delta = 30\%$, il job è considerato idoneo e viene diretto alla workstation successiva, implicando una correzione efficace che non necessita di ulteriore rilavorazione.
- Con probabilità residua $1 - \gamma - \delta$ pari a 30%, il job viene scartato e rimosso dal sistema, rappresentando difetti critici non recuperabili.

Tali valori di probabilità sono stati calibrati per bilanciare realisticamente le tre possibili uscite, privilegiando gli scenari in cui i difetti risultano risolvibili (rilavorazione o correzione parziale), ma mantenendo una quota significativa di scarti per riprodurre la presenza concreta di errori critici nel processo produttivo.

La definizione del throughput target e del limite iniziale di CONWIP è stata guidata da considerazioni sulla capacità complessiva del sistema e sulla necessità di bilanciare

saturazione delle risorse e stabilità operativa. Il sistema, composto da sei workstation in serie e da una stazione di controllo qualità, presenta tempi medi di processo pari a circa venti minuti per job nelle workstation principali, con la sola WSQC caratterizzata da tempi più lunghi e variabili. La presenza di due macchine in alcuni punti (in particolare in WS2 e WS5) e la necessità di lavorazioni multiple nella WS5 contribuiscono a ridistribuire i carichi, mentre i nastri trasportatori e i caroselli introducono tempi accessori di movimentazione e il rischio di congestione.

Per la definizione del TH target, è stato considerato il comportamento della risorsa più critica del sistema, corrispondente alla Workstation 5, che costituisce il principale collo di bottiglia del flusso produttivo.

La WS5 rappresenta la fase più critica del processo, poiché ogni job deve svolgere due lavorazioni complessive presso questa stazione. Le due macchine in serie (WS5.1 e WS5.2) sono identiche, e ciascuna lavorazione può essere eseguita indifferentemente su una delle due, con probabilità pari al 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:

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

Tale valore rappresenta il limite teorico del throughput del sistema, poiché la WS5 costituisce la principale risorsa di sincronizzazione del flusso. In condizioni di equilibrio, il sistema non può quindi superare un tasso di completamento medio pari a circa 3 job/ora, che viene assunto come TH target per la calibrazione del limite CONWIP e la valutazione delle politiche di controllo.

Al di sopra di tale soglia, un aumento del WIP non si tradurrebbe in un incremento del throughput, ma in un progressivo aumento dei tempi medi di attraversamento e del rischio di congestione nelle stazioni a valle.

Per impostare il valore del limite CONWIP si segue un ragionamento basato sulla quantificazione del tempo medio di attraversamento minimo del job e sull'applicazione della legge di Little.

Per quantificare il tempo medio di attraversamento del sistema produttivo, si parte considerando che il job attraversa la WS1, WS3, WS4, WS6 con un tempo medio di lavorazione pari a 20 min per job su ciascuna macchina.

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

La WS2 è composta da due macchine identiche in parallelo, ciascuna con tempo medio di lavorazione pari a 20 min. Poiché le due risorse operano in parallelo, il tempo medio equivalente per job si dimezza.

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

Il tempo di processo sulla WS5 è invece:

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

Quindi il tempo medio di processing atteso è:

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

La probabilità di invio alla WSQC è $\alpha = 0,10$. L'ispezione richiede, in media, 30 min. Se l'esito è rilavorazione con probabilità $\gamma = 0,40$, assumiamo che la rilavorazione aggiunga un'ulteriore operazione da 20 min (semplificazione). Il contributo medio è quindi:

$$\begin{aligned} T_{QC_atteso} &= \alpha \times (30 + \gamma \times 20) = 0,10 \times (30 + 0,40 \times 20) = 0,10 \times (30 + 8) \\ &= 3,8 \text{ min.} \end{aligned}$$

Si stima un movimento medio totale di:

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

Sommo i contributi:

$$FT_{min} = T_{proc} + T_{QC_atteso} + T_{mov} = 130 + 3,8 + 40 = 174 \text{ min} \approx 2,9 \text{ h.}$$

Questo valore è il lower bound del flow time (esclude i tempi di attesa in coda dovuti al WIP).

Per includere un margine per code tollerabili, si assume:

$$FT_{desired} = 3,5 \text{ h} = 210 \text{ min.}$$

Usando la legge di Little:

$$WIP = TH \times FT$$

si ottiene il limite CONWIP iniziale:

$$W^* = TH_{target} \times FT_{desired} = 3 \text{ job/h} \times 3,5 \text{ h} \approx 11 \text{ job.}$$

4.3 Impostazioni sperimentali

Il sistema produttivo analizzato si colloca nel contesto tipico di una simulazione all'equilibrio, essendo caratterizzato da un funzionamento continuativo e ininterrotto su lunghi intervalli temporali. Tale configurazione consente di osservare e analizzare le prestazioni del sistema in uno stato stazionario, ovvero in una condizione dinamicamente stabile in cui le variabili chiave, come throughput, livelli di WIP e tempi di attraversamento, si mantengono costanti nel tempo in media.

La simulazione in stato stazionario si prefigge dunque l'obiettivo di studiare il comportamento operativo del sistema nella sua funzionalità generale e permanente, evitando distorsioni dovute alle condizioni iniziali transitorie tipiche dello start-up.

Per garantire che le misure di performance siano rappresentative dello stato stazionario, è fondamentale escludere i dati relativi alla fase transitoria iniziale, durante la quale il sistema evolve dalla condizione iniziale allo stato di equilibrio. Il tempo impiegato per superare questa fase è denominato *warm-up time* t_d . L'andamento tipico della variabile di output durante la fase di simulazione è illustrato in Figura 10, dove si distingue chiaramente la fase transitoria di warm-up e il successivo regime stazionario.

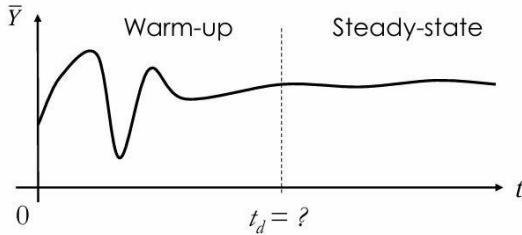


Figura 10: andamento tipico di una variabile di output nel tempo durante una simulazione: distinzione tra fase di warm-up (transitoria) e regime stazionario (steady-state).

Fonte: slide del corso *Simulazione dei sistemi gestionali*, prof.ssa Alfieri (Politecnico di Torino, 2025)

Questo grafico rappresenta l'andamento tipico di una variabile di output \bar{Y} nel tempo durante una simulazione a eventi discreti.

La scelta di un warm-up time adeguato consente di isolare le condizioni d'equilibrio, garantendo che i risultati statistici riflettano fedelmente il comportamento del sistema in regime permanente, senza distorsioni causate da eventi iniziali non rappresentativi [22].

La durata totale della simulazione è una scelta di progetto stabilita in base agli obiettivi di analisi e alla natura del sistema produttivo studiato, e non si calcola tramite formule standard. Nel presente studio, è stata impostata una durata di 730 giorni simulativi, con un periodo di warm-up di 365 giorni.

Questa scelta consente di analizzare il sistema per un intervallo operativo lungo e rappresentativo, superando abbondantemente il periodo transitorio e garantendo la possibilità di raccogliere dati statisticamente significativi in condizioni di equilibrio.

L'adozione di un periodo di warm-up pari a metà della durata della simulazione rappresenta una scelta prudentiale, che assicura che tutte le variabilità e i fenomeni di transizione siano superati prima della raccolta dei dati analitici. Questo consente di evitare che i risultati siano influenzati da condizioni iniziali non rappresentative e di focalizzarsi su performance effettive e stabili del sistema.

Un altro aspetto importante da stimare, oltre al periodo di warm-up, riguarda il numero di repliche necessarie per ottenere risultati statisticamente affidabili. Il numero di repliche rappresenta un elemento fondamentale per garantire la significatività dei risultati di una simulazione.

In generale, la letteratura raccomanda di effettuare un numero adeguato di repliche, tale da assicurare che i risultati non dipendano da singole realizzazioni casuali ma riflettano il comportamento medio del sistema, soprattutto nei casi in cui il modello presenti un'elevata

variabilità stocastica o una forte dipendenza dalle condizioni iniziali. Nel presente studio si è scelto di effettuare una sola replica per ogni scenario simulativo. Questa decisione implica che i risultati debbano essere analizzati principalmente da un punto di vista qualitativo piuttosto che quantitativo, in quanto non è possibile calcolare formalmente la varianza né costruire intervalli di confidenza statistici. [27].

4.4 Scenari ipotizzati per l'analisi del sistema CONWIP

Al fine di valutare l'efficacia delle logiche di controllo del WIP e delle regole di priorità, sono stati definiti quattro scenari principali di simulazione. Ciascun scenario rappresenta una configurazione concettuale del sistema, progettata per isolare e analizzare in modo mirato specifici meccanismi di controllo e instradamento del flusso produttivo.

- Scenario 1 – Controllo statico del WIP (benchmark di riferimento): il primo scenario costituisce la base di riferimento per l'intera analisi. In esso viene adottato un limite fisso al livello di Work In Process, mantenuto costante per tutta la durata della simulazione. Le workstation WS5 e WSQC operano secondo la regola FIFO, senza alcuna forma di priorità. Questo scenario consente di ottenere un punto di confronto “pulito” per le configurazioni successive, poiché permette di osservare gli effetti diretti della variazione del livello di WIP sul throughput e sulla probabilità di deadlock, in assenza di logiche adattive o dinamiche.
- Scenario 2 – Controllo dinamico del WIP (feedback periodico): il secondo scenario introduce un meccanismo di controllo dinamico del WIP basato su feedback periodici. A intervalli regolari (pari a un'ora di simulazione), il sistema verifica il livello complessivo di WIP e aggiorna di conseguenza il limite CONWIP, incrementandolo o riducendolo in funzione delle condizioni operative rilevate. Le workstation WS5 e WSQC operano sempre secondo la regola FIFO. L'obiettivo è verificare se una regolazione dinamica del WIP consente di migliorare le prestazioni rispetto al benchmark statico, sia in termini di incremento del throughput che di riduzione della probabilità di deadlock.
- Scenario 3 – Effetto delle regole di priorità in WS5: il terzo scenario estende la configurazione precedente introducendo una regola di priorità selettiva nella workstation WS5. In particolare, viene adottata, su entrambe le macchine della WS5, la logica Shortest Remaining Processing Time, secondo la quale entrambe le macchine danno priorità ai job con il minor lavoro residuo da completare, ossia quelli che hanno già eseguito una delle due operazioni previste nella workstation.

La motivazione di questa scelta risiede nella necessità di limitare i tempi di permanenza dei job nei caroselli soggetti a rilavorazione, riducendo così il rischio di congestione e potenziale deadlock. Lo scopo dello scenario è valutare se l'introduzione di tale priorità comporti un miglioramento del flusso complessivo, riducendo i tempi di attraversamento dei job rilavorati, senza compromettere in misura significativa il throughput globale del sistema.

- Scenario 4 – Il quarto scenario, sempre basato sulla medesima configurazione di controllo dinamico del WIP, adotta invece la logica Longest Remaining Processing Time, nella quale entrambe le macchine in serie danno priorità ai job con il maggior lavoro residuo, ossia quelli che non hanno ancora effettuato nessuna delle due lavorazioni previste.

Questa impostazione si basa sull'idea di anticipare l'avanzamento dei job "nuovi" nel ciclo produttivo, al fine di mantenere un livello di WIP distribuito più uniformemente tra le varie fasi e prevenire situazioni in cui un eccessivo numero di job "in rilavorazione" possa monopolizzare la workstation. In altre parole, questa logica tende a massimizzare l'utilizzo delle macchine e a mantenere elevata la produttività istantanea, privilegiando i job con maggiore carico di lavoro residuo.

Il confronto tra gli scenari 3 e 4 permette di valutare in che misura la scelta della regola di priorità, orientata rispettivamente alla minimizzazione dei tempi di completamento o alla massimizzazione dell'avanzamento dei job nuovi, influenzi le performance complessive del sistema in termini di throughput, tempi medi di attraversamento e stabilità del flusso.

Gli scenari analizzati e i relativi obiettivi sperimentali sono sintetizzati nella Tabella 1, che costituisce un riferimento introduttivo per l'interpretazione dei risultati attesi.

Tabella 1: sintesi degli scenari

Scenario	Tipo di controllo WIP	Regola di priorità WSS	Obiettivo principale
Scenario 1	CONWIP statico	FIFO	Definire un benchmark di riferimento
Scenario 2	CONWIP dinamico	FIFO	Analizzare gli effetti della regolazione dinamica del WIP
Scenario 3	CONWIP dinamico	SRPT (priorità ai job con minor lavoro residuo)	Valutare l'impatto dell'assegnazione di priorità ai job rilavorati
Scenario 4	CONWIP dinamico	SRPT (priorità ai job con maggior lavoro residuo)	Valutare l'impatto dell'assegnazione di priorità ai job nuovi

4.5 Indicatori di performance

Al fine di valutare in modo rigoroso e completo l'efficacia delle politiche di controllo del WIP e delle regole di priorità implementate, viene definito un insieme di indicatori di performance (*Key Performance Indicators, KPI*), scelti per offrire una visione bilanciata e multidimensionale delle prestazioni complessive del sistema. Questi indicatori rappresentano parametri di riferimento per l'analisi comparativa degli scenari proposti, includendo aspetti produttivi, temporali e di stabilità operativa.

Gli indicatori selezionati sono:

- Throughput (TH): rappresenta il numero di job completati ed usciti dal sistema per unità di tempo. Rappresenta il principale indicatore di produttività e costituisce una misura diretta della capacità del sistema di convertire il WIP in output utilizzabili. La massimizzazione del throughput è l'obiettivo primario perseguito dal meccanismo di controllo dinamico CONWIP, in quanto favorisce l'efficienza produttiva complessiva.
- Average Flow Time (tempo medio di attraversamento): misura il tempo medio trascorso da un job dall'ingresso all'uscita del sistema. Questo indicatore è complementare al throughput e, in accordo con la legge di Little, $WIP = TH \times Flow Time$, consente di valutare l'efficienza temporale complessiva del processo produttivo e la capacità di risposta del sistema.
- Probabilità di deadlock: calcolata come il rapporto tra il numero di repliche che terminano in uno stato di blocco irreversibile e il numero totale delle repliche

eseguite. Questo KPI risulta fondamentale in sistemi con loop e rilavorazioni, nei quali fenomeni di congestione localizzati possono propagarsi causando paralisi di intere sezioni produttive.

- Job in attesa di ingresso nel sistema (numero medio di unità nell'Hold CONWIP): rappresenta la quantità media di job che rimangono temporaneamente trattenuti prima di essere rilasciati nel sistema produttivo, in attesa che il livello di WIP scenda sotto la soglia consentita dal controllo CONWIP. Tale indicatore consente di valutare la pressione esercitata dal meccanismo di controllo sulla fase di rilascio e il grado di saturazione del sistema.
- Tempo medio di attesa in Hold CONWIP: misura la durata media della permanenza dei job nella coda di ingresso prima di poter accedere al processo produttivo. Un valore elevato di questo indicatore segnala un eccessivo livello di congestione o una soglia di WIP troppo restrittiva, mentre valori contenuti indicano una gestione più fluida e reattiva del rilascio dei job.

Il set di indicatori consente di valutare in modo completo le prestazioni del sistema, combinando aspetti di efficienza, stabilità e fluidità del flusso produttivo, e offrendo così una visione bilanciata tra produttività e robustezza operativa.

4.6 Semplificazione del modello e validità dei risultati

Il modello completo della linea, comprensivo di nastri trasportatori, stazioni di merge e split e loop di rilavorazione, è stato sviluppato integralmente per rappresentare in modo realistico la dinamica del sistema. Tuttavia, la complessità associata alla gestione della movimentazione automatizzata ha comportato difficoltà operative e di stabilità del modello, non risolvibili in modo efficiente entro i tempi previsti dal progetto.

Per poter comunque analizzare gli effetti delle politiche di controllo e delle regole di priorità sulle prestazioni complessive, è stata quindi realizzata una versione semplificata del modello, priva dei nastri trasportatori. Tale versione mantiene invariata la logica di flusso, la sequenza delle lavorazioni e le condizioni di controllo CONWIP, consentendo di ottenere risultati coerenti e significativi sul piano teorico e simulativo.

Questa scelta si inserisce in una consolidata tradizione della modellazione e simulazione di sistemi complessi, in cui approssimazioni e astrazioni rappresentano strumenti fondamentali per isolare e analizzare le dinamiche di interesse. In letteratura, infatti, si evidenzia come la semplificazione strutturale di un modello possa rendere più trasparente la comprensione dei

meccanismi di causa-effetto e consentire di esplorare rapidamente gli scenari di interesse [27] [28].

L'eliminazione dei nastri trasportatori dal modello consente di focalizzarsi sulle logiche di controllo della produzione e sul flusso dei job, evitando complessità legate alla simulazione dettagliata del trasporto fisico. Ciò permette di verificare con chiarezza l'impatto delle diverse politiche CONWIP e delle priorità operative sul throughput e sul livello medio di WIP.

Pur non riproducendo esattamente i tempi di movimentazione o le micro-dinamiche di congestione, il modello semplificato mantiene coerenti le tendenze attese: a parità di condizioni, gli scenari che producono un incremento del throughput o una riduzione dei tempi medi nel modello semplificato mostrano lo stesso andamento direzionale che ci si aspetterebbe nel modello completo.

In altre parole, le relazioni qualitative tra le variabili restano valide, anche se i valori assoluti differiscono a causa dell'assenza dei tempi di trasporto e delle interazioni fisiche tra le stazioni.

Naturalmente, un modello completo con nastri trasportatori consentirebbe di ottenere stime più accurate in termini di tempi di attraversamento, saturazione delle risorse e fenomeni di accumulo, grazie alla rappresentazione esplicita delle distanze e delle logiche di movimentazione. Tuttavia, nel contesto di questo studio, l'obiettivo principale consiste nell'analizzare l'effetto delle politiche di controllo piuttosto che la cinematica del trasporto. Per questo motivo, il modello semplificato risulta adeguato e pienamente coerente con gli obiettivi di ricerca, offrendo una base solida per la valutazione comparativa degli scenari.

Pertanto, questa semplificazione rappresenta una scelta metodologica consapevole e funzionale, che valorizza il rigore dell'analisi e garantisce al tempo stesso una maggiore stabilità e leggibilità dei risultati.

Il modello con conveyor rimane una base utile per futuri sviluppi orientati a un maggior dettaglio fisico, mentre la versione semplificata consente di concentrarsi sui meccanismi decisionali e di controllo che governano la performance complessiva del sistema.

5. ANALISI QUALITATIVA E DISCUSSIONE DEI RISULTATI

5.1 Obiettivi e impostazione dell'analisi

Il presente capitolo ha l'obiettivo di delineare le aspettative teoriche relative al comportamento del sistema simulato e di definire il quadro concettuale all'interno del quale vengono interpretati i risultati attesi. L'analisi mira a valutare l'impatto delle politiche di controllo del Work In Process e delle regole di priorità sulla produttività, la stabilità e l'efficienza complessiva del sistema manifatturiero modellato.

A supporto di tale analisi, vengono inoltre presentati e discussi i risultati ottenuti dal modello semplificato, sviluppato per consentire la valutazione sperimentale delle politiche di controllo del WIP e delle regole di priorità, e per verificare la coerenza delle dinamiche osservate con le previsioni teoriche.

Gli scenari sperimentali, descritti nel Capitolo 4, rappresentano quattro configurazioni alternative del medesimo impianto, differenziate in base alla logica di controllo del WIP (statica o dinamica) e alla strategia di sequenziamento adottata nella workstation soggetta a rilavorazioni. Tale impostazione consente di isolare progressivamente l'effetto di ciascun elemento di controllo – il meccanismo CONWIP e le regole di priorità – sul comportamento del sistema, favorendo un confronto diretto e coerente tra le diverse politiche operative.

Le prestazioni del sistema vengono valutate attraverso gli indicatori di performance definiti nel paragrafo 4.5, selezionati per offrire una prospettiva bilanciata tra produttività, efficienza temporale e stabilità operativa. In particolare, il confronto tra gli scenari si basa sull'analisi del throughput, del tempo medio di attraversamento, del WIP medio e della probabilità di deadlock.

5.2 Analisi teorico-simulativa del comportamento del sistema

L'analisi seguente si concentra sull'elaborazione di ipotesi relative al comportamento atteso del sistema, fondate sui principi teorici della gestione della produzione e sulle dinamiche operative illustrate nei capitoli precedenti. L'obiettivo principale è tracciare, in maniera ragionata e coerente con la struttura del modello, le dinamiche previste in funzione delle politiche di controllo adottate e delle regole decisionali applicate. A supporto di tali considerazioni teoriche, viene inoltre presentata una verifica sperimentale condotta sul modello semplificato, che consente di osservare in modo diretto l'effetto delle

diverse configurazioni operative e di validare le relazioni attese tra i principali indicatori di performance.

La discussione si sviluppa su due aree principali:

- La prima riguarda l'impatto delle politiche di controllo del WIP (Scenario 1 rispetto a Scenario 2), valutando in quale misura una regolazione dinamica possa contribuire a migliorare la produttività e la stabilità del sistema rispetto a un approccio statico.
- La seconda si concentra sull'effetto delle regole di priorità in WS5 (Scenario 3 rispetto a Scenario 4), con l'intento di analizzare come diverse strategie di scheduling possano influire sull'equilibrio tra throughput, tempi medi di attraversamento e livelli di congestione.

5.2.1 Effetto del controllo del WIP

L'effetto del controllo del WIP può essere interpretato alla luce della legge di Little, secondo la quale il numero medio di job presenti nel sistema è dato da:

$$WIP = TH \times FT$$

Da questa relazione emerge che, aumentando il WIP, il throughput tende inizialmente a crescere, poiché le workstation risultano meglio saturate e le fasi di inattività si riducono. Tuttavia, oltre una certa soglia, il sistema raggiunge il proprio limite di capacità produttiva: le macchine operano già a pieno regime e ulteriori incrementi del WIP non si traducono in un aumento significativo del throughput, bensì in un aumento consistente dei tempi medi di attraversamento.

Questo comportamento può essere descritto formalmente attraverso il concetto di utilizzazione media delle risorse:

$$U = \frac{\lambda}{\mu}$$

dove λ è il tasso di arrivo effettivo dei job e μ il tasso di servizio medio della workstation. Quando λ si avvicina a μ , l'utilizzo tende a 1, il sistema si satura e la stabilità operativa si riduce drasticamente. In questa condizione, anche piccoli incrementi del carico generano forti aumenti dei tempi di attesa.

Nel contesto del modello analizzato, il controllo statico del WIP mantiene costante il numero massimo di job presenti nel sistema, rischiando di fissare un livello troppo basso (riducendo

il throughput) o troppo alto (causando congestione e deadlock). Il controllo dinamico, invece, adatta periodicamente la soglia di WIP alle condizioni operative, cercando di mantenere il sistema nella zona di equilibrio in cui l'utilizzo è elevato ma inferiore a 1. Ci si aspetta quindi che tale strategia consenta di migliorare il throughput medio e ridurre la probabilità di blocco, a fronte di possibili oscillazioni nei tempi di attraversamento dovute alla regolazione continua del limite di WIP.

Inoltre, utilizzando un controllo dinamico del limite CONWIP, quando il flusso produttivo risulta regolare e le workstation non mostrano segni di congestione, il limite di WIP viene aumentato, consentendo l'ingresso di un maggior numero di job. Al contrario, quando il sistema si avvicina a condizioni critiche — con buffer saturi o job in attesa di rilavorazione — il limite CONWIP viene temporaneamente ridotto, così da diminuire la quantità di lavoro circolante e favorire lo smaltimento delle code.

Questa regolazione del limite di WIP agisce indirettamente sul tempo medio di inter-arrivo dei job, r_a , e quindi sul throughput complessivo, legato dalla relazione:

$$TH = \frac{1}{r_a}$$

Riducendo il limite CONWIP in presenza di congestione, il sistema aumenta di fatto r_a , ossia allunga il tempo medio tra un arrivo e l'altro. Ciò comporta una temporanea riduzione del throughput, ma consente al sistema di ritornare in condizioni operative stabili, evitando l'insorgere di blocchi permanenti. Viceversa, quando il limite CONWIP viene aumentato, r_a si riduce e gli arrivi diventano più frequenti, incrementando il throughput fino a quando il sistema rimane entro soglie di carico sostenibili.

Un approccio di questo tipo consente di migliorare la produttività complessiva e la stabilità del flusso attraverso il bilanciamento automatico del carico tra le diverse risorse produttive. L'adattamento continuo del limite di WIP riduce la probabilità di accumuli eccessivi nei buffer, mitigando il rischio di deadlock e di saturazione delle workstation critiche. Tuttavia, poiché la soglia di WIP varia nel tempo, è plausibile osservare oscillazioni nei tempi medi di attraversamento: quando il limite aumenta, il sistema tende a riempirsi, allungando i tempi di permanenza; quando si riduce, il flusso si svuota, generando fasi di maggiore reattività ma anche di potenziale instabilità.

A supporto di queste considerazioni teoriche, è stata condotta un'analisi sperimentale sul modello semplificato, al fine di verificare in modo preliminare gli effetti delle due modalità di controllo, statico e dinamico, sui principali indicatori prestazionali del sistema, ossia throughput, tempo medio di attraversamento, WIP medio, probabilità di deadlock e grado di utilizzo delle workstation.

Nel caso del controllo CONWIP statico, l'analisi dei risultati evidenzia un comportamento anomalo del sistema, riconducibile sia alla natura semplificata del modello sia all'assenza dei nastri trasportatori, che nel modello completo avrebbero gestito la movimentazione fisica e il ritorno dei job nei loop di rilavorazione. Mantenendo un limite di WIP fisso, il sistema tende progressivamente a saturarsi, fino a raggiungere una condizione in cui i job non riescono più a rientrare nel ciclo produttivo. Dopo un certo tempo operativo, il limite massimo di WIP risulta costantemente pieno e non si verificano più completamenti, bloccando gli ingressi successivi.

È importante chiarire che molti indicatori risultano pari a zero non per assenza di dati reali, ma per effetto del periodo di warm-up: durante questa fase iniziale, il modello registra correttamente alcune lavorazioni, ma successivamente il sistema si blocca. Poiché SIMAN calcola le statistiche solo dopo il periodo di warm-up, i dati validi raccolti successivamente risultano nulli, in quanto il sistema, già saturo, non produce più nuove osservazioni utili. In altre parole, il file di output risulta “vuoto” perché il periodo analizzato coincide con la fase di blocco completo del sistema, in cui nessun job esce e nessuna attività viene completata.

Questo comportamento mette in evidenza una criticità strutturale del controllo CONWIP statico nei sistemi con rilavorazioni o loop ciclici: la mancanza di adattività rende il sistema vulnerabile alla congestione e può condurre, come in questo caso, a un deadlock permanente, dove il throughput si annulla e l'analisi statistica perde significato.

Viceversa, nel controllo CONWIP dinamico, la soglia di WIP viene continuamente aggiornata in base allo stato operativo del sistema, in particolare al livello di congestione e alla saturazione dei buffer. Quando il flusso produttivo risulta regolare, il limite viene aumentato per consentire un maggior afflusso di job e mantenere alta l'utilizzazione delle risorse; al contrario, in condizioni di congestione, la soglia viene temporaneamente ridotta per consentire lo smaltimento delle code e ristabilire l'equilibrio del sistema.

Questa logica consente di confrontare solo alcuni indicatori selezionati:

- Probabilità di deadlock: trascurabile nel controllo dinamico, ma elevatissima nel controllo statico, che converge inevitabilmente verso una condizione di blocco permanente.
- Nel caso CONWIP dinamico, la variabile che rappresenta la coda di attesa all'ingresso del sistema, ossia l'insieme dei job che non possono ancora entrare nel ciclo produttivo perché è stato raggiunto il limite massimo di WIP imposto dal controllo CONWIP, assume un valore medio di 8750 unità, con un tempo medio di attesa di circa 2916 ore. Ciò suggerisce che, pur in presenza di code consistenti, il sistema riesca periodicamente a svuotarle, mantenendo una certa fluidità del flusso.
- Nel caso CONWIP statico, invece, la coda raggiunge un valore medio di 13129 unità, ma i job non riescono mai a rientrare effettivamente nel ciclo: il limite fisso di WIP blocca progressivamente il sistema fino a portarlo in deadlock permanente.

Il controllo dinamico del WIP consente quindi di ridurre la coda media all'ingresso del sistema di circa il 33% rispetto al caso statico, migliorando la fluidità del flusso produttivo e limitando l'accumulo di job in attesa di ingresso. La Figura 11 mostra il confronto tra la dimensione media della coda di attesa all'ingresso del sistema nei due scenari analizzati, evidenziando l'effetto del controllo adattivo del WIP sulla riduzione della congestione iniziale.

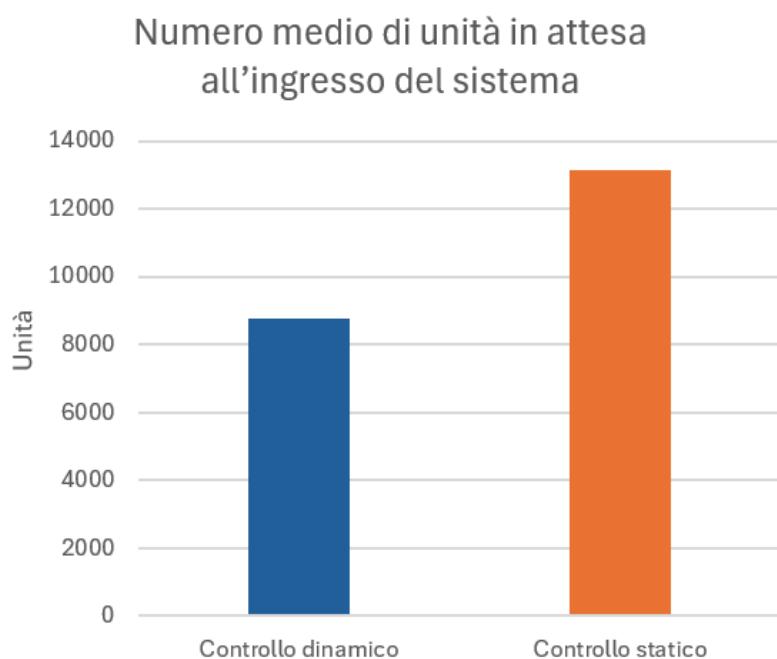


Figura 11: confronto tra la dimensione media della coda di attesa all'ingresso del sistema nei casi CONWIP statico e CONWIP dinamico.

Poiché la simulazione con un periodo di warm-up di un anno e una durata operativa di un altro anno non risultava adeguata, in quanto non era possibile confrontare in modo coerente i diversi scenari, si è deciso di ridurre l'orizzonte temporale di analisi, impostando un warm-up di 0 giorni e una durata di simulazione di 5 giorni.

Questa configurazione, pur non rappresentando un periodo sufficiente a raggiungere una condizione di regime stabile, consente di osservare il comportamento del sistema nelle fasi iniziali del flusso produttivo, confrontando la reattività delle logiche di controllo. È importante sottolineare che i risultati sono fortemente influenzati dai transitori iniziali, ovvero dalla fase in cui il sistema si popola progressivamente di job e le workstation non sono ancora tutte operative. In questa fase il flusso produttivo non ha ancora raggiunto la stabilità necessaria affinché i valori medi di throughput e tempi di attraversamento siano rappresentativi.

Nel controllo CONWIP statico si osserva una forte tendenza alla saturazione del sistema: quasi tutti i job rimangono in coda in attesa di essere rilasciati, circa 50 in media. Il numero massimo di entità in attesa raggiunge circa 110 unità, evidenziando una congestione significativa. Questo comportamento conferma la rigidità del limite di WIP fisso, che impedisce il ricambio dei job una volta raggiunta la soglia impostata, determinando una condizione di blocco parziale e una drastica riduzione della produttività.

Nel caso del controllo CONWIP dinamico, invece, il sistema mostra una maggiore fluidità operativa. Le unità in attesa di entrare nel sistema sono in media 30, con un massimo di 70 unità, mentre il tempo medio di attesa in coda, pari a circa 11 ore, suggerisce che il sistema regola attivamente il flusso di ingresso, evitando la congestione e permettendo un turnover più regolare dei job, a vantaggio della stabilità complessiva del processo produttivo.

Anche in questo caso, il numero medio di unità in coda, passando da 50 nel modello con controllo CONWIP statico a 30 nel modello con controllo dinamico, riflette una riduzione di circa il 40%, coerente con quella osservata nel modello a orizzonte annuale.

L'andamento temporale di queste grandezze, e in particolare la variazione del numero di entità in attesa nel tempo, è rappresentato nel grafico in Figura 12.

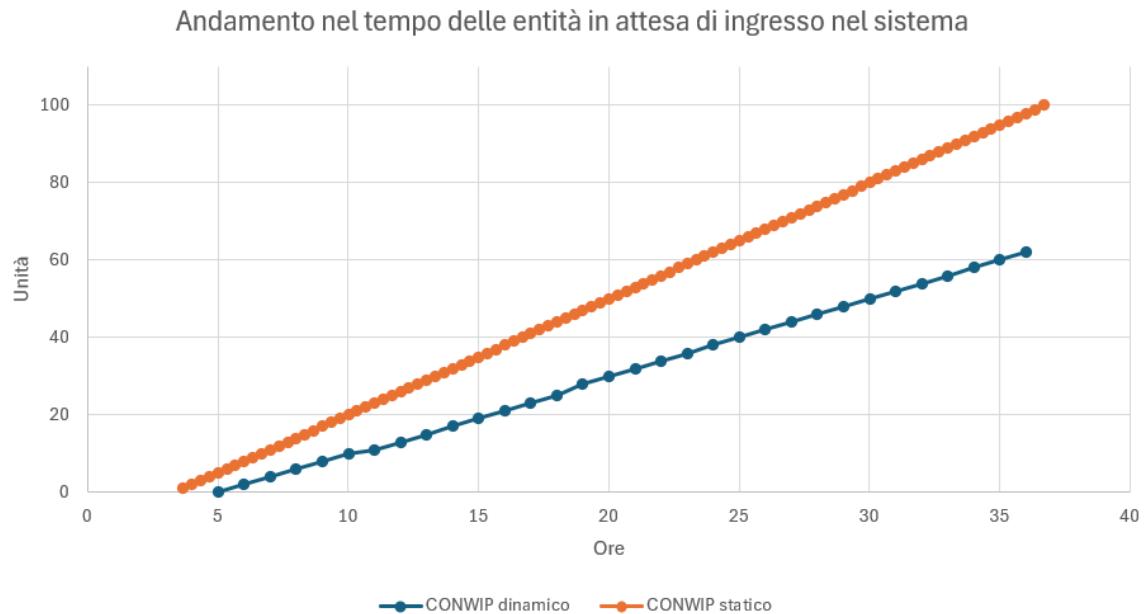


Figura 12: andamento nel tempo del numero di entità in attesa di ingresso nel sistema nei casi di controllo CONWIP statico e dinamico.

Per quanto riguarda invece i KPI del throughput e del flow time, il confronto tra i due scenari evidenzia l’effetto positivo del meccanismo di regolazione dinamica del WIP sulle prestazioni complessive del sistema.

In particolare, il throughput medio passa da 1.0765 job/ora nel caso del controllo CONWIP statico a 1.2870 job/ora nel caso del controllo dinamico, registrando un incremento di circa 19,5%. Questo risultato conferma che l’adattività del limite di WIP consente di mantenere elevata l’utilizzazione delle risorse, evitando condizioni di saturazione prolungata che nel controllo statico riducono progressivamente la produttività complessiva.

Analogamente, il flow time medio si riduce da 2.5179 ore nello scenario statico a 2.3238 ore nel dinamico, con un miglioramento di circa 7,7%. Tale riduzione, seppur più contenuta rispetto all’aumento del throughput, indica che il controllo dinamico favorisce una distribuzione più equilibrata del carico di lavoro tra le workstation, limitando la formazione di code e riducendo i tempi di attesa nei buffer intermedi.

Per completezza, la Tabella 2 riassume i principali risultati ottenuti nei due scenari di simulazione, mettendo a confronto le prestazioni del controllo CONWIP statico e CONWIP dinamico in termini dei KPI definiti precedentemente.

Tabella 2: sintesi dei risultati ottenuti

Indicatore	Scenario 1	Scenario 2	Osservazioni
Throughput medio	1.0765 job/h	1.2870 job/h	+19,5%: incremento della produttività grazie alla regolazione dinamica del WIP
Flow time medio	≈ 2.5179 ore	≈ 2.3238 ore	-7,7%: riduzione dei tempi medi di attraversamento
Probabilità di deadlock	Elevata (deadlock permanente)	Trascutibile	Il controllo dinamico evita il blocco completo del sistema
Job in attesa di ingresso	≈ 50 job	≈ 30 job	-40%: minore congestione all'ingresso del sistema
Tempo medio di attesa in coda	Blocco permanente	≈ 11.33 ore	Il sistema dinamico consente il rilascio periodico dei job in attesa

Dalla Tabella 2 emerge chiaramente come il controllo CONWIP dinamico migliori in modo significativo la stabilità e l'efficienza del sistema produttivo. L'adattamento continuo della soglia di WIP riduce la probabilità di deadlock, favorisce un flusso più regolare dei job e consente di mantenere un buon equilibrio tra produttività e tempi medi di attraversamento.

5.2.2 Effetto delle regole di priorità

Nel secondo ambito di analisi l'attenzione si concentra sull'effetto delle regole di priorità applicate nella workstation dotata di due macchine parallele, confrontando tre configurazioni operative: la regola FIFO, adottata come riferimento, e le logiche SRPT e LRPT, implementate nel contesto del controllo dinamico del WIP.

L'obiettivo è comprendere in che modo la scelta della regola di priorità influenzi le prestazioni del sistema, in particolare in termini di throughput, tempi medi di attraversamento e stabilità del flusso. Poiché il livello di WIP è già regolato dinamicamente, l'analisi si concentra esclusivamente sugli effetti dovuti alla gestione dell'ordine di lavorazione dei job, per valutare se politiche di scheduling più selettive possano migliorare la fluidità del sistema o introdurre, al contrario, nuove forme di variabilità.

Dal punto di vista teorico, l'adozione di diverse regole di priorità nelle workstation influisce direttamente sulla gestione delle code e, di conseguenza, sulle prestazioni complessive del

sistema. La scelta della regola di sequenziamento determina infatti l'ordine con cui i job vengono processati, modificando il tempo medio di attesa e, di conseguenza, il tempo medio di attraversamento.

Lo scenario di riferimento considerato nell'analisi è quello basato sulla regola FIFO, ovvero il primo elemento che entra nella coda è anche il primo ad essere servito. Questa scelta rappresenta la politica di sequenziamento più semplice e intuitiva, spesso utilizzata come configurazione di default nei sistemi produttivi automatizzati. Tale approccio non richiede informazioni aggiuntive sui job né una stima dei tempi residui di lavorazione, risultando quindi facilmente implementabile e 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, poiché i job vengono processati nell'ordine di arrivo, senza alcuna distinzione basata sulla durata o sulla complessità del lavoro. Questa caratteristica garantisce un comportamento ordinato e privo di favoritismi tra le lavorazioni, ma non sempre conduce alla massima efficienza del sistema. In presenza di variabilità nei tempi di lavorazione, infatti, può manifestarsi il cosiddetto effetto "convoy", ossia la formazione di code in cui i job con tempi più lunghi rallentano l'avanzamento di quelli più rapidi. Tale fenomeno comporta un aumento del tempo medio di attraversamento e, in generale, una minore reattività del sistema a fronte di fluttuazioni nella domanda o nei tempi di processo.

Questo scenario di riferimento viene confrontato con due regole di priorità alternative, SRPT e LRPT, selezionate in quanto rappresentano due approcci opposti nella gestione delle priorità operative. Il confronto tra queste due strategie consente quindi di analizzare il trade-off tra efficienza locale e bilanciamento globale del flusso produttivo, evidenziando come differenti criteri di priorità influenzino le prestazioni complessive del sistema in termini di throughput e tempi medi di attraversamento.

Con la regola SRPT viene data priorità ai job con minor lavoro residuo. Questo approccio favorisce il completamento rapido delle lavorazioni brevi, riducendo la congestione e migliorando la fluidità del flusso. In generale, si osserva una diminuzione del flow time, poiché i job tendono a permanere meno tempo nel sistema. Per quanto riguarda il throughput, la regola SRPT tende a mantenerlo stabile o incrementarlo leggermente rispetto a FIFO, grazie a un uso più efficiente delle risorse: la rapida chiusura

dei job brevi libera più frequentemente le workstation, permettendo l'avvio di nuove lavorazioni e migliorando il ritmo complessivo di produzione.

La regola LRPT, al contrario, assegna priorità ai job con il lavoro residuo maggiore. Questa logica tende a peggiorare le prestazioni complessive del sistema rispetto a FIFO: i job più brevi vengono costantemente posticipati, con conseguente aumento del flow time medio, e la lavorazione di job lunghi finisce per saturare le risorse. Ciò può determinare una riduzione del throughput o comunque un rendimento non superiore a quello della FIFO, poiché il sistema risulta meno reattivo e più incline alla congestione [27] [29] [30] [31].

In sintesi, dal confronto teorico ci si aspetta che:

- SRPT → riduca il flow time e mantenga o aumenti il throughput rispetto a FIFO.
- LRPT → aumenti il flow time e riduca il throughput rispetto a FIFO.

Nella simulazione realizzata, il confronto tra le tre logiche di priorità è stato condotto analizzando principalmente i KPI di performance: throughput e flow time. Non sono stati invece confrontati altri indicatori, come il numero medio di job in coda o la probabilità di deadlock, poiché risultano pressoché identici nei tre scenari. Questo comportamento è dovuto al fatto che le condizioni operative e il carico di lavoro complessivo rimangono invariati: i job arrivano con la stessa frequenza e il limite CONWIP impone la medesima soglia di WIP, mantenendo costante la quantità media di job presenti nel sistema.

I risultati ottenuti sono i seguenti:

- FIFO: TH = 0,84762 job/h, FT = 2,5237 h
- SRPT: TH = 0,84762 job/h, FT = 2,5237 h
- LRPT: TH = 0,84513 job/h, FT = 2,5245 h

Il confronto tra le tre logiche di priorità in termini di Flow Time è illustrato nel grafico di Figura 13, mentre il Throughput medio associato a ciascuna strategia è riportato nel grafico di Figura 14.

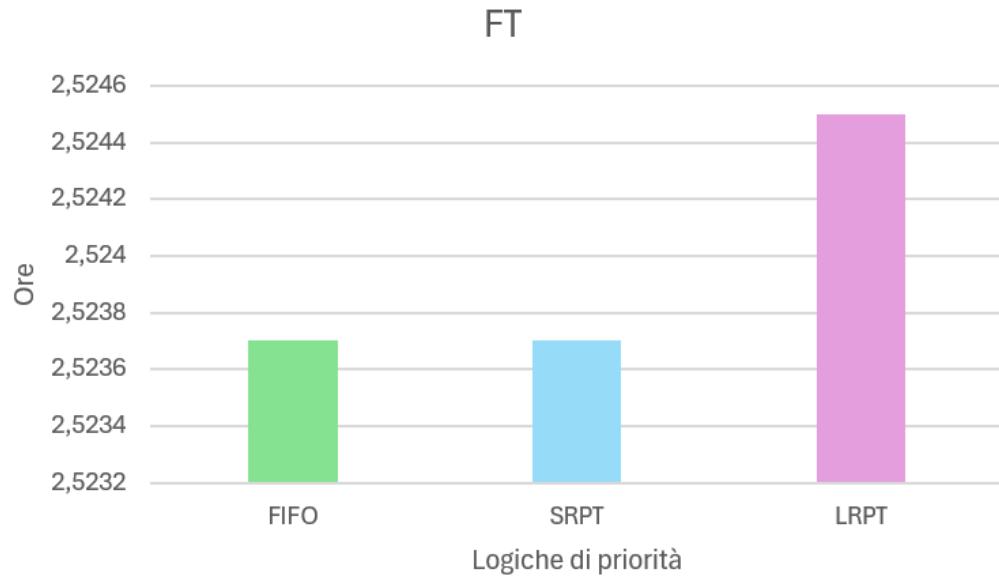


Figura 13: confronto del Flow Time medio per le tre logiche di priorità (FIFO, SRPT e LRPT)

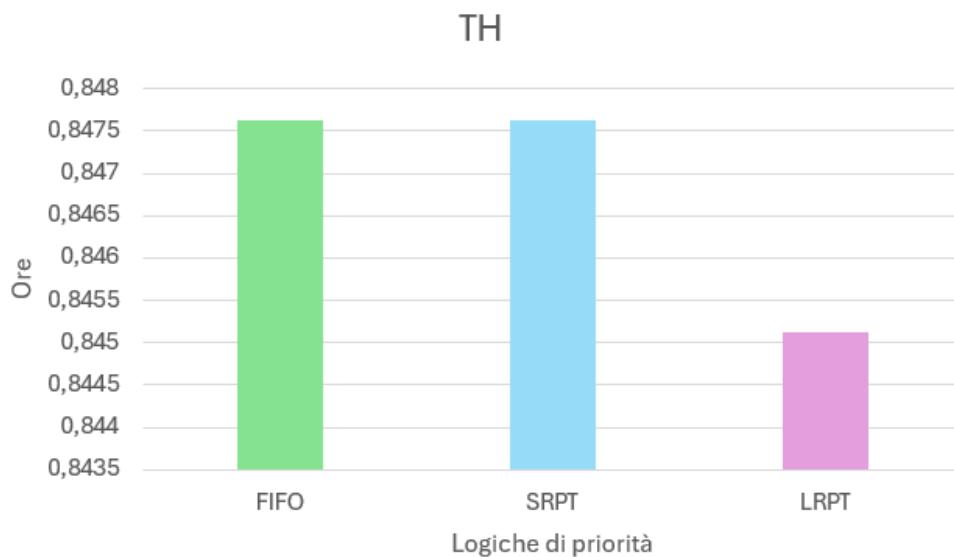


Figura 14: confronto del Throughput medio per le tre logiche di priorità (FIFO, SRPT e LRPT)

Come si osserva, le logiche FIFO e SRPT forniscono risultati praticamente identici.
Questo può accadere in simulazione quando:

- la variabilità del lavoro residuo tra i job è ridotta, ossia le durate delle lavorazioni sono simili, per cui dare priorità a chi ha meno lavoro residuo non porta benefici rispetto al servire nell'ordine di arrivo (FIFO);
- il sistema non presenta colli di bottiglia marcati né code sufficientemente lunghe da rendere efficace una regola di priorità più selettiva.

La regola LRPT, invece, mostra un leggero peggioramento: il throughput diminuisce, mentre il flow time cresce marginalmente.

Questo effetto è coerente con la teoria: assegnando priorità ai job con più lavoro residuo, la LRPT tende a privilegiare le lavorazioni più lunghe, che saturano le risorse e aumentano il tempo medio di permanenza nel sistema, riducendo di conseguenza la capacità produttiva effettiva [32] [33] [34].

In sintesi:

- FIFO e SRPT offrono prestazioni equivalenti in scenari con job simili e code moderate.
- LRPT comporta un lieve peggioramento di TH e FT, coerente con la logica che privilegia job lunghi e meno efficienti dal punto di vista del flusso produttivo. Questi risultati confermano che le differenze tra le politiche di priorità emergono in modo significativo solo in sistemi caratterizzati da alta variabilità dei tempi di processo o da forti congestioni.

[35] [36].

6. DISCUSSIONE MANAGERIALE DEI RISULTATI

6.1 Implicazioni per la definizione di metodologie di controllo di sistemi produttivi pull

L'analisi dei risultati ottenuti consente di formulare considerazioni significative e utili alla progettazione e gestione di sistemi produttivi pull. L'obiettivo principale è approfondire come le diverse configurazioni analizzate influenzino la stabilità del sistema e la fluidità del flusso produttivo, valutando al contempo la capacità di adattamento alle variazioni operative e alla domanda reale.

Il confronto tra i controlli CONWIP statico e dinamico evidenzia chiaramente che un approccio adattivo può offrire una gestione più efficiente e reattiva dei livelli di WIP, elemento centrale nei sistemi pull. Mentre il controllo statico si presenta come una soluzione semplice e di immediata applicazione, esso tende comunque a causare saturazioni e blocchi nei flussi, specialmente in presenza di rilavorazioni o loop ciclici che incrementano complessità e incertezza operativa. D'altro canto, il controllo CONWIP dinamico si dimostra più robusto e resiliente, grazie alla regolazione flessibile e continua del rilascio dei job in base alle condizioni operative attuali del sistema, riducendo la congestione e stabilizzando il throughput a lungo termine. Questa osservazione suggerisce che, nella progettazione di sistemi complessi con elevata variabilità nei tempi di processo e notevole interdipendenza tra le fasi produttive, il controllo dinamico del WIP rappresenta una scelta preferibile.

Riguardo alle regole di priorità, il confronto tra le politiche FIFO, SRPT e LRPT mette in luce differenze meno marcate ma comunque rilevanti in termini di performance operative. Nei modelli simulati, le politiche FIFO e SRPT ottengono risultati simili in condizioni di bassa variabilità dei tempi di lavorazione, indicando che la scelta della regola in questi casi ha un impatto relativamente limitato sulle prestazioni generali. Tuttavia, dal punto di vista operativo, la regola SRPT risulta più efficace ed efficiente poiché dà priorità ai job con minor lavoro residuo; questo accelera i completamenti, riduce i tempi medi di attraversamento e migliora la reattività del sistema. Al contrario, la politica LRPT, favorendo i job con carichi più lunghi, tende a rallentare il flusso produttivo, generando squilibri nella saturazione delle risorse e compromettendo stabilità e regolarità.

In conclusione, un sistema produttivo pull efficace deve bilanciare semplicità gestionale e adattività operativa per rispondere alle reali esigenze produttive. La combinazione di controlli dinamici del WIP con regole di priorità volte alla riduzione dei tempi di completamento può migliorare notevolmente la stabilità del processo produttivo, evitare

congestioni e garantire prestazioni affidabili e sostenibili nel lungo periodo. Questi aspetti sono cruciali nei moderni contesti industriali, dove flessibilità e capacità di rispondere ai cambiamenti improvvisi della domanda rappresentano fattori decisivi per il successo competitivo.

6.2 Benefici e limiti del controllo dinamico del WIP

Questo studio ha evidenziato come il controllo dinamico del WIP costituisca una delle innovazioni più significative nella gestione dei sistemi produttivi pull, caratterizzati dalla capacità di adattarsi in modo flessibile a condizioni operative in costante evoluzione.

Tra i principali benefici che emergono dall'adozione di un approccio dinamico uno dei più importanti è la capacità di mantenere il livello di WIP costantemente ottimizzato, anche quando si verificano cambiamenti nel sistema. Questo consente di evitare sia un accumulo eccessivo di scorte sia problematiche come il blocco o la congestione del flusso produttivo. Inoltre, si favorisce una maggiore stabilità operativa, garantendo un throughput calibrato, regolare e prevedibile nel tempo tramite un rilascio ben bilanciato dei job.

Un ulteriore punto di forza è rappresentato dalla capacità del sistema di reagire automaticamente a variazioni della domanda o a imprevisti, quali rilavorazioni e modifiche nei tempi di processo. Questa flessibilità, che consente un adeguamento continuo dei livelli di WIP, incarna uno dei principi chiave della produzione snella, dove il monitoraggio e il controllo in tempo reale risultano cruciali per limitare gli sprechi e massimizzare l'efficienza nell'utilizzo delle risorse. Rispetto ai modelli statici, spesso vulnerabili a inefficienze o sovraccarichi, il controllo dinamico offre una struttura più solida e reattiva.

Un altro beneficio significativo di questo approccio è la riduzione della necessità di frequenti interventi manuali, rendendo possibile una gestione più fluida delle priorità e delle code interne. Questo si traduce in un netto miglioramento del processo decisionale, particolarmente per quanto riguarda l'allocazione delle risorse.

Nonostante i numerosi vantaggi, l'approccio dinamico presenta anche alcune difficoltà pratiche. La sua implementazione richiede infatti tecnologie per il monitoraggio e la regolazione in tempo reale. Questi prerequisiti possono risultare impegnativi, soprattutto per le imprese meno strutturate o con risorse limitate. Inoltre, il successo del modello dipende strettamente dalla qualità e dall'affidabilità dei dati raccolti: ritardi o errori nella loro rilevazione possono influenzare negativamente l'efficacia del controllo, generando oscillazioni indesiderate nelle prestazioni operative.

Un aspetto critico da considerare è la necessità di sincronizzazione tra le diverse fasi della produzione. L'elevato livello di coordinamento richiesto dal controllo dinamico può complicarsi in caso di guasti, imprevisti o rapide variazioni nella domanda. Sebbene il sistema sia progettato per mantenere una buona capacità di risposta, rimane vulnerabile a queste condizioni. Per questo motivo, è fondamentale adottare strategie compensative e prevedere margini di sicurezza per mitigare i rischi. La sfida principale resta quella di individuare un equilibrio ottimale tra flessibilità operativa e stabilità complessiva.

Complessivamente, questo approccio di controllo dinamico si configura come una soluzione avanzata e promettente per ottimizzare i sistemi produttivi pull. Se implementato correttamente, può apportare vantaggi concreti in termini di produttività, qualità e adattabilità alle mutevoli condizioni di mercato. Tuttavia, il successo dipende dalla capacità dell'impresa di investire nelle infrastrutture tecnologiche adatte, garantire un'integrazione efficace delle varie fasi produttive e gestire consapevolmente la complessità operativa dell'intero sistema.

6.3 Riflessioni sulla gestione della flessibilità e congestione

La gestione della flessibilità e della congestione rappresenta un tema cruciale nei sistemi produttivi di tipo pull, che si caratterizzano per la capacità di adattarsi rapidamente alle variazioni della domanda e alle condizioni operative variabili. La flessibilità, intesa come la capacità del sistema di adeguare il ritmo produttivo, riorganizzare le sequenze operative e adattarsi efficientemente a cambiamenti rapidi o inattesi, è un elemento essenziale per mantenere un vantaggio competitivo nelle realtà industriali odierne, sempre più soggette a dinamiche di mercato volatili e imprevedibili. Nei sistemi pull, tale flessibilità si traduce nella possibilità di controllare in tempo reale il rilascio dei job e di effettuare una redistribuzione dinamica dei carichi di lavoro. Questi aspetti risultano cruciali per mitigare fenomeni come l'accumulo eccessivo in magazzino o l'insorgere di colli di bottiglia produttivi che ostacolano fluidezza ed efficienza.

D'altro canto, la congestione rappresenta una delle principali criticità da monitorare e gestire attentamente. L'accumulo di lavoro in attesa nei punti critici del sistema, come stazioni produttive sovraccaricate o risorse non perfettamente bilanciate, può condurre rapidamente a gravi inefficienze che compromettono la linearità del flusso produttivo. La problematica risiede nella capacità di progettare e configurare il sistema in modo da mantenere un equilibrio ottimale tra la capacità produttiva disponibile e le richieste in entrata, evitando che i nodi critici diventino fonti di rallentamento o di blocco. Per contrastare tali rischi, si ricorre

sempre più frequentemente a tecniche avanzate, come il controllo dinamico del Work In Progress e l'applicazione di regole di priorità. Tali strumenti permettono una distribuzione razionale degli incarichi tra le risorse disponibili, aiutando a prevenire accumuli indesiderati e garantendo una gestione fluida delle operazioni.

I risultati ottenuti mostrano come l'introduzione di meccanismi dinamici di controllo del WIP e di regole di priorità flessibili abbiano un impatto positivo nel contrastare gli inconvenienti legati sia alla congestione interna sia alla variabilità della domanda. Una caratteristica fondamentale di questi sistemi è la capacità di regolare automaticamente il rilascio dei job in funzione dello stato reale del processo produttivo, contribuendo così a migliorare diversi indicatori operativi. Tra questi spiccano la riduzione dei tempi medi di attesa, una maggiore prevedibilità del throughput e una gestione più efficiente del livello di saturazione delle risorse, mantenendolo entro soglie sostenibili. Tuttavia, per raggiungere tali obiettivi senza introdurre instabilità nel sistema, la flessibilità gestionale deve essere ben bilanciata con un'adeguata robustezza operativa. Un'eccessiva sensibilità alle variazioni può indurre fluttuazioni dannose nei livelli di WIP e nei tempi di attraversamento complessivi, compromettendo la stabilità dell'intero sistema.

Un aspetto rilevante è la necessità di una visibilità e monitoraggio continuo del sistema, che consenta di rilevare tempestivamente segnali di congestione o variabilità e di intervenire con strategie di aggiustamento mirate. L'integrazione di strumenti analitici e di monitoraggio in tempo reale favorisce decisioni più informate e rapide, migliorando la capacità di risposta e la resilienza complessiva. La gestione della flessibilità e congestione deve quindi basarsi su un approccio proattivo e coordinato, che faciliti la comunicazione tra le varie fasi produttive e la sincronizzazione delle attività.

Dal punto di vista decisionale, queste considerazioni evidenziano l'importanza di sviluppare meccanismi di controllo integrati e adattivi, in grado di prevenire le situazioni di congestione anziché limitarsi a gestirle retroattivamente. In questo contesto, l'utilizzo della simulazione si conferma come una risorsa strategica per valutare diversi scenari di carico e identificare configurazioni in grado di ottimizzare la produttività mantenendo un'elevata flessibilità operativa.

L'analisi evidenzia inoltre la necessità di trovare un equilibrio tra flessibilità e stabilità. Un eccesso di flessibilità può generare variabilità indesiderata, mentre un approccio eccessivamente rigido potrebbe compromettere la capacità di adattamento. Il segreto del

successo risiede nella progettazione di sistemi in grado di garantire un controllo del flusso produttivo che sia granulare, adattativo e modulato, con una gestione integrata delle congestioni supportata da politiche dinamiche di rilascio e priorità. Un simile approccio permette non solo di ottimizzare le prestazioni operative, ma anche di creare sistemi più sostenibili e resilienti nel lungo termine.

6.4 Suggerimenti per implementazione industriale

L'implementazione industriale di un controllo dinamico del WIP nei contesti produttivi con approccio pull richiede una metodologia integrata che coniughi sapientemente competenze tecnologiche, organizzative e gestionali. Sebbene i risultati ottenuti attraverso simulazioni evidenzino in modo inequivocabile i vantaggi di una gestione adattiva dei flussi produttivi, l'applicazione concreta di tali soluzioni in un ambiente reale necessita di una pianificazione accurata, unitamente a specifiche misure pratiche e strategiche.

Un elemento imprescindibile per il successo di tale implementazione è rappresentato dalla raccolta di dati affidabili e accessibili in tempo reale. La qualità e la tempestività delle informazioni disponibili costituiscono infatti la base consolidata su cui costruire qualsiasi modello di controllo dinamico. Per garantire un monitoraggio continuo ed efficace delle operazioni, è essenziale un'integrazione stretta tra sensori, sistemi MES (*Manufacturing Execution System*) e piattaforme digitali avanzate. Questi elementi devono operare in sinergia per fornire un quadro costantemente aggiornato delle condizioni di processo, comprendendo parametri come il livello attuale di WIP, i tempi di ciclo e il grado di utilizzo delle risorse. Una rete informativa robusta e ben strutturata costituisce pertanto il fondamento indispensabile per una gestione reattiva ed efficiente.

Oltre all'infrastruttura tecnologica, è necessario definire e calibrare attentamente le logiche di regolazione del flusso e i relativi parametri. Questi devono essere sviluppati tenendo conto della capacità operativa delle risorse, della variabilità nei tempi di processo e della fluttuazione della domanda del mercato. Nei contesti produttivi complessi, adottare una regolazione eccessivamente sensibile o reattiva può condurre a destabilizzazioni del flusso operativo, mentre un approccio eccessivamente rigido rischia di annullare i benefici attesi da un sistema dinamico. Per questa ragione, il percorso di implementazione deve includere una fase iniziale di sperimentazione controllata e successivo affinamento progressivo, fondandosi su dati storici e simulazioni analitiche per ottimizzare gradualmente il comportamento del sistema.

Da un punto di vista organizzativo, introdurre un controllo dinamico del WIP non significa solo apportare modifiche tecniche ai processi, ma richiede anche un cambio significativo nell'approccio gestionale. Gli operatori e i responsabili delle varie funzioni produttive devono essere adeguatamente formati per interpretare correttamente le indicazioni fornite dal sistema, intervenendo soltanto quando strettamente necessario. Il controllo automatico dovrebbe infatti gestire autonomamente la maggior parte delle situazioni operative, limitando al minimo le interruzioni manuali. Il coordinamento tra personale tecnico, ingegneristico e informatico diventa quindi essenziale per ottimizzare il funzionamento complessivo del sistema e garantirne la piena operatività.

È consigliabile avviare la sperimentazione del controllo dinamico del WIP in aree pilota o in linee produttive specifiche piuttosto che estenderlo immediatamente all'intera fabbrica. Questo approccio modulare permette di monitorare inizialmente i risultati in contesti circoscritti, facilitando l'identificazione di eventuali criticità locali e consentendo un'ottimizzazione incrementale dei parametri operativi. Inoltre, offre l'opportunità di trasferire progressivamente la conoscenza acquisita agli operatori, rafforzando le competenze interne necessarie prima di estendere il modello a livello globale.

Per massimizzare i benefici derivanti dall'adozione del controllo dinamico del WIP, esso dovrebbe essere integrato con altre pratiche della lean manufacturing, quali la riduzione dei tempi di setup, la standardizzazione delle procedure operative e l'implementazione continua di miglioramenti nei processi. Tale approccio sinergico permette non solo di amplificare l'efficacia dei risultati sul breve termine, ma contribuisce anche a rendere l'intero sistema produttivo più stabile, trasparente e pronto a reagire agli scenari mutevoli del mercato.

Nel complesso, l'adozione industriale del controllo dinamico del WIP rappresenta un passo significativo verso la fabbrica intelligente, in cui le decisioni operative vengono guidate da dati in tempo reale e algoritmi di ottimizzazione. Tuttavia, il successo di questa trasformazione dipende non solo dalla tecnologia, ma anche dalla capacità dell'organizzazione di adattare processi, competenze e cultura aziendale a un modello produttivo più flessibile, collaborativo e data-driven.

7. CONCLUSIONI

7.1 Conclusioni generali

Il lavoro svolto si è concentrato sull'analisi e comprensione del comportamento di un sistema produttivo regolato attraverso la logica CONWIP, con l'obiettivo di valutare l'impatto del controllo dinamico del Work In Process e delle diverse regole di priorità sulle prestazioni complessive del sistema.

Grazie all'attività di modellazione e simulazione, è stato possibile approfondire le interazioni tra variabili operative fondamentali, quali il throughput, i tempi medi di attraversamento e i livelli di WIP, delineando le dinamiche che influenzano stabilità ed efficienza del flusso produttivo.

L'analisi teorica, supportata dai risultati del modello semplificato, ha evidenziato che una regolazione dinamica dei limiti di WIP rappresenta una strategia efficace per migliorare le prestazioni complessive del sistema. In particolare, il controllo adattivo permette di incrementare la produttività in situazioni di bassa congestione e di ridurre i rischi di saturazione e deadlock nelle fasi più critiche. Questo approccio consente di ottenere un equilibrio ottimale tra efficienza operativa e stabilità del sistema.

Nel confronto tra le diverse regole di priorità, FIFO e SRPT hanno dimostrato prestazioni sostanzialmente equivalenti nei risultati simulati, suggerendo una bassa variabilità nei tempi di lavorazione all'interno del sistema. Tuttavia, in termini teorici, la regola SRPT dovrebbe offrire una gestione più efficiente del flusso produttivo, favorendo il completamento rapido dei lavori di breve durata e riducendo complessivamente la congestione del sistema. Al contrario, l'applicazione della regola LRPT appare meno vantaggiosa, comportando un aumento dei tempi medi di attraversamento e una lieve riduzione del throughput complessivo.

Nel loro insieme, i risultati ottenuti e le evidenze teoriche raccolte dimostrano l'efficacia del controllo dinamico del WIP come metodo di ottimizzazione per i sistemi manifatturieri discreti. Inoltre, sottolineano l'importanza di definire regole di priorità adeguate per garantire un flusso produttivo equilibrato, stabile e in grado di rispondere efficacemente alle variazioni operative.

7.2 Contributi della tesi

Il principale contributo di questa tesi consiste nello sviluppo di un modello concettuale di linea di produzione controllata con logica CONWIP, focalizzato sull'analisi dell'interazione

tra il controllo del WIP e l'adozione di strategie di priorità, al fine di comprendere il loro impatto combinato sui principali indicatori di prestazione di un sistema manifatturiero discreto.

Nonostante le difficoltà incontrate nell'esecuzione completa della simulazione, il lavoro ha permesso di consolidare un approccio metodologico rigoroso alla modellazione di sistemi complessi, integrando aspetti di movimentazione automatizzata, controllo del flusso e gestione dei colli di bottiglia. L'impostazione concettuale elaborata rappresenta un punto di partenza ben definito per future analisi e studi, poiché fornisce un quadro interpretativo solido che allinea in modo coerente vari ambiti teorici, quali regole di priorità, i fondamenti delle leggi sulla produzione (ad esempio, la legge di Little) e le logiche avanzate di controllo adattivo.

Un ulteriore valore aggiunto è dato dall'individuazione di un insieme strutturato di indicatori di prestazione specifici, tra cui throughput, tempo medio di attraversamento, probabilità di deadlock e unità in coda per entrare. Questi indicatori permettono una valutazione esaustiva sia dell'efficienza complessiva del sistema sia della sua stabilità operativa, offrendo uno strumento quantitativo per il monitoraggio e il miglioramento del processo produttivo.

Infine, l'introduzione e lo studio di un modello semplificato, che conserva intatta la logica fondamentale del flusso produttivo originale, hanno svolto un ruolo cruciale nella validazione qualitativa delle ipotesi teoriche elaborate.

Il lavoro complessivo, dunque, si configura come un contributo nello scenario della ricerca applicata ai sistemi produttivi avanzati.

7.3 Limiti del lavoro svolto

Il lavoro svolto presenta un limite significativo, rappresentato dalla mancata esecuzione completa della simulazione del modello integrale, dovuta a errori di configurazione legati alla gestione dei nastri trasportatori e alle logiche di sincronizzazione nei loop di movimentazione. Questi problemi, principalmente di natura computazionale, hanno impedito di ottenere risultati quantitativi completi, circoscrivendo l'analisi alla validazione teorica e concettuale del comportamento del sistema.

Nonostante queste difficoltà operative, l'approccio metodologico adottato si dimostra pienamente valido. La definizione delle logiche operative, accompagnata dalla strutturazione rigorosa degli scenari sperimentali e dalla scelta accurata dei KPI, costituisce una base metodologica forte e replicabile. Tali elementi potranno essere immediatamente

riutilizzati una volta risolti i problemi legati alla simulazione, consentendo un'applicazione pratica e più completa del modello.

Un ulteriore limite riguarda le semplificazioni introdotte nel modello alternativo realizzato senza i nastri trasportatori, il quale, pur semplificando significativamente le complessità del sistema, non permette di catturare accuratamente le dinamiche reali di congestione e accumulo tipiche di sistemi automatizzati complessi. Sebbene questa versione del modello sia stata utile per condurre un'analisi coerente delle logiche di controllo e priorità, essa compromette l'aderenza alle condizioni effettive dei processi produttivi reali.

In entrambi i modelli, generale e semplificato, sono state inoltre applicate alcune assunzioni che riducono la complessità del sistema rispetto a un contesto produttivo reale. Nello specifico, non sono stati presi in considerazione guasti, variabilità nei tempi di set-up o interruzioni operative. Anche i tempi di lavorazione delle macchine sono stati rappresentati attraverso una distribuzione triangolare semplificata. Per un maggiore realismo, sarebbe opportuno ampliare la raccolta di dati sperimentali, acquisendo informazioni più rappresentative sulle prestazioni delle risorse, comprese le fasi di rilavorazione e i parametri specifici della produzione.

Anche le stime relative alle probabilità di invio dei job alla stazione di controllo qualità (WSQC) e agli esiti delle operazioni di rilavorazione sono state effettuate in modo approssimativo, introducendo un ulteriore livello di semplificazione nel modello. Per superare questa limitazione, futuri sviluppi dovrebbero prevedere analisi di sensibilità più estese su tali variabili stocastiche, accompagnate dall'integrazione di ulteriori fattori dinamici. Tra questi, ad esempio, potrebbero essere considerati la manutenzione programmata, l'impiego di logiche predittive e la gestione delle variazioni nei turni lavorativi. Questi elementi contribuirebbero a migliorare significativamente la robustezza complessiva del modello, accrescendone anche la capacità di rappresentare fedelmente le reali condizioni operative.

Un altro elemento da sottolineare riguarda il numero limitato di scenari analizzati, che riduce la possibilità di generalizzare i risultati ottenuti. Sebbene la scelta di utilizzare un insieme circoscritto di regole di priorità e parametri operativi sia giustificata per una fase preliminare o dimostrativa, essa non è sufficiente per rappresentare la varietà potenziale delle configurazioni del sistema in condizioni più ampie e variabili.

In prospettiva futura, sarebbe opportuno estendere l'analisi mediante una valutazione più approfondita delle variabili stocastiche e l'introduzione di ulteriori aspetti dinamici. In particolare, considerare elementi come la manutenzione programmata e una gestione più dettagliata della variabilità nei turni lavorativi potrebbe consentire di rendere il modello ancora più robusto e aderente alla realtà operativa. Tali interventi andrebbero a colmare le attuali lacune, offrendo un supporto altamente affidabile per analisi più accurate e applicazioni pratiche.

7.4 Prospettive di estensione

Una possibilità di estensione riguarda l'introduzione di job caratterizzati da priorità di lavorazione differenziata, con l'obiettivo di simulare e analizzare scenari produttivi più realistici. In questi contesti, l'ordine con cui le attività vengono eseguite sarebbe determinato da una serie di vincoli che comprendono l'urgenza delle operazioni, l'importanza strategica e le scadenze temporali da rispettare. Un simile approccio metodologico offrirebbe l'opportunità di approfondire l'impatto che il controllo del Work in Progress esercita quando viene integrato con politiche di scheduling multicriterio. Ciò consentirebbe, inoltre, di verificare in modo più dettagliato come le regole di assegnazione delle priorità possano influenzare non solo i livelli di produttività complessiva del sistema, ma anche la sua stabilità nel tempo, contribuendo a ottimizzare processi e risorse in maniera più efficace e mirata.

Sarebbe anche utile ampliare l'analisi includendo le regole di priorità applicate alla stazione di controllo qualità, dove la gestione della coda dei job difettosi ha un impatto significativo sui tempi di rilavorazione e sulla fluidità del flusso complessivo. Approfondire scenari più complessi, nei quali differenti politiche di controllo del WIP si combinano con regole di priorità variabili nelle diverse workstation, consentirebbe di ottenere una visione più completa delle interdipendenze operative all'interno del sistema.

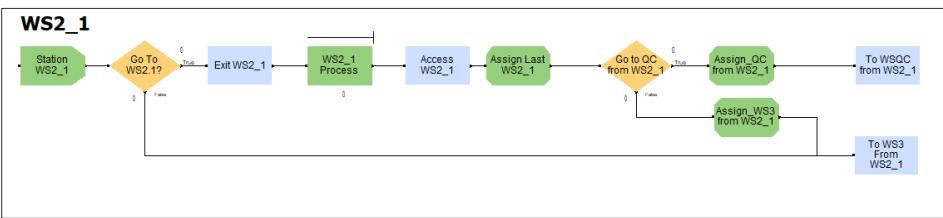
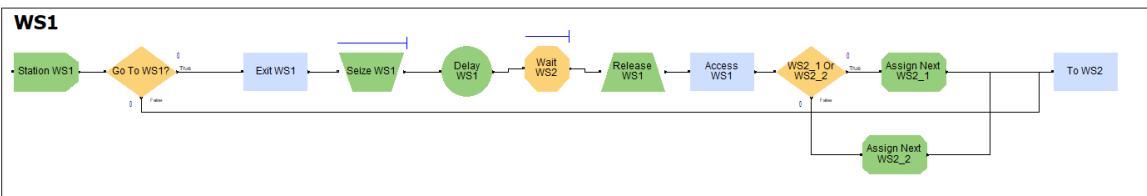
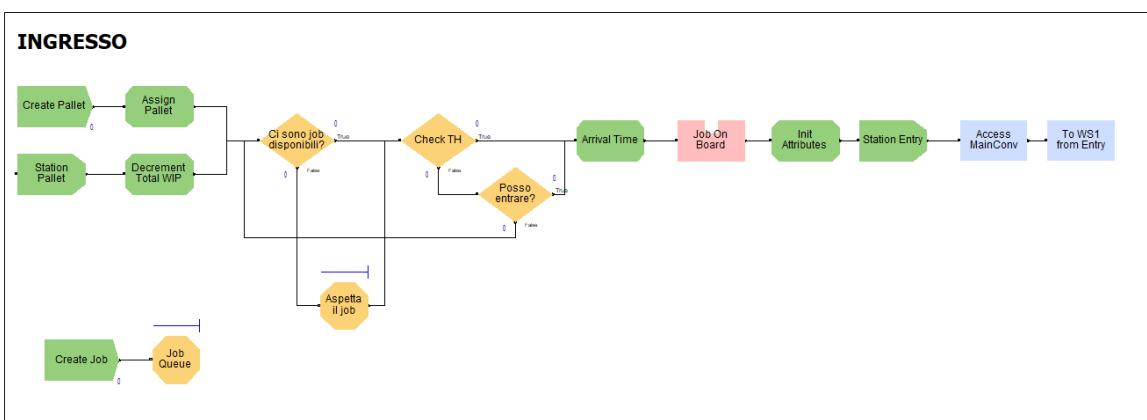
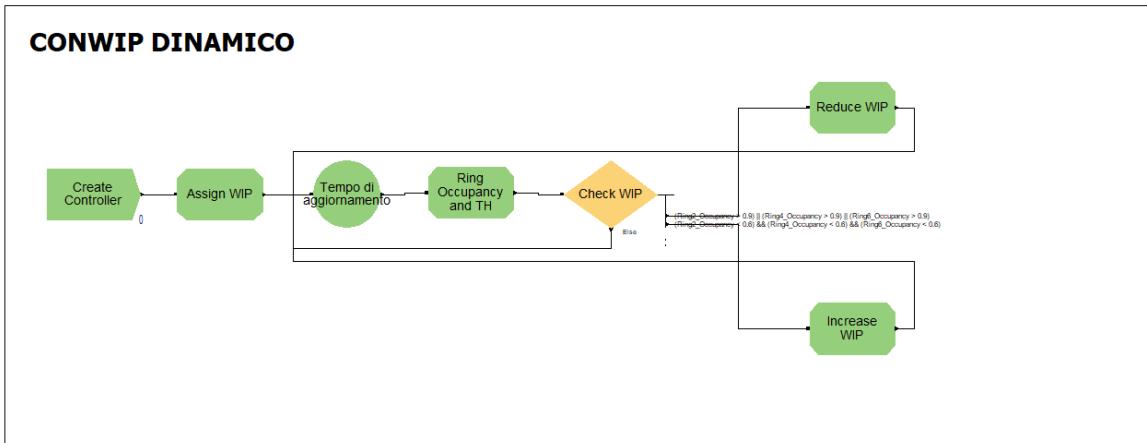
Un'ulteriore prospettiva interessante riguarda l'implementazione di controlli più avanzati, utilizzando tecnologie all'avanguardia come l'intelligenza artificiale avanzata e le logiche predittive. Questi strumenti tecnologici offrono la possibilità di modificare in maniera dinamica e adattiva i parametri del sistema, rispondendo in modo immediato e preciso alle condizioni operative che si manifestano in tempo reale. L'impiego di tali soluzioni non solo incrementerebbe la capacità del modello di reagire con maggiore prontezza ai cambiamenti, ma contribuirebbe anche ad anticipare potenziali situazioni critiche, come congestioni o blocchi del sistema, evitando che tali problematiche possano effettivamente verificarsi.

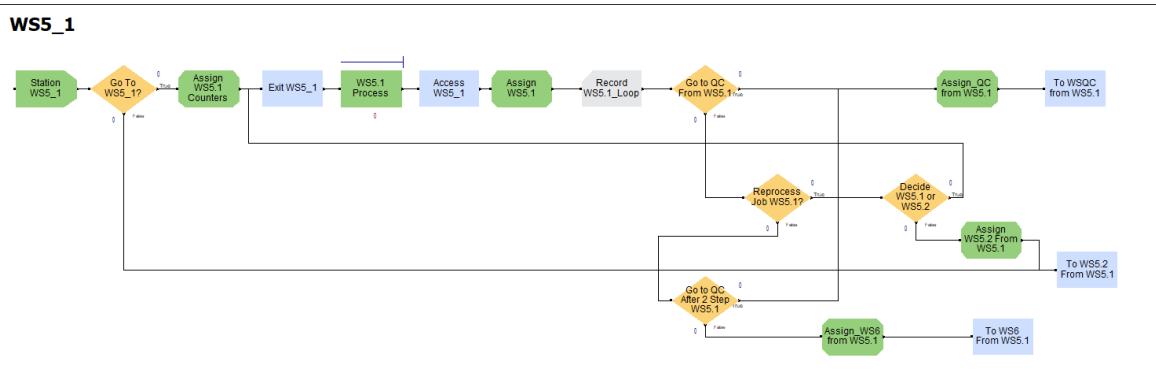
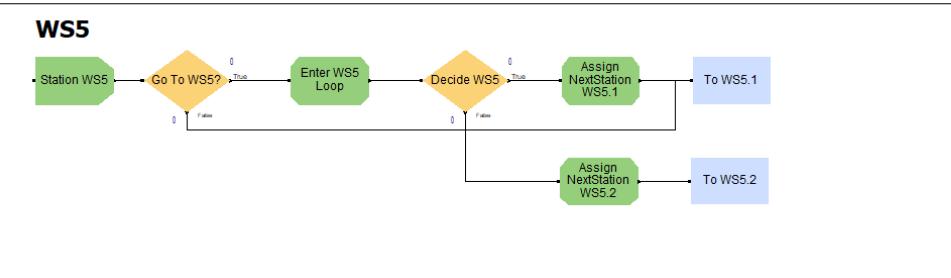
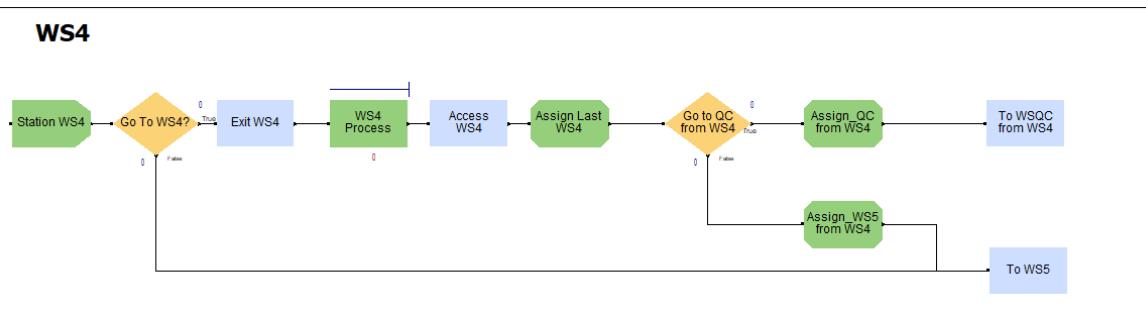
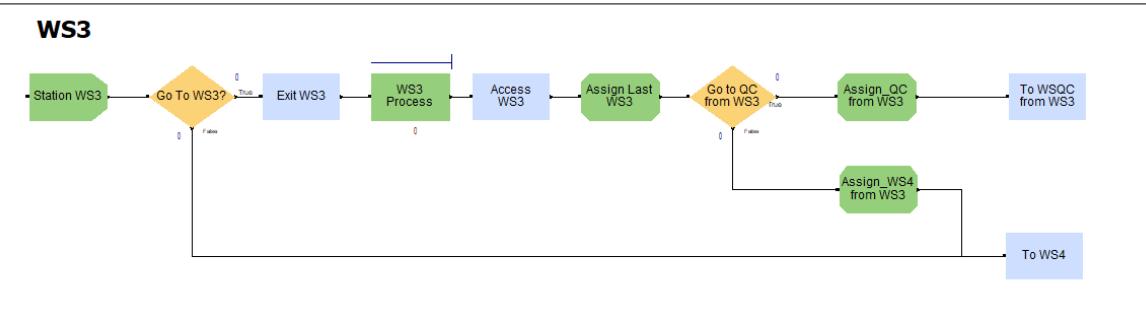
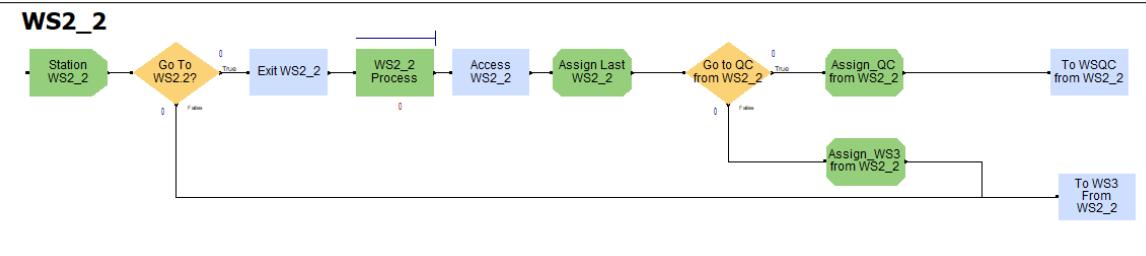
In aggiunta a ciò, sarebbe strategicamente vantaggioso considerare l'integrazione del monitoraggio del WIP e della gestione delle priorità con ulteriori aspetti organizzativi e gestionali, quali la pianificazione preventiva delle attività di manutenzione e l'ottimizzazione dei turni macchina. Una tale integrazione, considerando i vincoli operativi esistenti, consentirebbe di ottenere una visione globale, più dettagliata e realistica dell'intero processo produttivo. Questo approccio migliorerebbe la capacità dell'azienda di adattarsi alle circostanze variabili, riflettendo con maggiore fedeltà le dinamiche effettive della produzione e migliorando l'efficienza complessiva delle operazioni.

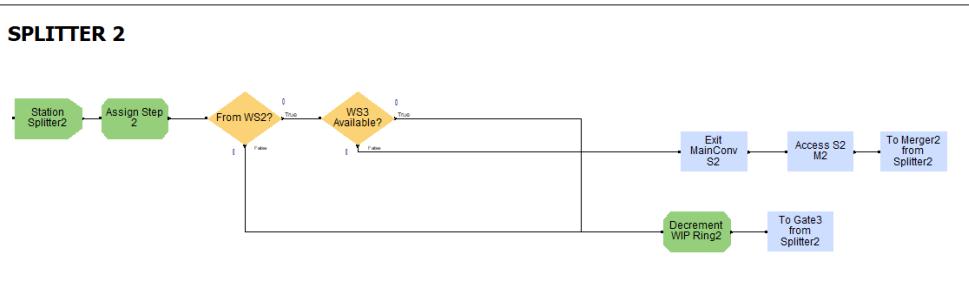
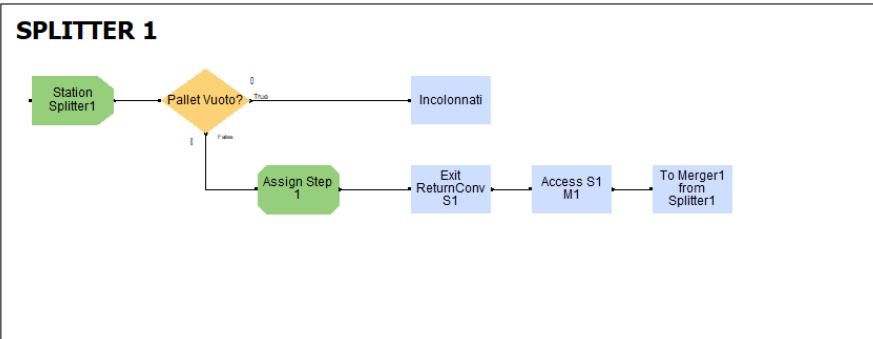
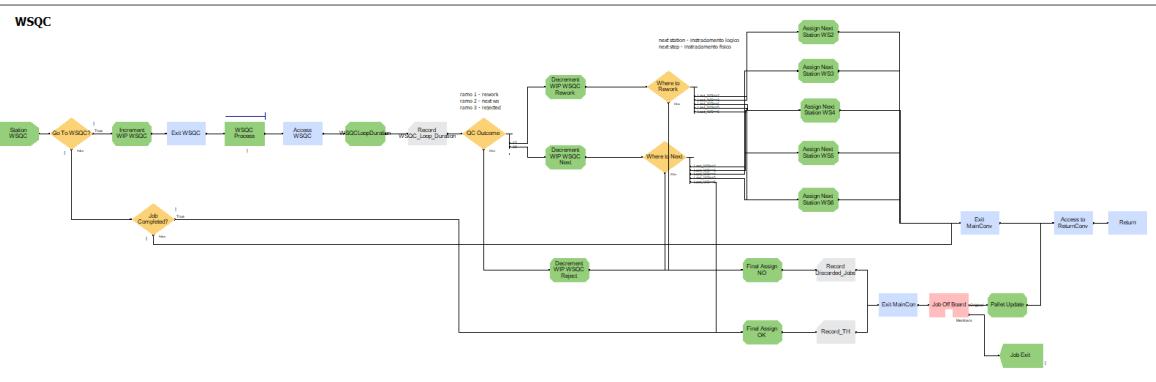
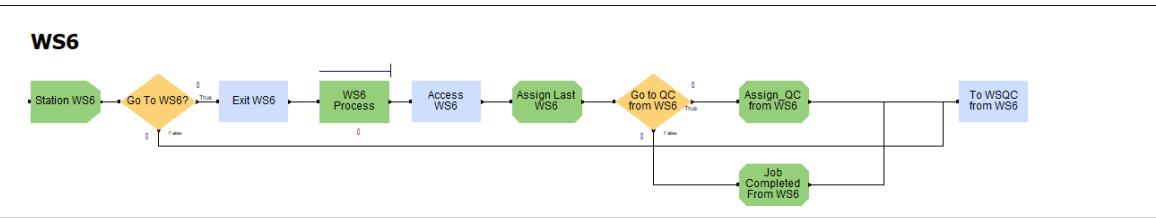
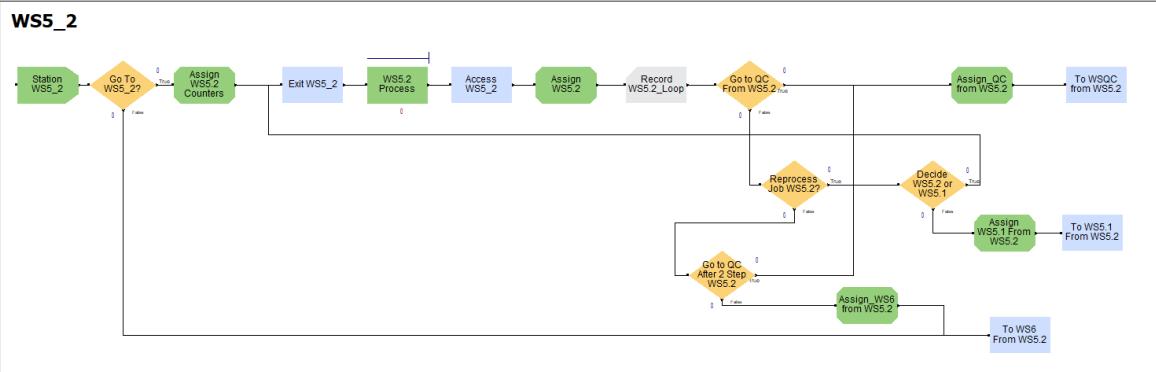
APPENDICI

Appendice A – Modello con nastri su ARENA Simulation

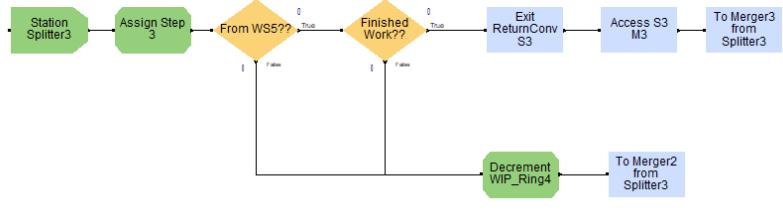
Questa appendice riporta le schermate principali del modello realizzato in ARENA Simulation, comprendenti la struttura dei moduli, i blocchi logici e le connessioni tra workstation, conveyor e stazione di controllo qualità. Le immagini sono fornite a fini di completezza e documentazione del modello implementato.



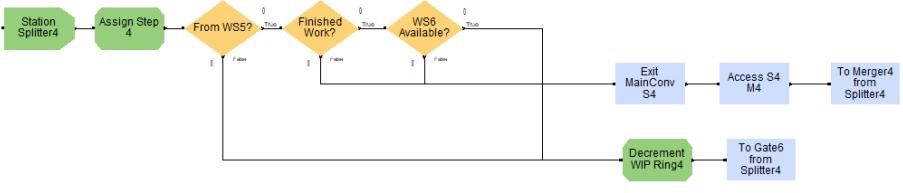




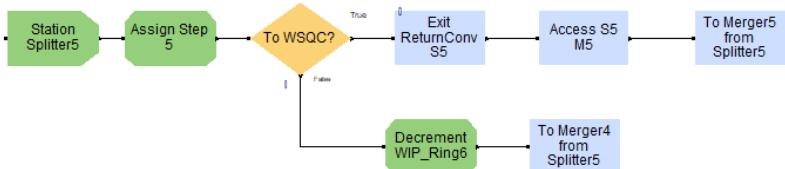
SPLITTER 3



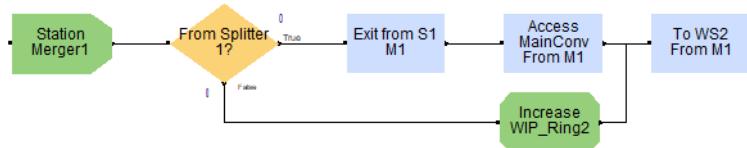
SPLITTER 4



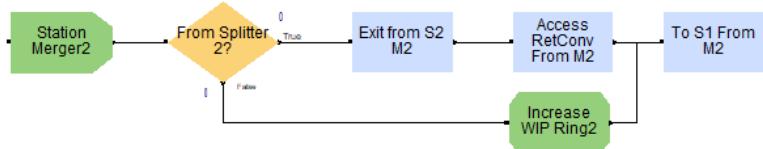
SPLITTER 5



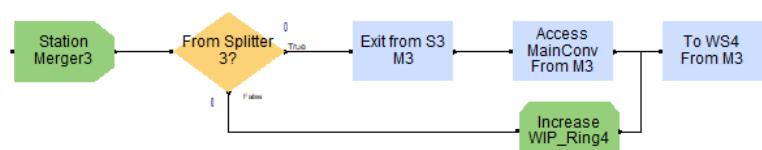
MERGER 1



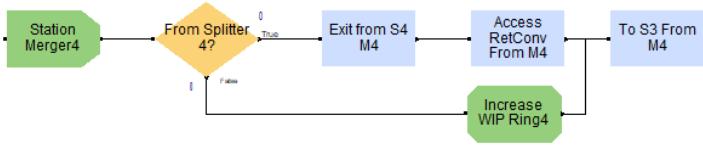
MERGER 2



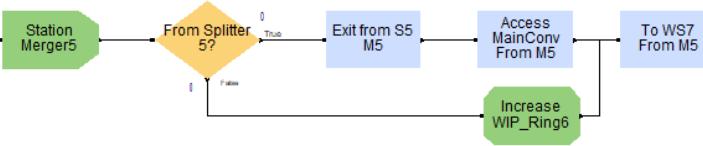
MERGER 3



MERGER 4



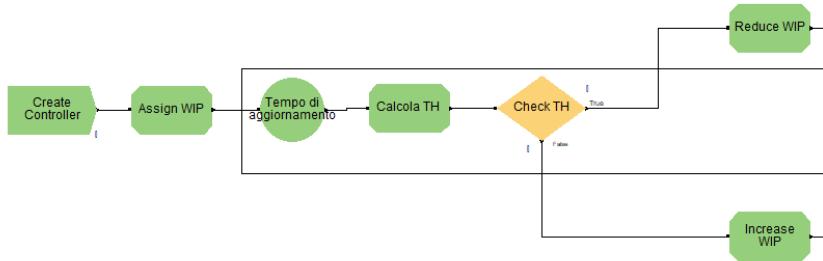
MERGER 5



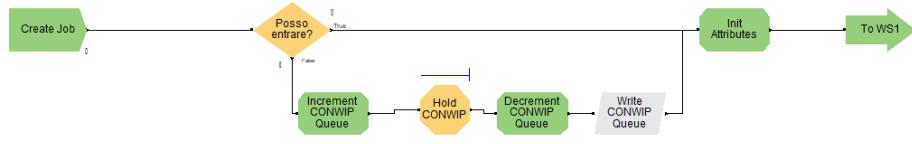
Appendice B – Modello senza nastri CONWIP dinamico su ARENA Simulation

Questa appendice presenta il modello di simulazione senza nastri trasportatori, sviluppato per analizzare il comportamento del sistema CONWIP con controllo dinamico del WIP. Le schermate riportate illustrano la struttura logica del modello, le regole operative implementate e i principali moduli utilizzati per la gestione adattiva del flusso produttivo.

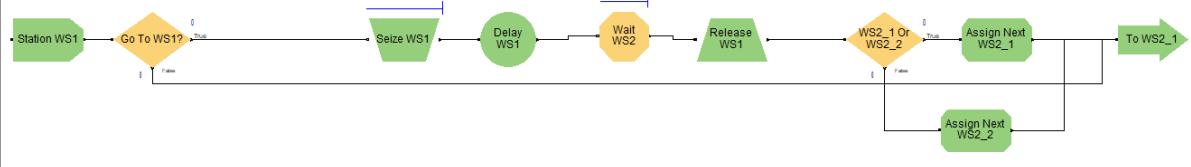
CONWIP DINAMICO

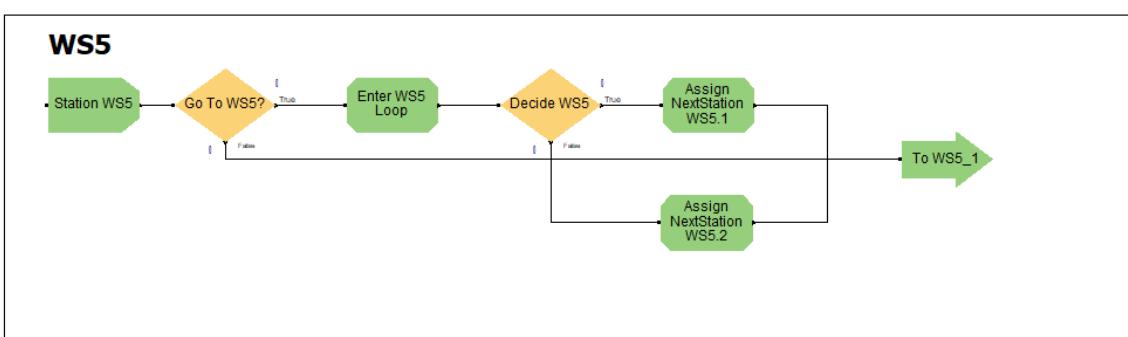
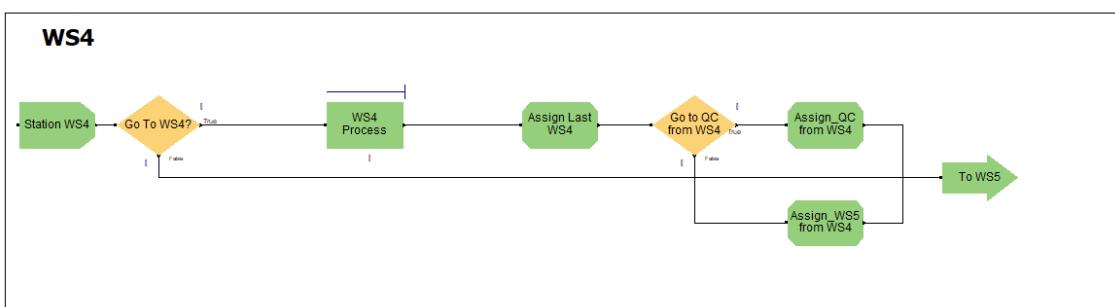
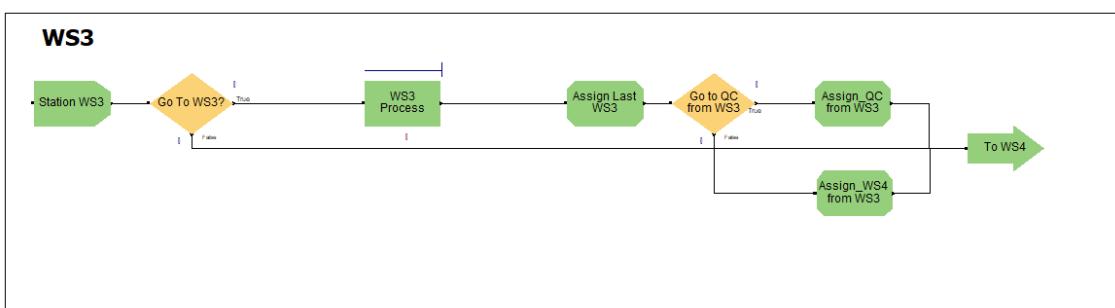
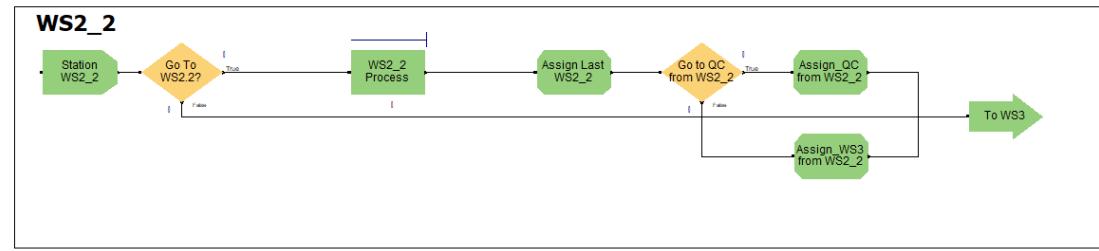
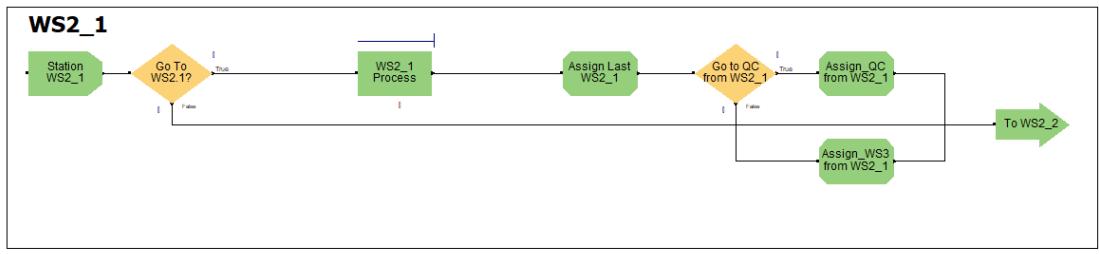


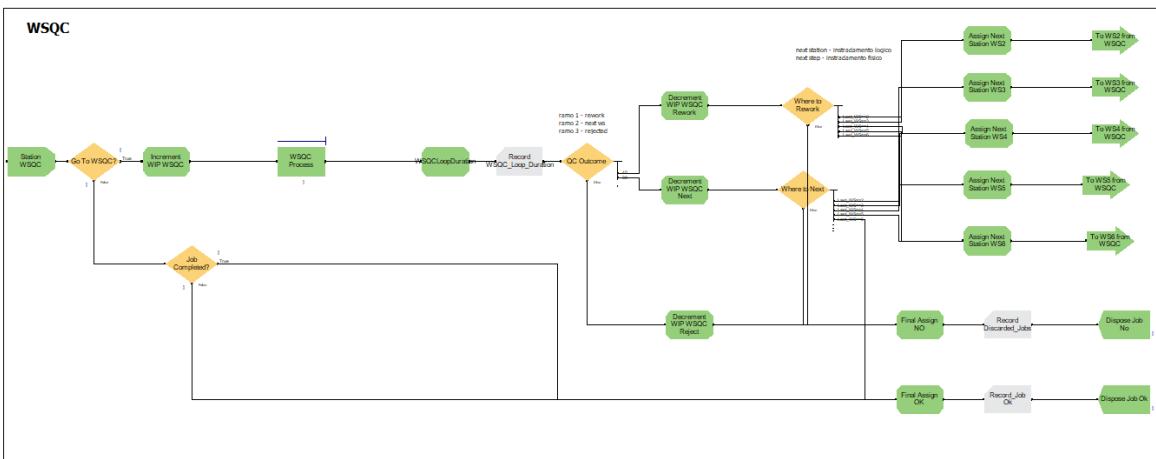
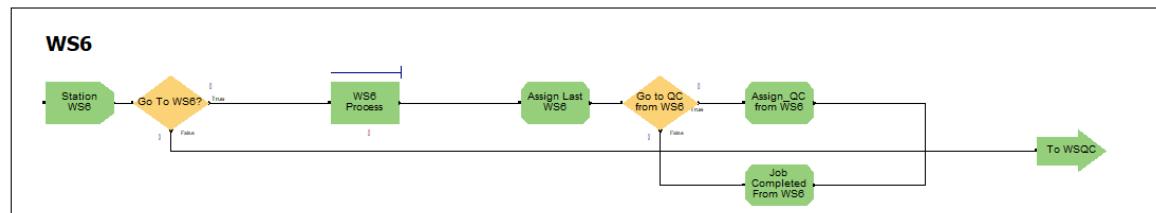
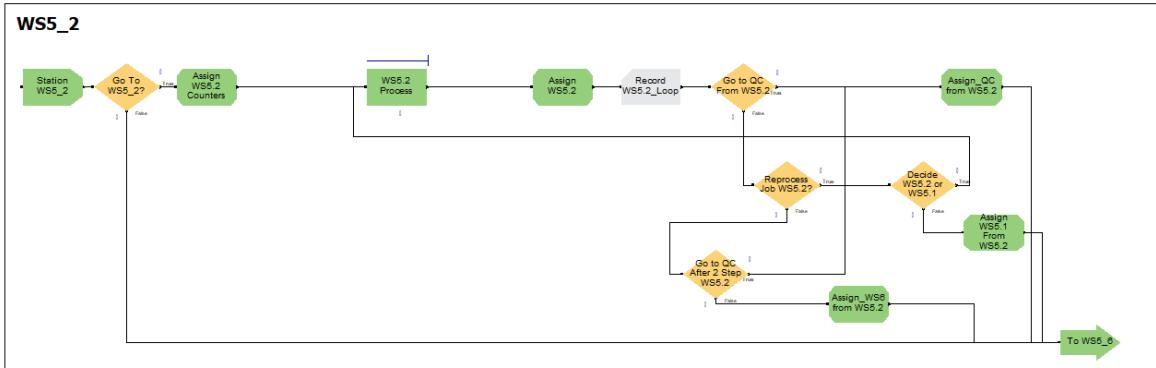
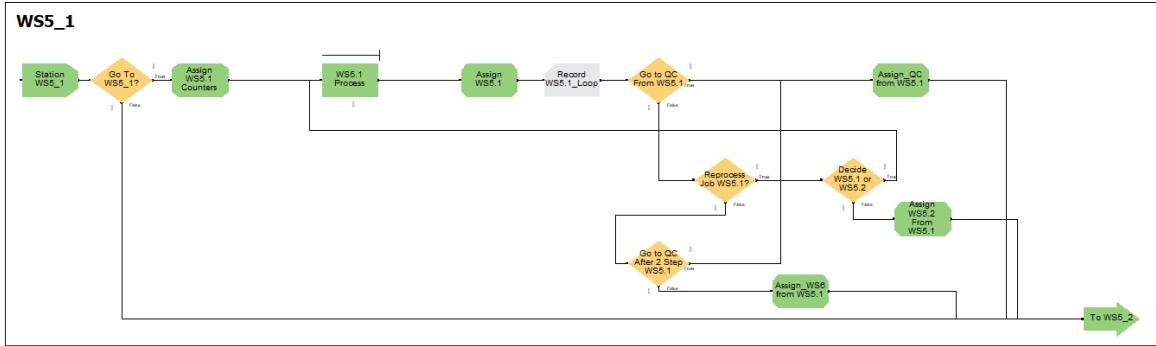
INGRESSO



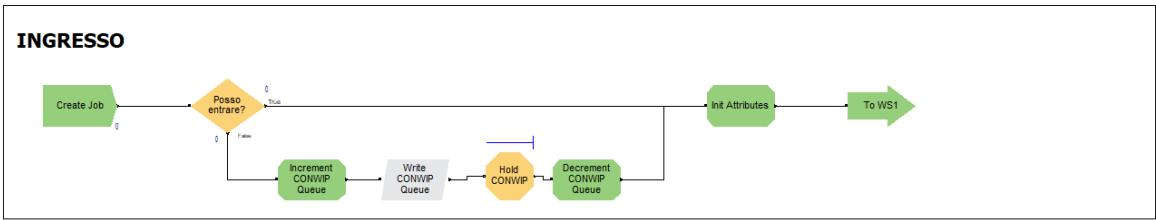
WS1







Appendice C – Modello senza nastri CONWIP statico su ARENA Simulation
 In questa appendice è riportato lo screenshot relativo alla sezione di ingresso del modello CONWIP statico, unica parte differente rispetto alla versione dinamica illustrata in Appendice B. Il resto della struttura logica e dei moduli di simulazione rimane invariato.



BIBLIOGRAFIA

- [1] Alaswad, Suzan, et al. "The effect of demand variability on supply chain performance." *2019 8th International Conference on Modeling Simulation and Applied Optimization (ICMSAO)*. IEEE, 2019.
- [2] Germain, Richard, Cindy Claycomb, and Cornelia Dröge. "Supply chain variability, organizational structure, and performance: the moderating effect of demand unpredictability." *Journal of operations management* 26.5, 2008, pp. 557-570.
- [3] Småros, Johanna, et al. "The impact of increasing demand visibility on production and inventory control efficiency." *International journal of physical distribution & logistics management* 33.4, 2003, pp. 336-354.
- [4] Wolniak, Radosław. "The concept of operation and production control." *Production engineering archives* 27.2, 2021, pp.100-107.
- [5] Alavian, Pooya, Peter Denno, and Semyon M. Meerkov. "Multi-job production systems: definition, problems, and product-mix performance portrait of serial lines." *International Journal of Production Research* 55.24, 2017, pp. 7276-7301.
- [6] Hemalatha, C., K. Sankaranarayanasamy, and N. Durairaaj. "Lean and agile manufacturing for work-in-process (WIP) control." *Materials Today: Proceedings* 46, 2021, pp. 10334-10338.
- [7] G. Herrera Vidal, J.R. Coronado Hernández, “*Complexity in manufacturing systems: a literature review*”, Production Engineering, 2021.
- [8] Deka, R., P. P. Borthakur, E. Baruah, P. Sarmah, and M. Saikia. "A Comprehensive Review on Mechanical Conveyor Systems: Evolution, Types, and Applications." Department of Mechanical Engineering, Dibrugarh University, Assam, India, 2024.
- [9] Tsimashenka, I., W. J. Knottenbelt, and P. G. Harrison. "Controlling Variability in Split-Merge Systems and Its Impact on Performance." Springer, 2014.
- [10] Khojasteh-Ghamari, Yaghoub. "A Performance Comparison Between Kanban and CONWIP Controlled Assembly Systems." *Journal of Intelligent Manufacturing*, vol. 20, no. 6, 2009, pp. 751-760.
- [11] Framinan, José M., Pedro L. González, and Raúl Ruiz-Usano. "Dynamic Card Controlling in a CONWIP System." *International Journal of Production Economics*, vol. 99, 2006, pp. 102–116.
- [12] Ip, W. H., M. Huang, K. L. Yung, D. Wang, and X. Wang. "CONWIP-Based Control of a Lamp Assembly Production Line." *Journal of Intelligent Manufacturing*, vol. 18, no. 2, 2007, pp. 261–271.
- [13] Khojasteh-Ghamari, Y. "A Performance Comparison between Kanban and CONWIP Controlled Assembly Systems." *Journal of Intelligent Manufacturing*, 2009 (forthcoming).
- [14] Dallery, Yves, and George Liberopoulos. "A Unified Framework for Pull Control Mechanisms in Multi-Stage Manufacturing Systems." *Annals of Operations Research*, 2000.

- [15] Khojasteh-Ghamari, Y. “Developing a Framework for Performance Analysis of a Production Process Controlled by Kanban and CONWIP.” *Springer Science+Business Media, LLC*, 2009.
- [16] “The Analysis of Performances of an Assembly Line in Synchronous Supply Managed with Kanban and Conwip Methods.” *Applied Mechanics and Materials*, vol. 657, Trans Tech Publications, 2014, pp. 966–970.
- [17] Xanthopoulos, A. S., G. Chnitidis, and D. E. Koulouriotis. “Reinforcement Learning-Based Adaptive Production Control of Pull Manufacturing Systems.” *Journal of Industrial and Production Engineering*, vol. 36, no. 3, 2019, pp. 155–170.
- [18] Wu, J., S. Yang, and Y. Chen. “Deadlock Control of Automated Manufacturing Systems Based on Petri Nets—A Literature Review.” *IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews*, vol. 42, no. 4, July 2012.
- [19] Minini, Andrea. “*Le reti di Petri*.” Andrea Minini, andreaminini.org/sistemi/reti-di-petri/. Accessed 1 Nov. 2025.
- [20] Castelnuovo, A., L. Ferrarini, and L. Piroddi. “An Incremental Petri Nets Approach to the Modeling of Manufacturing Systems.” *IEEE Transactions on Automation Science and Engineering*, vol. 4, no. 3, July 2007, pp. 424–434.
- [21] Chao, D. Y. “Computation of Elementary Siphons for Deadlock Control.” *The Computer Journal*, vol. 49, no. 4, 2006, pp. 470–479.
- [22] Ahmad, F., Huang, H., and Wang, X.-L. “Petri Net Modeling and Deadlock Analysis of Parallel Manufacturing Processes with Shared Resources.” *Journal of Systems and Software*, vol. 83, no. 4, 2010, pp. 675–688.
- [23] Badi, Ibrahim, Mouhamed Bayane Bouraima, Yanjun Qiu, and Željko Stević. “Optimizing Priority Sequencing Rules in Parallel Machine Scheduling: An Evaluation and Selection Approach Using Hybrid MCDM Techniques.” *Decision Analytics Journal*, vol. 10, Elsevier, 2024, p. 100354.
- [24] Moxoff. “Controllo qualità: Cos’è e Come Automatizzarlo.” *Moxoff* (blog).
- [25] Menghini, L., and L. Morato. *Confronto di due sistemi di ORR in sistemi produttivi Two-Stage Assembly Shop all'aumento di variabilità del routing e al variare del carico di lavoro di sistema: La limitazione del carico di lavoro e il bilanciamento del flusso*. Tesi di laurea, Politecnico di Milano, A.A. 2009/2010.
- [26] Shetwan, A. G., V. I. Vitanov, and B. Tjahjono. “Allocation of Quality Control Stations in Multistage Manufacturing Systems.” *Computers & Industrial Engineering*, vol. 60, no. 4, 2011, pp. 473–484.
- [27] Law, Averill M., and W. David Kelton. *Simulation Modeling and Analysis*. 5th ed., McGraw-Hill, New York, 2015.

- [28] Banks, Jerry, John S. Carson II, Barry L. Nelson, and David M. Nicol. *Discrete-Event System Simulation*. 5th ed., Pearson Education, Upper Saddle River, 2010.
- [29] Vespoli, S., G. Mattera, M. G. Marchesano, L. Nele, and G. Guizzi. "Adaptive Manufacturing Control with Deep Reinforcement Learning for Dynamic WIP Management in Industry 4.0." *Computers & Industrial Engineering*, 2025.
- [30] Gong, Mingwei, and Carey Williamson. "Quantifying the properties of SRPT scheduling." *11th IEEE/ACM International Symposium on Modeling, Analysis and Simulation of Computer Telecommunications Systems, 2003. MASCOTS 2003*. IEEE, 2003.
- [31] Nisar, Muhammad Usman, et al. "A Model for Evaluating the Impact of Priority Rules on Flow Time and Wait Time In A Job Shop Scheduling System: A Single Machine Case." *Proceedings of the 4th International Conference on Informatics, Technology and Engineering 2023 (InCITE 2023)*. Vol. 21. Springer Nature, 2023.
- [32] Lixin, Wang, Francis Tay Eng Hock, and Lee Loo Hay. "Scheduling MEMS manufacturing." *2000 Winter Simulation Conference Proceedings (Cat. No. 00CH37165)*. Vol. 2. IEEE, 2000.
- [33] Davydow, Alex, et al. "SRPT-based congestion control for flows with unknown sizes." *2021 IFIP Networking Conference (IFIP Networking)*. IEEE, 2021.
- [34] Arisha, Amr, Paul Young, and Mohie El Baradie. "Flow shop scheduling problem: A computational study." (2002).
- [35] Alenany, Emad, and M. Adel El-Baz. "Modelling a hospital as a queueing network: Analysis for improving performance." *Int. J. Ind. Manuf. Eng* 11.5 (2017): 1181-1187.
- [36] Zheng, Yousi, Prasun Sinha, and Ness B. Shroff. "Performance analysis of work-conserving schedulers for minimizing total flow-time with phase precedence." *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2012.