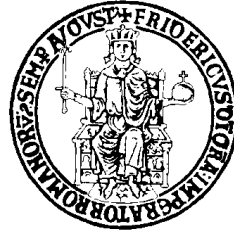


UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II



**SCUOLA DI DOTTORATO
IN
TECNOLOGIE E SISTEMI DI PRODUZIONE**

CICLO XXVII – TRIENNIO ACCADEMICO 2012/2015

**DIPARTIMENTO DI INGEGNERIA CHIMICA DEI
MATERIALI E DELLA PRODUZIONE INDUSTRIALE**

TESI DI DOTTORATO

**DUALITÀ DELLE LOGICHE LEAN-AGILE NEL
DIMENSIONAMENTO DELLE RISORSE PER L'EROGAZIONE DEI
SERVIZI OSPEDALIERI AD ALTO CONTENUTO TECNOLOGICO**

COORDINATORE

Ch.mo Prof. Ing. L. Carrino

TUTOR

Ch.ma Prof.ssa Ing. L.C. Santillo

CANDIDATO

Dott. Ing. G. Converso, PhD

ANNO ACCADEMICO 2013 – 2014

INDICE

INTRODUZIONE	14
---------------------	----

PARTE PRIMA **ELEMENTI STRUTTURALI, ANALITICI E MODELLISTICI**

CAPITOLO 1 – IL SISTEMA SANITARIO NOSOCOMIALE

1.1 - Il Sistema Sanitario Nazionale	17
1.2 - Strutture ospedaliere e standard assistenza	21
1.3 - Il concetto di efficienza in sanità	27
1.4 - Gestione per processi	32
1.5 - Definizione e caratteristiche dei processi	35

CAPITOLO 2 - LEAN SERVICE & AGILE PRODUCTION

2.1 - Lean Manufacturing e Lean Service	39
2.1.1 - Il Lean Thinking	43
2.1.2 - Strumenti della Lean Manufacturing	43
2.1.3 - Miglioramento Continuo	57
2.2 - Agile Production (e Manufacturing)	61
2.2.1 – Caratteristiche dell'Agile Supply Chain	69

CAPITOLO 3 – LO STRUMENTO MARKOVIANO NELL'ANALISI

3.1 - Concetti introduttivi	73
3.2 - I processi stocastici	76
3.3 - Filtraggio e Stopping Times	78
3.4 - Simulare una catena di Markov	81
3.5 - La proprietà di Markov	82
3.6 - Teoria Matriciale	84
3.6.1 - Una rappresentazione grafica	86
3.7 - Il teorema di base limite delle Catene di Markov	88
3.8 - Distribuzione limite per una C.M.	90
3.8.1 - Classi di comunicazione e irriducibilità	90
3.8.2 - Esempi	92
3.9 - Ricorrenza e Distribuzioni stazionarie	93
3.9.1 - Ricorrenza e Transienza	93
3.9.2 - Tempo di ritorno ad uno stato	95
3.9.3 - Distribuzione Stazionaria Limite	97
3.9.4 - Connessione tra $E(\tau_{jj})$ e π_j	99
3.9.5 - Calcolo algebrico di π	101
3.9.6 - Caso dello Spazio di Stato Finito	104
3.9.7- Stazionarietà di Catene Ricorrenti Positive	105
3.9.8 - Convergenza a π in senso forte e Aperiodicità	107

CAPITOLO 4 - MODELLAZIONE E LOGICHE DI SIMULAZIONE

4.1 - La simulazione dei sistemi	110
4.1.1 - I vantaggi della simulazione	111
4.2 – Discret Event Simulation	112
4.3 - System Dynamics	114
4.3.1 - Gli strumenti della System Dynamics	116
4.4 – Confronto: SD vs DES	122
4.5 - Elaborazione del modello dinamico di simulazione	124
4.6 - I software per la costruzione dei modelli SD	129

PARTE SECONDA

I SISTEMI PRODUTTIVI SANITARI: UNA LITERATURE REVIEW

CAPITOLO 5 – IL DEA: ARCHITETTURE E PROBLEMATICHE

5.1 - La gestione del P.S. in letteratura	133
5.2 - L'abbandono dal reparto di pronto soccorso	138
5.3 - Stato dell'arte delle tecniche risolutive	140
5.3.1 - Tecnologia RFID	142
5.3.2 - Bar Code e Braccialetto elettronico	145
5.3.3 - Altre soluzioni tecnologiche	147
5.4 – L'abbandono come rischio.	148

CAPITOLO 6 – LE LOGICHE LEAN E AGILE IN AMBIENTE SANITARIO

6.1 – Definizione e principali caratteristiche del servizio	152
6.2 - Lean HealthCare	157
6.2.1 – Modello per intensità di cure	167
6.2.2 – Lean: literature review	168
6.3 - Agile Healthcare	180
6.3.1 – Agile: literature review	183

CAPITOLO 7 - ELEMENTI ORIGINALI DI MODELLAZIONE AGILE

7.1 – Matrice delle caratteristiche: Lean vs Agile	192
7.1.1 – Lean Manufacturing: calcolo analitico del Takt Time	194
7.1.2 – Scelta della grandezza Agile d' interesse	195
7.2 – Agile Manufacturing: formulazione della caratteristica	197
7.2.1 - Definizione di α e ρ	202
7.2.2 – Osservazione sulla validità delle condizioni	204

PARTE TERZA

DUALITÀ DELLE LOGICHE LEAN AGILE IN SANITÀ

CAPITOLO 8 – LA LOGICA MARKOVIANA APPLICATA ALL’ANALISI DEL FENOMENO DELL’ ABBANDONO DEI PRONTO SOCCORSO

8.1 - Il reparto di pronto soccorso	210
8.1.1 - Requisiti strutturali minimi	211
8.2 - Descrizione delle principali aree di pronto soccorso	213
8.3 - Rappresentazione del processo sanitario	219
8.4 - Dia- gramma di flusso del processo di Pronto Soccorso	227
8.4.1 - Diagramma di flusso del processo di Pronto Soccorso	227
8.4.2 - Modello del processo tramite l’approccio Markoviano	229
8.4.3 - La Catena di Markov “comportamentale”	232
8.5 - Validazione dello studio presso A.O.R.N. “A . Cardarelli”	239
8.5.1 - Descrizione A.O.R.N. “A. Cardarelli	239
8.5.2 - Il processo di Pronto Soccorso nell’A.O.R.N. Cardarelli	243
8.5.3 - Validazione del modello	256
8.5.4 - Comportamento della catena nelle condizioni attuali	258
8.6 - Analisi di sensitività	261

CAPITOLO 9 – LA MODELLAZIONE DELLE U.O. DI EMERGENZA

9.1 - L’analisi di una U.O. di Pronto Soccorso di DEA	266
9.2 - Stock & Flow Diagr per la simulazione di una U.O. di P.S.	271
9.2.1 - L’approccio matriciale	272
9.3 Il modello di Simulazione	275
9.4 - La modellazione degli arrivi per la eterminazione del carico	286
9.4.1 - I processi di poisson	288
9.5 - Validazione del modello	290
9.5.1 - Analisi dei risultati ottenuti in configurazione “as is”	295
9.6 - Il modello in configurazione Agile	299
9.7 – Osservazioni sulle risultanze	307

CAPITOLO 10 – LA MODELLAZIONE DELLE U.O. DI URGENZA DEI D.E.A.

10.1 - Sviluppo del modello di simulazione per l'Urgenza	311
10.1.1 - Costruzione Causal Loop Diagram	313
10.1.2 - L'architettura del modello	318
10.2 - Descrizione della struttura di urgenza dell' A.O.R.N	334
10.3 - Implementazione del modello	336
10.4 - Onda di carico e condizione di stazionarietà degli impianti	338
10.5 - Validazione del modello	344
10.6 - Soluzioni proposte: Scenario TO BE	352
10.7 - Sviluppi futuri	374
CONCLUSIONI	378
BIBLIOGRAFIA	383

INDICE DELLE FIGURE

Figura 1: <i>Il processo nelle aziende sanitarie</i>	31
Figura 2: <i>Le sette categorie di sprechi</i>	42
Figura 3: <i>La TPS come una casa</i>	44
Figura 4: <i>Time Line</i>	45
Figura 5: <i>Da una produzione a lotti al one-piece flow</i>	49
Figura 6: <i>Current state map e relativa future state map</i>	52
Figura 7: <i>Simboli per il VSM</i>	53
Figura 8: <i>Il sistema Kanban</i>	55
Figura 9: <i>Il ciclo di Deming</i>	59
Figura 10: <i>I 4 elementi chiave dell'Agile Manufacturing</i>	67
Figura 11: <i>Diagramma di Kano</i>	156
Figura 12: <i>Quadro concettuale del Lean Thinking</i>	158
Figura 13: <i>Gestione tipica delle attività in un processo reale</i>	163
Figura 14: <i>Natura delle variabilità in ambito sanitario</i>	166
Figura 15: <i>Esempio dell'andamento della domanda in ambiente Lean Manufacturing</i>	194
Figura 16: <i>Profilo di domanda nel tempo</i>	198
Figura 17: <i>Caratteristica di sollecitazione puntuale del sistema</i>	199
Figura 18: <i>Andamento pdf distribuzione esponenziale</i>	201
Figura 19: <i>Grafico tramite cui è possibile calcolare il ritmo del sistema in funzione del livello di servizio</i>	207
Figura 20: <i>Esempio di causal loop diagram</i>	117
Figura 21: <i>Esempio di circuito a retroazione positivo</i>	118
Figura 22: <i>Esempio di circuito a retroazione negativo</i>	119
Figura 23: <i>Comportamento dinamico caratterizzato da oscillazioni</i>	120
Figura 24: <i>Comportamento dinamico caratterizzato da oscillazioni</i>	120
Figura 25: <i>Costrutti elementari dei modelli in ambiente Powersim</i>	132

Figura 26: <i>Causal Loop Diagram</i>	314
Figura 27: <i>Modello di simulazione</i>	321
Figura 28: <i>Costrutto che consente il caricamento dei dati relativi ai pazienti provenienti da altre unità operative o altre strutture ospedaliere</i>	324
Figura 29: <i>Costrutto che consente il caricamento dei dati relativi agli accessi presso il Pronto Soccorso</i>	325
Figura 30: <i>Costrutto clessidra-tempo random per la visita medica</i>	327
Figura 31: <i>Costrutto orologi verde/rosso associati ai pazienti</i>	329
Figura 32: <i>Costrutto clessidra per il laboratorio di analisi</i>	330
Figura 33: <i>Costrutto rappresentante il loop d'urgenza</i>	331
Figura 34: <i>Costrutto della gestione dei farmaci</i>	332
Figura 35: <i>Estensione territoriale e suddivisione in pianta della struttura ospedaliera "A. CARDARELLI"</i>	336
Figura 36: <i>Andamento della domanda in una giornata ritenuta significativa</i>	342
Figura 37: <i>Andamento degli accessi nel turno di lavoro 8-14</i>	343
Figura 38: <i>Andamento degli accessi nel turno di lavoro 14-20</i>	343
Figura 39: <i>Andamento degli accessi nel turno di lavoro 20-8</i>	344
Figura 40: <i>Costrutto che consente il calcolo dell'andamento del coefficiente di riempimento dei posti letto</i>	345
Figura 41: <i>Confronto tra output reali e output simulati</i>	351
Figura 42: <i>Confronto tra i valori del coefficiente di riempimento nei due scenari</i>	355
Figura 43: <i>Andamento del coefficiente di riempimento per 30 run nei due scenari</i>	Errore. Il segnalibro non è definito.
Figura 44: <i>Costrutto per valutare la permanenza media presso il DEA</i>	358
Figura 45: <i>Grafico di confronto dei tempi di permanenza nei due scenari</i>	359
Figura 46: <i>Costrutto del massimo tempo di attesa per essere sottoposti ad intervento chirurgico</i>	361

Figura 47: *Grafico dei max tempi di attesa per l'intervento registrati su 30 run nei 2 scenari*.....363

Figura 48: *Grafico dei tempi medi di refertazione registrati in 30 run nei due scenari*365

INDICE DELLE TABELLE

Tabella 1: <i>Dimensionamento del sistema sanitario rispetto al bacino di utenza</i>	24
Tabella 2: <i>Principali differenze tra beni e servizi</i>	153
Tabella 3: <i>Caratteristiche Lean Production vs Agile Manufacturing</i>	192
Tabella 4: <i>Calcolo di k e Δk negli intervalli di tempo statisticamente significativi</i>	341
Tabella 5: <i>Posti letto presenti nelle unità operative costituenti il DEA</i>	346
Tabella 6: <i>Confronto tra dati reali e dati ottenuti dalla simulazione relativamente al</i>	349
Tabella 7: <i>Valori del Coefficiente di riempimento dei posti letto nei due scenari</i>	354
Tabella 8: <i>Confronto tempi di permanenza media presso il DEA nei due scenari</i>	356
Tabella 9: <i>Confronto massimo tempo di attesa intervento nei due scenari</i> .	359
Tabella 10: <i>Confronto tra le medie dei tempi di refertazione nei due scenari</i>	361
Tabella 11: <i>Valori del Coefficiente di riempimento dei posti letto nei due scenari (TO BE 2)</i>	363

INTRODUZIONE

L'espressione "diritto alla salute" sintetizza il diritto di ogni persona ad ottenere prestazioni sanitarie adeguate. Sebbene evidente, si ritiene doveroso sottolineare l'importanza della Sanità quale argomento di pubblico dominio che, soprattutto negli ultimi decenni, è stato ampiamente trattato e posto al centro di dibattiti socio-politici del Paese, in particolare per l'ingente consumo e spreco in termini di risorse soprattutto finanziarie e con, in molti casi, difficoltà nel raggiungimento di standard qualitativi prefissati.

I nosocomi offrono spunti gestionali di estremo interesse modellistico, in quanto concentrano al proprio interno strutture organizzative complesse e ad elevato contenuto tecnologico, che presentano severe problematiche di ottimizzazione, sia dal punto di vista economico che dal punto di vista di risorse impiegate, al fine di aumentare l'efficienza dei servizi erogati nonché la percezione del paziente sulla qualità delle prestazioni sanitarie. Pertanto, in questo quadro generale, si introduce l'importanza di approfondire le moderne tecniche di gestione dei processi produttivi adattandole all'erogazione dei servizi ed in particolare al sistema sanitario. L'obiettivo del presente lavoro di Tesi è quello di analizzare i processi esistenti in un Dipartimento d'Emergenza e Accettazione – D.E.A. -, evidenziandone le criticità così da proporre un miglioramento di tali processi attraverso una riconfigurazione delle risorse presenti. In particolare sono state valutate due politiche di gestione, la Lean Service e l'Agile Production, per determinare quale di queste riesce a garantire il raggiungimento di migliori performance sia in ottica di un miglioramento del livello di servizio offerto al cliente sia in termini di minori tempi per l'esecuzione delle attività. La Lean Service è una

metodologia che si pone come obiettivo, strategico e gestionale, la massimizzazione del valore attraverso la minimizzazione dello spreco.

Si definisce Agile Production (o Manufacturing, per gli impianti per la produzione di beni), invece, l'insieme delle strategie organizzative che si concentra sul rafforzamento del sistema di risposta ai clienti attraverso il coordinamento delle risorse e una migliore flessibilità, ridisegnando strutture organizzative, sistemi informativi, processi logistici, ed euristiche decisionali di gestione.

L'obiettivo del presente elaborato è stata raggiunto realizzando una articolata serie di modelli analitici fondati sull'ausilio dei principi e delle tecniche più avanzate dell'analisi dei sistemi, quali le catene Markoviane e l'approccio simulativo dinamico, che consentono di comprendere la correlazione esistente tra le cause e gli effetti delle politiche gestionali adottate in azienda e l'evoluzione dinamica di tutti i fenomeni aziendali.

Questo lavoro di tesi nasce dall'esperienza di Dottorato condotto, per la parte di implementazione sperimentale dei modelli, presso l'Azienda Ospedaliera di Rilievo Nazionale "A. Cardarelli" di Napoli.

L'elaborato si divide in tre parti e dieci capitoli.

Nella Parte Prima sono forniti tutti i fondamenti teorici, normativi, funzionali e modellistici poi utilizzati nel seguito del lavoro. In modo particolare viene illustrato il concetto di "sistema sanitario", le sue caratteristiche e gli obiettivi che esso si prefigge, nonché le metodologie con cui un sistema sanitario viene gestito. Successivamente si definisce il contesto teorico di riferimento per la gestione produttiva di simili impianti: la Lean Service e l'Agile Production. Infine nei capitoli 3 e 4, si espongono gli strumenti di analisi utilizzati nella parte terza

La seconda parte dell'elaborato è dedicato per il suo sviluppo quasi integrale, ad una approfondita *literature review* dei modelli di riferimento, fino ad oggi adottati per la trattazione dei problemi di produzione industriale a sollecitazione "livellata" ovvero variabile, con particolare riferimento a quelli orientati ai sistemi sanitari. Il Capitolo 7, ultimo di questa parte dell'elaborato, ha lo scopo di definire un primo corpo modellistico in chiave analitica rivolto al superamento (per quanto possibile a chi scrive) dello stato dell'arte emerso a valle della ricerca bibliografica condotta.

Nella terza ed ultima parte, infine, sono realizzati e presentati i modelli di simulazione e valutazione atti a descrivere le problematiche produttive di un Dipartimento di Emergenza e Accettazione di un ospedale moderno, distinguendo gli impianti di Emergenza (Pronto Soccorso) da quelli di Urgenza (terapie intensive e medicine operatorie di urgenza).

Si è proceduto, ovviamente, alla validazione dei modelli in configurazione "as is", sulla scorta degli andamenti emersi dall'analisi dell'A.O.R.N "A.Cardarelli" di Napoli. Successivamente sono stati considerate e analizzate le risposte dei sistemi all'introduzione di alcune soluzioni (riconfigurazione del modello in versione "to be"), che rispondono ad una visione Agile del sistema produttivo rispetto alla precedente visione Lean, in cui le risorse sono dimensionate in forma (almeno parzialmente) stazionaria. In ultimo, valutati e commentati i risultati operativi per un sistema ospedaliero DEA, sono state riconsiderate le caratteristiche strutturali del modello analitico Agile presentato in chiusura della parte seconda riformulandolo opportunamente alla luce delle conclusioni esposte.

CAPITOLO 1

Sistema sanitario nosocomiale

1.1. – Il Sistema Sanitario Nazionale

Il Sistema Sanitario è l'insieme organizzato delle persone, delle istituzioni e delle risorse umane e materiali il cui fine è la promozione, il recupero e il mantenimento della salute della popolazione.

Frenk definisce il Sistema Sanitario come *“un sistema complesso che non deve essere analizzato in termini di singole componenti, ma anche delle relazioni che tra essi intercorrono; inoltre nella valutazione e nella descrizione di un tale sistema non si dovrebbero considerare unicamente gli aspetti istituzionali ma anche le caratteristiche della popolazione a cui esso si rivolge”*.

Il Servizio Sanitario Nazionale –SSN-, nell'ordinamento giuridico italiano, identifica il complesso delle funzioni, delle attività e dei servizi assistenziali gestiti ed erogati dallo stato italiano. Attraverso di esso viene data attuazione all'art. 32 della Costituzione italiana che sancisce il “diritto alla salute” di tutti gli individui. Si pone dunque come un sistema pubblico di carattere “universalistico”, tipico di uno stato sociale, che garantisce l'assistenza sanitaria a tutti i cittadini. È finanziato dallo Stato attraverso la fiscalità generale e le entrate dirette, percepite dalle aziende sanitarie locali, attraverso ticket sanitari (cioè delle quote con cui l'assistito contribuisce alle spese) e prestazioni a pagamento.

Esso è costituito sostanzialmente dai vari servizi sanitari regionali, da enti e istituzioni di rilievo nazionale e dallo Stato, volte a garantire l'assistenza sanitaria ovvero la tutela o salvaguardia della salute dei cittadini.

Secondo una ricerca dell'OMS, l'Organizzazione Mondiale della Sanità, risalente al 2000, l'Italia aveva il secondo Sistema Sanitario migliore del mondo in termini di efficienza di spesa e accesso alle cure pubbliche per i cittadini, dopo la Francia.

Nel 2014, secondo una classifica elaborata da Bloomberg, risultava terza nel mondo per efficienza della spesa.

In base al principio di sussidiarietà, il servizio sanitario è articolato secondo diversi livelli di responsabilità e di governo, che sono:

- livello centrale: lo Stato ha la responsabilità di assicurare a tutti i cittadini il diritto alla salute, mediante un forte sistema di garanzie, attraverso i Livelli Essenziali di Assistenza (LEA);
- livello regionale: le Regioni hanno la responsabilità diretta della realizzazione del governo e della spesa per il raggiungimento degli obiettivi di salute del Paese.

Le Regioni hanno competenza esclusiva nella regolamentazione ed organizzazione di servizi e di attività, destinate alla tutela della salute, e dei criteri di finanziamento delle Aziende Sanitarie Locali e delle Aziende Ospedaliere, anche in relazione al controllo di gestione e alla valutazione della qualità delle prestazioni sanitarie nel rispetto dei principi generali fissati dalle leggi dello Stato.

Il Servizio Sanitario Nazionale non è dunque un'unica amministrazione, ma un insieme di enti ed organi che concorrono al raggiungimento degli obiettivi di tutela della salute dei cittadini. Lo compongono infatti:

- il Ministero della Salute, che coordina il piano sanitario nazionale, ferme le competenze costituzionalmente garantite delle Regioni;

ed una serie di enti e organi a livello nazionale, quali:

- il Consiglio Superiore di Sanità (CSS);

- l’Istituto Superiore di Sanità (ISS);
- l’Istituto Superiore per la Prevenzione e Sicurezza del Lavoro (ISPESL) incorporato nel 2013 dall’ INAIL;
- l’Agenzia Nazionale per i Servizi Sanitari Regionali (Age.na.s.);
- gli istituti di ricovero e cura a carattere scientifico (IRCCS);
- gli Istituti Zooprofilattici Sperimentali;
- l’Agenzia Italiana del Farmaco (AIFA);
- i “servizi sanitari regionali”, che a loro volta comprendono:
 - le regioni e le province autonome di Trento e Bolzano;
 - le aziende sanitarie locali (ASL) e le aziende ospedaliere (AO), attraverso le quali le regioni e le province autonome assicurano l’assistenza sanitaria.

Il finanziamento del Fondo Sanitario Nazionale trova le proprie fonti in:

- entrate proprie convenzionali e ricavi delle aziende sanitarie;
- compartecipazione da parte delle regioni a statuto speciale;
- IRAP, Imposta Regionale Attività Produttive;
- IRPEF, Imposte di Reddito sulle Persone Fisiche.

Il Servizio Sanitario Nazionale è caratterizzato da un sistema di “programmazione sanitaria” che si articola:

- nel Piano sanitario nazionale;
- nei piani sanitari regionali.

Il “Piano sanitario nazionale” ha durata triennale (anche se può essere modificato nel corso del triennio) ed è adottato dal Governo, su proposta del Ministero della Salute, sentite le commissioni parlamentari competenti, nonché le confederazioni sindacali maggiormente rappresentative, tenendo conto delle proposte trasmesse dalle regioni.

Esso indica:

- le aree prioritarie di intervento, anche ai fini di una progressiva riduzione delle diseguaglianze sociali e territoriali nei confronti della salute;
- i livelli essenziali di assistenza sanitaria da assicurare per il triennio di validità del Piano;
- la quota capitaria di finanziamento assicurata alle regioni, per ciascun anno di validità del Piano, e la sua disaggregazione per livelli di assistenza;
- gli indirizzi finalizzati a orientare il Servizio Sanitario Nazionale verso il miglioramento continuo della qualità dell'assistenza, anche attraverso la realizzazione di progetti di interesse sovra regionale;
- i progetti-obiettivo, da realizzare anche mediante l'integrazione funzionale e operativa dei servizi sanitari e dei servizi socio-assistenziali degli enti locali;
- le finalità generali e i settori principali della ricerca biomedica e sanitaria, prevedendo altresì il relativo programma di ricerca;
- le esigenze relative alla formazione di base e gli indirizzi relativi alla formazione continua del personale, nonché al fabbisogno e alla valorizzazione delle risorse umane;
- le linee guida e i relativi percorsi diagnostico-terapeutici allo scopo di favorire, all'interno di ciascuna struttura sanitaria, lo sviluppo di modalità sistematiche di revisione e valutazione della pratica clinica e assistenziale e di assicurare l'applicazione dei livelli essenziali di assistenza;
- i criteri e gli indicatori per la verifica dei livelli di assistenza assicurati in rapporto a quelli previsti.

Il “piano sanitario regionale” rappresenta, invece, il piano strategico degli interventi per gli obiettivi di salute e il funzionamento dei servizi, per soddisfare le esigenze specifiche della popolazione regionale anche in

riferimento agli obiettivi del Piano sanitario nazionale. Le regioni, entro centocinquanta giorni dalla data di entrata in vigore del Piano sanitario nazionale, adottano o adeguano i piani sanitari regionali.

I tipici punti di debolezza delle strutture sanitarie sono:

- ✓ Processi decisionali non chiaramente definiti, lenti e farraginosi;
- ✓ Disfunzioni e lentezze burocratiche;
- ✓ Problematiche organizzative in relazione alla definizione di ruoli e responsabilità dei livelli gerarchici;
- ✓ Difficoltà di gestione del decentramento produttivo.

La necessità di riorganizzare i servizi e le strutture sanitarie aumentandone il grado di efficienza ed efficacia, è riscontrata nel Piano Sanitario Nazionale in vigore dal 2011 al 2013, nel quale si sottolinea l'importanza di garantire l'equità del sistema e monitorare i livelli di spesa, nel rispetto dell'erogazione dei servizi secondo i livelli di assistenza determinati.

Uno degli obiettivi definiti dal piano è il conseguimento di una “gestione più efficiente dei servizi sanitari, eliminando liste d'attesa, anche attraverso la predisposizione di percorsi facilitati per le cronicità” [Ministero della salute, 2011].

1.2. – Strutture ospedaliere e standard assistenza

I significati cambiamenti registrati in questi anni in tema di assistenza sanitaria, e in particolare in quella ospedaliera, richiedono un ammodernamento del Sistema Sanitario Nazionale partendo da alcune tematiche prioritarie come l'implementazione della *Clinical Governance* e la sicurezza delle cure, la ricerca e l'innovazione.

Tutte le strutture sanitarie, che concorrono a garantire gli obiettivi assistenziali, debbono operare secondo il principio dell'efficacia, della qualità e sicurezza delle cure, dell'efficienza, della centralità del paziente e dell'umanizzazione, nel rispetto della dignità della persona.

I sistemi sanitari più avanzati si sono attivati per dare risposte concrete ai nuovi bisogni di salute determinati dagli effetti delle tre transizioni, epidemiologica, demografica e sociale, che hanno modificato il quadro di riferimento negli ultimi decenni. Un tale cambiamento strutturale e organizzativo implica una redistribuzione delle risorse che può essere effettuata attraverso la valutazione dei volumi e strategicità delle prestazioni, delle performance e degli esiti clinici.

È possibile classificare le strutture ospedaliere secondo livelli di complessità crescente. Si distingue tra:

- Presidi ospedalieri di base;
- Presidi ospedalieri di primo livello;
- Presidi ospedalieri di secondo livello.

I presidi ospedalieri di base sono strutture dotate di Pronto Soccorso con la presenza di un numero limitato di specialità che sono medicina interna, chirurgia generale, ortopedia, anestesia e servizi di supporto, di pronta disponibilità 24 ore su 24, di radiologia, laboratorio di analisi chimico-cliniche e emoteca. Devono essere dotati inoltre di letti di osservazione di breve intensiva (OBI).

Per Pronto Soccorso si intende la struttura complessa dedicata all'attività diagnostica e terapeutica d'emergenza e d'urgenza, funzionante in ospedale 24 ore su 24 ore. Compito primario del Pronto Soccorso è quello di gestire le emergenze e le urgenze mediche, stabilizzando i pazienti con alterazioni delle funzioni vitali per poterli poi affidare all'area di degenza di competenza.

I presidi ospedalieri di primo livello sono strutture sede di DEA di primo livello dotate delle seguenti specialità: medicina interna, chirurgia generale, anestesia e rianimazione, ortopedia e traumatologia, ostetricia e ginecologia, pediatria, cardiologia con unità di terapia intensiva cardiologica, neurologia, psichiatria, oculistica, otorinolaringoiatra e urologia. Devono essere presenti inoltre, h24, servizi di radiologia con TAC ed ecografia, laboratorio di analisi chimico-cliniche e servizio immunotrasfusionale. Il presidio deve disporre, inoltre, di letti di OBI e di terapia intensiva in ragione di almeno 2 letti ogni 10.000 accessi di cui almeno 1/3 ad alta intensità, comunque modulabili in base alla dotazione complessiva di posti letto dell'ospedale.

Tale struttura svolge tutti gli interventi previsti per gli ospedali sede di Pronto Soccorso e svolge funzioni di accettazione di emergenza urgenza per patologie di maggiore complessità.

Per DEA si intende il Dipartimento di Emergenza Urgenza e Accettazione.

Il DEA rappresenta “l'aggregazione funzionale di più Strutture Complesse, che, pur mantenendo la propria autonomia e responsabilità clinico assistenziale, riconoscono la reciproca interdipendenza adottando un comune codice di comportamento assistenziale, al fine di assicurare una risposta rapida, completa e, ove occorre, sequenziale, in collegamento con le strutture operanti sul territorio” (cfr. Atto d'Intesa Stato-Regioni di linee Guida per l'emergenza G.U. 17 maggio 1996 in applicazione del DPR 27 marzo 1992).

I DEA afferiscono a due livelli di complessità, in base alle Unità operative che li compongono: DEA di I livello e DEA di II livello.

I presidi ospedalieri di secondo livello sono strutture dotate di DEA di secondo livello. Tali presidi sono istituzionalmente riferibili alle Aziende Ospedaliere, alle Aziende Ospedaliere universitarie e a presidi di grande dimensione della ASL. Tali presidi sono dotati di tutte le strutture previste per l'ospedale di primo livello nonché delle strutture che attengono alle discipline più complesse, tra cui cardiologia con emodinamica interventistica,

neurochirurgia, cardiocirurgia, chirurgia vascolare, chirurgia toracica, chirurgia maxillo-facciale, chirurgia plastica ed altre discipline di alta specialità.

Il dimensionamento delle strutture appena descritte dipenderà dal bacino di utenza della struttura stessa, come si evince in Tabella 1.

Tabella 1: Dimensionamento del sistema sanitario rispetto al bacino di utenza

Fonte: “Definizione degli standard qualitative, strutturali, tecnologici e quantitativi relativi all’assistenza ospedaliera”, Ministero della Salute

Disciplina	Bacino di utenza per dimensionare strutture (milioni di abitanti)	
	Bacino max	Bacino min
Cardiocirurgia infantile	6	4
Cardiocirurgia	1.2	0.6
Cardiologia	0.3	0.15
Chirurgia generale	0.2	0.1
Chirurgia maxillo-facciale	2	1
Chirurgia pediatrica	2.5	1.5
Chirurgia plastica	2	1
Chirurgia toracica	1.5	0.8
Chirurgia vascolare	0.8	0.4
Ematologia	1.2	0.6
Malattie endocrine	1.2	0.6
Geriatrics	0.8	0.4
Malattie infettive e tropicali	1.2	0.6
Medicina Generale	0.15	0.08
Nefrologia	1.2	0.6

Neurochirurgia	1.2	0.6
Neurologia	0.3	0.15
Neuropsichiatria infantile	4	2
Oculistica	0.3	0.15
Odontoiatria	0.8	0.4
Ortopedia	0.2	0.1
Ostetricia e Ginecologia	0.3	0.15
Otorinolaringoiatra	0.3	0.15
Pediatria	0.3	0.15
Psichiatria	0.3	0.15
Urologia	0.3	0.15
Grandi ustionati	6	4
Nefrologia	4	2
Terapia intensiva	0.3	0.15
Unità coronarica cardiologica	0.3	0.15
Medicina e chirurgia urgenza	0.3	0.15
Dermatologia	1.2	0.6
Emodialisi	0.6	0.3
Recupero e riabilitazione funzionale	0.07	0.04
Gastroenterologia	0.8	0.4
Lungodegenti	0.15	0.08
Neonatologia	1.2	0.6
Oncologia	0.6	0.3
Oncoematologia pediatrica	4	2
Pneumatologia	0.8	0.4
Reumatologia	1.2	0.6
Terapia intensiva neonatale	1.2	0.6

Neurochirurgia pediatrica	6	4
Nefrologia pediatrica	6	4
Urologia pediatrica	6	4

La variabilità dei bacini di utenza deve tener conto dei tempi di percorrenza dei cittadini, calcolata anche con la georeferenziazione, e quindi le regioni dovranno utilizzare i bacini minimi in presenza di territori a bassa densità abitativa e quelli massimi in caso opposto.

È possibile definire degli standard relativi all'assistenza ospedaliera.

Questi si dividono in:

- *Standard generali di qualità:* è necessario promuovere e attivare standard organizzativi secondo il modello di Clinical Governance, per il cambiamento complessivo del sistema sanitario, e fornire strumenti per lo sviluppo delle capacità organizzative necessarie a erogare un servizio di assistenza di qualità, sostenibile, responsabile (accountability), centrato sui bisogni della persona. Le strutture ospedaliere applicano le dimensioni della Clinical Governance secondo linee di indirizzo e profili organizzativi, fissati dalle Regioni, entro 6 mesi dall'emanazione del regolamento, che comprendano una serie di programmi tra cui, per i presidi di primo livello, la gestione del rischio clinico, l'Evidence Based Medicine e l'Health Technology Assessment, la valutazione e il miglioramento continuo delle attività cliniche, la documentazione sanitaria, comunicazione, informazione e partecipazione del cittadino/paziente, la formazione continua del personale. Quelli di secondo livello, oltre a quanto indicato per gli altri, devono svolgere un ruolo di promozione e sviluppo di metodi, strumenti e programmi da diffondere e rendere disponibili ad altre strutture di I/II livello in ambito regionale e nazionale.

- *Standard organizzativi, strutturali e tecnologici generali:* la prima regola è che nei presidi ospedalieri il rapporto percentuale tra il numero del personale del ruolo amministrativo e il numero totale del personale non può superare il valore del 7%. Non facilmente individuabile è, invece, il fabbisogno di Operatori Socio Sanitari (OSS) che comunque costituiscono una risorsa insostituibile nella presa in carico del paziente soprattutto in condizioni di emergenza. Ogni struttura poi ha l'obbligo del rispetto delle norme nazionali e regionali in materia di sicurezza in particolare: protezione antisismica; antincendio; sicurezza per i pazienti, degli operatori e soggetti ad essi equiparati; rispetto della privacy sia per gli aspetti amministrativi che sanitari; monitoraggio periodico dello stato di efficienza e sicurezza delle attrezzature biomedicali; graduale sostenibilità energetico-ambientale in termini di riduzione dei consumi energetici; smaltimento dei rifiuti; controlli periodici per gli ambienti che ospitano aree di emergenza, sale operatorie, rianimazione e terapie intensive e medicina nucleare; monitoraggio periodico dello stato di efficienza e sicurezza degli impianti tecnici e delle attrezzature biomedicali; controllo periodico della rispondenza delle opere edilizie alle normative vigenti.

1.3. – Il concetto di efficienza in sanità

Non risulta, ancor oggi, totalmente definito il rapporto di interrelazione tra il concetto economico di efficienza e la sanità.

È necessario segnare i confini entro i quali il concetto di efficienza debba inserirsi, quale obiettivo a capo degli interventi nelle strutture sanitarie per migliorarne le prestazioni.

Nell'ambito dell'efficienza microeconomica si parla di:

- 1) Efficienza di tipo allocativo, cioè la distribuzione delle risorse disponibili tra la prevenzione, l'assistenza primaria e la medicina curativa in modo da massimizzare i benefici per la collettività in termini di efficacia del sistema;
- 2) Efficienza di tipo tecnico, cioè la scelta dei processi produttivi che riescano a minimizzare i costi, evitando le duplicazioni e lo spreco di risorse;
- 3) Efficienza di tipo gestionale, cioè quell'insieme di azioni volte a limitare al massimo la burocrazia inutile e a creare una rete di controlli al fine di prevenire sprechi ed inefficienze.

Non ha senso però parlare di efficienza se non si tiene nella giusta considerazione anche l'efficacia. L'efficienza misura l'impiego economico di risorse nel processo produttivo ed è definita come $\frac{\text{output}}{\text{input}}$. Nel concreto, l'efficienza è rappresentata dal numero di prestazioni realizzate da un'unità di fattore produttivo impiegato, per esempio il numero di visite per ora di lavoro oppure il numero di ricoveri annuali per posti letto.

L'efficacia, invece, misura il contributo dei servizi sanitari al miglioramento dello stato di salute ed è definita come $\frac{\text{prestazioni}}{\text{salute}}$. L'efficacia è misurata, quindi tramite il miglioramento dello stato di salute in seguito al consumo di una prestazione sanitaria, come per esempio la riduzione del tasso di infezione in seguito all'assunzione di una terapia antibiotica.

Per valutare la bontà di una prestazione bisogna tener presenti entrambi questi parametri. Se a seguito di un intervento il paziente non migliora vorrà dire che quell'operazione non è stata efficace. Parimenti, se per un intervento che potrebbe eseguirsi in regime di day hospital il paziente necessita invece

di una settimana di ricovero, vorrà dire che la gestione del paziente è stata inefficiente.

Alla medicina spetta il compito di fissare e raggiungere i propri obiettivi (efficacia), all'economia e alla politica sanitaria quello di applicare i propri principi (efficienza).

La collaborazione tra le due scienze è necessaria al fine di massimizzare il benessere collettivo.

La creazione di valore si ottiene attraverso la delicata sintesi di efficacia ed efficienza e una serie di fattori tra i quali:

- La capacità dell'impresa di essere orientata al mercato e ai bisogni dei consumatori;
- La presenza di processi efficienti in grado di realizzare i prodotti/servizi richiesti dal mercato con il dispendio minimo di risorse;
- Il dinamismo, in quanto l'ambiente in cui l'impresa vive è in continua trasformazione.

Il miglioramento delle cure prescinde dal controllo sul budget, in quanto la condizione di scarsità finanziaria è diventata permanente. Le richieste di risorse per i servizi sanitari saranno sempre maggiori delle risorse disponibili.

Si rende necessario distinguere quali siano le voci di spesa, e quindi gli obiettivi da raggiungere, di priorità più alta da quelli a bassa priorità, in modo tale da allocare le risorse nel miglior modo possibile. Il controllo della spesa è una manovra a posteriori, quando la gestione ormai consolidata evidenzia insufficienza di risorse.

Una gestione efficiente comporterebbe manovre preventive orientate verso un concetto di *sostenibilità economica, sociale e accettabile per i cittadini*. Attraverso politiche di efficienza preventiva nell'uso delle risorse, è

facilmente comprensibile il vantaggio di introdurre una visione economico-gestionale in sanità.

A questo punto è lecito chiedersi come sia possibile determinare il giusto livello di spesa per la sanità, in quali ambiti sia doveroso non limitare le risorse e dove invece sia indispensabile agire secondo criteri di efficienza.

Per molti aspetti possiamo affermare che la sanità è paragonabile ad una grande azienda; in questo caso la risposta per individuare i livelli di spesa rispondenti ad una gestione efficace ed efficiente sarebbe abbastanza semplice, basterebbe adottare tutte quelle tecniche di gestione dei processi produttivi, analisi dei costi, budgeting, che oggi supportano le grandi imprese nella gestione del loro business. Purtroppo la risposta non è così ovvia, in quanto la sanità risponde sì per molti fattori al modello di un'azienda produttiva, ma presenta delle caratteristiche proprie del settore che la fanno risultare un sistema particolare e non facilmente governabile.

Un'azienda ospedaliera è un sistema nel quale intervengono molteplici *input* (materie prime, personale, impianti, risorse finanziarie) che innescano una serie di processi formati da attività e procedure, volti alla produzione di *output* (prestazioni sanitarie) che rispondono alla soddisfazione dei bisogni degli utenti. I bisogni in questo caso assumono i connotati di bisogni di salute (ripristino o tutela della stessa). Se il processo è gestito al meglio, le attese, ovvero i bisogni di salute dell'utenza, troveranno riscontro in esiti di salute positivi.

Questo sistema possiede un preciso *assetto organizzativo interno*, nel quale sono racchiuse le competenze specifiche regolate da una mappa di ruoli e responsabilità. Visto sotto questa prospettiva, l'ospedale sembra avere tutte le caratteristiche per essere considerata una vera e propria azienda di produzione.

Una differenza significativa tra imprese e aziende sanitarie emerge nella costruzione della catena economica dei correlati flussi: possiamo rilevare infatti che, non a caso, il beneficiario dei servizi sanitari si chiama *utente* e non cliente.

Il cliente è infatti colui che paga un prezzo per beneficiare del prodotto o servizio, a differenza dell'utente di un servizio sanitario che non paga (almeno direttamente) il servizio stesso. L'utente pagherà indirettamente la prestazione attraverso prelievo fiscale che consentirà il finanziamento del SSN.

In sostanza la differenza evidenziata sta a significare che lo scambio di prestazioni sanitarie non avviene, come per le imprese, in condizioni di mercato, ma secondo modalità diverse, infatti, mentre nel mercato i prezzi sono determinati dalla libera contrattazione delle parti, ciò non avviene nella sanità.

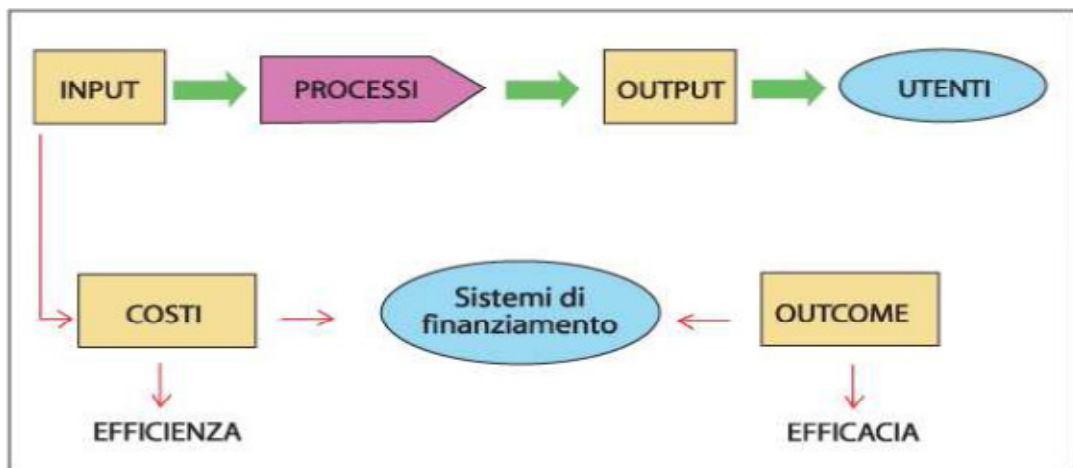


Figura 1: Il processo nelle aziende sanitarie

1.4. – Gestione per processi

Alla luce di quanto descritto precedentemente, è chiara l'esigenza di individuare delle modalità di gestione, delle aziende sanitarie, volte al raggiungimento di obiettivi di efficienza ed efficacia nonché al miglioramento delle cure e della qualità del servizio offerto.

A tal proposito, la letteratura gestionale propone un approccio di “gestione per processi”.

La gestione in ambito sanitario è diversa dalla gestione in ambito produttivo anche se in entrambi i casi si tratta di gestire una serie di attività in sequenza, ovvero dei processi.

Tale concetto è esposto da Fillingham nel libro *Lean Healthcare* in cui scrive “Caring for patients is not the same as manufacturing products” (Assistere pazienti non è lo stesso che produrre beni), ma aggiunge “this is true...but it is a process” (è vero... ma è un processo).

La sanità pubblica italiana non può più prescindere oggi dall'adozione di un approccio focalizzato sui processi, attraverso il quale agevolare il cambiamento organizzativo, traducibile in termini di maggior efficienza ed efficacia nella gestione delle attività.

L'adozione di una tale logica significa affrontare il tema più generale dell'organizzazione. Ragionando per processi infatti non si punta tanto ad ottimizzare le funzioni aziendali, quanto ad individuare quali sono i processi più importanti per l'organizzazione sanitaria e su quali di questi attivare una fase di miglioramento continuo.

L'intervento mirato su una singola funzione – struttura semplice o complessa – rappresenta un'azione limitata che può generare un impatto positivo solo su qualche aspetto marginale dell'organizzazione. Un *approccio globale* per processo permette invece d'identificare chiaramente le finalità dello stesso

processo ed il grado di utilità di un'attività, portando a benefici ben più generali ed importanti.

Si passa da una visione per funzione aziendale ad una visione per processi, con l'obiettivo di rendere l'azienda molto più flessibile rispetto alle mutevoli condizioni del contesto e vicina alle esigenze del mercato.

Si passa, quindi, dall'idea di azienda come sommatoria di funzioni, all'idea di azienda come insieme di processi di business trasversali alle divisioni.

L'uso della logica per processi come strumento di sviluppo organizzativo negli enti pubblici trae origine dall'applicazione di una metodologia manageriale denominata Business Process Reengineering – BPR. Tale metodologia ha trovato diffusione, soprattutto nel mondo anglosassone, all'inizio degli anni 90 nel settore delle imprese private. È di quegli anni una delle prime definizioni di process reengineering ed è stata elaborata da Hammer in un articolo comparso sulla Harward Business Rewiew. In tale definizione l'autore definisce questa tecnica come “il ripensamento sostanziale ed il ridisegno radicale dei processi aziendali al fine di ottenere risultati rilevanti, in termini di miglioramento della qualità e dei servizi all'utente e di incrementi nella produttività (riduzione dei costi a fronte di un aumento dei volumi di prestazioni erogate)”. Successivamente, intorno metà anni novanta, alcune applicazione di tale approccio hanno cominciato a riscontrarsi anche presso enti pubblici nel settore sanitario.

La gestione per processi implica, quindi, una riprogettazione dell'organizzazione dell'azienda, non più strettamente legata a ruoli statici, ma al contrario, dinamica e orientata a obiettivi comuni di massimizzazione dell'output e creazione di valore aggiunto per il cliente.

Le risorse dovranno essere coordinate, collaborative e interfunzionali, cercando di superare le problematiche poste dalla struttura gerarchica.

Qualità, efficienza ed efficacia divengono obiettivi primari non solo per le aziende private ma anche per il sistema pubblico, e soprattutto per il settore sanitario.

Il sistema organizzativo per processi appare più idoneo a favorire il raggiungimento di questi obiettivi, attraverso il trasferimento delle responsabilità verso i ruoli operativi, che effettivamente eseguono le attività e possiedono le competenze specifiche, e l'individuazione del responsabile di processo, il *process owner*, il quale dovrà pianificare le attività e coordinare le risorse garantendo il corretto svolgimento dell'intero processo.

Le singole attività sono parte di un processo integrato di risposta alle esigenze del paziente, il quale si trova a svolgere il ruolo di fulcro del processo di erogazione del servizio.

Nel caso specifico del settore sanitario, l'utente è direttamente coinvolto nel processo, per questo motivo sarà attento non solo al risultato finale ma anche alla modalità con cui è stato ottenuto, alla qualità del servizio, e all'aspetto relazionale derivato dalle interazioni con il personale da cui è stato assistito¹.

Gli obiettivi prioritari dell'organizzazione per processi in sanità sono tre:

1. il miglioramento dei processi aziendali, che porta ad una maggiore efficienza della struttura, attraverso l'ottimizzazione dell'uso e dell'allocazione di risorse;
2. la soddisfazione del paziente/cliente, raggiungibile tramite un alto livello di qualità del servizio offerto, sia in termini di qualità di prestazione medica sia in termini di miglioramento del rapporto medico-paziente;
3. la riduzione del rischio clinico.

¹ De Risi (2002), "Introduzione alla gestione per processi nelle organizzazioni", Università degli studi di Pisa.

In conclusione, quindi, la gestione per processi offre soluzioni flessibili, in cui emergono ruoli di coordinamento e collegamento manageriale, e team di progetto in grado di sviluppare la collaborazione lungo una dimensione orizzontale dell'organizzazione, ovvero tra quelle attività che formano processi finalizzati all'erogazione di specifiche prestazioni o servizi.²

La gestione per processi permette di ottenere una visione completa delle problematiche organizzative aziendali.

La scelta del modello organizzativo più idoneo per la gestione del sistema sanitario è al giorno d'oggi un tema molto dibattuto a causa delle molteplici e notevoli criticità che ci si ritrova ad affrontare.

In questo elaborato facciamo riferimento a due metodologie di produzione e di gestione, la Lean Production e l'Agile Manufacturing.

1.5. – Definizione e caratteristiche dei processi

Per processo si intende una sequenza di attività, logicamente correlate, svolte secondo una determinata sequenzialità e/o simultaneità, che ha un'origine (un punto di partenza) e che permette di raggiungere un determinato risultato finale (punto di arrivo) attraverso l'impiego di risorse (persone, macchine, materiale).

Tale sequenza è caratterizzata da:

- input misurabile;
- attività con valore aggiunto;
- output misurabile;

² E. Vignati, P. Bruno (2003) “*Organizzazione per processi in sanità*”, Franco Angeli.

- attività ripetitive.

Gli input provengono dai fornitori (interni e/o esterni) e gli output sono destinati ai clienti.

Biroli definisce i processi come “catene di fornitori/clienti ed, in questa logica, ogni fase del processo deve conoscere i bisogni sia del cliente finale che del cliente a valle”.

L’efficienza e l’efficacia del processo sono obiettivi primari misurabili attraverso sistemi di monitoraggio e controllo del processo.

I 4 elementi fondamentali che costituiscono un processo sono:

- ✓ Input: risorse tangibili o intangibili che innescano il processo;
- ✓ Vincoli: regole, condizioni, tempi che influenzano le attività;
- ✓ Risorse: persone e mezzi utilizzati nel processo;
- ✓ Output: sono i risultati del processo.

L’identificazione e la mappatura dei processi aziendali di un’azienda, permettono di identificare quale o quali sono i *processi core* e quali *di supporto*.

Il principale modello di riferimento per l’analisi dei processi è la Catena del Valore di M. Porter, che identifica la struttura di un’azienda come un insieme interrelato di processi, volti ad un obiettivo comune.

Il modello di Porter distingue due tipologie di processi, quelli che contribuiscono direttamente alla creazione dell’output (prodotti e servizi) di un’organizzazione (processi core), e quelli che non intervengono direttamente alla produzione di beni o erogazione di servizi ma sono altrettanto indispensabili affinché questo avvenga (processi di supporto).

Nel settore sanitario il processo di cura e assistenza del paziente è sicuramente il processo primario delle strutture ospedaliere, al quale sono affiancati diversi processi di supporto come il processo logistico, il processo

di approvvigionamento dei materiali o il processo di gestione delle risorse umane. Individuare quali sono e da quali elementi sono formati, aiuta l'azienda a controllare e misurare l'andamento dei processi.

Tra gli indicatori più significativi nella valutazione di un processo ritroviamo:

- L'efficienza, che si misura dal rapporto tra output e input utilizzati;
- L'efficacia, misurabile attraverso la valutazione degli obiettivi effettivamente raggiunti rispetto a quelli prefissati;
- La capability, identificata nella capacità di riprodurre nel lungo periodo, lo stesso prodotto o servizio senza alterarne il risultato;
- La flessibilità, che indica la capacità del processo di adattarsi ai cambiamenti interni all'organizzazione o esterni come il mutamento delle condizioni del mercato o la variazione dei requisiti della clientela.

Il fine ultimo del processo è la creazione di valore, sia per il cliente sia per l'organizzazione stessa.

È necessario focalizzare l'attenzione sulla difficoltà nel determinare l'effettiva creazione di valore, soprattutto per i processi di produzione di servizi che non sono tangibili quindi non misurabili. In questo caso infatti è molto più difficile valutare e controllare il valore prodotto dal processo e, soprattutto, il valore percepito dall'utente.

Possiamo affermare che il processo di erogazione di un servizio si differenzia per le seguenti caratteristiche:

- *La difficoltà di standardizzazione:* tipicamente il processo di erogazione di un servizio non è statico, ma varia fortemente a seconda delle situazioni;
- *La contestualità:* il servizio viene fruito dal cliente nello stesso luogo e nello stesso momento in cui è prodotto. Non c'è possibilità di

“sostituire” il servizio difettoso, poiché non è previsto un “magazzino”;

- *La partecipazione del cliente al processo:* molto spesso lo stesso cliente partecipa al processo di erogazione e ne influenza fortemente il risultato. Il cliente non percepisce solo la qualità del servizio finale erogato, ma anche la qualità di tutto il processo di erogazione (*qualità del processo e non solo del prodotto*).
- *La fondamentale importanza delle risorse umane:* la qualità di un servizio è in genere strettamente dipendente dalla professionalità di chi lo eroga.

Se ben gestiti, i processi possono avere un impatto strategico nell'azienda su quattro versanti:

- ❖ Costi;
- ❖ Ricavi;
- ❖ Investimenti;
- ❖ Competenze.

L'analisi del processo affina i dettagli operativi per assicurarsi che gli obiettivi siano raggiunti. Parte dell'analisi del processo consiste nel capire come il processo è attualmente strutturato e nel configurare i compiti e l'allocazione della capacità nelle attività che lo compongono.

CAPITOLO 2

Lean Service & Agile Production

2.1-Lean Manufacturing

La Lean Manufacturing, nota anche come Toyota Production System (TPS), è una metodologia utilizzata per la prima volta in Toyota da Sakichi Toyoda, Kiichiro Toyoda, ed in particolare dal giovane ingegnere Taiichi Ohno tra il 1948 e il 1975.

Come suggerisce il nome “Lean” il suo scopo principale è *snellire* le attività del processo di produzione eliminandone gli sprechi (*muda*). Lo strumento cardine è l’analisi del processo produttivo e l’obiettivo è massimizzare il valore minimizzando lo spreco, ciò che è inutile.

In sintesi, il sistema produttivo della casa automobilistica Toyota, contrariamente all’approccio della produzione di massa tipica della Ford, verteva sul desiderio di produrre secondo un flusso continuo, attraverso cicli produttivi che non fossero particolarmente prolungati, con il fine ultimo del conseguimento di una maggiore efficienza: in questo modo Toyota riusciva a produrre una vasta gamma di prodotti in piccoli volumi.

Gli obiettivi che la Lean Manufacturing si pone sono:

- massimizzare il valore per il cliente mediante l’implementazione di un sistema produttivo che pulsa con il mercato. Per valore si intende la capacità di soddisfare le esigenze del cliente in un certo momento e ad un dato prezzo, tenendo conto che il cliente combina la sua percezione di

qualità, del prodotto o servizio, con quella del prezzo pagato ottenendo così quello che per lui è il valore;

- minimizzare gli sprechi (muda) ottenendo così l'ottimizzazione della qualità, dei costi e del lead time del servizio/prodotto offerto. Per spreco si intende qualsiasi attività o azione svolta da un'azienda che assorbe risorse ma non apporta valore.
- massimizzare il coinvolgimento attivo dell'operatore nello svolgimento dell'attività.

In particolare Ohno identifica sette categorie di sprechi che, come egli stesso ha asserito, ammontano a circa il 95% dei costi complessivi caratterizzanti gli ambienti produttivi non Lean.

Le tipologie di muda sono:

- Muda di sovrapproduzione: si verifica quando la produzione non segue la domanda e quindi l'impresa decide di realizzare pezzi non richiesti dal mercato o in quantità superiori o in periodi in cui non c'è richiesta. La sovrapproduzione, oltre a non essere funzionale alla creazione di un "sistema che pulsa con il mercato", genera anche un aggravio dei costi, collegati allo stoccaggio dei prodotti in eccesso in magazzino.
- Muda per trasporto: i materiali dovrebbero essere consegnati direttamente laddove devono essere utilizzati, piuttosto che essere spediti dal fornitore al luogo di ricezione per poi essere processati ed immagazzinati e solo successivamente trasportati sulla linea di lavorazione/assemblaggio dove sono richiesti. Questa tipologia di sprechi è dovuta principalmente a: un errato layout delle fabbriche; spazi occupati dalle linee di produzione eccessivi rispetto alle reali necessità; organizzazione del lavoro che non prevede precise sequenze

di prelievo; attrezzature non specificatamente dedicate all'ottimizzazione dei trasporti interni.

- Muda per scorte: con il termine “scorta” è indicato tutto ciò che giace fermo in attesa di un evento, come una lavorazione successiva (per le materie prime e i semi-lavorati) o la vendita (per i prodotti finiti). Durante questa attesa non viene aggiunto al pezzo alcun valore, ma al contrario, i costi aumentano.
- Muda per attese: si manifesta quando un operatore non svolge alcun lavoro, rimanendo in attesa di un evento successivo, ad esempio per mancata consegna del materiale, rottura di un macchinario o manutenzione.
- Muda di processo: in questo caso gli sprechi si nascondono all'interno delle diverse fasi del processo produttivo e sono quindi legati ad una progettazione imprecisa delle fasi stesse e ad una tecnologia degli impianti inadeguata. Per eliminare tali muda vanno eliminate le attività che non aggiungono valore: fasi che, pur modificando il prodotto, non sono riconosciute come valore aggiunto dal cliente finale; fasi aggiuntive per porre rimedio a fasi precedenti non ottimizzate; fasi che non sono necessarie per l'ottenimento delle prestazioni tecniche richieste al prodotto.
- Muda per prodotti difettosi: la presenza di difetti nei prodotti rallenta la produzione, aumenta il lead time e i costi (per riparazione e movimentazione dei pezzi, ma anche per gestione dei reclami). Non va inoltre sottovalutata anche la perdita di immagine connessa con la consegna al cliente di un output difettoso.
- Muda per movimentazione: è utile distinguere tra lavoro e movimento. Il primo è un movimento che produce valore e si distingue perciò dal movimento improduttivo. Quest'ultimo, a sua volta, può essere suddiviso in: spostamenti (che si rendono necessari a causa di layout mal disegnati o strutture inutilmente sovradimensionate) e azioni

improduttive (dovute al mancato studio ergonomico dei posti di lavoro).

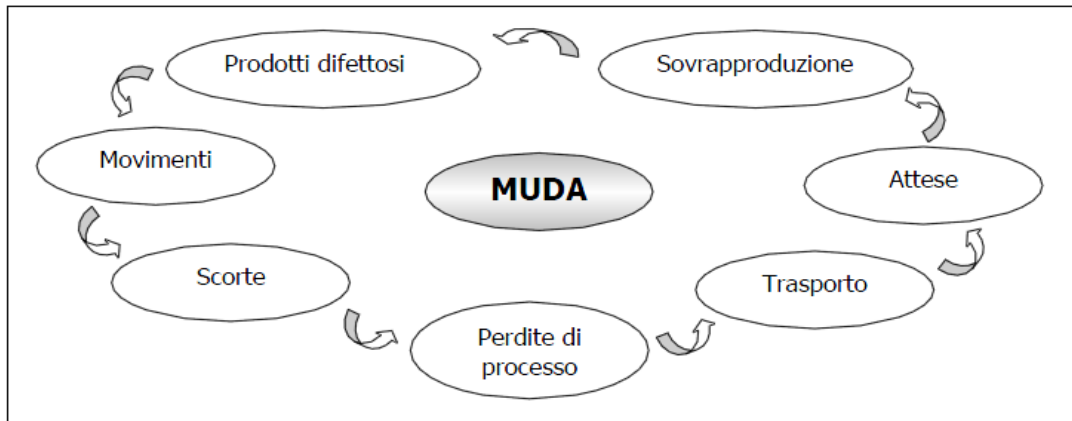


Figura 2: Le sette categorie di sprechi

Fonte: LEAN MANUFACTURING IN TOYOTA *Un'evoluzione continua* Convegno Unindustria Treviso, 15.11.2005

Ultimamente si è parlato anche di un'ottava categoria di spreco: essa considera l'intelletto, ovvero lo spreco di idee vincenti e risolutive. Spesso le persone non vengono consultate per nuove idee o per migliorare i metodi di lavoro e non c'è investimento nell'addestramento e nella creazione di una cultura di impresa. Questo implica che non vengano utilizzate, o addirittura espresse, idee innovative o migliorative e non vengano ottimizzate le capacità degli operatori.

La Lean Manufacturing consente continui miglioramenti al sistema produttivo, senza aggiungere risorse.

2.1.1 – Il Lean Thinking

L'approccio ai problemi, e lo snellimento aziendale ed industriale, hanno luogo con successo solo se guidati da un pensiero e da una mentalità in grado di fornire un approccio razionale e corretto, identificando e comprendendo ciò che è *valore* e ciò che non lo è.

Questa mentalità è il motore della rivoluzione ed è detta Pensiero Snello: *Lean Thinking*.

Si tratta di un approccio razionale che influenza tutte le aree di azione perché guida il “pensiero”.

Il pensiero snello Lean Thinking è solo una parte del risultato, perché rappresenta il “motore iniziale” della rivoluzione lean, la quale, se da un lato inizia attraverso il pensiero snello, d'altro canto necessita di strumenti e di riferimenti operativi estremamente concreti, per evitare che un approccio teorico rimanga validissimo ma privo di riscontri concreti tangibili. Il Lean Thinking è il pensiero snello che alimenta l'azione e guida il modo di vedere, capire ed interpretare le esigenze dell'azienda, del cliente per poter operare in maniera diretta, efficiente, razionale, immediata, consumando il minor numero di energie e garantendo la massima resa.

Tale pensiero, con la sua enfasi sulla standardizzazione, cerca di eliminare le scorte e migliorare i processi riducendo, inoltre, il tempo tra il momento in cui il cliente richiede un servizio e il momento in cui lo riceve.

2.1.2 – Strumenti della Lean Manufacturing

L'adozione della logica di Lean Manufacturing coinvolge l'intera struttura produttiva dell'impresa e il suo rapporto con il mercato, dati i numerosi elementi e principi che essa racchiude. La Lean Manufacturing può essere rappresentata come una “casa” in cui gli aspetti basilari (il tetto) sono il focus sulla time line ed il sistema produttivo pulsante, i due pilastri invece

rappresentano i cardini, punti chiave del sistema produttivo ed infine le fondamenta rappresentano gli strumenti utilizzati.

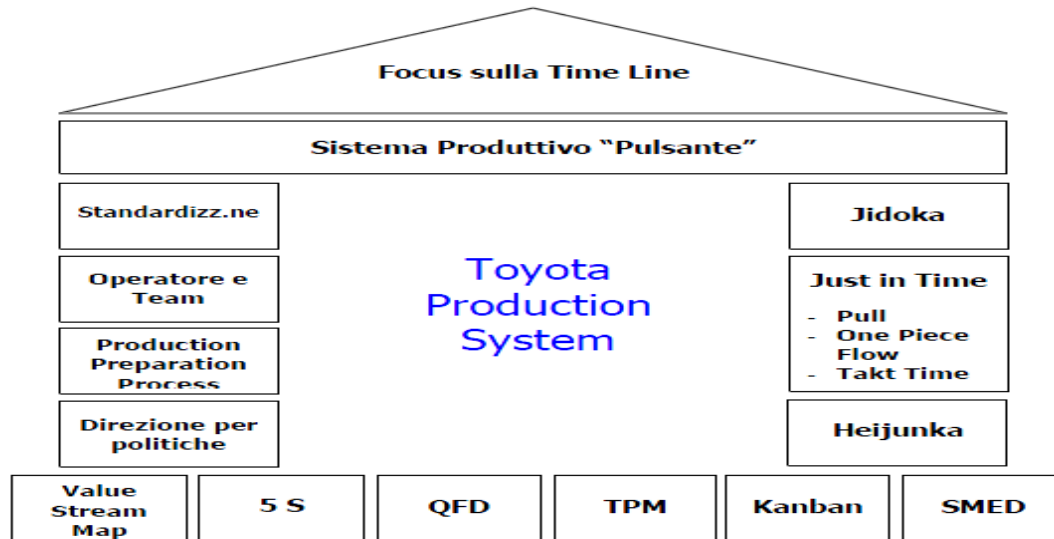


Figura 3: La TPS come una casa

Fonte: LEAN MANUFACTURING IN TOYOTA *Un'evoluzione continua* Convegno Unindustria Treviso, 15.11.2005

Filo conduttore dell'intero sistema sono i **2 aspetti chiave** che hanno rivoluzionato il sistema:

- ✓ Bisogna essere in presenza di un sistema produttivo pulsante: data l'alta variabilità del mercato ed il continuo mutamento delle esigenze del consumatore, risulta necessario all'azienda eseguire dei cambiamenti in modo tale da poter soddisfare il mercato "pulsando" con esso. L'azienda, in particolare, per essere costantemente in linea con le richieste della clientela, deve organizzare un sistema produttivo altamente flessibile e affidabile, in grado di "pulsare" in sintonia con le esigenze del mercato.
- ✓ Il secondo aspetto è il focus sulla Time Line: essa rappresenta la linea del tempo che va dal momento in cui il cliente inoltra l'ordine

all'impresa, a quello in cui quest'ultima riceve la somma di denaro corrispondente all'ordine. Focalizzando l'attenzione su tale linea, si cerca continuamente di ridurre la lunghezza, attivando i flussi e rimuovendo i muda. Per questo motivo essa diventa la molla principale per la caccia agli sprechi.

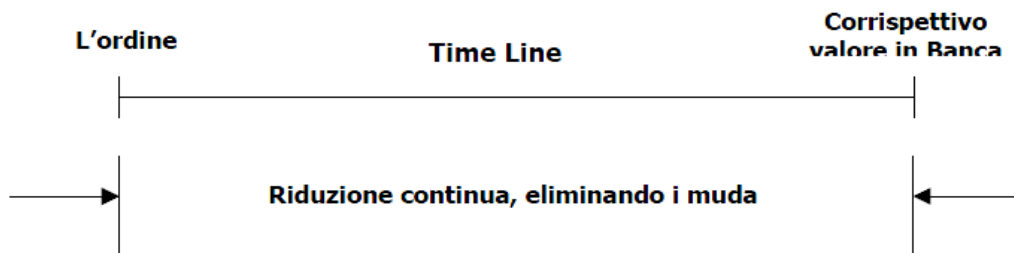


Figura 4: Time Line

Fonte: LEAN MANUFACTURING IN TOYOTA *Un'evoluzione continua* Convegno Unindustria Treviso, 15.11.2005

I pilastri del modello TPS possono essere così sintetizzati:

STANDARDIZZAZIONE: lavorare sulla base di operazioni standardizzate è un prerequisito fondamentale per realizzare un miglioramento continuo. Il lavoro standard rappresenta uno strumento per mantenere nel tempo la produttività, la qualità, e la sicurezza ad alti livelli evitando di incorrere in perdite di efficienza. Tali standard riguardano l'esatto svolgimento del lavoro, in termini di corretta esecuzione delle singole attività e definizione dei tempi e metodi e permettono inoltre di poter individuare delle misure correttive in caso di scostamento dai valori nominali.

OPERATORE E TEAM: gli operatori sono visti come membri di una squadra coordinata e controllata da un team leader, il quale lavora come gli altri operatori, ma ha in più la responsabilità sul comportamento dell'intero team. Il team opera in un contesto di flessibilità, fornendo utili suggerimenti e ricevendo periodici addestramenti. L'obiettivo fondamentale del team è

fornire al cliente un prodotto di qualità elevata con costi ridotti, attraverso un sistema produttivo il più possibile esente da sprechi.

PRODUCTION PREPARATION PROCESS (3P): tale metodologia comprende sia l'ideazione che l'implementazione di nuovi prodotti e consente di soddisfare le esigenze del cliente finale, nel rispetto dei volumi effettivamente richiesti dal mercato e soprattutto nel rispetto dei tempi di offerta.

DIREZIONE PER POLITICHE: sono necessarie delle politiche di coordinamento tra le diverse aree aziendali che si basano sul concetto di Qualità Totale secondo cui il miglioramento continuo e la ricerca dell'efficienza deve essere perseguita ad ogni livello aziendale, a partire da quello manageriale fino a quello operativo. Si cerca il coinvolgimento di tutti i livelli aziendali che devono essere tra loro allineati e coordinati mediante un'efficace comunicazione e circolazione di informazioni.

JIDOKA: metodo basato sul controllo autonomo eseguito dall'operatore che viene posto pertanto al centro del processo, quale garante del risultato finale. L'operatore, infatti, è autorizzato a fermare la linea ed evitare il proliferare di anomalie nel caso in cui noti un difetto e viene coinvolto nella risoluzione dei problemi attraverso l'elaborazione e la costruzione di sistemi Poka Yoke (a prova di stupido), semplici accorgimenti che consentono all'operatore di verificare la correttezza dell'operazione che si sta per compiere o che si è appena terminata. L'operatore viene incentivato ed abituato ad applicare il metodo dei 5 perché, che consente di scoprire qual è la causa alla radice di un determinato inconveniente, in modo da eliminarla totalmente.

JUST IN TIME: è l'insieme degli accorgimenti e di tecniche che consentono al sistema produttivo di pulsare come pulsa il mercato e, nello stesso tempo, di ottenere il minimo livello di sprechi e una time line sempre più breve.

Gli obiettivi del JIT si possono tradurre in 5 principi cardine che rappresentano il concetto portante di produrre:

- Solo il necessario;
- Quando necessario;
- Senza attese o accumuli;
- Con qualità perfetta;
- Senza sprechi.

Tali obiettivi operativamente corrispondono alla realizzazione del *pezzo Giusto*, nel *posto Giusto*, al *momento Giusto*, nella *quantità Giusta* e nella *qualità Giusta*, da cui deriva il termine (italianizzato) del “5G”. In perfetta linea con l'imperativo di eliminare ogni tipo di spreco, della Lean Manufacturing, il JIT parte dalla caccia allo spreco madre, la sovrapproduzione, prefiggendosi la riduzione massima dei lead time attraverso la riduzione dell'entità dei lotti produttivi.

Requisiti vincolanti per la sua applicazione possono essere ad esempio:

- ✓ Affidabilità dei fornitori rispetto ai volumi, alla qualità e ai tempi di consegna;
- ✓ Modularità dei prodotti e standardizzazione dei componenti a garanzia di consumi regolari e set up ridotti;
- ✓ Layout di processo definito su principi di cell design e group technology;
- ✓ Sistemi di produzione semplici e “in tiro” attraverso strumenti quali il kanban;
- ✓ Standardizzazione delle unità di movimentazione;
- ✓ Affidabilità, disponibilità ed efficienza degli impianti;
- ✓ Formazione e coinvolgimento del personale a tutti i livelli.

Il Just in Time si compone di 3 sotto elementi:

- ✓ SISTEMA PULL: attraverso tale sistema l'avanzamento del flusso produttivo è guidato dai clienti. Questo significa che nessuno, a monte, dovrebbe produrre beni o servizi fino al momento in cui il cliente a valle (interno o esterno) non li richiede. Con questa logica, la produzione non potrà essere realizzata seguendo dei piani, ma monitorando i consumi dei clienti. E' fondamentale che esista una fidelizzazione del fornitore (che deve essere preciso e affidabile, deve lavorare anch'esso con logica JIT e deve assicurare alta qualità e brevi tempi di consegna).

- ✓ SISTEMA ONE PIECE FLOW: la produzione è organizzata prevedendo l'avanzamento del materiale un pezzo alla volta, con un flusso continuo. In questo modo, i singoli pezzi passano da una fase produttiva all'altra senza accumuli tra le macchine, contribuendo alla riduzione della Time Line (il materiale attraversa i reparti nel modo più rapido), all'ottenimento della massima flessibilità, all'abbattimento in misura importante delle scorte intermedie (Work in Process – WIP), e al recupero di spazio fisico all'interno della linea, grazie all'impiego di macchinari più piccoli, che vengono avvicinati tra loro per la presenza di piccoli lotti.

La Lean mira a rendere il flusso il più possibile continuo e regolare grazie all'implementazione di un approccio di tipo PULL.



Figura 5: Da una produzione a lotti al one-piece flow

- ✓ TAKT TIME (T_{takt}): indica il tempo in cui deve essere ottenuta un'unità di prodotto. E' pertanto uno strumento che serve a legare la produzione ai clienti finali, uniformando il ritmo della stessa a quello delle vendite. Il T_{takt} rappresenta, dunque, il ritmo della produzione ovvero il tempo medio di produzione necessario per soddisfare la domanda del cliente. Il Takt time non deve essere confuso con il tempo ciclo. Quest' ultimo è il tempo che trascorre tra l'inizio e la fine di una lavorazione, per fare un prodotto o fornire un servizio. Indica il contenuto di lavoro che è necessario fare per eseguire una lavorazione. Il tempo ciclo si può anche definire come indicatore dell'efficienza odierna dell'azienda. Il takt time, invece, è scandito dal cliente (è il cosiddetto battito cardiaco dell'azienda) essendo definito come tempo massimo entro il quale bisogna produrre un prodotto o effettuare un servizio per poter soddisfare la domanda del cliente.

Il calcolo del Takt Time si effettua attraverso i seguenti passi:

1. definizione dell'orizzonte temporale per il quale si vuole calcolare il Takt Time;

2. determinazione del volume di vendita previsto nel periodo di riferimento;
3. individuazione del tempo lavorativo a disposizione.

E' possibile, dunque, calcolare il Takt Time come:

$$T_{\text{takt}} = \frac{\text{Tempo lavorativo a disposizione}}{\text{Volume di vendita previsto}}$$

Produrre al Takt Time significa:

- Realizzare il prodotto e rispondere ai problemi entro il Takt Time: il tempo ciclo e la reazione agli imprevisti (guasti, scarti, ecc) deve avvenire entro il Takt Time;
- Eliminare e controllare le cause di fermo o attesa o difetto non pianificate: va perseguita la piena regolarità operativa;
- Minimizzare i tempi di set – up.

Altra grandezza di riferimento, oltre al Takt Time, in ambiente Lean Production è l'indice di flusso così definito:

$$I = \frac{\text{Lead Time}}{\text{Tempo a valore aggiunto}}$$

L'indice di flusso rapporta il Lead Time al tempo a valore aggiunto per valutare di quanto è maggiore il tempo di attraversamento rispetto all'effettivo tempo a valore.

Il Lead Time, anche detto tempo di attraversamento, rappresenta il tempo totale che intercorre tra il momento in cui si verifica la necessità di un prodotto, determinato, ad esempio da un ordine del cliente, e quello in cui il prodotto si rende disponibile. Più basso è il Lead Time, maggiori sono la reattività e l'efficienza del sistema.

E' possibile distinguere tra il: "Lead Time di produzione", che rappresenta il tempo necessario per fabbricare un certo prodotto nel reparto produzione, dal momento dell'ingresso delle materie prime all'uscita del prodotto finito, e il "Lead Time di approvvigionamento" (o *procurement time*) che è il tempo che intercorre tra l'inoltro dell'ordine di acquisto e l'arrivo delle merci.

HEIJUNKA (livellamento della produzione): consiste nel produrre nel breve periodo tutte le varietà di prodotti richieste dal cliente nel medio periodo e nelle stesse proporzioni quantitative (mantenendo quindi i volumi costanti), con l'obiettivo di ridurre al minimo le scorte e di migliorare la capacità di risposta alle variazioni della domanda. Questo comporta una produzione contemporanea sulla linea di diversi modelli, con lotti anche unitari, e non più una produzione per lotti.

Per quanto riguarda **le fondamenta** e quindi le tecniche con cui si può perseguire l'ottenimento di un ambiente di tipo Lean se ne descrivono in seguito le caratteristiche principali.

VSM (Value Stream Mapping): è una tecnica che, tramite una mappatura del flusso di valore, analizza i flussi fisici, informativi e di persone all'interno del processo logico-produttivo con lo scopo di individuare e migliorare le attività a valore aggiunto.

Lo scopo viene perseguito seguendo il percorso di fabbricazione di un prodotto dal cliente al fornitore, offrendo così una rappresentazione visiva del flusso fisico, informativo e della time line (flusso temporale), descritta su due livelli per distinguere il tempo a valore aggiunto dal tempo impiegato per attività non a valore.

La realizzazione di questa mappatura avviene in due fasi: una prima di "current state map", che descrive lo stato attuale del sistema e definisce ogni fase che coinvolge materiali o informazioni, ed una seconda di "future state

map” che descrive gli interventi di miglioramento che si vogliono effettuare laddove si è riscontrata un’inefficienza e quindi uno spreco.

La definizione della future state map è basilare in un’ottica di miglioramento continuo poiché, individuando le aree in cui intervenire per un miglioramento delle attività a valore aggiunto e indicando il modo in cui si vuole far fluire il value stream (insieme delle azioni volte a creare valore per il cliente), risulta lo strumento più efficace per la lotta agli sprechi e per una conversione Lean dell’azienda.

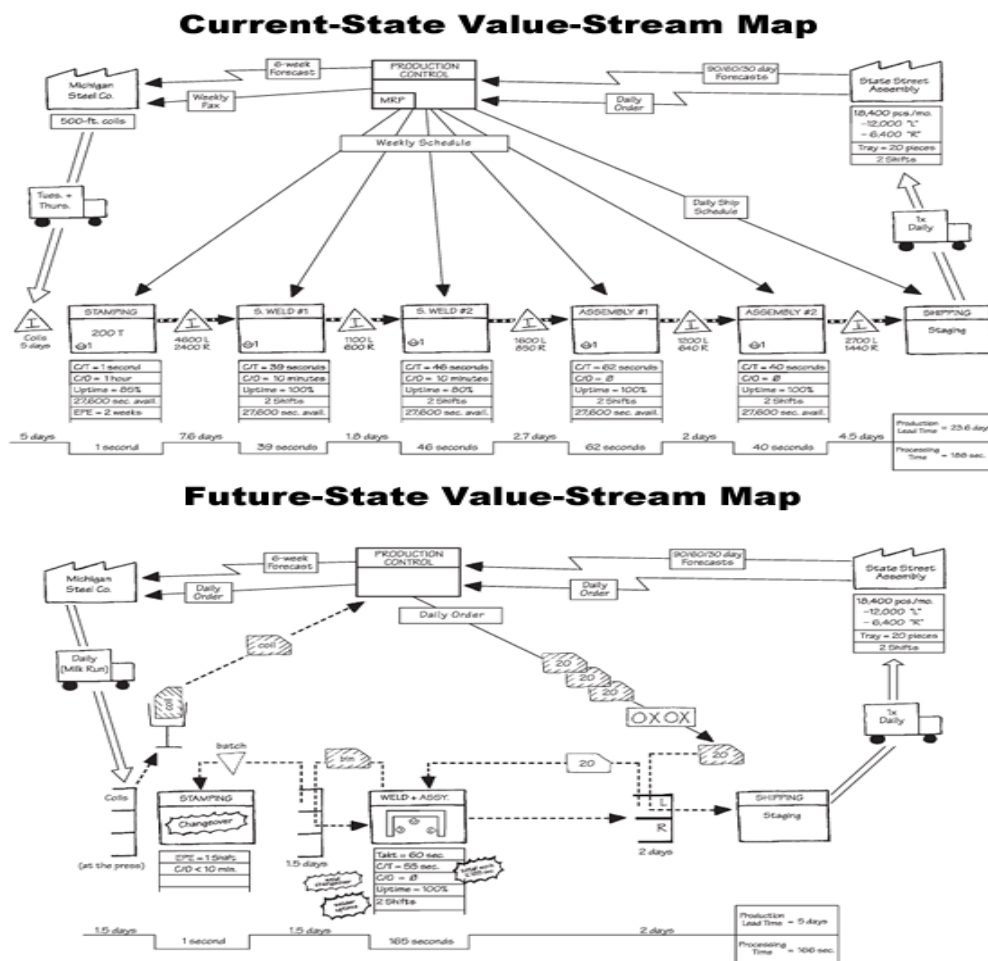


Figura 6: Current state map e 52esi sta future state map

Fonte: Lean Lexicon: A Graphical Glossary for Lean Thinkers di John Shook e Chet Marchwinski (2014)

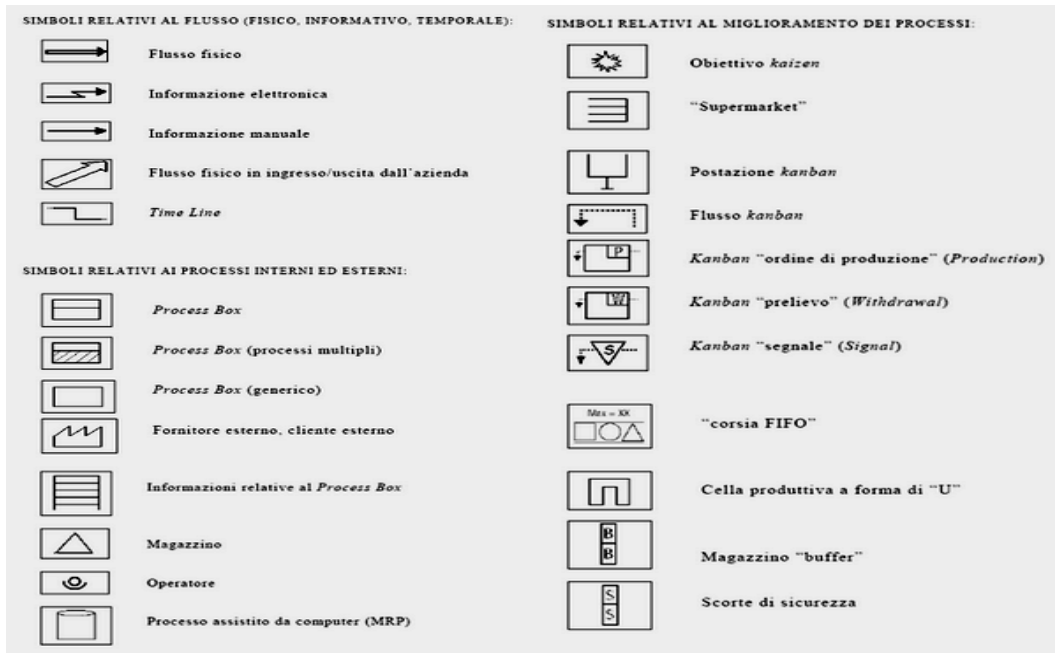


Figura 7.a: Simboli per il VSM

Fonte: <http://www.leanmanufacturing.it/valuestreammapping.htm>

5S: è un approccio finalizzato a mettere in ordine il posto di lavoro (reparti, linee, uffici).

Il nome deriva dalle iniziali di 5 parole giapponesi, che indicano le 5 fasi di implementazione di un sistema 5S:

1. Seiri (sgomberare): separare le cose utili da quelle inutili così da poter eliminare quelle inutili. Per eliminare tutto ciò che non serve è importante operare una razionale classificazione degli oggetti (utensili, attrezzi, materiali) presenti nell'area di lavoro e nello stesso tempo agire sulla fonte delle cause che generano le inutilità.
2. Seiton (ordinare): mettere in ordine le cose utili in modo che tutti possano utilizzarle facilmente e capire rapidamente qual è il loro posto.
3. Seiso (pulire): mantenere il posto di lavoro pulito così da ridurre guasti e irregolarità.

4. Seiketsu (standardizzare): standardizzare le attività del posto di lavoro e comunicare le modalità operative corrette, nel modo più semplice ed efficace.
5. Shitsuke (sostenere): creare un posto di lavoro tale da poter rispettare gli standard definiti. Il sostenimento della disciplina consiste nel mantenimento dei risultati raggiunti. Attraverso l'analisi continua dei problemi e dell'identificazione delle contromisure, unito al continuo monitoraggio delle prestazioni, è possibile fissare sempre nuovi obiettivi secondo il principio del miglioramento continuo.



Figura 7.b: Le 5S

Fonte: <http://blog.crit-research.it/?tag=5s>

L'obiettivo, quindi, è la definizione e la standardizzazione delle condizioni ottimali dei posti di lavoro, così da rendere ovvie tutte le anomalie rispetto agli standard definiti.

TPM (Total Productive Maintenance): è un approccio alla manutenzione ideato per minimizzare le fermate indesiderate degli impianti e massimizzare il loro impiego. E' una filosofia di miglioramento continuo e di lavoro in team poiché richiede il coinvolgimento attivo e la responsabilizzazione di

tutti gli operatori, per garantire il corretto funzionamento dei macchinari e ottenere così la massima efficienza degli impianti. Si preferisce la manutenzione preventiva rispetto a quella correttiva di modo da ridurre i costi.

QFD (Quality function deployment): è una metodologia per lo sviluppo di nuovi prodotti in grado di assicurare la qualità a partire già dalla fase di progettazione. Attraverso la sua applicazione, l'azienda è quindi in grado di tradurre gli specifici bisogni dei clienti in specifiche tecniche per la produzione del bene o del servizio.

KANBAN: è un sistema di programmazione, controllo e regolazione della produzione. La parola Kanban significa “cartellino” che contiene informazioni (articolo, quantità da produrre ecc.) e circola periodicamente tra fornitore e cliente fornendo ad entrambi le notizie necessarie alla gestione quotidiana dei materiali in produzione.

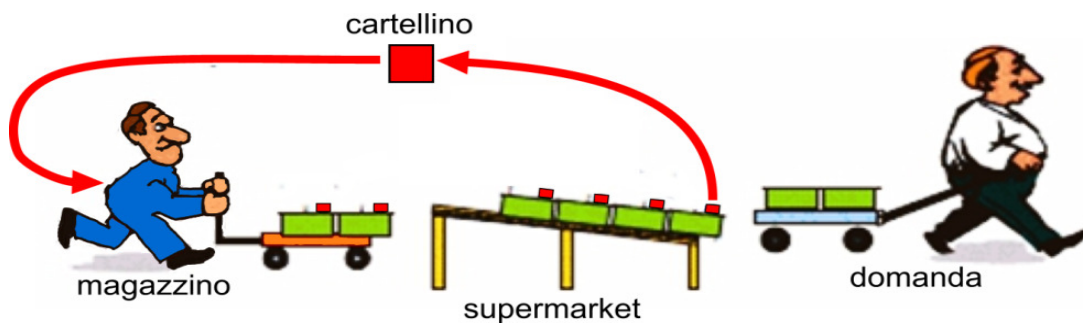


Figura 8: Il sistema Kanban

Fonte: <http://www.umbertosantucci.it/?p=1189>

Esso consente di avere il materiale al posto giusto nel momento giusto, in modo automatico e rappresenta un sistema pull che si attiva solo a fronte della richiesta del cliente finale. L'avanzamento della produzione è inoltre verificabile con una rapida osservazione dei Kanban e dei materiali stessi e,

le piccole fluttuazioni della domanda, vengono gestite autonomamente e automaticamente senza necessità di interventi da parte di una pianificazione centrale.

Esistono due tipi di cartellini: di produzione e di movimentazione. Nella stazione cliente vi sono dei contenitori con degli item; quando l'operatore utilizza tali oggetti spedisce poi il contenitore vuoto alla stazione fornitrice insieme al kanban di movimentazione; nella stazione fornitrice vi sono dei contenitori in "attesa" con gli stessi item e forniti di un kanban di produzione; all'arrivo del contenitore vuoto l'operatore scambia i cartellini e invia il contenitore "pieno" alla seconda stazione. I kanban di produzione vengono invece posti su di una lavagna e la loro sequenza costituirà il piano di produzione reale, continuamente aggiornato, per la stazione di lavoro.

I principi fondamentali su cui si fonda la tecnica Kanban sono rappresentati al meglio dalla *Regola dei 6 zero*. Tale regola viene percepita come una tendenza e punta a:

- ❖ Zero difetti: non è permesso a nessun pezzo difettoso di avanzare lungo il flusso produttivo, rischiando così di interrompere la continuità che è alla base del successo del Kanban;
- ❖ Zero guasti: operare in modo che il sistema risulti altamente affidabile, segnalando guasti e facendo un'autodiagnosi continua;
- ❖ Zero scorte: si istaura una lotta alle scorte per minimizzare gli sprechi;
- ❖ Zero tempi di set up: un'intera famiglia di prodotti deve poter passare tutta la linea senza provocare significativi tempi di set up;
- ❖ Zero carta: si libera il processo dalla burocrazia. Non è necessario gestire un pesante flusso informativo dato che esso nasce e muore sulla linea.
- ❖ Zero addetti: una volta che sono state ridotte le attività all'essenziale, sarà necessario un minor numero di addetti.

SMED (Single Minute Exchange of Die): è un metodo sviluppato per la riduzione dei tempi di setup (intervallo di tempo tra fine produzione di un item e inizio di uno diverso sulla stessa macchina). L'idea alla base della metodologia SMED è di trasformare le operazioni di setup in qualcosa di elementare così da renderne impossibile l'esecuzione scorretta. L'espressione SMED significa "cambio stampo in un solo digit" ossia consente l'esecuzione delle operazioni di set up in un lasso di tempo inferiore a dieci minuti. Ridurre i tempi significa eliminare una componente che non dà valore aggiunto al prodotto finito: i tempi morti di attrezzaggio macchina.

Con l'utilizzo dello SMED, lo switch tra un lotto e un altro non rappresenta un problema, con la necessaria conseguenza di poter produrre Just in Time con l'obiettivo finale di consentire al sistema produttivo di produrre solo quando effettivamente richiesto dal mercato. Si tratta quindi di una produzione tirata dal cliente.

La metodologia è composta da 4 fasi: analisi della situazione iniziale; separazione tra setup interni ed esterni; conversione dei setup interni in setup esterni; miglioramento delle procedure di setup.

Il set-up interno è composto da tutti gli elementi del processo di attrezzaggio che possono essere svolti solo quando la macchina è ferma.

Il set-up esterno, invece, è composto da tutti gli elementi del processo di attrezzaggio che possono essere svolti solo quando la macchina è in azione.

2.1.3 – Miglioramento Continuo

La Lean Manufacturing consente continui miglioramenti al sistema produttivo, senza aggiungere risorse.

Il miglioramento continuo (Kaizen in giapponese) consta in una strategia di azione in cui “non passa giorno senza che si produca qualche miglioramento all’interno dell’azienda”³.

Lavorare in un’ottica di miglioramento continuo richiede necessariamente il coinvolgimento attivo, non solo del manager e dei dirigenti, ma di tutto il personale coinvolto nel processo produttivo, l’unico a conoscere a fondo il proprio lavoro.

Il concetto di miglioramento continuo trova le sue radici più profonde nelle idee di Deming, il quale teorizzò che qualsiasi processo può essere visto come un ciclo, noto come il ciclo di Deming o PDCA, costituito, come mostrato in Figura 9, dalle quattro fasi di Plan → Do → Check → Act.

Il Ciclo PDCA serve per promuovere una cultura della qualità che è tesa al miglioramento continuo dei processi e all’utilizzo ottimale delle risorse. L’essenza della qualità totale consiste nelle applicazioni ripetute della ruota di Deming fino al raggiungimento dell’obiettivo.

Plan: si determinano gli obiettivi e i metodi necessari per raggiungerli in accordo con i desideri del cliente e le politiche dell’organizzazione.

Do: viene svolto il lavoro in accordo con gli standard prefissati.

Check: si monitorano i processi e i prodotti per rilevare eventuali non conformità. E’ una fase di analisi e controllo.

Act: si adottano azioni per migliorare le prestazioni dei processi e per eliminare le eventuali non conformità rilevate nella fase di controllo.

³ Imai Masaaki, *Kaizen (1997) – “Lo Spirito Giapponese del Miglioramento”*, Il Sole 24 ore Libri.

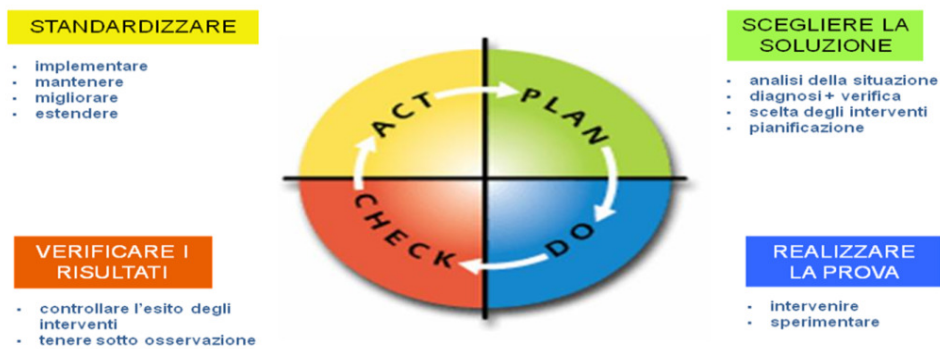


Figura 9: Il ciclo di Deming

Fonte: *leanacademy.it*

Gli strumenti impiegati per l'implementazione del miglioramento continuo in ottica Lean Manufacturing sono essenzialmente i seguenti:

- Value Stream Mapping
- Kaizen Event

KAIZEN EVENT

Il Kaizen Event è lo strumento per svolgere in maniera programmata le attività inerenti il Miglioramento Continuo.

L'intervento per la realizzazione di un Kaizen Event, la cui durata è tipicamente di cinque giorni, si divide in tre fasi:

1. Fase Preparatoria (1° giorno): in tale step si procede alla scelta dell'area che dovrà essere oggetto dell'intervento, si definiscono quantitativamente gli obiettivi del miglioramento (mediante il ricorso

ad opportuni indicatori) e si compone adeguatamente il gruppo di lavoro.

2. Fase Esecutiva (2°-4° giorno): tale fase, in cui si concentra il lavoro di analisi, problem solving e progettazione della soluzione, si apre con un briefing iniziale durante il quale si procede ad organizzare il lavoro nelle diverse giornate, ad individuare le idee Kaizen ed a redigere il Giornale Kaizen.
3. Follow-up dei risultati ottenuti (5° giorno): si tratta di una fase di verifica strutturata dei risultati conseguiti, svolta secondo quanto indicato nel Giornale Kaizen. Qualora i target stabiliti non siano stati raggiunti, si procede all'individuazione dei passi successivi da implementare.

Tra i vantaggi relativi all'adozione di una strategia Kaizen si ricordano, innanzitutto, la sua rapidità di implementazione ed il costo contenuto (è, infatti, possibile contare sul personale dell'azienda senza dover ricorrere a dispendiose consulenze) ed inoltre la possibilità di ottenere risultati significativi dato che in questa fase si attua la "caccia agli sprechi".

È possibile definire degli strumenti di pianificazione e controllo in ambiente Lean Manufacturing che sono espressi come di seguito:

$Tpitch = Takt * \text{numero pezzi contenuti nell'imballo del prodotto finito}$

$EPEI = \text{tempo internal changeover} / \text{tempo disponibile}$

L'EPEI indica il livello di flessibilità che l'azienda possiede per poter produrre tutto ciò che il cliente richiede.

2.2 – Agile Manufacturing

La capacità di poter soddisfare le richieste dei clienti per la consegna in tempi sempre più brevi, e assicurare che l'approvvigionamento e la produzione possono essere sincronizzati per soddisfare i picchi e le depressioni della domanda, è di fondamentale importanza in questa epoca di concorrenza basata sul tempo.

Per diventare più rispondente alle esigenze del mercato si richiede più velocità. Si richiede inoltre un elevato livello di manovrabilità che oggi è stato definito agilità.

L'Agility, è definita come la funzionalità a livello di business che abbraccia strutture organizzative, sistemi informativi, processi logistici, e, in particolare, mentalità.

Essa aiuta a fornire il prodotto giusto, al momento giusto al consumatore, che è l'obiettivo principale di ogni filiera.

L'Agile Manufacturing è visto come il passo successivo alla Lean Production nell'evoluzione della metodologia di produzione.

Si definisce come l'insieme di tecniche e metodologie che permettono, oltre che una produzione snella, anche una pronunciata capacità di risposta alle continue evoluzioni di un mercato estremamente variabile ed imprevedibile. Proprio per questo motivo, la Supply Chain Agile è definita *market sensitive* ovvero capace di leggere e rispondere alle esigenze dei clienti in tempo reale.

Il termine Agile, come concetto di produzione, è stato coniato nel 1991 da un gruppo di ricercatori presso la Iaccoca Institute, Lehigh University, per descrivere delle iniziative, aventi luogo in USA, Europa Occidentale e Giappone, che miravano a creare un clima industriale che garantisse la competitività globale nell'ambito della produzione. La produzione Agile

rappresentava il potenziale per gli Stati Uniti per riprendere il ruolo di leader di mercato nella produzione.

Una caratteristica fondamentale di un'organizzazione Agile è la flessibilità. Inizialmente, si pensava che la via verso la produzione flessibile era attraverso l'automazione, per consentire un rapido cambio (cioè, ridotto tempo di set-up) e, quindi, una maggiore reattività alle variazioni di mix di prodotto o di volume. Più tardi, questa idea di flessibilità produttiva è stata estesa in un contesto più ampio ed è nato il concetto di agilità, inteso come orientamento organizzativo.

I sistemi di produzione Agile consentono: risposte rapide e convenienti rispetto alla mutevole domanda di prodotto, e di sostenere il lancio, non pianificato, di prodotti su misura per soddisfare i desideri del cliente.

Goldman (1995) definisce Agile *“le aziende di produzione che sono in grado di operare con profitto, in un ambiente competitivo e imprevedibilmente, cambiando le opportunità dei clienti e rispondendo efficacemente alle mutevoli tendenze della domanda”*.

Analogamente, Gunasekaran (1998) definisce Agile *“la capacità di sopravvivere e prosperare in un ambiente competitivo di cambiamento continuo e imprevedibile, e di reagire rapidamente ed efficacemente ai cambiamenti del mercato, guidati da prodotti e servizi progettati per i clienti”*.

Inoltre, Lee (2004) definisce Agile *“un insieme di strategie che risolve il problema dell'incertezza e della variabilità della domanda aumentando la flessibilità del sistema”*.

L'Agile Manufacturing è quindi un approccio alla produzione che si concentra sulla soddisfazione delle esigenze dei clienti, pur mantenendo elevati standard di qualità e di controllo dei costi complessivi coinvolti nella

produzione di un particolare prodotto. Questo approccio è orientato verso aziende che operano in un ambiente altamente competitivo, in cui piccole variazioni nelle prestazioni e nella consegna del prodotto possono fare una grande differenza nel lungo termine per la sopravvivenza di un'azienda e la reputazione tra i consumatori.

In Lean Manufacturing, l'azienda mira a tagliare tutti i costi che non sono direttamente collegate alla produzione di un prodotto. L'Agile Manufacturing può includere questo concetto, ma aggiunge anche una dimensione ulteriore, l'idea che le richieste dei clienti devono essere soddisfatte rapidamente ed efficacemente.

In situazioni in cui le aziende integrano entrambi gli approcci, è possibile parlare di “produzione agile e snella.”

Le tre dimensioni critiche che consentono di scegliere l'approccio più opportuno tra Lean e Agile sono: varietà, variabilità (o prevedibilità) e volume.

L'ambiente di produzione considerato, per una agile 63esi s chain, ma anche in ambiente Lean, è il make to order.

Per Make To Order (M.T.O) si intende un processo di produzione che inizia solo dopo aver acquisito l'ordine del cliente. Le sole attività di progettazione/ingegnerizzazione possono essere anticipate rispetto al momento dell'acquisizione dell'ordine.

Il make to order rappresenta un processo di produzione di tipo pull, ovvero tirato dal cliente. Questo crea un tempo di attesa supplementare per il consumatore per ricevere il prodotto, ma consente una maggiore personalizzazione rispetto all'acquisto dagli scaffali dei rivenditori (M.T.S).

Le aziende che utilizzano un approccio produttivo Agile tendono ad avere forti rapporti con i fornitori e le società collegate, insieme a numerose

squadre di cooperative che operano all'interno della società per fornire in modo efficace i prodotti. Esse possono riorganizzare strutture in fretta, negoziare nuovi accordi con fornitori e altri partner in risposta alle mutevoli forze di mercato, e adottare altre misure per soddisfare le richieste dei clienti.

La caratteristica vincente di una Supply Chain Agile è quindi quella di garantire brevi Lead Time e grande differenziazione dei prodotti con un altissimo livello di servizio.

In una Supply Chain guidata da politiche Agile il livello minimo di magazzino sarà certamente più elevato, rispetto a quello riscontrabile in sistemi Lean, così da garantire una maggiore flessibilità nel rispondere alla domanda dei clienti. Sempre per questo motivo potrebbero essere necessarie alcune attività non value adding, intollerabili in un'ottica Lean.

In un sistema Lean Production, l'eliminazione di tutte le attività a non valore aggiunto comporta una riduzione del costo del prodotto e questo rende competitivo il sistema.

I sistemi Agile Manufacturing pongono molta enfasi sulla rapidità dei changeovers in modo da garantire un'elevata flessibilità al sistema. I muda devono essere eliminati, ma non a discapito della flessibilità che rimane l'obiettivo più importante.

Processi di produzione Lean e Agile differiscono anche in termini di come facilmente i prodotti possono essere personalizzati in ogni sistema. La modifica di qualsiasi parte di un prodotto in produzione snella, come ad esempio la personalizzazione del funzionamento o l'aspetto, richiede una riprogettazione dei componenti interni ed esterni, nonché la generazione di prototipi per verificarne la corretta funzionalità. Prodotti su misura sono estremamente costosi a causa dei costi elevati di ricerca e progettazione. Inoltre, la linea di produzione viene interrotta per produrre il prodotto

personalizzato, e ciò impatta negativamente sui tempi di produzione e sui costi.

Al contrario, la produzione Agile è in grado di ospitare ordini di prodotti personalizzati in quanto la costruzione modulare può essere modificata rapidamente. La linea di produzione deve semplicemente adattare o aggiungere nuovi moduli per il prodotto esistente. Di conseguenza, il consumatore può acquistare un prodotto personalizzato a prezzi competitivi, senza ostacolare la normale linea di produzione aziendale.

Entrambi i processi di produzione, Lean e Agile, sono in grado di soddisfare le esigenze dei clienti, ma con un effetto diverso sul costo del prodotto finale.

Le principali caratteristiche dell'Agile Manufacturing sono: reattività, competenza, flessibilità, e velocità.

La produzione Agile si propone come una strategia estremamente efficace per i seguenti motivi:

- i consumatori amano la gratificazione immediata e sono spesso disposti a pagare per essa;
- i consumatori amano scegliere. Essi preferiscono ottenere un prodotto esattamente come lo vogliono, senza compromessi;
- i consumatori sono volubili. I loro interessi si spostano e si muovono in modo imprevedibile.

L'Agile Manufacturing è efficace proprio perché affronta direttamente questi problemi. Essa riconosce la realtà del mercato moderno e la trasforma in un vantaggio competitivo dando voce al mercato.

Ci sono quattro elementi chiave per la produzione Agile:

- *Modular Product Design*: progettazione di prodotti in maniera modulare che consente, ai prodotti stessi, di servire come piattaforme

per la variazione facile e veloce. Utilizzando un design modulare del prodotto, come strategia di sviluppo di nuovi prodotti, è possibile ridurre il time to market, aumentare il numero di varianti di prodotto e diminuire il numero di pezzi unici nell'architettura del prodotto. Ne consegue un sistema di produzione snella e flessibile nonché l'ottimizzazione della supply chain;

- *Information Technology*: automatizzazione dell'informazione per renderne rapida la diffusione in tutta l'azienda e per ottenere una veloce risposta agli ordini. Rappresenta una risorsa essenziale delle organizzazioni, all'interno delle quali diventa sempre più importante riuscire a gestire in maniera rapida, efficace ed efficiente il volume crescente di informazioni. L'Information Technology va considerata come un'arma strategica, in grado di mettere a disposizione dati e informazioni qualitativamente migliori nell'ambito dell'organizzazione e, grazie alla diffusione della tecnologia e dell'interconnettività, può aiutare le organizzazioni a ridefinire i propri rapporti con clienti, fornitori e altre organizzazioni;

- *Corporate Partners*: creazione di alleanze virtuali a breve termine, con altre aziende o con parti dell'azienda stessa, che consentono un migliore time to market per i segmenti di prodotti selezionati. Tale concetto, quindi, non riguarda solo i rapporti che l'azienda detiene con l'esterno ma si ripercuote anche all'interno dell'azienda stessa. La Supply Chain deve essere vista come una confederazione di partner collegati tra loro attraverso una rete di collaborazione. Ciò implica la pianificazione centralizzata, cioè la progettazione e realizzazione di strutture di gestione cooperative. Operativamente, la pianificazione centralizzata richiede alle diverse unità della catena di produzione di adottare decisioni di gestione e obiettivi di prestazione per massimizzare le prestazioni dell'intera catena, e non le prestazioni di

ciascuna unità di produzione individualmente. Ogni unità deve prendere in considerazione le unità adiacenti al momento di prendere decisioni di produzione. La pianificazione centrale crea una prospettiva comune e garantisce la massimizzazione delle prestazioni complessive;

- *Knowledge Culture*: investire nella formazione dei dipendenti per raggiungere una cultura che sostiene il cambiamento rapido e l'adattamento in corso.

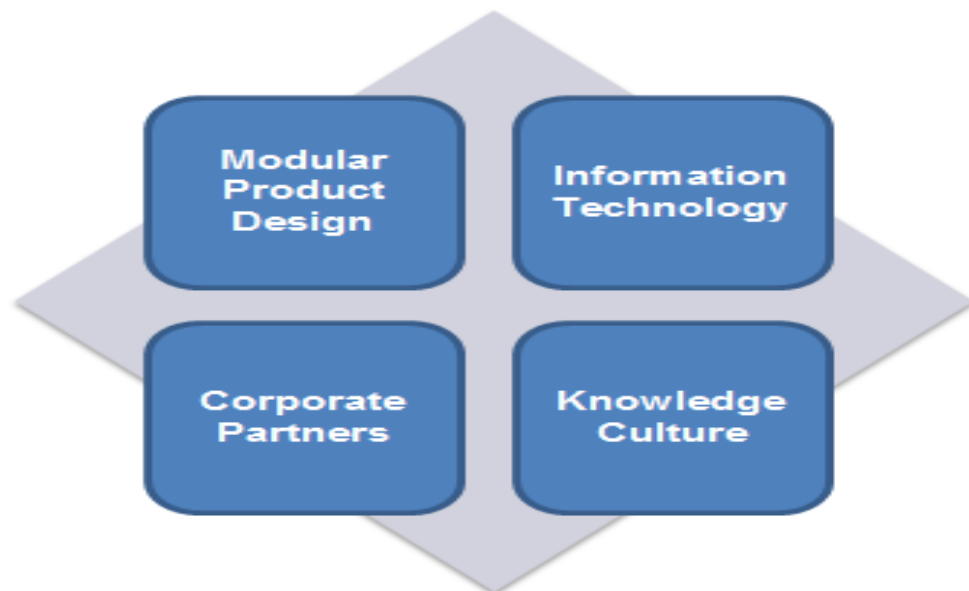


Figura 10: I 4 elementi chiave dell'Agile Manufacturing

Fonte: <http://www.leanproduction.com/>

L'obiettivo di un produttore Agile è quello di trovare una soluzione specifica per i bisogni dei clienti e non solo un prodotto generico.

Quindi riassumendo le principali differenze, nei risultati di performance, tra Lean Production e Agile Manufacturing, sono correlati ai costi e alla flessibilità, in modo tale che, la produzione snella ha un impatto significativo

sulle prestazioni di costo e che la produzione agile ha un impatto più forte sulla flessibilità del mix di prodotti.⁴

Un sistema di produzione Agile consente ad un'azienda di ri-allocare la capacità della linea di produzione per i prodotti che sono superiormente richiesti rispetto alla domanda prevista, e di lanciare rapidamente nuovi prodotti (non precedentemente concepiti quando il sistema di produzione è stato progettato), e tuttavia mantenere la capacità di produzione di altri prodotti con richiesta inferiore rispetto alla domanda prevista.

In tal modo è possibile gestire meglio la volatilità della domanda, con conseguente aumento dei ricavi e quote di mercato.

Il sistema di lavorazione Agile è quindi un sistema che può cambiare rapidamente e facilmente per produrre una prevista gamma di modelli in una classe di prodotti, ed essere rapidamente e convenientemente riconfigurato per rispondere alle nuove introduzioni di modello.

In alcuni settori le industrie stanno attualmente applicando Sistemi Agile perché desiderano la versatilità dei sistemi flessibili, ma vogliono anche il minimo costo aggiuntivo di investimenti per introdurre nuovi modelli.

I sistemi flessibili permettono l'introduzione di nuovi modelli di prodotto (non programmati), ma con un significativo aumento di tempo e penalità dei costi sostenuti. Gli strumenti e la gestione di un sistema flessibile sono più costosi, perché hanno utilità per una varietà di applicazioni di lavorazione.

Sistemi di lavorazione Agile permettono, invece, di realizzare diversi modelli di prodotto, all'interno di una classe di prodotto, sulla stessa linea, con rapidi cambi da modello a modello.

Sistemi Agile sono più efficienti nella gestione del cambio di modello e perdono poco o niente in termini di tempo di produzione per l'introduzione

⁴ Jan Olhager, (2009) "*Lean and agile manufacturing: external and internal drivers and performance outcomes*", International Journal of Operations & Production Management.

del nuovo modello ma non offrono un'elevata riutilizzabilità delle attrezzature per le altre applicazioni di lavorazione.

La produzione Agile è, quindi, una tecnica di produzione e assemblaggio di prodotti nuovi e rivoluzionari.

Nella catena evolutiva delle tecnologie di produzione l'Agile Manufacturing è il passo successivo rispetto ai suoi predecessori che sono: produzione artigianale, produzione di massa e produzione snella.

Ciò che deve essere fatto per preparare con successo la strada per l'attuazione della produzione Agile, richiede cambiamenti in cinque aree: regolamento del governo, cooperazione tra imprese, tecnologia dell'informazione, reengineering, e flessibilità dei dipendenti.

2.2.1 – Caratteristiche dell'Agile Supply Chain

L'agile supply chain, essendo definita come evoluzione della lean supply chain, avrà con quest'ultima molteplici peculiarità in comune soprattutto in termini di caratteristiche strutturali e funzionali nonché in termini di logica ed ambiente di produzione.

Un totale di 15 variabili sono state individuate per lo sviluppo di un quadro di riferimento per una agile supply chain.

Le variabili individuate sono:

- Market sensitive: si intende che la supply chain è in grado di leggere e rispondere alla domanda reale. La sensibilità di mercato di una supply chain è influenzabile attraverso il livello di collaborazione tra i suoi partner, e la sua capacità di utilizzare strumenti informatici.
- Delivery speed: definito come capacità di soddisfare rapidamente l'obiettivo. Si riferisce alla capacità di fornire prodotti o servizi più

velocemente di altri concorrenti. Questa definizione comprende anche il tempo per portare nuovi prodotti sul mercato, il tempo per la produzione di un prodotto, e il tempo esistente per fornire un prodotto al cliente.

- Data accuracy: essa è definita come l'accuratezza dei dati utilizzati dai diversi partner per prendere decisioni di pianificazione. Con dati poco precisi si rischia di prendere decisioni errate.
- New product introduction (NPI): l'introduzione di nuovi prodotti è emerso come la strategia vincente. E' particolarmente importante in un business in cui il ciclo di vita del prodotto è relativamente breve. L'introduzione di un nuovo prodotto sul mercato può certamente portare vantaggi significativi, tra cui una maggiore quota di mercato e una maggiore redditività.
- Centralized and collaborative planning (CCP): l'efficace integrazione della supply chain e la sincronizzazione tra i partner consente di eliminare le scorte in eccesso, ridurre il lead time, aumentare le vendite e migliorare il servizio al cliente. In tal modo è possibile aumentare la soddisfazione del cliente e la competitività.
- Process integration (PI): la condivisione di informazioni tra i partner della supply chain è raggiunta mediante l'integrazione dei processi. Per "integrazione dei processi" si intende la collaborazione tra acquirenti e fornitori, lo sviluppo congiunto del prodotto e informazioni condivise.
- Use of IT tools (UIT): l'uso delle tecnologie per condividere dati tra acquirenti e fornitori consente la creazione di una supply chain virtuale.
- Lead time reduction (LTR): la riduzione del lead time interno è il meccanismo per essere competitivi in una concorrenza basata sul tempo. La gestione del tempo è l'immagine speculare della gestione della qualità, dei costi, dell'innovazione e della produttività.

- Service level improvement (SLI): il miglioramento del livello di servizio è sempre citato come un obiettivo chiave di gestione della supply chain poiché esso porta al miglioramento delle prestazioni di quest'ultima.
- Cost minimization (COM): le supply chain devono sforzarsi di gestire i costi associati con i loro prodotti / sistema di erogazione di servizi. L'obiettivo però dovrebbe essere sia aumentare la soddisfazione del cliente e che la minimizzazione dei costi della supply chain.
- Customer satisfaction (CUS): per migliorare le prestazioni, le metriche della supply chain devono essere collegate alla soddisfazione del cliente. Questa misura è necessaria per integrare le richieste del cliente nelle logiche di produzione, per impostare le dimensioni della qualità, per il controllo dei costi, e come feedback per il controllo del processo. La soddisfazione del cliente rappresenta la reazione del cliente alla propria percezione del valore ricevuto come risultato dell'utilizzo di un particolare prodotto o servizio.
- Quality improvement (QI): il miglioramento della qualità è riconosciuto come d'obbligo per avere successo in un mercato internazionale competitivo. Il coinvolgimento dei fornitori è fondamentale per migliorare la qualità e soddisfare le specifiche del cliente. Non è possibile massimizzare il valore percepito dai clienti se i componenti acquisiti dai fornitori sono difettosi e in ritardo. A tal fine è necessario sviluppare relazioni con i fornitori. Migliorare la qualità di tutti i processi della supply chain provoca la riduzione dei costi, un migliore utilizzo delle risorse e il miglioramento efficienza del processo.
- Minimizing uncertainty (MU): l'incertezza si propaga attraverso la supply chain e conduce a inefficienze e ad attività non a valore aggiunto. Assume grande importanza la minimizzazione

dell'incertezza per ottenere prestazioni competitive a livello internazionale.

- Trust development (TD): il requisito relazionale primario per una migliore capacità di risposta è lo sviluppo di maggiori livelli di fiducia tra l'azienda e i fornitori. La fiducia tra i partner migliora la comunicazione e il dialogo e crea visioni strategiche comuni.
- Minimizing resistance to change (MRTC): la resistenza al cambiamento è stata riconosciuta come un fattore estremamente importante che può influenzare il successo o meno di uno sforzo di cambiamento organizzativo.

CAPITOLO 3

Lo strumento Markoviano

3.1 - Concetti introduttivi

Una catena di Markov è un insieme di variabili aleatorie (v.a.) con un ordinamento temporale aventi la proprietà che, dato il valore attuale, il valore futuro non dipende da quello passato.

Questo concetto, che può essere visto sotto la forma della Proprietà di Markov, sarà poi approfondito in seguito, ma ciò su cui ora si vuole focalizzare l'attenzione è la caratteristica per cui tale insieme giace a metà tra un insieme di variabili aleatorie indipendenti e un insieme completamente generico che potrebbe essere molto complicato da analizzare.

Andrei Andreivich Markov cominciò l'analisi di tali insiemi di v.a. nel 1907, e tale campo di ricerca è ancora in espansione al giorno d'oggi. Lo studio delle catene di Markov è, infatti, uno dei grandi risultati della teoria delle probabilità. Nel suo primo lavoro Andrei Nikolaevich Kolmogorov osservò che “storicamente, l'indipendenza tra esperimenti e v.a. rappresenta il concetto matematico fondamentale che ha dato alla probabilità il suo stampo peculiare”.

Nonostante ciò, vi sono molte situazioni in cui è necessario considerare sequenze di v.a. che non possono essere considerate indipendenti. Kolmogorov proseguì osservando che Markov et al frequentemente sbagliavano nell'assumere la completa indipendenza, tuttavia rivelò anche l'importanza di dover assumere condizioni analoghe più deboli per poter ottenere risultati significativi. La sopracitata proprietà di Markov,

caratteristica distintiva dell'omonima catena, è una condizione analoga ma più debole rispetto all'indipendenza di v.a. e si è dimostrata essere abbastanza forte per poter ottenere risultati importanti ma anche sufficientemente debole da permetterle di poter ricomprendere diversi casi interessanti.

La maggior parte degli sviluppi che si sono avuti negli ultimi anni del secolo scorso nel campo della teoria delle probabilità consistono nello studio di sequenze di v.a. che non sono interamente indipendenti. Due condizioni più deboli si sono mostrate molto utili: la proprietà di Markov e la martingala. In parole povere la martingala è una sequenza di v.a. in cui il valore atteso in ogni istante futuro, dato il valore presente e passato, è uguale al suo valore attuale. Esiste una vasta e profonda letteratura a riguardo, tuttavia non sarà oggetto del presente studio; ci si focalizzerà sulle catene di Markov e le sue proprietà.

Durante il secolo scorso un gran numero di esperti ha approfondito lo studio sulla teoria delle catene di Markov impiegandone i concetti in una tale molteplicità di ambiti da non poter enumerare la totalità delle applicazioni.

Esistono due ragioni principali per cui le catene di Markov giocano un ruolo così importante nella moderna teoria delle probabilità:

1. La prima ragione è che forniscono un quadro potente ma trattabile con cui descrivere, caratterizzare ed analizzare una vasta gamma di v.a. che trovano applicazione in numerosi ambiti: dal trasporto di particelle agli automi a stati finiti fino alla teoria della codifica dei geni.
2. La seconda è che tramite la simulazione di particolari C.M. sono stati sviluppati un insieme di potenti algoritmi computazionali in grado di fornire campioni provenienti da complicate distribuzioni di probabilità: queste catene di Markov “metodo Monte Carlo” sono ora diffuse in tutti i campi in cui è necessario ottenere campioni di

distribuzioni di probabilità anche complesse e ciò ha trascinato la recente ricerca nel campo delle C.M..

Le aree in cui compaiono le C. M. sono troppe per essere elencate qui, ma eccone alcuni tipici esempi:

- Ogni insieme di v.a. crea una C.M.: in questo caso, dato il valore attuale, il valore futuro è indipendente dalla storia passata e dal presente.
- La ben nota “symmetric random walk over the integers” fornisce un classico esempio: il valore successivo nella catena è una unità in più o in meno rispetto al valore attuale con stessa probabilità di accadimento, indipendentemente dal percorso con cui il valore attuale è stato raggiunto. Nonostante la sua semplicità, questo esempio, e altre semplici generalizzazioni possono mostrare proprietà molto interessanti.
- Molti giochi da tavola popolari sono rappresentabili con una C.M.: ad es. in “Snakes and Ladder” vi sono 100 possibili stati per ogni segnalino e lo stato successivo per ogni segnalino può essere uno dei sei stati raggiungibili da quello attuale con identica probabilità. Per cui, lo stato successivo è funzione dello stato presente e di una v.a. esterna da esso indipendente che corrisponde al valore ottenuto con il lancio di un dado.
- Un esempio più pratico di C.M. può essere quello che considera la quantità di acqua contenuta in un serbatoio: il volume di acqua immagazzinato dipende dal volume attuale e da due quantità aleatorie: la quantità di acqua in uscita dal serbatoio e quella in ingresso. Varianti più sofisticate di questo modello sono utilizzate in numerose aree, in particolare nel campo della teoria delle code (in cui la quantità di acqua è sostituito dal numero di clienti in attesa del servizio).

- Infine un tipico esempio dell'utilizzo delle C.M. si ha con gli automi a stati finiti: in particolare l'evoluzione di un automa a stati finiti può essere vista come l'evoluzione (normalmente deterministica) di una C.M..

Si è soliti pensare alle C.M. per descrivere le traiettorie di oggetti dinamici; in alcune circostanze, infatti, esiste un sistema dinamico associato all'insieme di v.a. caratterizzate da una particolare struttura di indipendenze. Ad esempio la "random walk" discussa precedentemente può essere vista come un percorso in cui ci si sposta da una posizione alla successiva in cui l' n -simo elemento della C.M. associata corrisponde alla posizione occupata nel tempo discreto nell'istante n -simo. Poiché la distribuzione di ogni v.a. nella sequenza dipende solo dal valore del precedente elemento nella sequenza, si può dotare ogni insieme di v.a. di questo tipo di una struttura dinamica (assunto che si possano ordinare gli elementi di questo insieme, in modo tale che la definizione di C.M. qui fornita sia verificata). In modo molto semplice, si potrà poi vedere la distribuzione di ogni elemento dato il valore precedente come la probabilità di spostarsi tra quegli stati in quell'istante. Questa interpretazione può fornire una comprensione e una descrizione più semplice del comportamento di insiemi di v.a. del tipo descritto precedentemente negli esempi. Fondamentalmente è l'immagine di una catena di stati in cui uno stato è connesso al successivo che suggerisce il termine "catena" di Markov.

3.2- I processi stocastici

Per poter fornire una definizione formale di una Catena di Markov, è prima necessario precisare cosa si intende per insieme di v.a. aventi un ordinamento

temporale. Un tale insieme di v.a. può essere meglio rappresentato da un processo stocastico.

Un processo “E-Valued” è una funzione $X: I \rightarrow E$ che a valori appartenenti a un insieme I fa corrispondere valori in un altro spazio E . L’evoluzione del processo è descritta considerando la variazione di $X_i := X(i)$ con i . Un processo stocastico “E-Valued” può essere visto come un processo in cui, per ogni $i \in I$, X_i è una v.a. con valori in E . Sebbene esista una vasta letteratura su casi più generali, in questo lavoro si considererà unicamente il caso di processi stocastici tempo discreti in cui l’insieme degli indici I corrisponde all’insieme dei numeri naturali N . La notazione X_i è utilizzata per indicare il valore del processo al tempo i (N.B. non c’è bisogno di alcuna connessione tra l’insieme degli indici e il tempo reale, ma questa terminologia è sia standard che conveniente). Si noti che la proprietà di Markov può essere estesa a processi tempo continui in cui il set degli indici coincide con l’insieme dei numeri reali positivi, ciò porterebbe a un insieme di processi noti anche come Processi di Markov o C.M. tempo continue. Tali processi non saranno oggetto del presente lavoro, poichè sono in qualche modo meno importanti nel campo dell’informatica e delle applicazioni ingegneristiche.

Esiste una ricca letteratura su questi processi, e molti dei risultati disponibili nel caso del tempo discreto hanno dei casi analoghi nel tempo continuo; in realtà alcuni risultati possono essere ottenuti più naturalmente nel caso tempo continuo.

A questo punto, è necessario specificare un pò di terminologia. Originariamente il termine “Catena di Markov” era utilizzato per descrivere ogni processo stocastico che godesse della proprietà di Markov e che avesse uno spazio a numero di stati finiti. Alcuni riferimenti ancora utilizzano tale definizione. In ogni caso, nel campo informatico, ingegneristico e statistico è divenuto più usuale utilizzare tale termine per riferirsi a ogni processo stocastico tempo discreto che goda della proprietà di Markov, senza alcun

riferimento allo spazio degli stati; tale è la definizione di cui si farà uso nel presente lavoro.

Processi tempo continui che godono della proprietà di Markov saranno chiamati Processi di Markov ed si farà riferimento ad essi solo in parte. L'utilizzo di processi continui è dovuto a considerazioni che scaturiscono dalla valutazione dei metodi Monte Carlo applicati alle catene di Markov, ampiamente presenti in letteratura.

3.3 - Filtraggio e Stopping Times

Questo paragrafo riguarda alcuni dettagli tecnici che, sebbene non essenziali a una comprensione di base delle catene di Markov (come processi stocastici), sono fondamentali e si possono ritrovare in ogni lavoro che riguardi questi argomenti.

Una struttura più tecnica è generalmente richiesta nel considerare processi stocastici piuttosto che semplici variabili aleatorie. Sebbene dettagli tecnici saranno evitati il più possibile in questo lavoro, i seguenti concetti sono necessari per comprendere la maggior parte della letteratura riguardante le catene di Markov.

Per avere a che fare con semplici v.a. è sufficiente considerare uno spazio di probabilità (ω, \mathcal{F}, P) in cui ω è un insieme di eventi, \mathcal{F} è un σ -algebra corrispondente a un insieme di output misurabili (ad esempio l'insieme dei sottoinsiemi ω di \mathcal{F} a cui è possibile assegnare una probabilità, tipicamente l'insieme di tutti i sottoinsiemi di ω nel caso discreto) e P è la misura di probabilità, che ci dice qual è la probabilità con cui ogni elemento di \mathcal{F} contenga l'evento che si realizza, ed è tale che: $P: \mathcal{F} \rightarrow [0,1]$.

Nel considerare processi stocastici conviene definire uno spazio di probabilità filtrato $(\mathcal{F}, \{\mathcal{F}_i\}_{i \in \mathbb{N}}, P)$. L'insieme dei sottoinsiemi σ -algebra, $\{\mathcal{F}_i\}_{i \in \mathbb{N}}$, detto Filtraggio, ha una struttura particolare:

$$\mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots \subset \mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \dots \subset \mathcal{F}$$

e la sua proprietà più importante è che, $\forall n$, l'insieme di variabili X_1, X_2, \dots, X_n deve essere misurabile rispetto a \mathcal{F}_n .

Sebbene si possa ancora generalizzare, normalmente è sufficiente considerare il filtraggio naturale di un processo: che è quello generato dal processo stesso. Dato un qualunque insieme di v.a. presi da uno stesso spazio di probabilità, in esso esiste il più piccolo σ -algebra rispetto a cui le v.a. sono congiuntamente misurabili. Il filtraggio naturale è il filtraggio generato ponendo ogni \mathcal{F}_n uguale al più piccolo σ -algebra rispetto a cui le X_1, \dots, X_n sono misurabili. In questa trattazione sarà considerato solo questo tipo di filtraggio. Una interpretazione intuitiva di questo filtraggio (che fornisce una suddivisione dello spazio di probabilità sempre più raffinata) è che \mathcal{F}_n ci dice quanta informazione può essere ottenuta dalla conoscenza dei valori delle prime n v.a.: ovvero ci dice quali eventi possono realizzarsi data la conoscenza delle X_1, \dots, X_n .

Nel considerare processi di questo tipo è naturale porsi domande riguardo istanti di tempo aleatori: c'è qualcosa che ci impedisce di definire delle ulteriori v.a. che si possono interpretare come indici che rappresentano un particolare istante nell'evoluzione del processo? In generale, bisogna avere un minimo di attenzione nel verificare l'utilità di tali v.a. aggiuntive: se la struttura temporale è reale, allora è necessario essere in grado di determinare (considerata nota la realizzazione del processo fino a quel dato istante) se l'istante che è stato raggiunto è l'istante di interesse o meno. Informalmente si potrebbe richiedere che a $\{\tau = n\}$ sia attribuita una probabilità unitaria (data la conoscenza dei primi n stati) per ogni n . In effetti questa condizione è

più forte ma fornisce una semplice interpretazione che sopperisce a molteplici scopi. Formalmente se τ è un istante randomico e l'evento $\{\omega: \tau(\omega)=n\} \in F_n$ per tutti gli n , allora τ è noto come *Stopping Time*. Si noti che questa condizione corrisponde con la condizione che l'evento $\{\omega: \tau(\omega)=n\}$ è indipendente da tutti gli stati successivi della catena X_{n+1}, X_{n+2}, \dots noti gli stati X_1, \dots, X_n . L'esempio più comune di stopping time è il tempo di impatto τ_A di un insieme A :

$$\tau_A := \inf \{n : X_n \in A\},$$

che corrisponde al primo istante in cui il processo entra nell'insieme A . Si noti che l'apparentemente simile:

$$\tau^* A := \inf \{n : X_{n+1} \in A\}$$

non è uno stopping time (in ogni grado di generalità) poichè lo stato della catena all'istante $n+1$ non è necessariamente conosciuto nei termini dei primi n stati.

Si noti che questa distinzione non è una distinzione artificiale o futile. Si consideri la catena generata ponendo $X_n = X_{n-1} + W_n$ dove $\{W_n\}$ è un insieme di v.a. indipendenti che corrispondono al valore delle vincite, in dollari, di un giocatore d'azzardo all' n -sima puntata indipendente che gioca.

Se $A = [10000, \infty)$ allora τ_A corrisponderà all'evento di vincere \$10000 e, in realtà sarà possibile fermare la catena quanto tale evento accadrà. Al contrario se $A = (-\infty, -10000]$, allora $\tau^* A$ corrisponderà all'istante precedente a quello in cui \$10000 vengono persi. Sebbene molte persone preferirebbero essere in grado di smettere di scommettere immediatamente prima di perdere i soldi, non è possibile sapere che si perderà subito dopo una sequenza di giochi indipendenti.

Sia dato uno stopping time τ , è possibile definire il processo stopped, $X_1^\tau, X_2^\tau, \dots$ associato al processo X_1, X_2, \dots che ha la seguente definizione:

scrivendo $m \wedge n$ per indicare il più piccolo tra m e n , si definisce $X_n^\tau = X_{m \wedge n}$. Ovvero il processo “stopped” corrisponde al processo stesso tutte le volte che l’istante risulta pari allo stopping time, dopodichè assume il valore che aveva prima di quell’istante. Nel caso di τ_A , ad esempio, il processo “stopped” rispecchia il processo originale fino a che non entra in A , dopodichè conserva il valore che aveva prima di entrare in A per tutti gli istanti successivi.

3.4 - Simulare una catena di Markov

Si descrive una C.M. nel seguente modo: abbiamo un insieme di stati $S = \{s_1, s_2, \dots, s_n\}$. Il processo inizia in uno di questi stati e si muove successivamente da uno stato a un altro. Ogni mossa è detta Step. Se la catena è correntemente nello stato s_i , allora si muoverà verso s_j al passo successivo con una probabilità denotata con p_{ij} , e questa probabilità non dipende da in quale stato la catena fosse prima di quello corrente.

Le probabilità p_{ij} sono dette probabilità di transizione. Il processo può rimanere nello stato in cui si trova con una probabilità p_{ii} . Una distribuzione iniziale di probabilità, definita su S , specifica lo stato di partenza. Normalmente lo stato di partenza è uno stato particolare.

R.A.Howard ci fornisce una pittoresca descrizione di una catena di Markov come una rana che salta su una serie di ninfee. La rana comincia su una delle foglie e dopo salta da una all’altra con una data probabilità di transizione. Vi sono tre elementi da specificare nel caratterizzare una C.M.:

- Lo spazio di Stato S : S è un insieme di stati finito o numerabile, ovvero lo sono i valori che le v.a. X_i possono assumere. Per approfondire ma senza perdere di generalità, etichettiamo gli stati nel seguente modo: $S = \{1, 2, \dots, N\}$ per un N finito, o $S = \{1, 2, \dots\}$ per il caso in cui $N = \infty$.

- La distribuzione iniziale π_0 . Questa è la distribuzione di probabilità della catena di Markov al tempo zero. Per ognuno degli stati $i \in S$ si indicherà con $\pi_0(i)$ la probabilità $P\{X_0=i\}$ con cui la catena di Markov si inizializza nello stato i . Formalmente, π_0 è una funzione che da S ha valori nell'intervallo $[0,1]$ tale che $\pi_0 \geq 0 \forall i \in S$ e

$$\sum_{i \in S} \pi_0(i) = 1.$$

Equivalentemente, invece di considerare π_0 come una funzione che da S va in $[0,1]$, lo si può pensare come un vettore la cui i -esima componente è pari alla probabilità $\pi_0(i) = P\{X_0=i\}$.

- La regola di probabilità di transizione. Viene specificata fornendo una matrice $P=(P_{ij})$. Se S contiene un numero N di stati, allora P sarà una matrice $N \times N$. La P_{ij} può essere interpretata come la probabilità condizionale: ovvero la probabilità che la catena si sposti verso lo stato j all'istante $n+1$ dato che si trovava nello stato i all'istante n ; ovvero $P_{ij} = P\{X_{n+1}=j | X_n=i\}$.

Verrà utilizzata equivalentemente la notazione $P(i,j)$. Si noti che è stata scritta questa probabilità come una funzione dei soli i e j , ma ovviamente potrebbe dipendere anche dal particolare "n". La restrizione di omogeneità temporale menzionata è solo un'assunzione per cui questa probabilità non dipende dall'istante temporale n , ma resta costante nel tempo. Formalmente, la matrice di transizione è una matrice quadrata $N \times N$ i cui elementi sono tutti non-negativi e le cui righe hanno somma unitaria. Infine, ci si potrebbe chiedere perchè organizzare queste probabilità all'interno di una matrice: nel successivo paragrafo ne sarà fornita una spiegazione.

3.5 - La proprietà di Markov

Nel caso precedente se avessimo avuto i valori $X_0=3$, $X_1=1$ e $X_2=2$, allora la distribuzione di probabilità condizionale per X_3 è:

$$P\{X_3=j | X_0=3, X_1=1, X_2=2\} = \begin{cases} \frac{1}{3} & \text{per } j = 1 \\ 0 & \text{per } j = 2 \\ \frac{2}{3} & \text{per } j = 3 \end{cases}$$

che rappresenta anche la distribuzione di probabilità condizionale per X_3 data l'unica informazione per cui $X_2=2$. In altre parole, dato che $X_0=3$, $X_1=1$ e $X_2=2$, l'unica informazione rilevante alla distribuzione per X_3 è l'informazione che $X_2=2$, e potremmo ignorare l'informazione che $X_0=3$ e $X_1=1$. Ciò è chiaro dalla descrizione di come simulare la catena. Per cui,

$$P\{X_3=j | X_0=3, X_1=1, X_2=2\} = P\{X_3=j | X_2=2\} \quad \forall j.$$

Questo è un esempio di proprietà di Markov.

Definizione: Un processo X_0, X_1, \dots soddisfa la proprietà di Markov se

$$\begin{aligned} P\{X_{n+1}=i_{n+1} | X_n=i_n, X_{n-1}=i_{n-1}, \dots, X_0=i_0\} = \\ = P\{X_{n+1}=i_{n+1} | X_n=i_n\} \end{aligned}$$

$$\forall n \text{ e } \forall i_0, \dots, i_{n+1} \in S.$$

La problematica che scaturisce dalla proprietà di Markov è la struttura di dipendenza tra le v.a.. La forma più semplice di dipendenza per X_0, X_1, \dots è l'indipendenza totale. La proprietà di Markov potrebbe essere enunciata per "catturare la struttura di dipendenza successiva": nel generare il processo X_0, X_1, \dots sequenzialmente, lo stato "successivo" X_{n+1} dipende solo dal valore attuale X_n e non dai valori passati $X_0 \dots X_{n-1}$. La proprietà di Markov ci permette di considerare processi più generici e interessanti rispetto alla restrizione a processi di v.a. indipendenti, senza però che la trattazione matematica diventi intrattabile.

- Il significato della proprietà di Markov può essere racchiuso in una semplice frase:

"Dato il presente, il futuro non dipende dal passato"

dove $P(i,j) = P\{X_{n+1}=j|X_n=i\} = P\{X_1=j|X_0=i\}$. La legge di probabilità totale ci fornisce

$$\begin{aligned}\pi_{n+1}(j) &= P\{X_{n+1}=j\} = \\ &= \sum_1^N P\{X_n = i\} P\{X_{n+1} = j|X_n = i\} = \\ &= \pi_n(i) P(i,j),\end{aligned}$$

che in notazione matriciale fornisce l'equazione:

$$\pi_{n+1} = \pi_n P.$$

Si noti che qui si pensa a π_n e a π_{n+1} come a dei vettori riga: ovvero, per esempio:

$$\pi_n = (\pi_n(1), \dots, \pi_n(N)).$$

Per cui otteniamo

$$\pi_1 = \pi_0 P$$

$$\pi_2 = \pi_1 P = \pi_0 P^2$$

$$\pi_3 = \pi_2 P = \pi_0 P^3,$$

e così via, per cui per induzione

$$\pi_n = \pi_0 P^n.$$

D'ora in poi $P^n(i,j)$ denoterà l'elemento (i,j) della matrice P^n .

Per cui, in linea di principio, possiamo trovare la risposta a ogni domanda riguardo il comportamento probabilistico di un C.M. semplicemente applicando dell'algebra matriciale, trovando potenze di matrici, etc. In ogni caso, ciò che è fattibile in pratica è un altro conto. Per esempio, lo spazio di stato per una C.M. che descrive il mescolamento ripetuto di un mazzo di 52 carte contiene $52!$ elementi; questo valore è molto elevato: circa 80 milioni di milioni di milioni di milioni di milioni di milioni di milioni di milioni di

milioni di milioni di milioni. La matrice di probabilità di transizione che descrive l'effetto di un singolo mescolamento è una matrice quadrata $52! \times 52!$. Per cui "tutto ciò che dobbiamo fare" per rispondere a domande riguardo il mescolamento è effettuare potenze di questa matrice, trovarne gli autovalore e così via. In un senso pratico, la semplice riformulazione delle questioni probabilistiche in calcoli matriciali spesso fornisce solo minime chiarificazioni in domande concrete come "quanti mescolamenti sono richiesti affinché il mazzo sia ben mischiato?". Ragionamenti probabilistici possono portare a intuizioni e risultati che potrebbero essere difficili da estrapolare pensando a questi problemi solo come problemi di teoria matriciale.

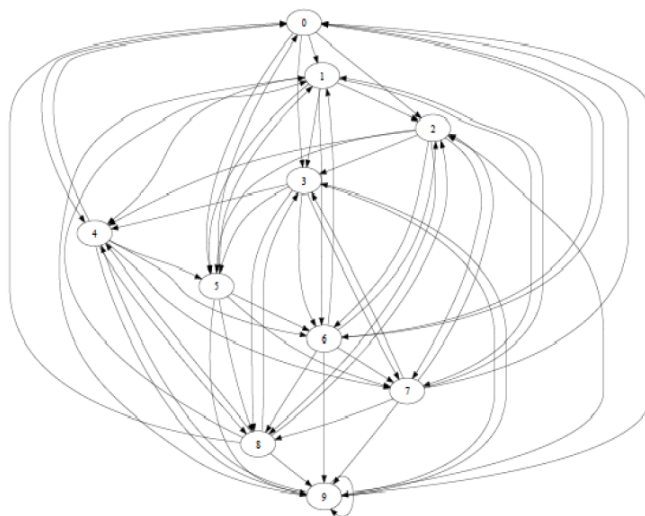
3.6.1 Una rappresentazione grafica

E' molto comune rappresentare C.M. omogenee e con uno spazio di stati finito in modo grafico. Un grafico orientato con degli archi a cui è associato un valore, è sufficiente a descrivere completamente la matrice di transizione di una C.M..

Insieme alla distribuzione dello stato iniziale questa matrice caratterizza completamente la catena. I nodi del grafo corrispondono agli stati e gli archi esistenti rappresentano quali siano le mosse ammissibili. E' solito etichettare gli archi con la probabilità associata alla mossa che essi rappresentano, a meno che le mosse possibili sono tutte equiprobabili. La rappresentazione matriciale può quindi essere piuttosto complicata da interpretare; un semplice esempio lo può mostrare. Si consideri una C.M. ottenuta nello spazio $S = \{0,1,\dots,9\}$ in cui lo stato successivo è ottenuto considerando il numero corrispondente al lancio di un dado non truccato e aggiungendolo allo stato presente a meno che il lancio dia 6 e lo stato attuale sia 9, in tal caso la catena permane nello stesso stato. Questa catena ha una rappresentazione matriciale semplice, ma piuttosto "ingombrante":

$$K = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

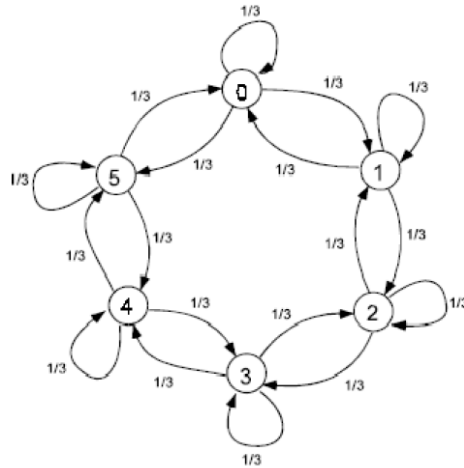
La figura mostra una rappresentazione grafica dello stesso “schema” delle transizioni di Markov; le probabilità di transizione sono omesse in questo caso, poichè tutte uguali. Sebbene questo grafico possa sembrare di non più semplice interpretazione rispetto alla matrice di transizione, ad uno sguardo più attento risulta chiaro che è possibile determinare facilmente senza alcun tipo di calcolo quale stato sia possibile raggiungere da un altro o in quale stato è possibile spostarsi con un dato numero di mosse. Queste proprietà permettono una rappresentazione più semplice da interpretare anche in caso di C.M. con un enorme numero di stati, caso in cui la rappresentazione matriciale diventa rapidamente molto difficile da manipolare. Si noti come raffigurato nel *Mod.1*, il loop che permette di rimanere nello stato 9: è equivalente alla presenza nella matrice di transizione di una diagonale principale con elementi non nulli.



Mod.1 - Rappresentazione grafica di una catena di Markov

3.7 - Il teorema di base limite delle Catene di Markov

Come indica lo stesso nome il teorema oggetto del presente paragrafo ha un ruolo fondamentale nella teoria sulle C.M.. Si consideri un esempio in cui è possibile osservare intuitivamente di cosa si tratta (*Mod.2*).



Mod.2 - Random walk

Per semplicità di scrittura e rappresentazione grafica, si consideri un orologio con 6 numeri: 0,1,2,3,4,5; si supponga di realizzare un cammino casuale con mosse orarie,antiorarie o permanendo nello stesso stato. Ad ogni istante n la probabilità di ciascuna delle 3 mosse è pari a $1/3$. Ovvero:

$$P(i, j) = \begin{cases} \frac{1}{3} & \text{se } j = i - 1 \pmod{6} \\ \frac{1}{3} & \text{se } j = i \\ \frac{1}{3} & \text{se } j = i + 1 \pmod{6} \end{cases}$$

Si supponga di cominciare in $X_0=2$. Si ottiene:

$$\pi_0 = (\pi_0(0), \pi_0(1), \dots, \pi_0(5)) = (0, 0, 1, 0, 0, 0).$$

Da cui ovviamente

$$\pi_1 = (0, 1/3, 1/3, 1/3, 0, 0),$$

ed è facile infine calcolare

$$\pi_2=(1/9,2/9,1/3,2/9,1/9,0)$$

e

$$\pi_3=(3/27,6/27,7/27,6/27,3/27,2/27).$$

Si noti come la probabilità si disperda dalla sua concentrazione iniziale sullo stato 2. Potremmo continuare a calcolare π_n per altri valori di n , ma è intuitivamente chiaro cosa accadrà: la probabilità continuerà a disperdersi e π_n si avvicinerà alla distribuzione uniforme:

$$\pi_n \rightarrow (1/6,1/6,1/6,1/6,1/6,1/6)$$

per $n \rightarrow \infty$. Si provi ad immaginare: se la catena cominciasse nello stato 2 all'istante 0, dopo si chiudino gli occhi mentre il cammino casuale fa 10000 passi; successivamente ci viene chiesto di indovinare in quale stato la catena si troverà all'istante di tempo 10000, quale si penserebbe che siano le probabilità dei diversi stati? Si potrebbe rispondere: “ X_{10000} è per tutti gli scopi pratici uniformemente distribuito sui 6 stati”. Al tempo 10000, il cammino casuale ha praticamente “dimenticato” che ha cominciato il percorso nello stato 2 al tempo zero, ed ha quasi la stessa probabilità di essere ovunque.

Ora si osservi che lo stato iniziale 2 non fosse speciale; saremmo potuti partire da qualunque stato, e lungo il tempo le probabilità si sarebbero disperse dal punto iniziale e si sarebbero avvicinate alla stessa distribuzione limite. Per cui, π_n si avvicina a un limite che non dipende dalla distribuzione iniziale π_0 .

Teorema[Teorema Basic Limit]: Siano X_0, X_1, \dots una C.M. irriducibile e aperiodica con una distribuzione stazionaria $\pi(\cdot)$. Abbia X_0 una distribuzione π_0 , una distribuzione iniziale arbitraria. Allora $\lim_{n \rightarrow \infty} \pi_n(i) = \pi(i)$ per tutti gli stati i .

Bisogna a tal punto definire i termini “irriducibile”, “aperiodica” e “distribuzione stazionaria”.

3.8 - Distribuzione limite per una C.M.

In questo paragrafo viene spiegato il comportamento limite (ovvero il comportamento per $n \rightarrow \infty$) delle catene di Markov. In particolare, sotto alcune condizioni adeguate e facili da verificare, è dimostrato che una C.M. possiede una distribuzione di probabilità limite $\pi = (\pi_j)_{j \in S}$ e se tale catena iniziasse il suo cammino con tale distribuzione allora sarebbe un processo stocastico stazionario. In seguito sarà mostrato che è possibile trovare tale distribuzione π semplicemente risolvendo un set di equazioni lineari.

1.9.1 Classi di comunicazione e irriducibilità delle catene di Markov

Si consideri per una C.M. con spazio di stato S una coppia di stati (i, j) . Si dirà che j è *raggiungibile* da i , e si scriverà $i \rightarrow j$, se \exists un intero $n \geq 0$ tale che $P_{ij}^n > 0$. Ciò significa che partendo dallo stato i c'è una probabilità positiva (ma non necessariamente uguale a 1) che la catena si troverà nello stato j all'istante n (ovvero n passi dopo): $P(X_n=j | X_0=i) > 0$. Se j è raggiungibile da i , e i è raggiungibile da j , allora si dice che gli stati i e j *comunicano* e si indicherà con $i \leftrightarrow j$. La relazione definita dalla *comunicazione* soddisfa le seguenti condizioni:

1. Tutti gli stati comunicano con sè stessi: $P_{ii}^0 = 1 > 0^5$.
2. Simmetria: Se $i \leftrightarrow j$ allora $j \leftrightarrow i$.
3. Transitività: Se $i \leftrightarrow k$ e $k \leftrightarrow j$ allora $i \leftrightarrow j$.

Le condizioni sopra menzionate implicano che la comunicazione è un esempio di relazione di equivalenza, ovvero ne condivide le proprietà con la più comune relazione di eguaglianza “=”:

$i=i$; se $i=j$ allora $j=i$; se $i=k$ e $k=j$ allora $i=j$.

⁵ $P_{ii}^0 = P\{X_0=i | X_0=i\} = 1$

Solo la terza condizione ha bisogno di una spiegazione, per cui ne forniamo la dimostrazione per completezza. Si supponga che esistano degli interi n, m tali che $P_{ik}^n > 0$ e $P_{kj}^m > 0$; sua $l = n + m$, si può concludere che $P_{ij}^l \geq P_{ik}^n P_{kj}^m > 0$ dove abbiamo implicitamente utilizzato le equazioni di Chapman-Kolmogorov. Ciò significa che la catena può andare da i a j (con una probabilità positiva) andando in n passi da i a k e poi (indipendentemente dal cammino passato) da k a j con m passi aggiuntivi.

Se considerassimo un topo in un labirinto aperto, possiamo facilmente osservare che l'insieme degli stati $C_1 = \{1, 2, 3, 4\}$ comunicano tutti tra di loro, ma lo stato 0 comunica solo con sè stesso (poichè è uno stato "assorbente"). Nonostante lo stato 0 sia raggiungibile da tutti gli altri stati, $i \rightarrow 0$, nessun altro stato può essere raggiunto da 0. Possiamo concludere che lo spazio di stato $S = \{0, 1, 2, 3, 4\}$ può essere suddiviso in due insiemi disgiunti $C_1 = \{1, 2, 3, 4\}$ e $C_2 = \{0\}$ la cui unione è uguale a S e sono tali che ognuno dei sottoinsiemi ha la proprietà che tutti gli stati ad essi appartenenti comunicano tra di loro. Disgiunti significa che la loro intersezione non contiene elementi: $C_1 \cap C_2 = \emptyset$.

Riflettendo su questa assunzione si può affermare che questo tipo di disgiunzione può essere fatta con qualunque C.M.:

Proposizione 3.1 *Per ogni C.M. esiste un'unica scomposizione dello spazio di stato S in una sequenza di insiemi disgiunti C_1, C_2, \dots*

$$S = \bigcup_{i=1}^{\infty} C_i$$

in cui ogni sottoinsieme possiede la proprietà che tutti gli stati in esso comunicano tra loro. Ognuno di tali sottoinsiemi è detto classe di comunicazione della Catena di Markov.

Se ora riconsideriamo il topo nel labirinto chiuso $S = \{1, 2, 3, 4\}$, possiamo osservare che esiste un'unica classe di comunicazione $C = \{1, 2, 3, 4\} = S$:

tutti gli stati sono comunicanti. Questo è un esempio di quella che è chiamata Catena di Markov *irriducibile*.

Definizione 3.1 Una catena di Markov per cui esiste un'unica classe di comunicazione è detta catena di Markov *irriducibile*: tutti gli stati comunicano.

1.9.2 Esempi

2. *La camminata casuale è irriducibile.* $S = \{\dots, -1, 0, 1, \dots\}$ Poichè $0 < p < 1$ possiamo sempre raggiungere ogni stato da ogni altro stato passo dopo passo utilizzando il fatto che $P_{i,i-1} = 1-p$. Per esempio $-4 \rightarrow 2$ poichè $P_{-4,2}^6 \geq p^6 > 0$ e $2 \rightarrow -4$ poichè $P_{2,-4}^6 \geq (1-p)^6 > 0$ per cui $-4 \leftrightarrow 2$. In generale $P_{ij}^n > 0$ per $n = |i-j|$.
3. *La camminata casuale dal problema del giocatore in rovina è non riducibile.* Qui la camminata casuale è ristretta allo spazio di stati finiti $\{0, 1, \dots, N\}$ e $P_{00} = P_{NN} = 1$. $C_1 = \{0\}$, $C_2 = \{1, \dots, N-1\}$, $C_3 = \{N\}$ sono classi di comunicazione.
4. Si consideri la C.M. con $S = \{0, 1, 2, 3\}$ e matrice di transizione data da:

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1/3 & 1/6 & 1/6 & 1/3 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Si noti come i soli stati 0,1, nonostante comunichino tra di loro, nessun altro stato è raggiungibile da loro (insieme formano un insieme "assorbente") $C_1 = \{0, 1\}$ è una classe di comunicazione. Nonostante ogni stato è raggiungibile dallo stato 2, non è possibile raggiungerlo da ogni altro stato; un'altra classe è $C_2 = \{2\}$. Infine lo stato 3 è assorbente e otteniamo l'ultima classe $C_3 = \{3\}$. Questo esempio illustra il metodo generale per dedurre le classi di comunicazione analizzando la matrice di transizione.

4.9 - Ricorrenza e Distribuzioni stazionarie

4.9.1 Ricorrenza e Transienza

Sia τ_{ii} il *tempo di ritorno* allo stato i dato $X_0=i$:

$$\tau_{ii} = \min\{n \geq 1: X_n=i | X_0=i\}, \tau_{ii} =_{def} \infty \text{ se } X_n \neq i, n \geq 1.$$

Esso rappresenta il tempo (inteso in numero di passi) fino a che la catena ritorna nello stato i dato che è stata inizializzata nello stato i . Si noti come sia possibile il “non ritorno” nella definizione ponendo $\tau_{ii} = \infty$; per cui un avviene un ritorno solo e solo se $\tau_{ii} < \infty$.

$f_i =_{def} P(\tau_{ii} < \infty)$ è la probabilità di ritorno allo stato i dato che la catena è partita dallo stato i . Uno stato i è detto *ricorrente* se $f_i=1$, *transiente* se $f_i < 1$. Partendo dalla proprietà di Markov, una volta rivisitato lo stato i , il futuro è indipendente dal passato ed è come se la catena ricominciasse sempre dallo stato i per la prima volta. Ogni volta che lo stato i è visitato sarà rivisitato con la stessa probabilità f_i indipendentemente dal passato. In particolare se $f_i=1$ allora la catena ritornerà continuamente nello stato i un numero infinito di volte. Questo è il motivo del termine “ricorrente”. Se lo stato i , invece è *transiente* ($f_i < 1$) allora sarà rivisitato solo un numero finito (casuale) di volte, dopo di cui solo è rimanenti stati $j \neq i$ può essere visitato dalla catena. Contando lungo il tempo il numero totale di visite allo stato i dato che $X_0=i$, esso sarà dato da una sequenza infinita di indicatori:

$$N_i = \sum_{n=0}^{\infty} I\{X_n = i | X_0 = i\} \quad (1)$$

con una distribuzione geometrica:

$$P(N_i = n) = f_i^{n-1} (1 - f_i), n \geq 1.$$

(N.B. contiamo lo stato iniziale $X_0=i$ come prima visita).

Il numero previsto di visite sarà così dato da $E(N_i) = 1/(1 - f_i)$, da cui possiamo concludere che:

Uno stato i è ricorrente ($f_i=1$) se e solo se $E(N_i)=\infty$,

o equivalentemente

Uno stato i è transiente ($f_i<1$) se e solo se $E(N_i)<\infty$.

Considerando che il valore atteso nella (1) vale:

$$E(N_i)=\sum_{n=0}^{\infty} P_{i,i}^n$$

poichè $E(I\{X_n=i|X_0=i\})=P(X_n=i|X_0=i)=P_{i,i}^n$, si ottengono le seguenti proposizioni:

Proposizione 3.2 *Uno stato è ricorrente se e solo se*

$$\sum_{n=0}^{\infty} P_{i,i}^n = \infty$$

altrimenti è transiente.

Proposizione 3.3 *Per ogni classe di comunicazione C , tutti gli stati in C sono o tutti ricorrenti o tutti transienti. Ovvero: se i e j comunicano e i è ricorrente, lo sarà anche j . Equivalentemente se i e j comunicano e i è transiente, lo sarà anche j . In particolare, per una catena di Markov irriducibile, tutti gli stati sono tutti ricorrenti o tutti transienti.*

Dimostrazione: Si supponga che i due stati i e j (diversi tra loro) comunichino; si scelga un n appropriato tale che $p=P_{i,j}^n>0$. Ora se i è ricorrente, lo dovrà essere anche j perchè ogni volta che si visita i c'è la stessa probabilità positiva p (probabilità di "successo") che j sarà visitato n passi dopo. Ma dire che i sia ricorrente significa dire che sarà visitato continuamente, un numero infinito di volte; per cui vedendolo come una sequenza di Bernoulli, concludiamo che con il tempo ci sarà un successo. (Formalmente stiamo utilizzando il teorema di Borel-Cantelli).

Definizione 3.2 *Diremo che una Catena di Markov è ricorrente se tutti i suoi stati sono ricorrenti, altrimenti sarà transiente.*

Il topo nel labirinto chiuso rappresenta una catena di Markov ricorrente; il topo nel labirinto aperto invece è una catena di Markov che non è irriducibile, ci sono due classi di comunicazione $C_1=\{1,2,3,4\}$ che è transiente, $C_2=\{0\}$ che invece è ricorrente. Chiaramente se lo spazio di stato per una data C.M. è finito allora non è possibile che tutti gli stati siano transienti (poichè altrimenti dopo un numero finito di passi, ovvero in un tempo finito, la catena dovrebbe lasciare ogni stato senza mai ritornarci, ma dove andrebbe?). Per cui possiamo concludere:

Proposizione 3.4 *Una catena di Markov irriducibile con uno spazio di stati finito è sempre ricorrente: tutti gli stati sono ricorrenti.*

Infine si osservi (dalla constatazione che se due stati comunicano e uno è ricorrente allora lo sarà anche l'altro) che per una catena irriducibile ricorrente anche se inizializziamo la catena in qualche altro stato $X_0 \neq i$, essa visiterà i un numero infinito di volte: *Per una catena di Markov irriducibile e ricorrente, ogni stato j sarà visitato un numero infinito di volte a prescindere dallo stato iniziale X_0 .*

Per esempio se il topo nel labirinto chiuso cominciasse nella cella 3 ritornerà comunque infinite volte nella cella 1.

4.9.2 Tempo di ritorno ad uno stato: ricorrenza positiva e ricorrenza nulla

Uno stato ricorrente j è detto *ricorrente positivo* se il tempo atteso previsto di ritorno allo stato j (dato che la catena è stata inizializzata in j) ha un momento del primo ordine finito:

$$E(\tau_{jj}) < \infty.$$

Uno stato j per cui $E(\tau_{jj}) = \infty$ è detto *ricorrente nullo*.

In generale anche per $i \neq j$ si definisce $\tau_{ij} =_{\text{def}} \min \{ n \geq 1 : X_n = j \mid X_0 = i \}$, tempo (dopo l'istante 0) prima di raggiungere lo stato j dato che $X_0 = i$.

Proposizione 3.5 *Si supponga che i e j siano entrambi ricorrenti ($i \neq j$). Se i e j comunicano e se i è ricorrente positivo ($E(\tau_{ij}) < \infty$), allora j è ricorrente positivo ($E(\tau_{jj}) < \infty$) e anche $E(\tau_{ij}) < \infty$. In particolare tutti gli stati una classe di comunicazione ricorrente sono o tutti ricorrenti positivi o tutti ricorrenti nulli.*

Dimostrazione: Si assuma che $E(\tau_{ij}) < \infty$ e che i e j comunicano. Si scelga il più piccolo $n \geq 1$ tale che $P_{j,i}^n > 0$. Sia $X_0 = j$ e $A = \{X_k \neq j, 1 \leq k \leq n, X_n = i\}$; $P(A) > 0$. Allora $E(\tau_{jj}) \geq E(\tau_{jj}|A)P(A) = (n + E(\tau_{ij}))P(A)$; per cui $E(\tau_{ij}) < \infty$ (altrimenti $E(\tau_{jj}) = \infty$, una contraddizione).

Sia $X_0 = j$ e sia $\{Y_m : m \geq 1\}$ un insieme di variabili indipendenti ed equamente distribuite; τ_{jj} denoterà il tempo di interarrivo tra le visite allo stato j . Per cui l' n -sima rivisita della catena allo stato j avverrà nell'istante $t_n = Y_1 + \dots + Y_n$ e $E(Y) = E(\tau_{jj}) < \infty$. Sia inoltre

$$p = P(\{X_n\} \text{ visita lo stato } i \text{ prima di ritornare allo stato } j | X_0 = j).$$

$p \geq P(A) > 0$, dove A è come definito sopra. Ogni volta che la catena rivisita lo stato j c'è, indipendentemente dal passato, tale probabilità p che la catena visiterà lo stato i prima di rivisitare lo stato j . Sia N il numero di rivisite che la catena fa allo stato j prima di visitare la prima volta lo stato i , possiamo allora vedere che N ha una distribuzione geometrica con probabilità di "successo" p , per cui $E(N) < \infty$. N è uno "stopping time" rispetto a $\{Y_m\}$ e

$$\tau_{ij} \leq \sum_{m=1}^N Y_m,$$

per cui dalla equazione di Wald $E(\tau_{j,i}) \leq E(N)E(Y) < \infty$. E infine

$$E(\tau_{i,i}) \leq E(\tau_{i,j}) + E(\tau_{j,i}) < \infty.$$

La proposizione 3.3 e la 3.5 insieme portano alla:

Proposizione 3.6 *Tutti gli stati in una classe di comunicazione C sono o tutti ricorrenti positivi, o tutti ricorrenti nulli o transienti. In particolare, per una*

catena di Markov irriducibile tutti i suoi stati devono essere ricorrenti positivi, o ricorrenti nulli o transienti.

Definizione 3.3 Se tutti gli stati in una catena di Markov irriducibile sono ricorrenti positivi, allora diremo che la catena di Markov è ricorrente positiva. Se tutti gli stati in una catena di Markov irriducibile sono ricorrenti nulli, allora diremo che la catena di Markov è ricorrente nulla. Se tutti gli stati in una catena di Markov irriducibile sono transienti, allora diremo che la catena di Markov è transiente.

4.9.3 Distribuzione Stazionaria Limite

Quando esistono i limiti, π_j indicherà la porzione di tempo nel lungo periodo che la catena trascorre nel tempo j :

$$\pi_j = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n I\{X_m = j\} \text{ w.p. } 1 \quad (2)$$

Considerando anche la condizione iniziale $X_0=i$ è più precisamente definito come:

$$\pi_j = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n I\{X_m = j | X_0 = i\} \text{ w.p. } 1 \quad (3)$$

per tutti gli stati iniziali i .

Considerati i valori attesi ($E(I\{X_m=j|X_0=i\}) = P(X_m=j|X_0=i) = P_{ij}^m$) possiamo vedere che se π_j esiste allora può essere calcolato alternativamente (tramite il teorema di convergenza limitata) come:

$$\pi_j = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n P(X_m = j | X_0 = i) =$$

$$= \lim_{n \rightarrow \infty} \sum_{m=1}^n P_{ij}^m \quad (4)$$

per tutti gli stati iniziali i .

Per semplicità di notazione assumiamo d'ora in avanti che lo spazio di stato $S=N=\{0,1,2,\dots\}$ o un qualunque sottoinsieme di N .

Definizione 3.4 *Se per ognuno degli stati $j \in S$, π_j esiste come definito nella (3) ed è indipendente dallo stato iniziale ed inoltre $\sum_{j \in S} \pi_j = 1$, allora la distribuzione di probabilità $\pi = (\pi_0, \pi_1, \dots)$ sullo spazio di stato S è chiamata distribuzione limite o stazionaria o a stati fissi della Catena di Markov.*

Ricordandosi che P_{ij}^m rappresenta la (ij) -esima componente della matrice P^m , possiamo concludere che la (4) può essere espressa in forma matriciale nel seguente modo:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n P^m = \begin{pmatrix} \pi \\ \pi \end{pmatrix} = \begin{pmatrix} \pi_0 & \pi_1 & \dots & \\ \vdots & \ddots & & \vdots \\ \dots & \dots & & \dots \end{pmatrix} \quad (5)$$

Ovvero, quando facciamo la media delle matrici di transizione a m passi, ogni riga converge al vettore delle probabilità stazionarie $\pi = (\pi_0, \pi_1, \dots)$. L' i -esima riga si riferisce alla condizione iniziale $X_0=i$ nella (4), e per ognuna delle righe fissate i il j -esimo elemento delle medie converge a π_j .

Un modo semplice di interpretare π è *il seguente*: se si osserva lo stato di una catena di Markov in un istante casuale nel futuro, allora π_j è la probabilità che lo stato osservato sia proprio j . Per osservarlo: abbia N (il nostro tempo di osservazione randomico) una distribuzione uniforme lungo gli interi $\{1,2,\dots,n\}$ e sia indipendente dalla catena; $P(N=m)=1/n, m \in \{1,2,\dots,n\}$. Ora si assuma che $X_0=i$ e che n sia molto grande, allora ponendo $N=m$ otteniamo:

$$P(X_N = j) = \sum_{m=1}^n P(X_m = j | X_0 = i) P(N = m) =$$

$$= \frac{1}{n} \sum_{m=1}^n P_{i,j}^m \sim \pi_j$$

dove nell'ultima riga è stata utilizzata la (4).

4.9.4 Connessione tra $E(\tau_{jj})$ e π_j

La proposizione seguente è intuitiva e al contempo molto utile:

Proposizione 3.7 *Se $\{X_n\}$ è una catena di Markov ricorrente positiva allora esiste un'unica distribuzione stazionaria π ed è data da:*

$$\pi_j = \frac{1}{E(\tau_{jj})} > 0, \quad \forall j \in S.$$

Se la catena è ricorrente nulla o transiente allora i limiti nella (2) sono tutti 0 wpl; non esistono distribuzioni stazionarie.

Ne consegue che in media la catena visita lo stato j una volta ogni $E(\tau_{jj})$ istanti di tempo, ovvero: $\pi_j = \frac{1}{E(\tau_{jj})}$.

Dimostrazione: per prima cosa otteniamo immediatamente il risultato nel caso transiente poichè per definizione per ogni stato fissato j i ritorni allo stato j costituiscono punti ricorrenti di rigenerazione per la catena; per cui questo risultato è una conseguenza della teoria standard riguardo i processi rigenerativi; ma anche se il lettore non è pienamente consapevole di tale teoria, la dimostrazione è comprensibile. Per gli altri casi: si assuma che $X_0=j$. Siano $t_0=0$, $t_1=\tau_{jj}$, $t_2=\min\{k > t_1 : X_k=j\}$ e in generale $t_{n+1}=\min\{k > t_n : X_k=j\}$, $n \geq 1$. Questi t_n sono gli istanti consecutivi in cui la catena visita lo stato j . Siano $Y_n=t_n-t_{n-1}$ i tempi di interevento, allora rivisiteremo lo stato j per l' n -sima volta al tempo $t_n=Y_1+\dots+Y_n$. Qui l'idea è di "rompere" l'evoluzione della catena di Markov in *cicli* indipendenti e identicamente distribuiti dove un ciclo comincia ogni volta che la catena

visita lo stato j . Y_n è la lunghezza dell' n -simo ciclo. Dalla proprietà di Markov, la catena ricomincia di volta in volta ed è indipendente dal passato ogni volta che entra nello stato j (formalmente questo segue dalla proprietà di Markov in senso forte). Questo significa che le *lunghezze* dei cicli $\{Y_n : n \geq 1\}$ formano una sequenza indipendente ed equamente distribuita con una distribuzione comune a quella della lunghezza del primo ciclo τ_{jj} . In particolare $E(Y_n) = E(\tau_{jj})$ per tutti gli $n \geq 1$.

Ora si osservi che il numero di rivisite allo stato j è precisamente n visite al tempo $t_n = Y_1 + \dots + Y_n$, ovvero la porzione nel lungo periodo di visite allo stato j nell'unità di tempo può essere calcolata come:

$$\begin{aligned} \pi_j &= \lim_{n \rightarrow \infty} \frac{1}{m} \sum_{k=1}^m I\{X_k = j\} = \\ &= \lim_{n \rightarrow \infty} \frac{n}{\sum_{i=1}^n Y_i} = \\ &= \frac{1}{E(\tau_{jj})}, w.p. 1, \end{aligned}$$

dove l'ultima eguaglianza scaturisce dalla Legge dei Grandi Numeri. Per cui nel caso ricorrente positivo, $\pi_j > 0 \forall j \in S$, mentre nel caso ricorrente nullo $\pi_j = 0 \forall j \in S$. Infine, se $X_0 = i \neq j$, allora possiamo prima aspettare fino a quando la catena entra nello stato j (e lo farà certamente data la ricorrenza) e dopo possiamo procedere con la dimostrazione appena fornita. (L'unicità segue dalla rappresentazione unica di $\pi_j = \frac{1}{E(\tau_{jj})}$).

Il risultato appena ottenuto è molto utile nel calcolo di $E(\tau_{jj})$ se è già stato trovato π ; per esempio si consideri il topo nel labirinto chiuso. Dato che il topo comincia nella cella 1, qual è il numero atteso di passi fino a che il topo ritorni nella cella 1? La risposta è semplicemente $1/\pi_1$. Ma come calcoliamo π ? Sarà spiegato in seguito.

4.9.5 Calcolo algebrico di π

Teorema 3.1 *Si supponga che $\{X_n\}$ sia una catena di Markov irriducibile con matrice di transizione P . Allora $\{X_n\}$ è ricorrente positivo se e solo se esiste una soluzione (non-negativa, sommata a 1) $\pi=(\pi_0, \pi_1, \dots)$ al set di equazioni lineari $\pi=\pi P$; in questo caso π è l'unica distribuzione stazionaria per la catena di Markov.*

Per esempio si consideri lo spazio di stato $S=\{0,1\}$ e la matrice

$$P = \begin{pmatrix} 0.5 & 0.5 \\ 0.4 & 0.6 \end{pmatrix},$$

che è chiaramente irriducibile. Per $\pi=(\pi_0, \pi_1)$, $\pi=\pi P$ comporta due equazioni

$$\begin{cases} \pi_0 = 0.5 \pi_0 + 0.4 \pi_1 \\ \pi_1 = 0.5 \pi_0 + 0.6 \pi_1 \end{cases}$$

Possiamo anche utilizzare la condizione di “probabilità” per cui $\pi_0 + \pi_1 = 1$. Risolvendo otteniamo $\pi_0=4/9$ e $\pi_1=5/9$. Possiamo concludere che la catena è ricorrente positiva con una distribuzione stazionaria $(4/9, 5/9)$. Nel lungo periodo la catena visiterà in proporzione lo stato 0 per $4/9$ del tempo e lo stato 1 per $5/9$ del tempo. Inoltre poichè $\pi_j = \frac{1}{E(\tau_{jj})}$, possiamo concludere che il numero atteso di passi fino a che la catena rivisita lo stato 1 è uguale a $9/5$.

Come utilizzare il teorema 3.1. Il teorema può essere utilizzato nel modo seguente: se si ha a che fare con una catena di Markov irriducibile, allora si può provare a risolvere il set di equazioni :

$$\begin{cases} \pi = \pi P \\ \sum_{j \in S} \pi_j = 1 \end{cases}$$

Se lo si risolve per un qualche π , allora questa soluzione π è unica ed è la distribuzione stazionaria; inoltre la catena è ricorrente positiva. Potrebbe

anche non essere necessario risolvere il set di equazioni per ottenere π : si supponga di avere un candidato a essere la distribuzione stazionaria (per esempio per tentativo), allora basterà solo sostituire tale distribuzione e verificare se soddisfa $\pi = \pi P$. Se lo fa, allora il candidato è effettivamente la distribuzione stazionaria e la catena è ricorrente positiva⁶.

Dimostrazione del teorema 3.1: Sia assuma che la catena sia irriducibile e ricorrente positiva. Allora noi sappiamo dalla proposizione 3.7 che π esiste (come definito nelle equazioni (3) e (4)), è rappresentato da $\pi_j = \frac{1}{E(\tau_{jj})} > 0$, $j \in S$, ed è unico. D'altra parte se noi moltiplichiamo (da destra) ogni membro dell'equazione (5) per P , otteniamo:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n P^{m+1} = \begin{pmatrix} \pi \\ \pi \\ \dots \end{pmatrix} P.$$

Ma d'altra parte sappiamo che:

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n P^{m+1} &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n P^m + \lim_{n \rightarrow \infty} \frac{1}{n} (P^{n+1} - P) = \\ &= \begin{pmatrix} \pi \\ \pi \\ \dots \end{pmatrix} + \lim_{n \rightarrow \infty} \frac{1}{n} (P^{n+1} - P) = \quad (\text{dalla (5)}) \\ &= \begin{pmatrix} \pi \\ \pi \\ \dots \end{pmatrix}, \end{aligned}$$

poichè $\forall k \geq 1 \lim_{n \rightarrow \infty} \frac{1}{n} P^k = 0$ (poichè $p_{ij}^k \leq 1 \forall i, j$).

Per cui, per ogni riga otteniamo:

$$\begin{pmatrix} \pi \\ \pi \\ \dots \end{pmatrix} = \begin{pmatrix} \pi \\ \pi \\ \dots \end{pmatrix} P \quad \text{ovvero } \pi = \pi P$$

⁶ $0=(0,0,\dots,0)$ è sempre una soluzione a $\pi = \pi P$, ma non è una probabilità. Inoltre, per ogni soluzione π e per ogni costante c , $c\pi = c\pi P$ per linearità, per cui $c\pi$ è una soluzione.

Riassumendo: se una catena di Markov è ricorrente positiva, allora distribuzione stazionaria π esiste come definita nelle equazioni (3) e (4) ed è data da $\pi_j = \frac{1}{E(\tau_{jj})}$, $j \in S$ e deve soddisfare $\pi = \pi P$. Proviamone anche il viceversa: per una catena di Markov irriducibile si supponga che $\pi = \pi P$ abbia una soluzione π non negativa, di modulo unitario e sia una soluzione di probabilità. Dobbiamo quindi dimostrare che la catena sia ricorrente positiva e che questa soluzione π è una distribuzione stazionaria come definita nella (4). Per prima cosa dimostreremo che la catena non può essere nè transiente nè ricorrente nulla, per cui dovrà essere per forza ricorrente positiva data la proposizione 3.6. A questo scopo si supponga che la catena non sia transiente nè ricorrente nulla. Dalla proposizione 3.7 sappiamo che i limiti nella (4) sono identicamente nulli, ovvero per $n \rightarrow \infty$

$$\frac{1}{n} \sum_{m=1}^n P^m \rightarrow 0 \quad (6)$$

Ma se $\pi = \pi P$, allora (moltiplicando a destra entrambi i membri per P) $\pi = \pi P^2$ e più in generale $\pi = \pi P^m$ $m \geq 1$, per cui:

$$\pi \left[\frac{1}{n} \sum_{m=1}^n P^m \right] = \frac{1}{n} \sum_{m=1}^n \pi P^m = \pi, \quad n \geq 1$$

dalla (6) $\pi_0 = \pi$ o $\pi = (0, \dots, 0)$, che è contraddittorio con il fatto che π è una distribuzione di probabilità. Avendo escluso i casi transiente e ricorrente nullo, dobbiamo concludere dalla proposizione 3.6 che la catena deve essere ricorrente positiva. Per semplicità di notazione, si supponga che $\pi' = \pi' P$ rappresenti la soluzione non negativa di modulo unitario e che π denoti la distribuzione stazionaria della (4) data da $\pi_j = \frac{1}{E(\tau_{jj})}$ $j \in S$; ora dimostreremo che $\pi' = \pi$. A questo scopo, moltiplicando da sinistra entrambi i membri della (5) per π' , otteniamo:

$$\pi' = \pi' \begin{pmatrix} \pi \\ \pi \\ \dots \end{pmatrix} = \pi' \begin{pmatrix} \pi_0 & \pi_1 & \dots \\ \pi_0 & \pi_1 & \dots \\ \dots & \dots & \dots \end{pmatrix}$$

Dato che $\sum_{i \in S} \pi_i' = 1$ giungiamo alla conclusione che $\pi_j' = \pi_j \quad j \in S$, ovvero $\pi' = \pi$, come voleva essere dimostrato.

4.9.6 Caso dello Spazio di Stato Finito

Quando lo spazio di stato di una catena di Markov è finito, allora la teoria è ancora più semplice:

Teorema 3.2 *Ogni catena di Markov irriducibile con uno spazio di stato finito è ricorrente positiva e per questo motivo ha una distribuzione stazionaria (unica soluzione di probabilità a $\pi = \pi P$).*

Dimostrazione: Dalla proposizione 3.4, sappiamo che la catena è ricorrente. Ora mostreremo che non può essere ricorrente nulla, per cui dovrà essere ricorrente positiva a seguire dalla proposizione 3.6. A questo scopo, si noti che per ogni m fissato ≥ 1 , le righe di P^m (matrice di transizione a m passi) devono avere somma unitaria, ovvero:

$$\sum_{j \in S} P_{i,j}^m = 1, \quad i \in S \quad (7)$$

Inoltre se ricorrente nulla, sappiamo dalla proposizione 3.7, che per tutti gli $i \in S$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n P_{i,j}^m = 0, \quad i \in S \quad (8)$$

Sommando la 8 per tutti gli stati j si ottiene:

$$\sum_{j \in S} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n P_{i,j}^m = 0, \quad i \in S$$

Ma poichè lo spazio di stato è finito, possiamo scambiare la sommatoria finita esterna con il limite:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j \in S} \sum_{m=1}^n P_{i,j}^m = 0, \quad i \in S$$

Possiamo quindi intercambiare l'ordine della sommatoria ed utilizzare la (7), per poi ottenere una contraddizione:

$$0 = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j \in S} \sum_{m=1}^n P_{i,j}^m = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n \sum_{j \in S} P_{i,j}^m = 1, \quad i \in S$$

Per cui la catena deve essere per forza ricorrente positiva.

Questo risultato è molto utile. Per esempio, ci dice che il topo nella catena di Markov “labirinto”, quando è chiuso dall'esterno, è ricorrente positivo e abbiamo bisogno solo di risolvere le equazioni $\pi = \pi P$ per calcolare la distribuzione stazionaria.

4.9.7 Stazionarietà di Catene Ricorrenti Positive

L'utilizzo della parola “stazionarietà” qui ha il significato di “ non cambia con il tempo” ed in seguito viene mostrato perchè descrive π .

Per una data distribuzione di probabilità $v = (v_j)$ sullo spazio di stato S , utilizziamo la notazione $X \sim v$ per indicare che X è una v.a. con distribuzione v : $P(X=j) = v_j, j \in S$.

Data una qualunque distribuzione di questo tipo $v = (v_j)$ e una catena di Markov con matrice di transizione P , si noti che se $X_0 \sim v$ allora il vettore vP è la distribuzione di X_1 , ovvero la j -sima componente del vettore vP è $\sum_{i \in S} P_{i,j} v_i$ che precisamente è $P(X_1=j) \forall j \in S$:

$$\begin{aligned}
P(X_1 = j) &= \sum_{i \in S} P(X_1 = j | X_0 = i) P(X_0 = i) = \\
&= \sum_{i \in S} P_{i,j} v_i.
\end{aligned}$$

In altre parole se la distribuzione iniziale della catena è v , allora la distribuzione al passo successivo è data da vP e così via.

Lemma 3.1 *Se $X_0 \sim v$ allora $X_1 \sim vP$, $X_2 \sim vP^2, \dots, X_n \sim vP^n$ con $n \geq 1$.*

Proposizione 3.8 *Per una catena di Markov ricorrente positiva con una distribuzione stazionaria π , se $X_0 \sim v$ allora $X_n \sim \pi$ per tutti $n \geq 0$. In parole: iniziando la catena con una distribuzione stazionaria, la catena permane con la stessa distribuzione; X_n ha la stessa distribuzione $\forall n$. Questo è il significato della stazionarietà ed è per questo che π è detto distribuzione stazionaria per la catena.*

Dimostrazione: Dal teorema 3.1 sappiamo che π soddisfa l'equazione $\pi = \pi P$, da qui moltiplicando da destra entrambi i membri per P otteniamo $\pi = \pi P^2$ e così via fino a $\pi = \pi P^n$ con $n \geq 1$. Per cui dal lemma 3.1 concludiamo che se $X_0 \sim \pi$ allora $X_n \sim \pi$ con $n \geq 1$.

Definizione 3.4 *Un processo stocastico $\{X_n\}$ è detto processo stazionario se $\forall k \geq 0$ il processo $\{X_{k+n} : n \geq 0\}$ ha la stessa e identica distribuzione di $\{X_n\}$. "Stessa distribuzione" in particolare significa che tutte le distribuzioni di ogni dimensione sono le stesse: per ogni intero $l \geq 1$ e ogni intero $0 \leq n_1 < n_2 < \dots < n_l$ la distribuzione congiunta del vettore $(X_{k+n_1}, X_{k+n_2}, \dots, X_{k+n_l})$ ha la stessa e identica distribuzione per ogni k di $(X_{n_1}, X_{n_2}, \dots, X_{n_l})$.*

E' importante comprendere che (ponendo $l=1$ nella definizione di cui sopra) che per un processo stazionario si può affermare che X_n ha la stessa distribuzione per ogni n fissato, in particolare la stessa di X_0 ; ma questa è

solo la distribuzione marginale del processo. La stazionarietà significa molto di più ed in generale è una condizione molto più forte. Significa che se cambiamo l'istante iniziale 0 con un qualunque istante k nel futuro, allora l'intero processo continuerà ad avere ancora la stessa distribuzione che aveva con 0 istante iniziale. Per le catene di Markov, in ogni caso, la stazionarietà si intende solo nel caso delle distribuzioni marginali:

Proposizione 3.9 *Una catena di Markov ricorrente positiva con una distribuzione stazionaria π , se $X_0 \sim \pi$ è un processo stazionario.*

Dimostrazione: Il risultato segue immediatamente dalla proposizione 3.8 data la proprietà di Markov (dato X_k il futuro è indipendente dal passato); l'intera distribuzione dipende solo dalla distribuzione di X_k e dalla matrice di transizione P . Per cui se X_k ha la stessa distribuzione per ogni k allora nel futuro $\{X_{k+n} : n \geq 0\}$ ha la stessa distribuzione per ogni k .

Il risultato appena discusso è abbastanza intuitivo: otteniamo π scegliendo di osservare la catena in un istante nel futuro distante infinitamente. Dopo aver fatto ciò, muovendosi di k istanti di tempo oltre nel futuro, nulla cambia (siamo già nel futuro "infinito" andare oltre è ancora "infinito").

4.9.8 Convergenza a π in senso forte e Aperiodicità

Per una catena ricorrente positiva il significato per cui è stato mostrato che convergesse alla sua distribuzione stazionaria π è nella *media temporale* o "*Cesaro*" (si richiamino alla mente le equazioni (3), (4)). Se si volesse avere la convergenza senza considerarla in media, $\pi_j = \lim_{n \rightarrow \infty} P(X_n = j)$, $j \in S$ (senza considerare la condizione iniziale $X_0 = i$) allora è necessaria alla catena una nuova condizione conosciuta come *aperiodicità*.

Dato che $P(X_n = j | X_0 = i) = P_{ij}^n$, dobbiamo capire quando è verificato che per ogni $j \in S$ $P_{ij}^n \rightarrow \pi_j$ per tutti gli $i \in S$.

Per semplicità di notazione in ciò che segue assumiamo che $S=\{0,1,2,\dots\}$ o un suo sottoinsieme. E' facile vedere che per ogni distribuzione di probabilità ν , la matrice

$$M = \begin{pmatrix} \pi \\ \pi \\ \dots \end{pmatrix} = \begin{pmatrix} \pi_0 & \pi_1 & \dots \\ \pi_0 & \pi_1 & \dots \\ \dots & \dots & \dots \end{pmatrix},$$

soddisfa l'equazione $\nu M = \pi$. Per cui dal Lemma 3.1 possiamo vedere che allo scopo di ottenere $\pi_j = \lim_{n \rightarrow \infty} P(X_n = j)$, $j \in S$ a prescindere dalle condizioni iniziali, equivalentemente necessitiamo che $P^n \rightarrow M$ per $n \rightarrow \infty$. Già sappiamo che le medie convergono ($\frac{1}{n} \sum_{m=1}^n P^m \rightarrow M$) ed è facile da vedere che nel caso generale la convergenza forte non è verificata ($P^n \rightarrow M$). Per esempio prendiamo $S=\{0,1\}$ con

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Se $X_0 = 0$ allora $\{X_n : n \geq 0\} = \{0,1,0,1,\dots\}$; se $X_0=1$ allora $\{X_n : n \geq 0\} = \{1,0,1,0,\dots\}$. Chiaramente se $X_0=0$ allora $P(X_n=0)=1$ se n è pari ma è $P(X_n=0)=0$ se n è dispari; $P(X_n=0)$ non converge per $n \rightarrow \infty$. In termini di P si può vedere notando che per ogni $n \geq 1$

$$P^{2n} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Mentre

$$P^{2n+1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Per cui, sebbene $\pi = (1/2, 1/2)$ e in realtà $\frac{1}{n} \sum_{m=1}^n P^m \rightarrow M = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$,

non è verificato che $P^n \rightarrow M$.

La condizione extra (di aperiodicità) è spiegata in seguito.

Per uno stato $j \in S$, si consideri l'insieme $Q = \{ n \geq 1 : P_{i,j}^n > 0 \}$. Se esiste un intero $d \geq 2$ tale che $Q = \{ kd : k \geq 1 \}$, allora lo stato j è detto *periodico di periodo d* . Questo implica che $P_{i,j}^n = 0$ nonostante n non sia un multiplo di d . Non esiste tale d , allora lo stato j è *aperiodico*. Può essere mostrato che la periodicità ha la proprietà di essere una classe: tutti gli stati in una classe di comunicazione sono o tutti aperiodici o sono tutti periodici di stesso periodo d . Per cui una catena di Markov si potrà dire periodica o aperiodica. Nel caso del precedente esempio con due stati, $d=2$; la catena è periodica di periodo $d=2$. Chiaramente, se la catena è periodica, allora scegliendo una qualunque sottosequenza di interi $n_m \rightarrow \infty$ per $m \rightarrow \infty$ tale che $n_m \in Q$, otteniamo che per ogni stato j che $P(X_{n_m} = j | X_0 = j) = P_{i,j}^{n_m} = 0$, per cui la convergenza a $\pi_j > 0$ non è possibile lungo la sottosequenza; di qui la convergenza di $P(X_n = j)$ a π_j non regge. L'inverso è vero; la seguente proposizione sarà fornita senza dimostrazione:

Proposizione 3.10 *Una catena di Markov ricorrente positiva converge a π ($\pi_j = \lim_{n \rightarrow \infty} P(X_n = j)$, $j \in S$) se e solo se la catena è aperiodica.*

Nota 1 Per ragioni storiche, in letteratura una catena di Markov ricorrente positiva e aperiodica è a volte detta catena *ergodica*. La parola ergodica ha un preciso significato in matematica e questo termine non ha nulla a che fare con l'aperiodicità! Infatti ogni catena di Markov ricorrente positiva è ergodica in questo preciso senso matematico. (Per cui storicamente l'uso del termine ergodico nel contesto di catene di Markov aperiodiche è erroneo e fuorviante).

Nota 2 Il tipo di convergenza $\pi_j = \lim_{n \rightarrow \infty} P(X_n = j)$, $j \in S$ è detto *convergenza debole*, per cui di conseguenza si dirà che X_n converge debolmente a π .

CAPITOLO 4

Modellazione e logiche di simulazione

4.1 - La simulazione dei sistemi

Per poter parlare di simulazione è necessario prima definire 2 concetti.

Si definisce sistema un insieme di elementi interagenti che funzionano insieme per perseguire un determinato obiettivo, generando quindi un comportamento globale.

Per modello si intende, invece, la rappresentazione semplificata del sistema reale. Esso quindi aiuta nella comprensione del sistema stesso.

La simulazione permette di anticipare, attraverso la costruzione del modello, il comportamento del sistema reale. Inoltre permette di valutare la risposta del sistema in specifiche circostanze.

La simulazione è uno strumento sperimentale molto potente che si avvale della possibilità di calcolo informatico: trattasi di una trasposizione in termini logico-matematico-procedurali di un "modello concettuale" della realtà.

Può essere definito anche come l'insieme di processi che hanno luogo nel sistema valutato e il cui insieme permette di comprendere le logiche di funzionamento del sistema stesso.

Diventa, quindi, fondamentale ricorrere ad una metodologia di analisi in grado di esplicitare la complessità del sistema, rendendola più chiara e trasparente.

La dinamica di un sistema si definisce come la successione dei suoi stati nel tempo, dove per stato del sistema si intende un insieme di grandezze misurabili e considerate rilevanti per l'osservatore.

I principali approcci simulativi sono:

- Modellizzazione discreta in cui il tempo avanza quando nel sistema si verifica un evento ed è denominata Discret Event Simulation;
- Modellizzazione continua in cui il tempo avanza in modo continuo. Tale modellizzazione si basa sulla System Dynamics e permette di evidenziare le interazioni esistenti tra fattori chiave. Lo stato del sistema, in questo caso, sarà esprimibile attraverso equazioni differenziali.

4.1.1 - I vantaggi della simulazione

Negli ultimi anni è aumentata notevolmente l'attenzione nei confronti delle tecniche di simulazione, riconosciute quali strumenti indispensabili per analizzare contesti aziendali altamente complessi e dinamici e risolvere problematiche aziendali.

Queste tecniche rappresentano un supporto per i manager, facilitandoli non solo nella comprensione dei contesti aziendali, ma anche, nell'assunzione di decisioni critiche per la sopravvivenza dell'azienda. Le tecniche di simulazione sono utilizzate, quindi, per analizzare il comportamento di sistemi dinamici, per testare criteri di gestione, per confrontare soluzioni alternative, per validare scelte progettuali consentendo, quindi, di valutare gli effetti di un insieme di decisioni, senza doverle necessariamente implementare nella realtà.

La simulazione fornisce ai decisori aziendali, manager ed analisti, uno strumento per comprendere la correlazione esistente tra le cause e gli effetti delle politiche gestionali adottate e studiare l'evoluzione dinamica di tutti i fenomeni aziendali. Nei moderni impianti di produzione, sta acquisendo una rilevanza strategica quale supporto alle decisioni strategiche aziendali, facilitando il raggiungimento degli obiettivi aziendali.

4.2 – Discret Event Simulation

La Simulazione ad eventi discreti -Discret Event Simulation (DES)- è fondata su un meccanismo di ordinamento dinamico degli eventi nel tempo.

Il sistema evolve attraverso una successione di salti temporali, in corrispondenza dei quali si verifica un evento che modifica lo stato del sistema. Taluni eventi sono programmati all'inizio della simulazione, altri sono generati durante l'esecuzione. L'esperimento simulato consiste nella riproduzione sequenziale dei cambiamenti di stato.

Una simulazione di processo può essere descritta graficamente attraverso il ricorso alle Petri Nets (Reti di Petri), cioè un grafo bipartito, i cui nodi sono i posti e le transizioni sono le relazioni. Sfruttando tale rappresentazione grafica è possibile costruire modelli con un ridotto uso della programmazione.

Le Reti Di Petri sono un potente strumento grafico e matematico per la modellazione, l'analisi e la simulazione di un'ampia varietà di sistemi. Dal punto di vista grafico, è un grafo composto da due tipologie di nodi: i posti e le transizioni. I posti sono rappresentati da cerchi e le transizioni da linee rette. Per collegare posti e transizioni si utilizzano gli archi.

I nodi dello stesso tipo non sono mai collegati direttamente ad un arco; i posti possono essere collegati direttamente solo alle transizioni, e le transizioni solo ai posti. Se un posto è connesso ad una transizione con un arco che conduce dal posto alla transizione, il posto viene definito come di ingresso della transizione; mentre un posto di uscita di una transizione è collegato alla transizione con un arco che conduce dalla transizione al posto.

I posti e le transizioni possono contenere zero o più simboli circolari denominati token (detti anche marche o gettoni). Durante l'esecuzione di una rete di Petri, i token possono viaggiare da un nodo all'altro lungo gli archi diretti che li collegano. Le transizioni consumano token che subiscono un

flusso in entrata, e creano token che subiscono un flusso in uscita. La distribuzione dei token nei nodi di una rete di Petri è definita marcatura, e rappresenta lo stato corrente della rete di Petri. La primissima marcatura di una rete di Petri, prima dell'esecuzione, è definita marcatura iniziale.

Due regole governano l'esecuzione di una rete di Petri. La regola dell'abilitazione determina se e quando una transizione è abilitata, cioè una transizione è abilitata se e solo se, ognuno dei punti di ingresso contiene almeno un token per ogni arco che collega il posto di ingresso alla transizione, e sta ad indicare la possibilità che un evento accada. Poi vi è la regola dello scatto che determina il modo in cui i token sono rimossi dai posti di ingresso e aggiunti ai posti di uscita: quando una transizione è abilitata, può scattare.

Le Reti di Petri sono predominanti rispetto agli altri modelli ad eventi discreti, per una serie di motivi:

- Forniscono un formalismo grafico e matematico per la modellazione dei DES;
- Permettono di dare una rappresentazione compatta in termini di spazio di stato (rappresentano DES con un numero finito di stati, mediante un grafo con un numero finito di nodi);
- Permettono di rappresentare attività che possono venire svolte parallelamente;
- Consentano una rappresentazione modulare; cioè se un sistema è composto da più sottosistemi che interagiscono tra loro, è generalmente possibile rappresentare ciascun sottosistema come una sottorete e poi, mediante operatori di rete, unire le varie sotto-reti per ottenere il modello del sistema complessivo.

4.3 - System Dynamics

La System Dynamics (SD) è una metodologia che si basa sull'elaborazione di modelli matematici simulati tramite supporti informatici. La prospettiva di analisi è orientata all'identificazione, studio e comprensione dei sistemi caratterizzati da complesse relazioni di retroazione, anche dette relazioni di feedback, tra le differenti variabili, coniugando in un unico approccio la ricostruzione formale della struttura causale di un determinato sistema e la riproduzione e l'osservazione del comportamento dinamico di quella stessa struttura, generato attraverso la simulazione su computer. Conoscere la struttura causale, di conseguenza, risulta presupposto indispensabile alla vera e piena comprensione degli andamenti osservati all'interno del sistema indagato ed irrinunciabile supporto cognitivo alla razionale assunzione di decisioni volte alla gestione delle risorse presenti in tale sistema.

Un circuito di retroazione esiste ogni volta che lo stato di un sistema stimola una decisione che risulta in un'azione e quest'ultima determina un cambiamento dello stato originale del sistema creando, così, le premesse per le decisioni future.

L'analisi condotta nell'ambito della SD, quindi, si ancora a due ipotesi fondamentali. Da una parte, la SD postula che, dall'intreccio di processi decisionali, flussi informativi e relazioni interpersonali all'interno delle aziende, emergano strutture costituite da circuiti di retroazione concatenati. In secondo luogo, la SD ipotizza che i comportamenti dei sistemi siano la conseguenza delle caratteristiche strutturali che assumono tali aggregazioni di circuiti di retroazione, che regolano il "tasso" (rate) di accumulazione o erosione delle variabili livello (stock) in essi compresi.

Il metodo di ricerca System Dynamics ha una logica estremamente pragmatica. L'enfasi non è tanto sulla capacità del modello di prevedere particolari stati puntuali del sistema o sul rigore con cui le ipotesi del modello sono state testate empiricamente, quanto sulla possibilità che il modello offre

di comprendere la logica con cui le variabili rilevanti interagiscono, il ruolo che ciascuna di essa gioca, i punti in cui il sistema è sensibile agli interventi e gli scenari che emergono come conseguenza di ipotesi alternative circa lo stato iniziale del sistema.

La metodologia prevede una serie di fasi che possono essere riassunte nei seguenti punti:

- Identificazione di un problema;
- Sviluppo di ipotesi dinamiche per spiegare la causa del problema;
- Creazione della struttura di base di un diagramma causale;
- Arricchimento del diagramma causale con informazioni aggiuntive;
- Traduzione del diagramma causale in un modello SD.

Gli aspetti critici nell'applicazione della metodologia SD sono l'esplicazione dei legami esistenti tra le diverse variabili del sistema e il ragionare in termini di relazione causa-effetto tra tutti gli aspetti del problema.

E' necessario specificare anche come sono collegate le due variabili, ovvero definire la polarità. La polarità indica il tipo di legame causale che si realizza tra le due variabili e viene rappresentata con un segno "+" oppure un segno "-".

Il segno "+", a volte sostituito dalla lettera "S" (same), è indice di una causalità diretta tra le due variabili; le due variabili si muoveranno nella stessa direzione, ossia se il valore della prima variabile aumenta, il valore della variabile ad essa collegata assumerà un valore al di sopra di quello assunto. Nello stesso modo, una diminuzione del valore della prima variabile determinerà un decremento del valore della variabile ad essa connessa.

Il segno "-", a volte sostituito dalla lettera "O" (opposite), è indice di una causalità inversa tra le due variabili; le due variabili si muoveranno in direzioni opposte ossia se il valore della prima variabile aumenta, il valore

della variabile ad essa collegata, assumerà un valore al di sotto di quello assunto.

La polarità ci permette di capire cosa accade alla variabile dipendente in seguito ad un cambiamento della variabile indipendente.

4.3.1 - Gli strumenti della System Dynamics

La System Dynamics mette a disposizione dell'analista un insieme di strumenti tra i quali si ricordano:

- *diagrammi causal loop* (DCL), detti anche diagrammi ad anelli causali;
- *diagrammi di flusso* (DF), detti anche diagrammi livelli-flussi;
- *grafici temporali* delle singole variabili o *grafici di dipendenza temporale* fra variabili, indicati genericamente come GDT

In prima approssimazione si può affermare che:

- i *DCL* sono semplici da tracciare ma sono difficili da tradurre direttamente in equazioni descrittive del modello; i DF sono più complicati da definire, richiedono una caratterizzazione delle variabili in tipi ma sono più facilmente traducibili in equazioni utilizzabili in una simulazione;
- i *GDT* determinano legami fra le variabili e variazioni di variabili nel tempo e sono difficili da tradurre in una delle due forme suddette sebbene la traduzione non presenti la stessa difficoltà nei due casi dato che, noti gli andamenti delle dipendenze fra un gruppo di variabili, è possibile ricavare un DCL fra tali variabili mentre la traduzione in un DF richiede conoscenze ulteriori che permettano la tipizzazione delle variabili;

- il passaggio da un DCL al corrispondente DF è impossibile se non sono noti le tipologie delle variabili;
- il passaggio da un DF al corrispondente DCL è sempre possibile, sebbene non sia sempre possibile ricavare dai legami fra le variabili espressi in un DF, il tipo di legame (di proporzionalità diretta o inversa) fra queste, a meno di non esaminare l'equazione che ne caratterizza il legame, per poterlo riprodurre, in modi che vedremo, in un DCL.

I causal loop diagrams -CLD- sono delle rappresentazioni (mappe) costituite da un insieme di variabili, tra loro interconnesse da frecce, che segnalano l'influenza casuale tra le variabili stesse.

I CLD riescono a fornire una prima interpretazione grafica del problema; favoriscono la comunicazione e la condivisione delle conoscenze tra i vari attori coinvolti nello studio; forniscono una rappresentazione semplice e schematica delle ipotesi elaborate, in relazione alle cause responsabili degli andamenti problematici, illustrando chiaramente i meccanismi di feedback.

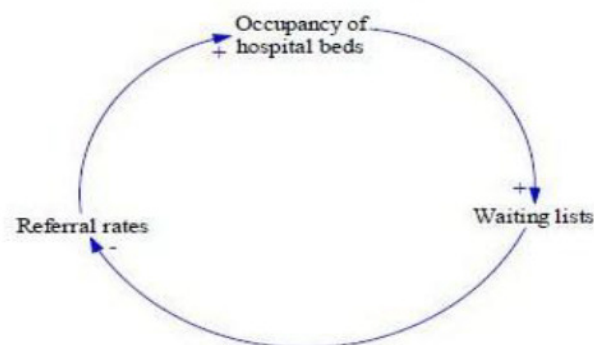
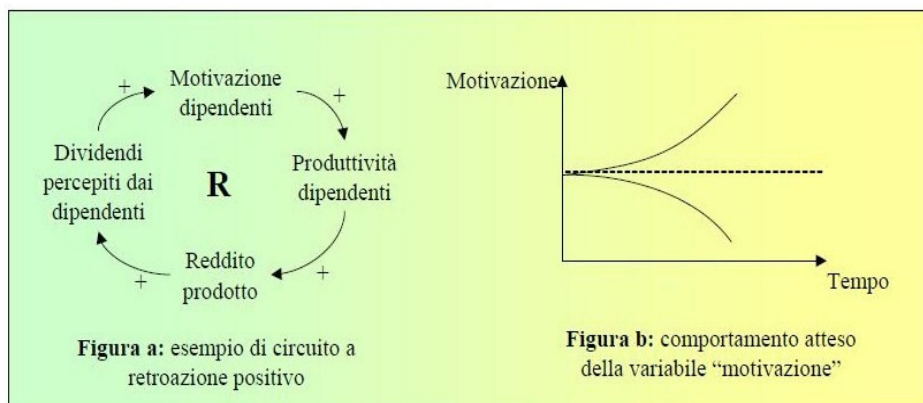


Figura 11: Esempio di causal loop diagram

Un circuito di retroazione può essere definito come una catena di due o più variabili che si influenzano reciprocamente. Ciascuna variabile compresa in un determinato circuito di retroazione può far parte di più circuiti contemporaneamente; tramite tali variabili, che svolgono un ruolo di connessione tra più circuiti, si formano strutture sistemiche più o meno complesse.

Vi possono essere due tipi di circuiti a retroazione:

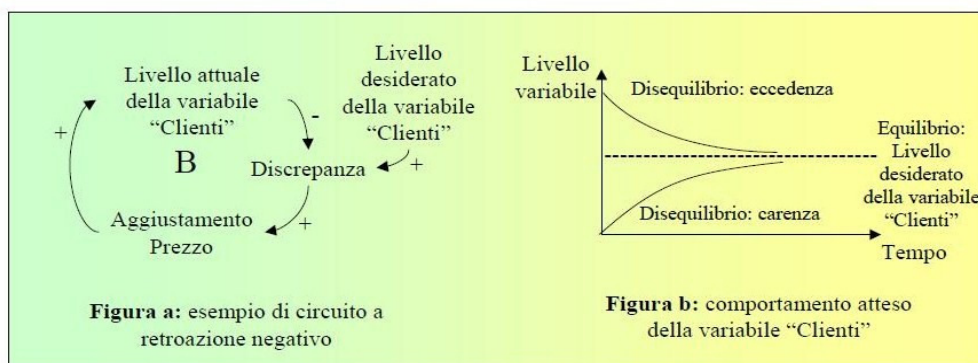
Circuiti a retroazione positiva: tendono ad amplificare in modo esponenziale gli stimoli esterni. Si parla di retroazione positiva quando si è in presenza di variabili legate da una casualità diretta. Nel caso in cui un sistema si trovi in equilibrio, gli stimoli che disturbano tale situazione di equilibrio saranno amplificati e il comportamento dinamico del sistema si discosterà, in modo esponenziale, dallo stato di equilibrio iniziale. Il circuito di retroazione positivo dà origine, quindi, ad un processo di auto rinforzo e viene indicato con la lettera “R” (Rinforzo) e si avrà o una crescita o una decrescita esponenziale.



(Fonte: adattamento da Mollona, 2000)

Figura 12: Esempio di circuito a retroazione positivo

- Circuiti a retroazione negativa: tendono, invece, ad attutire eventuali stimoli esterni. Nel caso in cui un sistema si trovi in equilibrio, gli stimoli che disturbano tale situazione di equilibrio saranno assorbiti e il comportamento dinamico del sistema sarà riportato allo stato di equilibrio iniziale. Il circuito di retroazione negativa dà origine, quindi, ad un processo di autoregolazione o auto bilanciamento e viene indicato con la lettera “B” (Bilanciamento).



(Fonte: adattamento da Mollona, 2000)

Figura 13: Esempio di circuito a retroazione negativo

Il concetto di struttura sistemica può, quindi, essere descritto come una serie di circuiti di causa-effetto concatenati che collegano un insieme di risorse ad un insieme di attività di raccolta di informazioni, decisioni, azioni.

Da un punto di vista dinamico, un circuito di feedback positivo può dare origine unicamente ad una crescita o decrescita di tipo esponenziale, come è possibile vedere in Figura 23.

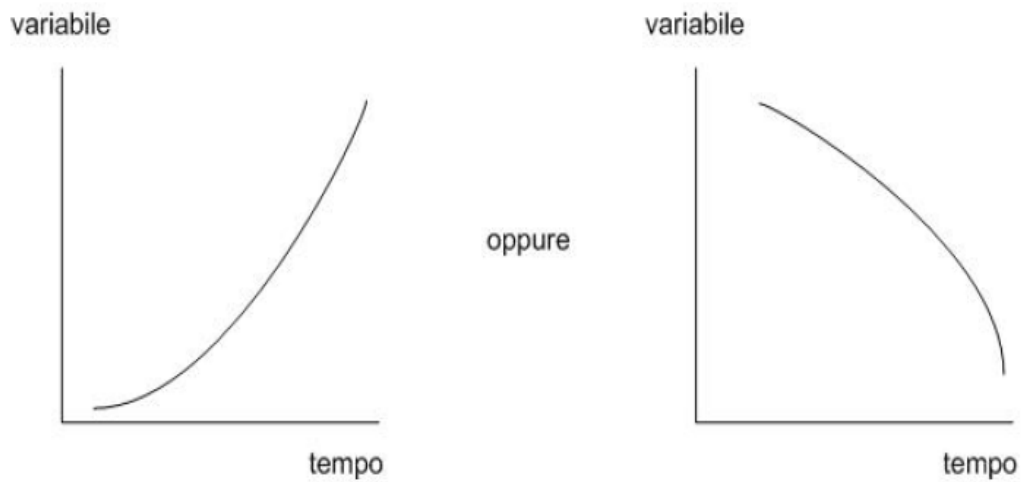


Figura 14: Comportamento dinamico caratterizzato da oscillazioni

Fonte: *Controllo di gestione*, Busca C., Giovannoni E., Riccarboni A., Milano, 2009

Nel caso di un feedback negativo, invece, ci si pone l'obiettivo di attuare azioni correttive che portino il gap riscontrato al livello desiderato.

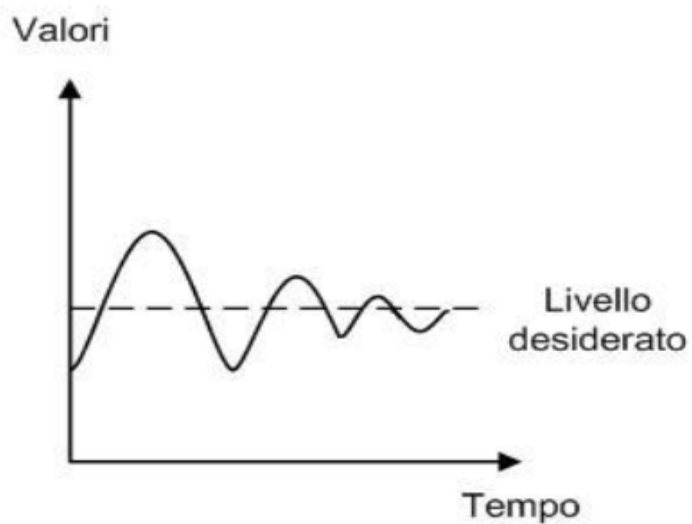


Figura 15: Comportamento dinamico caratterizzato da oscillazioni

Fonte: *Controllo di gestione*, Busca C., Giovannoni E., Riccarboni A., Milano, 2009

Si può affermare che la System Dynamics ha due aspetti distinti: uno qualitativo e uno quantitativo. L'aspetto qualitativo prevede la costruzione di un "Causal Loop Diagramm" (CLD) o "Diagramma di Influenza", che rappresenta, come già detto, graficamente la modalità in cui sono connessi gli elementi del sistema; in particolare un CLD consente di esplicitare e formalizzare le relazioni causali che legano un insieme di variabili, appartenenti allo stesso sistema di riferimento, identificando in questo modo quali meccanismi di feedback siano attivi all'interno di tale sistema e ne producano, conseguentemente, le relative dinamiche.

In pratica i CLD possono interpretare due ruoli importanti nella dinamica del sistema:

- fungono da schizzi preliminari d'ipotesi causali durante lo sviluppo del modello;
- sono in grado di semplificare la rappresentazione di un modello.

La modellazione quantitativa SD consiste, invece, nel convertire il CLD in un diagramma di *stock-flow*. I quattro elementi fondamentali della struttura di un modello System Dynamics sono:

- 1) Un insieme di *variabili livello (stock)*, rappresentate come contenitori, che descrivono lo stato di un sistema come accumulazione di azioni passate (ma costituiscono anche le risorse fondamentali sulla base delle quali sono possibili azioni future);
- 2) Un insieme di *variabili flusso (flow)*, rappresentate come valvole che, raccogliendo le informazioni che scaturiscono dalle variabili livello, contengono le indicazioni per cambiare lo stato di queste ultime;
- 3) Un tessuto di canali che trasportano le informazioni dalle variabili livello alle variabili flusso;

- 4) Un insieme di decisioni o di funzioni di decisioni che descrivono il modo in cui le informazioni, circa lo stato delle variabili livello, vengono utilizzate per azionare le variabili di flusso.

Immaginando di bloccare lo scorrere del tempo e di congelare ogni azione in un dato istante, si potrebbero osservare solo variabili di livello, perché le variabili di flusso, che sono legate all'esistenza di un'azione in corso, scomparirebbero; tali variabili possono, infatti, essere descritte solo con riferimento ad un determinato intervallo temporale. In termini matematici le variabili di flusso sono le derivate rispetto al tempo delle variabili di livello a cui sono collegate, mentre le variabili di livello sono gli integrali delle variabili di flusso. Il legame che unisce questi due tipi di variabili è il processo di integrazione o di differenziazione.

Infine il modello può essere eseguito o “simulato” e mostra come le variabili cambiano nel tempo, permettendo di monitorare e analizzare il loro comportamento.

4.4 – Confronto: SD vs DES

Le tecniche di modellazione utilizzate da SD e DES sono differenti, sia per le modalità di modellazione dei sistemi, sia per l'utilizzo di algoritmi alla base dei simulatori. Tradizionalmente la distinzione più importante è relativa allo scopo della modellazione. L'approccio DES consiste nel trovare il modello di dimensionamento delle risorse, di cui ha bisogno, il decisore come ad esempio il numero di camion, o come organizzare le risorse per evitare colli di bottiglia, o eccessive linee d'attesa, tempi di attesa elevati, ecc. Al contrario l'approccio SD consiste nel definire un processo decisionale, ad

esempio, come rispondere tempestivamente ad ogni cambiamento. In altre parole è caratterizzato da:

- Ricerca di soluzioni utili ai problemi reali, specialmente nei sistemi sociali (imprese, scuole, governi);
- Utilizzo di modelli di simulazione al computer per capire e migliorare tali sistemi;
- Basare i modelli di simulazione su modelli mentali, su conoscenze qualitative e informazioni numeriche;
- Ricerca di modi migliori per tradurre i risultati scientifici raggiunti nel miglioramento attuato.

Inoltre la SD è un approccio più adatto all'analisi strategica: aiuta a capire la direzione e la grandezza degli effetti (ad esempio dove il sistema ha bisogno di apportare le modifiche), è continua nel tempo e utilizza per lo più analisi deterministica. L'approccio DES, invece, tende a guardare i sottosistemi per un'analisi dettagliata ed è più adatto ad una re-ingegnerizzazione dei problemi: è un approccio predittivo (cioè ci dice di quante risorse abbiamo bisogno per raggiungere un certo obiettivo); è basato su un orizzonte temporale specifico ed utilizza un'analisi stocastica.

Nel seguente elaborato, per costruire il modello di simulazione, si preferisce utilizzare la System Dynamics poiché quest'ultima consente di trattare, mediante appositi costrutti, contemporaneamente, nello stesso laboratorio virtuale, sia fenomeni discreti che fenomeni continui.

4.5 - Elaborazione del modello dinamico di simulazione

Un sistema è un complesso costituito di più elementi interconnessi, in cui si possono distinguere grandezze soggette a variare nel tempo (variabili) che ne definiscono lo stato. Solitamente l'evoluzione di alcune variabili dipende da altre variabili, per cui si parla di ingressi (cause) ed uscite (effetti).

Un sistema dinamico è formato da tre elementi:

- Ingressi
- Stato
- Uscite

Gli ingressi rappresentano le variabili esogene che influenzano la dinamica del sistema, lo stato è una variabile interna, attraverso la quale è possibile conoscere la condizione in cui si trova il sistema in un dato istante, e le uscite sono determinate dagli ingressi e dallo stato.

La costruzione di un modello di simulazione del sistema è fondamentale per la risoluzione di un problema. I modelli si distinguono in statici o dinamici, lineari o non lineari, tempo continuo o discreto, a parametri concentrati o distribuiti, a seconda delle caratteristiche del sistema che modellizzano. I modelli di simulazione utilizzati nell'ambito della System Dynamics sono modelli dinamici. Qualora l'evoluzione del sistema avvenga in modo continuo, il modello del sistema viene descritto da un insieme di equazioni differenziali, al contrario, se l'evoluzione avviene tramite una successione di istanti di tempo discreti, il modello viene descritto da equazioni alle differenze, in cui lo stato al tempo $t + 1$ dipende dallo stato al tempo t .

Il vantaggio principale della simulazione è che gli esperimenti possono essere completamente controllati e le prestazioni di un sistema tutte osservate. La simulazione è uno strumento che aiuta a risolvere problemi, aiutando a prendere la "decisione giusta" evitando inutili e costosi tentativi di

realizzazione. Per la modellizzazione di un sistema dinamico, risulta indispensabile definire due tipologie di variabili: le variabili stock e le variabili di flusso, che costituiscono le strutture di retroazione (feedback structures).

Un modo semplice per distinguere stock e flussi è quello di chiedersi cosa accadrebbe se potessimo fermare il tempo e osservare il processo. Se la quantità non può essere misurata è un flusso, al contrario, si tratta di uno stock. Si può pensare agli stock come serbatoi contenenti del liquido ed ai flussi come valvole che controllano la velocità di scorrimento tra i serbatoi. Gli stock rappresentano lo stato del sistema, costituiscono le variabili sulle quali è possibile agire. I flussi, invece, incorporano le politiche gestionali dell'azienda, rappresentando l'unica modalità con la quale è possibile intervenire per modificare gli stock. Mentre gli stock sono delle accumulazioni nel tempo, i flussi rappresentano attività, movimenti che modificano gli stock nel corso del tempo. Per realizzare un corretto modello dinamico è necessario attribuire valori iniziali a ciascuno stock e definire le equazioni per ciascun flusso, equazioni che traducono, in termini matematici, le politiche da adottare in azienda per prendere delle decisioni.

4.5.1 - Operazioni di codifica necessarie alla costruzione di un modello

Le operazioni di codifica necessarie per costruire un modello vanno inquadrare in un processo di studio del sistema che segue diversi passi.

Il processo di studio del sistema si può suddividere nelle seguenti fasi:

- *Formulazione del problema*: viene definito il sistema da studiare e i suoi obiettivi.

- *Raccolta dati*: si raccolgono osservazioni e dati quantitativi sul sistema.
- *Costruzione del modello*: si formulano delle ipotesi e si costruisce il modello.
- *Validazione*: è necessario confrontare le caratteristiche del modello con quelle del sistema tenendo conto di dati sempre più approfonditi per far aumentare la credibilità del modello.
- *Costruzione del programma*: si scrive il programma in un linguaggio generico o di simulazione. L'impiego di un simulatore user-friendly può semplificare molto questa attività rispetto all'utilizzo di un linguaggio generico di programmazione.
- *Validazione*: si effettuano dei pilot run per vedere se gli output del modello si avvicinano agli output di un sistema già esistente avente le stesse caratteristiche del modello. Se gli output non sono in accordo allora si modificano le ipotesi e il modello stesso.
- *Definizione fase sperimentale*: si devono stabilire i valori ottimali del tempo di simulazione, delle condizioni iniziali, del warm-up period, e il numero di run di simulazione necessari.
- *Fase sperimentale*: si effettuano i run di simulazione.
- *Analisi dei risultati*: vengono elaborati i risultati utilizzando le tecniche della statica per costruire intervalli di confidenza per le misure delle prestazioni del sistema.
- *Presentazione dei risultati e loro implementazione*.

L'implementazione di un modello di simulazione discreta su di un elaboratore elettronico richiede l'impiego di un linguaggio di programmazione basato su particolari sintassi e di specifiche conoscenze di logica. I linguaggi disponibili possono essere classificati in tre categorie:

1. linguaggi generali (o general purpose);

2. estensione di linguaggi generali orientati alla simulazione;
3. linguaggi di simulazione.

Alla prima categoria appartengono quei linguaggi di programmazione ad alto livello quali Pascal, C, Fortran, etc., che non essendo dedicati specificatamente ad un'attività di simulazione, richiedono da parte del programmatore uno sforzo maggiore per l'implementazione di tutti i meccanismi e le strutture di dati necessari in un simulatore.

D'altra parte, proprio perché si tratta di linguaggi generali, e quindi molto conosciuti, essi non creano eccessivi problemi a livello di utilizzo. La seconda categoria è costituita da quei linguaggi ottenuti aggiungendo ai linguaggi generali esistenti, funzioni e costrutti tipici della simulazione sotto la forma di librerie di sottoprogrammi o procedure richiamabili dal programma di simulazione scritto in un linguaggio general purpose.

Esistono ormai sul mercato diversi linguaggi di simulazione in grado di facilitare il compito del programmatore, che con i suddetti linguaggi è in grado di scrivere con poche righe di programma interi modelli o sottomodelli che altrimenti richiederebbero tempi di stesura assai più lunghi con conseguente aumento della probabilità di errore.

A questi vantaggi, tuttavia, si possono contrapporre alcuni inconvenienti quali la minore efficienza (cosicché di solito il tempo di elaborazione aumenta) oppure la scarsa diffusione del linguaggio stesso dovuta al fatto che soltanto un numero ristretto di persone è in grado di padroneggiare efficacemente questo tipo di linguaggio.

Con riferimento all'evoluzione dinamica del sistema è possibile distinguere tra tre diverse metodologie di simulazione; si parla quindi di:

1. simulazione orientata agli eventi (*event-based*);
2. simulazione orientata ai processi (*process-based*);

3. simulazione orientata alle attività (*activity-based*).

La simulazione orientata agli eventi, tipica dei linguaggi di programmazione definiti in precedenza come generali, considera una serie di eventi che possono verificarsi nel corso dell'evoluzione del sistema; ciascun evento è descritto da una subroutine che può modificare o meno lo stato di tutti i componenti del sistema.

Nell'approccio orientato ai processi il simulatore è costituito da un insieme di procedure che descrivono tutto ciò che succede ad un'entità che progressivamente attraversa il sistema; il modello è pertanto descritto da un diagramma a blocchi, ognuno dei quali rappresenta un processo, collegati tra di loro da archi orientati, lungo cui fluiscono le entità.

In ogni linguaggio di simulazione orientato ai processi è presente un insieme di macro-istruzioni che traducono automaticamente un certo numero di situazioni che si presentano più frequentemente durante un modello di simulazione. L'approccio process-based sembra essere particolarmente adatto per la descrizione dei processi produttivi, ma può comportare complicazioni quando le operazioni coinvolgono diversi tipi di risorse. L'approccio orientato alle attività richiede la definizione delle condizioni necessarie per l'inizio e il termine di ciascuna attività prevista dal modello. Questo presuppone l'esistenza di un modulo del programma per definire ogni attività che impegna le entità; il modulo stesso include un test per determinare se l'attività può essere iniziata e se le azioni sono eseguibili. Questo tipo di approccio è il più facile per un non-specialista ed è indubbiamente il più semplice dal punto di vista della programmazione.

I linguaggi di simulazione basati sugli ultimi due approcci consentono di rappresentare il sistema in esame descrivendo la sequenza di processi (o attività) che ciascuna entità attraversa in sequenza dall'istante di ingresso nel sistema fino all'istante di uscita.

Ne consegue una notevole semplificazione nel lavoro di programmazione. Il sistema in esame può cioè essere schematizzato come una rete di nodi o blocchi costituiti da macchine, code e punti di diramazione, attraverso le quali transitano le entità seguendo percorsi differenziati a seconda degli specifici attributi assegnati a ciascuna di esse. Ad ogni tipo di nodo corrisponde, dal punto di vista del programma di simulazione, una macroistruzione.

Al programmatore resta unicamente il compito di scegliere le macroistruzioni che descrivono i nodi del sistema, assegnare per ognuna di esse i parametri che caratterizzano il funzionamento del nodo corrispondente e di stabilire i necessari collegamenti fra le diverse macroistruzioni (corrispondenti ai collegamenti esistenti tra i nodi della rete che descrive il sistema).

4.6 - I Software per la costruzione di modelli System Dynamics

Esistono quattro programmi software che sono stati progettati per facilitare la costruzione e l'uso di modelli System Dynamics: Dynamo, iThink/Stella, PowerSim e Vensim.

Dynamo: È il primo linguaggio di simulazione di sistemi dinamici e per lungo tempo il linguaggio e la materia sono stati considerati sinonimi. Originariamente sviluppato da Jack Pugh al Mit, il linguaggio fu reso disponibile in commercio da Pugh – Roberts nei primi anni del 196. Dynamo oggi funziona su PC con sistemi operativi Dos/Window. Esso fornisce un ambiente di sviluppo di equazioni per modelli System Dynamics.

IThink/Stella: Inizialmente introdotto sul Macintosh nel 1984, il software Stella fornisce un'interfaccia grafica per lo sviluppo di modelli di System Dynamics. I diagrammi di flussi e stock, usati in letteratura dalla System

Dynamics, sono direttamente supportati da una serie di strumenti per lo sviluppo del modello. La scrittura dell'equazione viene effettuata attraverso finestre di dialogo accessibili dai diagrammi di livelli/flussi. IThink è disponibile per i computer Macintosh e Windows.

PowerSim: A metà degli anni'80 il governo norvegese ha sponsorizzato la ricerca finalizzata a migliorare la qualità della scuola di istruzione superiore utilizzando modelli System Dynamics. Questo progetto ha portato allo sviluppo di Mosaic, un sistema finalizzato principalmente allo sviluppo di giochi di simulazione per l'istruzione. PowerSim è stato successivamente sviluppato come un ambiente Windows per lo sviluppo di modelli System Dynamics, che facilita anche la costruzione di giochi interattivi o ambienti di apprendimento.

Vensim: Originariamente sviluppato nella metà degli anni'80 per l'utilizzo in progetti di consulenza, Vensim è stato reso disponibile in commercio nel 1992. Si tratta di un ambiente integrato per lo sviluppo e l'analisi di modelli System Dynamics. Vensim può essere eseguito su computer Windows e Macintosh.

4.6.1 - Implementazione dei sistemi dinamici: Powersim Studio

Le caratteristiche della System Dynamics appena descritte sono generali, ma la loro implementazione richiede l'utilizzo di specifici software.

Un numero elevato di differenti pacchetti software è disponibile per implementare logiche di System Dynamics e, per lo sviluppo di questa tesi la scelta è stata orientata al pacchetto di modellazione *Powersim Studio*. Powersim Studio è un ambiente di modellizzazione basato appunto sulla scienza dei sistemi dinamici. Powersim permette di modellare sistemi, con

tutte le loro relazioni di causa ed effetto, cicli di retroazione e dilazioni, in maniera grafica ed intuitiva. Una volta inserito il modello concettuale nel computer, attraverso PowerSim Studio, e definite in maniera appropriata tutte le variabili e le equazioni che lo costituiscono, si può simulare il modello e vedere il suo comportamento nel tempo. La simbologia usata da Powersim è identica a quella adottata dalla System Dynamics in termini di stock e flow diagrams.

I *livelli* sono rappresentati da caselle e indicano le accumulazioni che possono aumentare o diminuire nel tempo. I livelli possono essere modificati solamente attraverso i flussi.

I *flussi* sono azioni o processi che aggiungono (afflusso) o sottraggono le accumulazioni nel livello. Le nuvole all'inizio e alla fine del flusso rappresentano la fonte e la fine della struttura.

La *variabile ausiliaria* è usata per combinare o riformulare informazioni; non ha una forma standard; è un calcolo algebrico di alcune combinazioni di livelli, flussi o altri ausiliari. Anche se le variabili ausiliarie possono sembrare delle accumulazioni, esse non hanno nessuna memoria diversamente dai livelli. Le ausiliarie sono usate per modellare informazioni, non il flusso fisico di beni, così esse cambiano senza dilazioni, istantaneamente. Possono costituire contributi ai flussi ma mai direttamente ai livelli, perché i flussi sono le uniche variabili che possono cambiare i livelli associati. Le variabili di flusso e le ausiliarie sono definiti esattamente alla stessa maniera; la differenza è che le prime sono connesse alla valvola di flusso e con ciò controllano direttamente il flusso.

Le *costanti* sono, diversamente dalle ordinarie variabili ausiliarie, costanti nel periodo di tempo della simulazione.




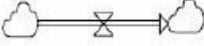

SIMBOLO	SIGNIFICATO IN BREVE
	Level (Livello): Variabile che accumula cambiamenti influenzata dai flussi
	Constant (Costante): Variabile che contiene un valore fisso che viene utilizzato nei calcoli dalle variabili ausiliarie. Corrispondono alle leve gestionali caratteristiche di uno scenario.
	Auxiliary (Ausiliaria): Variabile al cui interno si svolgono calcoli sulla base di altre variabili. Molto spesso sono collegate ai flussi.
	Flusso: influenza i livelli collegandoli tra loro. Le nuvole indicano i confini del sistema.
	Link: collega tra loro tutti i diversi tipi di variabili e determina le relazioni al loro interno. E' il mezzo per rappresentare il feedback. Esiste anche il link con ritardo temporale (Delayed Link)

Figura 16: Costrutti elementari dei modelli in ambiente Powersim

Le potenzialità del PowerSim, infine, sono le seguenti:

- Compatibilità per il trasferimento dei dati da Microsoft Excel e da SAP aziendali;
- Permette di generare scenari anche senza conoscere la struttura del modello (modalità simulatore);
- Utilizza una vasta libreria di funzioni matematiche per simulare comportamenti dinamici e complessi delle variabili.

CAPITOLO 5

Il D.E.A.: architettura di sistema e problematiche

5.1 - La gestione del P.S. in letteratura

In letteratura si riscontra un ampissimo numero di articoli che tratta il tema della gestione sanitaria: questo tema negli ultimi decenni ha suscitato un particolare interesse da un punto di vista politico-sociale, essendo argomento centrale di numerosi dibattiti riguardanti l'ingente consumo e spreco di risorse ad essa dedicate a fronte di bassi risultati in termini qualitativi. L'interesse verso questo tema si è trasferito anche nell'ambito accademico e della ricerca: gli ospedali ed in particolar modo i reparti di Pronto Soccorso offrono spunti gestionali di ampie vedute, in quanto costituiti da strutture organizzative complesse che possono essere analizzate, gestite e migliorate perseguendo gli obiettivi di efficacia ed efficienza e riducendo gli sprechi sia dal punto di vista economico che delle risorse impiegate.

Ci si è concentrati sull'analisi di quegli articoli che proponessero una soluzione ai problemi organizzativi che si riscontrano all'interno di un reparto di emergenza: le criticità che si riscontrano nel processo di pronto soccorso devono essere analizzate, gestite e risolte.

Il focus seguito in tutti gli articoli visionati si è rivelato essere l'ottenimento della qualità intesa come soddisfazione del paziente: i casi peculiari analizzati sono quelli relativi all'utilizzo di sistemi di miglioramento continuo della

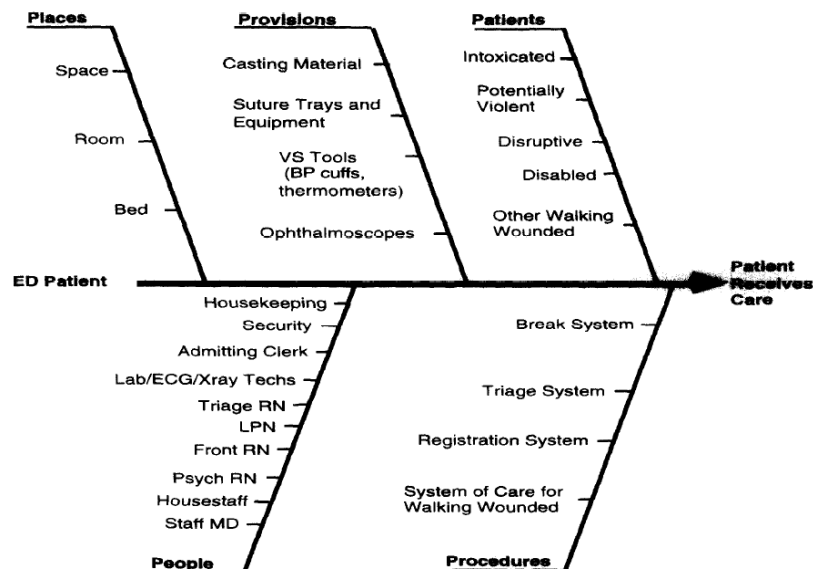
qualità e di strumenti di ricerca e sondaggi (come i questionari) per l'individuazione di aree di miglioramento della qualità.

Il miglioramento continuo della qualità è un processo che mira all'identificazione e soddisfacimento delle richieste del cliente; esso comprende anche tutte quelle azioni volte al miglioramento della qualità dei servizi e dei prodotti per i pazienti e altri fruitori del servizio. L'utilizzo di tecniche di miglioramento continuo della qualità aiuta non solo ad identificare le principali cause dei ritardi nella valutazione e nel trattamento dei pazienti di Pronto Soccorso, ma anche a valutare le soluzioni più efficaci per la riduzione degli stessi. In questi tempi caratterizzati da scarse risorse disponibili, il "Continuous Quality Improvement" è il migliore strumento per gestire al meglio l'assistenza sanitaria. In numerosi articoli vengono esaminate le funzioni del triage e delle aree "Fast-Track" (due delle aree più critiche del PS poichè caratterizzate da lunghi tempi di attesa e insoddisfazione del paziente), ma all'analisi di queste parti del processo deve essere affiancato l'utilizzo del CQI per ottenere una corretta pianificazione ed un giusto utilizzo delle risorse.

L'ospedale di St. Paul (Vancouver, Canada) ha attuato un progetto di miglioramento della qualità focalizzato sulle aree del triage e del Fast Track, reso necessario dall'insoddisfazione del paziente dovuta a lunghe attese e triage inadeguato: queste problematiche erano state evidenziate in passato da numerose denunce e tassi elevati di pazienti che abbandonavano il reparto senza aver ricevuto visita da un medico.⁷ Lo scopo del progetto era determinare se l'utilizzo di un sistema di CQI potesse aiutare a ridurre i tempi di attesa e migliorare la percezione della qualità da parte del paziente.

⁷ Fernandes, Christopher, and James M. Christenson. "Use of continuous quality improvement to facilitate patient flow through the triage and fast-track areas of an emergency department." *The Journal of emergency medicine* 13.6 (1995): 847-855.

E' stato organizzato un team di progetto i cui membri erano non solo operatori del reparto di emergenza, ma erano stati inseriti in esso tutti coloro che potevano avere un "impatto" sulla cura del paziente nelle aree analizzate. Il primo step del team e stato definire il processo in modo accurato, per poi stabilire un piano di "intenti". La definizione del processo ha portato alla realizzazione di un diagramma di flusso del processo in cui sono stati poi evidenziati i "colli di bottiglia". A ciò è seguita una rappresentazione di un diagramma "causa-effetto" (Mod.3) in riferimento alle problematiche riscontrate: sistemi di registrazione problematici, mancanza di spazio e letti, postazione inadeguata del triage, duplicazione di procedure, ruoli degli infermieri di triage, ritardi nei referti radiografici e di analisi di laboratorio, scarso numero di risorse mediche.



Mod.3 Cause-and-effect diagram for triage and Fast-Track areas

Attraverso un'analisi step-by-step del flusso e l'esame delle relazioni "customer-supplier" sono stati individuati i passi non a valore aggiunto e le aree di inefficienza. Il team ha poi lavorato alla proposizione di soluzioni per ognuna delle criticità esposte ed ha lavorato nei mesi successivi per l'applicazione di azioni correttive. I risultati ottenuti sono poi stati analizzati per valutare l'efficacia dell'adozione di un sistema CQI. In particolare lo

studio ha mostrato che l'utilizzo di questo strumento organizzativo ha permesso il superamento di due ostacoli: l'inefficienza della funzione infermieristica e l'inadeguatezza del sistema di ammissione al P.S.. Inoltre anche la lunghezza del soggiorno del paziente è diminuita significativamente ed l'ottimizzazione dei flussi ha contribuito ad aumentare la soddisfazione del paziente. Questo studio ha quindi dimostrato come il CQI è una tecnica importante e pratica per la gestione del Pronto Soccorso e può essere utilizzata per esaminare e migliorare il flusso dei pazienti del reparto d'urgenza e diminuire la durata della permanenza dei pazienti.

Un altro strumento spesso utilizzato nei dipartimenti di P.S. per la valutazione delle aree critiche da migliorare e per l'ottenimento della qualità, è l'utilizzo di questionari: essi permettono di valutare in che modo il paziente percepisce la qualità della cura e quali sono gli effetti dei tempi di attesa e delle modalità di attuazione della cura sulla sua soddisfazione. L'uso di uno strumento del genere può fornire informazioni preziose per il miglioramento della qualità nella pratica clinica.

Si può mostrare a titolo di esempio il caso dello studio avanzato sulla percezione della qualità del P.S. dell'ospedale universitario di Uppsala: Wilde Larsson et al. (2001) hanno realizzato un questionario per valutare la qualità dalla prospettiva del paziente (Quality Patient's Perspective). Il questionario contiene una serie di domande volte a misurare fattori nelle seguenti quattro dimensioni: competenza medico-tecnica, condizioni fisico-tecniche, approccio orientato all'identità del paziente e ambiente socio-culturale. Le quattro dimensioni sono interdipendenti ed la loro valutazione nell'insieme costituisce la qualità delle cure. Ogni domanda valuta la percezione della qualità del reparto e l'importanza soggettiva stimata dal paziente per ognuna delle dimensioni. Sono stati intervistati oltre 60000 pazienti ed il questionario è stato valutato valido ed affidabile (Larsson et al. 2000).

I dati ottenuti sono stati valutati e trasformati in indici di percezione della qualità e raccolti come nel Prospetto 1.

Factors	IQ (%)	BL (%)	BH (%)	EQ (%)	<i>n</i>
Medical-technical competence					
Waiting time	18	32	32	18	148
Medical care	17	19	48	16	138
Physical-technical conditions					
Care-room characteristic	28	36	30	8	83
Communication means	27	24	32	16	90
Nutrition	17	23	27	32	117
Care equipment	1	3	65	31	101
Identity-oriented approach					
Treatment	21	25	37	17	107
Doctor responsible	21	19	48	13	156
Information after procedures	16	20	41	22	123
Empathic and personal (doctors)	13	22	46	20	170
Commitment (nurses and assistant nurses)	13	31	40	16	146
Participation	10	27	45	17	142
Commitment (doctors)	9	34	37	19	164
Empathic and personal (nurses and assistant nurses)	8	28	41	22	163
Information before procedures	8	27	38	27	136
Respect (receptionist)	8	33	35	24	143
Respect (nurses and assistant nurses)	6	19	47	28	172
Respect (doctors)	4	15	57	24	185
Socio-cultural atmosphere					
General atmosphere	18	31	36	14	169
Relatives and friends	9	17	54	21	78
Privacy	7	22	52	20	88
Routines	7	33	34	26	123

Prosp.1 - Patients' perceptions of quality of care in the ED by factor, calculated as inadequate quality (IQ), balance low (BL), balance high (BH) and excess quality (EQ)

La percezione della qualità della cura nel reparto di emergenza è stato giudicato “abbastanza buono”, ma si sono identificate delle aree con necessità di miglioramento. I risultati più rilevanti sono stati: un’alta percentuale ha dichiarato che la qualità inadeguata era dovuta all’ambiente del reparto; il 20% degli intervistati ha dichiarato di non aver ottenuto sollievo al dolore; più del 20% ha asserito che gli infermieri non avessero mostrato interesse e non avessero fornito informazioni utili attese.

Queste informazioni sono di fondamentale importanza per poter “cambiare rotta” nella gestione del reparto di emergenza, poichè danno evidenza delle criticità da risolvere per poter incontrare la soddisfazione del paziente.

5.2 L’abbandono dal reparto di pronto soccorso

Pazienti che si presentano al Pronto Soccorso per valutazioni e cure ma che abbandonano il reparto senza aver ricevuto vista da un medico rappresentano un problema significativo per molte aziende sanitarie. E’ stato mostrato che la frequenza di abbandono varia moltissimo a seconda della particolare realtà sanitaria a cui ci si riferisce: negli USA il tasso medio varia tra il 1.4% e il 2.9%. Sempre per quanto riguarda il territorio americano il “General Accounting Office” ha riportato che oltre il 7% dei dipartimenti di emergenza presenta delle percentuali superiori al 5% toccando talvolta anche percentuali del 15%⁸.

Il tasso di abbandono volontario può essere considerato come uno dei più importanti indicatori dell’affollamento del P.S. oltre che della percezione della qualità in generale. Sono stati individuati un gran numero di fattori che possono influenzare questo tasso e tra questi sono inclusi: tempi di attesa o percezione del tempo di attesa; numero totale di pazienti trattati nel P.S., livello di qualità percepito dello staff medico, servizi forniti ai pazienti nelle sale d’attesa, ambiente confortevole.

Gli studi mostrano che la maggior parte degli abbandoni sono di pazienti etichettati come non urgenti, ovvero con codici di priorità inferiore; inoltre in base a recenti sondaggi i pazienti dichiarando di aver abbandonato il reparto per insoddisfazione (36%), ed in particolare la ragione di questa insoddisfazione è stata la lunga attesa. Dei pazienti intervistati che hanno

⁸ Chan, Theodore C., et al. "Impact of rapid entry and accelerated care at triage on reducing emergency department patient wait times, lengths of stay, and rate of left without being seen." *Annals of emergency medicine* 46.6 (2005): 491-497.

abbandonato il P.S. il 60% ha dichiarato la volontà a non tornare in futuro nella stessa azienda ospedaliera⁹.

Per cui il tasso di pazienti che lasciano l'ospedale senza una visita può essere considerato come un "indicatore clinico" che riflette la soddisfazione del paziente nei confronti della cura. In questo senso questo gruppo può essere studiato e "tracciato" periodicamente in modo da intraprendere le giuste misure correttive.

Poichè la maggior parte degli abbandoni è da parte di pazienti "non urgenti", se si riducessero i tempi di attesa per questa categoria si potrebbe ottenere una riduzione del tasso di abbandono: pazienti che aspettano di meno rispetto alle proprie aspettative saranno i più soddisfatti e quindi meno propensi all'abbandono. Molti sono gli studi presenti in letteratura per la realizzazione di aree "FastTrack": ovvero aree del pronto soccorso dedicate alla cura di pazienti non urgenti in cui si facilita lo scorrimento del flusso di pazienti in modo da ridurre i tempi totali di "processo" (di cura); molteplici sono inoltre gli studi che cercano la correlazione tra la riduzione dei tempi di attesa e la riduzione del tasso LWBS (patient who Leave Without Being Seen).

In sintesi si può affermare che la soddisfazione del paziente diminuisce e la frequenza del LWBS aumenta, quando il paziente attende più di quello che si aspetta per poter essere visitato da un medico nel reparto di emergenza. Per rispondere a tale problematica a San Diego (California) nel Pronto Soccorso di un centro accademico si è pensato di realizzare un processo di entrata rapida e cura accelerata al triage: il REACT (Rapid Entry and Accelerated Care at Triage)¹⁰. Esso consiste nella trasformazione di processi seriali in

⁹ Fernandes, Christopher, Ann Price, and James M. Christenson. "Does reduced length of stay decrease the number of emergency department patients who leave without seeing a physician?." *The Journal of emergency medicine* 15.3 (1997): 397-399.

¹⁰ Chan, Theodore C., et al. "Impact of rapid entry and accelerated care at triage on reducing emergency department patient wait times, lengths of stay, and rate of left without being seen." *Annals of emergency medicine* 46.6 (2005): 491-497.

processi paralleli: si intende facilitare il lavoro del medico d'urgenza nell'iniziare una serie di attività, interventi ed analisi prima ancora che sia disponibile il letto per la visita o che sia completata la registrazione del paziente nel sistema informativo. E' stato successivamente condotto uno studio per valutare in che modo il ridisegno del processo di triage avesse impattato sul tasso di abbandono: si è ottenuta una diminuzione nei tempi di attesa medi di 24 minuti dopo l'implementazione del REACT ed una diminuzione concomitante del tempo totale di durata del processo di pronto soccorso di 31 minuti; si è oltretutto ottenuta una diminuzione del LWBS del 3.2%.

In letteratura notevole è il numero di articoli che analizza la correlazione tra il tasso di abbandono e il sovraffollamento: ogni fattore che può essere inteso come indice di bassa qualità del reparto di emergenza è in realtà causa di insoddisfazione nel paziente che è spinto ad abbandonare l'ospedale e cercare cure altrove. Molti sono i pazienti che una volta abbandonato il Pronto Soccorso riescono a trovare una soluzione alla loro "urgenza" in altre strutture: dimostrazione del fatto che molti degli accessi sono in realtà impropri.

5.3 Stato dell'arte delle tecniche risolutive

Come espresso in precedenza tra le procedure più problematiche da gestire all'interno del processo di pronto soccorso si evidenziano quelle di identificazione, accettazione e monitoraggio della movimentazione del paziente che accede al reparto di emergenza. A tal proposito si è svolta una attività di ricerca riguardo lo stato dell'arte delle soluzioni tecnologiche proposte in letteratura in riferimento al problema della tracciabilità del paziente onde evitare l'allontanamento volontario dal P.S..

Il concetto di tracciabilità viene espresso come: "possibilità di identificare l'origine ed i diversi stadi di un processo di produzione e distribuzione di

beni di consumo insieme ad un documentazione che registri tutti i suddetti stadi”¹¹; questo concetto può essere facilmente trasferito al settore sanitario ed è particolarmente interessante nell’ambito del pronto soccorso, reparto che fa i conti giorno per giorno con un gran numero di pazienti. Si intende quindi risolvere la seguente problematica: seguire il percorso del paziente all’interno del PS, dal suo arrivo alla sua dimissione, registrando tempi di attesa e di cura ed eventuali allontanamenti dall’area di interesse.

Sistemi di tracciabilità sono stati implementati con successo nei processi commerciali e industriali; nell’ambito sanitario esiste una raccolta letteraria in lenta espansione su svariate applicazioni di sistemi tecnologici per la tracciabilità. Le soluzioni più comunemente riscontrate fanno riferimento all’utilizzo della tecnologia RFID (Radio Frequency IDentification) che permette l’identificazione, a mezzo radiofrequenza, automatica e remota di oggetti (in questo caso pazienti); inoltre si è riscontrato in letteratura anche l’adozione di strumenti tecnologici quali il c.d. braccialetto elettronico, i QRcode e l’utilizzo di una lavagna elettronica. Questi dispositivi e tecnologie sono sistemi elettronici utilizzati con lo scopo di agevolare la comunicazione tra i soggetti coinvolti ed il monitoraggio del paziente.

L’implementazione di processi di tracciabilità del paziente nei sistemi informativi sanitari non apporta in genere cambiamenti sostanziali nei metodi lavorativi nè aggiunge costi di implementazioni eccessivi.

Il lavoro di ricerca e analisi ha portato alla considerazione che esiste un’attenzione crescente riguardo l’utilizzo dell’Information Technology in ambito sanitario per abbattere costi, migliorare efficienza e sicurezza; esiste, infatti, una domanda crescente per i servizi legati alla gestione sanitaria dovuta per lo più all’ampio sviluppo tecnologico che si è avuto recentemente. Sistemi “intelligenti” possono giocare un ruolo cruciale nel fornire una

¹¹ Pérez, M. M., Cabrero-Canosa, M., Hermida, J. V., García, L. C., Gómez, D. L., González, G. V., & Herranz, I. M. (2012). Application of RFID Technology in Patient Tracking and Medication Traceability in Emergency Care. *Journal of medical systems*, 36(6), 3983-3993.

gestione efficace delle procedure di identificazione, accettazione e tracciabilità del paziente. Ci si aspetta che sistemi di monitoraggio “real-time” possano riscontrare ampio successo e utilizzo nelle aziende sanitarie in cui risulta sempre più indispensabile una corretta comunicazione ed una tracciabilità certa dei flussi (tra cui quello del paziente) in esse presenti.¹²

5.3.1 Tecnologia RFID

La tecnologia RFID ha mostrato una crescita esponenziale negli ultimi decenni, in relazione a numero di applicazioni realizzate, flussi economici, potenzialità tecnologiche. Tra i settori che più recentemente hanno assistito all'avvento della tecnologia RFID, vi è sicuramente il settore sanitario; la motivazione è da cercarsi nelle caratteristiche peculiari di un sistema RFID: l'affidabilità, la rapidità di trasmissione dei dati, la distanza di lettura.

L'utilizzo della tecnologia RFID è molto diffuso ed in particolare assolve ai seguenti compiti: localizzare il paziente nelle diverse aree; misurare i tempi di attesa e di cura; eliminare documenti in forma cartacea; costituire un “magazzino” di informazioni facilmente consultabile da medici e/o infermieri¹³.

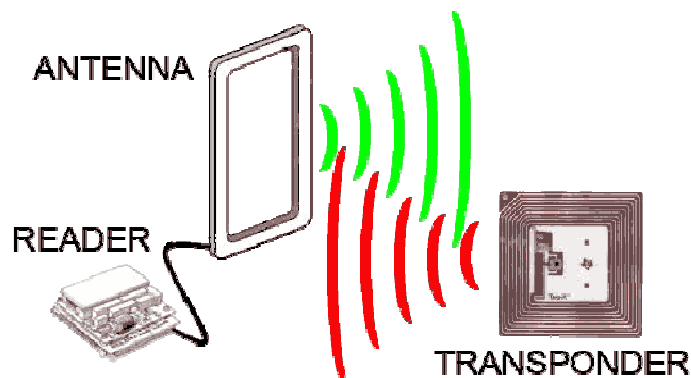
I sistemi RFID fanno parte della tecnologia di identificazione automatica a radiofrequenza o Auto-ID. Questa tecnologia presenta due caratteristiche basilari: riduce sia i tempi che i costi riguardanti l'inserzione manuale dei dati di interesse ed elimina i possibili errori nell'inserimento dei dati stessi. La tecnologia RFID utilizza quindi un sistema che permette l'acquisizione automatica di dati per l'identificazione e consente l'introduzione automatica, cioè senza l'ausilio di tastiera o di altre operazioni manuali, di questi dati di identificazione e di altri dati complementari all'interno di programmi dedicati

¹² Chowdhury, B., & Khosla, R. (2007, July). RFID-based hospital real-time patient management system. In *Computer and Information Science, 2007. ICIS 2007. 6th IEEE/ACIS International Conference on* (pp. 363-368). IEEE.

¹³ Pérez, M. M., Cabrero-Canosa, M., Hermida, J. V., García, L. C., Gómez, D. L., González, G. V., & Herranz, I. M. (2012). Application of RFID Technology in Patient Tracking and Medication Traceability in Emergency Care. *Journal of medical systems*, 36(6), 3983-3993.

presenti in un computer. Un sistema RFID si compone di tre elementi fondamentali:

- *Tag*: un transponder a radiofrequenza di piccole dimensioni, elemento principale del sistema. E' un'etichetta intelligente, di varie forme e dimensioni, che viene posta sugli oggetti da gestire permettendo la trasmissione di dati a corto o medio raggio senza contatto fisico;
- *Reader*: un ricetrasmittitore controllato da un microprocessore ed usato per interrogare e ricevere le informazioni in risposta dai tag. Ne esistono di varie dimensioni e possono essere statici o portatili;
- *Sistema di gestione*: un sistema informativo che, quando esiste, è connesso in rete con i reader. Tale sistema consente, a partire dai codici identificativi provenienti dai tag, di ricavare tutte le informazioni disponibili associate agli oggetti e di gestire tali informazioni per gli scopi dell'applicazione.



Schema1 - Il sistema RFID

Riassumendo, la tecnologia RFID si basa sulla lettura a distanza di informazioni contenute in un tag, usando dei reader e sistemi di gestione quando necessario. Il tag può essere posto su di un braccialetto fornito al paziente al momento dell'accettazione (costituendo così il c.d. braccialetto elettronico) e conterrà un'antenna ed un microchip; l'antenna quando rileva onde elettromagnetiche trasmette al lettore le informazioni contenute nel chip (ID del paziente, locazione, informazioni mediche etc.). Il lettore può essere

fisso o mobile ed in quest'ultimo caso può essere integrato con dispositivi mobili PDA (Personal Digital Assistant).¹⁴ Questa tecnologia si scontra con il classico modello “static information - dynamic clinician” (in cui i medici si spostano per reperire informazioni) ed è del tipo “dynamic information – static clinician”: i tag associati ai pazienti trasmettono informazioni riguardo la localizzazione e dati del paziente verso un board centrale informativo.¹⁵

Il più grande beneficio che viene raggiunto tramite l'implementazione di sistemi di tracciabilità per i pazienti di ospedale è l'ampia riduzione dell'occorrenza di cd “eventi avversi”: costituisce uno strumento utilissimo per la riduzione del rischio clinico. Applicazioni di sistemi Rfid in ambito sanitario sono molto frequenti e spesso correlati alla risoluzione di problematiche non direttamente connesse all'allontanamento volontario del paziente dal reparto di pronto soccorso.

Un'esempio di implementazione di sistema di tracciabilità è stato analizzato presso il Pronto Soccorso dell'ospedale di Coruña: in una precedente fase di analisi si era evidenziata la necessità di un tracciamento continuo del paziente nelle diverse zone del reparto di emergenza, in modo da poterne identificare in remoto ed in tempo reale l'identificazione e le attività in atto o da svolgersi. In una fase iniziale è stata realizzata una raccolta di informazioni riguardo l'intero processo di cura di urgenza, acquisendo ogni genere di documentazione atta alla realizzazione di un diagramma di flusso che rispondesse appieno all'esigenza di valutare in quali fasi si potevano rischiare allontanamenti o rischi clinici. Nelle fasi successive è stato poi progettato, basandosi sul flusso precedentemente analizzato, il sistema di tracciabilità RFID: esso è costituito da due sottosistemi, uno volta alla tracciabilità del paziente ed uno volta alla prevenzione dell'occorrenza di eventi avversi. In

¹⁴ Chowdhury, B., & Khosla, R. (2007, July). RFID-based hospital real-time patient management system. In *Computer and Information Science, 2007. ICIS 2007. 6th IEEE/ACIS International Conference on* (pp. 363-368). IEEE.

¹⁵ Dobson, I., Doan, Q., & Hung, G. (2013). A Systematic Review of Patient Tracking Systems for Use in the Pediatric Emergency Department. *The Journal of emergency medicine, 44*(1), 242-248.

Prospetto 2 si possono osservare le componenti hardware installate nei due sottosistemi.

Table 1 Subsystem of patient traceability: hardware components

Component	Manufacturer	Cost per unit	Task
Active tag with WIFI communication technology	Aeroscout	41€	To broadcast a signal for locating patients
Wireless access point (plus installation)	Cisco, Aironet 1130 AG model	1,000€	To interconnect WIFI communication devices and to serve as a reference point to calculate location of the patient.
Passive tag	Philips, size 15×15 mm and ISO 15693 approved model	0.3€	To broadcast RFID signal with patient identification data.

Prosp. 2 - Componenti dei sottosistemi per la tracciabilità del paziente

Alla fase di progettazione è poi seguita una fase di implementazione e test: si sono realizzati la rete wireless ed i punti di accesso in grado da garantire la copertura totale dell'area per poi valutare l'accuratezza della localizzazione. I risultati ottenuti sono stati poi valutati sottoponendo il personale medico ad un questionario sul livello di soddisfazione raggiunto in seguito all'utilizzo del sistema di tracciabilità: la valutazione è stata più che positiva e si sono ottenuti, tramite l'adozione di un sistema informatico ad hoc, l'identificazione in tempo reale del paziente fino a 4 metri di distanza dai punti di accesso, l'identificazione dei medicinali e si è assicurata la corretta associazione tra paziente e medicinale (con conseguente riduzione del rischio clinico).

5.3.2 Bar Code e Braccialetto elettronico

La tecnologia Bar Code può essere utilizzata per fornire soluzioni a molti dei problemi che ci si trova ad affrontare nel reparto di Pronto Soccorso, tra cui la tracciabilità del paziente. Si tratta di consegnare al paziente in fase di triage un braccialetto con apposta una etichetta che in lettura richiama i record elettronici. Questa soluzione permette di garantire una maggiore sicurezza,

un migliore sistema di comunicazione ed una migliore accuratezza nell'identificazione. Molteplici sono le applicazioni in campo sanitario del braccialetto elettronico: consiste in un braccialetto su cui è apposto un codice a barre in cui sono immagazzinati i dati del paziente relativi all'intero processo.



Schema.2 - Il braccialetto elettronico

Questa soluzione organizzativa è stata implementata al P.S. del St Joseph Hospital di Rhode Island (Massachusetts, USA) allo scopo di risolvere i problemi di identificazione ed eliminazione dei colli di bottiglia del processo e per garantire una rapida identificazione della localizzazione del paziente all'interno dell'area di riferimento¹⁶. Il sistema di tracciabilità funziona nel seguente modo: i pazienti vengono identificati e registrati a sistema al momento dell'accesso in pronto soccorso ed i dati vengono trasferiti sul bar code e automaticamente anche al sistema di tracciabilità del paziente. Il bar code viene successivamente apposto su un braccialetto che viene consegnato al malato. I dati immagazzinati nel codice possono essere successivamente analizzati per verificare quali sono stati i momenti critici nell'esecuzione delle attività di urgenza: in base all'analisi dei dati si potrà poi intervenire per la risoluzione delle criticità evidenziate. Le informazioni contenute nel bar-code sono "leggibili" da uno scanner in meno di 10 secondi e l'informazione viene trasmessa al sistema ed è immediatamente disponibile ad ogni operatore del reparto. L'implementazione del sistema di tracciabilità ha

¹⁶ Amber, Richard, and Valerie B. Everett. "Emergency department patient tracking: a cost-effective system using bar code technology." *Journal of Emergency Nursing* 22.3 (1996): 190-195.

richiesto un anno: non sono stati calcolati tempi e relative spese ma il sistema ha raggiunto ogni risultato sperato, oltre a produrre sostanziali miglioramenti in tutto il processo di P.S. ed aver incrementato la soddisfazione tanto del paziente che dello staff medico.

Negli ultimi anni l'utilizzo del braccialetto elettronico si è cominciato a diffondere anche in territorio nazionale: un esempio è quello del Pronto Soccorso civico di Palermo, in cui, a partire dall'Aprile dello scorso anno, si è cominciato a distribuire nel momento della valutazione triage un braccialetto identificativo che riporta i dati anagrafici e clinici del paziente. Per il commissario straordinario, Carmelo Pullara, "si tratta di un altro completamento di quelle dotazioni che costituiscono una percezione di buona accoglienza per i cittadini che si rivolgono a un ospedale e, in particolare, ad un'area d'emergenza. E' garanzia, inoltre, della tracciabilità del percorso del paziente".

5.3.3 Altre soluzioni tecnologiche

Processi di registrazione lenti, lunghi tempi di attesa, procedure di immissione dati manuali e comunicazioni interdipartimentali frammentate, sono problematiche comuni ad ogni pronto soccorso che possono condurre ad una percezione della qualità di cura non ottimale e a una riduzione della soddisfazione sia dello staff medico che del paziente. Per trovare una soluzione a queste criticità è stato sviluppato un nuovo strumento: una "lavagna elettronica"¹⁷ che facilita la tracciabilità del paziente, migliora la comunicazione e monitora il flusso del paziente lungo tutto il processo. Lo sviluppo e l'implementazione di questo strumento è stato il risultato dello sforzo collaborativo di infermieri, medici ed informatici oltre che di specialisti del sistema. La lavagna è un monitor di 60 pollici posizionato nel dipartimento di emergenza che opera come strumento di comunicazione per

¹⁷Vest, N., Rudge, N. K., & Holder, G. (2006). ED whiteboard: an electronic patient tracking and communication system. *Journal of Emergency Nursing*, 32(1), 8.

garantire aggiornamenti tanto per i medici quanto per i pazienti; esso permette di visualizzare il flusso del paziente in tempo reale e tracciare il suo percorso lungo tutto il processo e fornisce degli indicatori di allerta per il sopravvento di determinati rischi.

Il monitor ha quindi lo scopo di fornire informazioni, continuamente aggiornate, riguardo la sala d'attesa, l'area di trattamento e gli stati di dimissione. La "lavagna" si interfaccia con molteplici applicazioni del reparto di PS: sistema di accettazione, dimissione e trasferimento, applicazione di tracciabilità, sistemi elettronici di notifiche etc. Questa soluzione è stata sviluppata per facilitare la tracciabilità del paziente, migliorare la comunicazione e monitorare il flusso di pazienti all'interno del reparto di emergenza ed il suo utilizzo può fornire un aiuto sostanziale per prevenire l'allontanamento dal reparto poichè il paziente ottiene informazioni chiare su tempi di attesa ed il suo stato nella coda per l'accesso alla visita.

5.4 L'abbandono come rischio

Una volta analizzate le soluzioni organizzative proposte in letteratura per la risoluzione delle problematiche che inficiano il raggiungimento degli obiettivi di qualità nel reparto di emergenza, lo studio si è concentrato sulle modalità di valutazione delle criticità.

Come spesso evidenziato, il primo passo da compiere per migliorare e ottimizzare un processo, è l'analisi dello stesso. Analizzare un processo significa poterne valutare (anche in maniere quantitativa) le criticità che in esso possono occorrere. In particolare le aree di Pronto Soccorso sono soggette a rischi in termini sia in termini di rischio clinico, che rischio di malfunzionamento del processo.

L'analisi del processo di P.S. porta alla valutazione di fattori quali: tempi di attesa media, livello della qualità dello staff medico, adeguatezza del triage, valutazione del tasso di abbandono. In riferimento a quest'ultimo punto lo studio di questo elaborato, si propone di valutare questa "probabilità" facendo riferimento ai dati reperibili sui processi di Pronto Soccorso.

La valutazione quantitativa del tasso di abbandono può essere di notevole aiuto nella gestione sanitaria: essendo un fattore critico legato al concetto di insoddisfazione del cliente dovuta ad attese e sovraffollamento, la sua determinazione può portare ad analisi approfondite su azioni da compiere per la sua riduzione. Ogni reparto di P.S. è caratterizzato da un proprio tasso di probabilità di allontanamento volontario del paziente: è un indicatore di qualità del servizio che deve essere tenuto sotto controllo e su cui un'efficiente gestione deve porre obiettivi di qualità.

Ma in che modo valutare la probabilità di abbandono? L'abbandono può essere visto come un evento che si scatena in determinate condizioni sfavorevoli e di insoddisfazione: è un rischio. Il rischio è la potenzialità che un'azione o un'[attività](#) scelta (includendo la scelta di [non agire](#)) porti a una [perdita](#) o ad un evento indesiderabile: una scelta influenza il [risultato](#). Le stesse perdite potenziali possono anche essere chiamate "rischi". Per "rischio" possiamo indicare anche la distribuzione dei possibili [scostamenti](#) dai risultati attesi per [effetto](#) di eventi di incerta manifestazione, interni o esterni ad un [sistema](#).

In Pronto Soccorso l'abbandono può essere visto come rischio in questo senso: gli eventi di insoddisfazione, lunga attesa, mancata attenzione medica possono portare all'accadimento dell'allontanamento. Valutare la probabilità che il paziente raggiunga un tale livello di insoddisfazione da abbandonare il reparto, significa valutare un rischio (inteso come probabilità di una perdita o di un pericolo/minaccia).

Esistono molte definizioni di rischio in dipendenza del particolare contesto a cui ci si riferisce e delle applicazioni specifiche; in generale un indicatore di rischio è proporzionale all'effetto atteso e alla sua probabilità di accadimento. In letteratura i modelli di rischio sono spesso associati all'ambito finanziario in cui si focalizza l'attenzione sul rischio di credito per la creazione di un portafoglio investimenti. Più recentemente al rischio è stato associato un significato più ampio connesso al concetto di opportunità: ciò deriva dal significato della parola inglese "risk" che è meno restrittivo e preciso e quindi non collegato direttamente ad una minaccia; ciò nonostante la trattazione e la modellizzazione non cambiano sia nella accezione negativa che positiva.

Nel caso della valutazione del rischio di abbandono, ovviamente, il significato è inteso nel senso negativo del termine ed è legato alla "minaccia" del raggiungimento di un elevato livello di soddisfazione. Il lavoro di ricerca si è incentrato, a questo punto, sulle modalità con cui poter costruire un modello di rischio in cui gli eventi che generano insoddisfazione sono strettamente collegati all'accadimento del rischio.

Bisogna, infine, distinguere tra il concetto di rischio reale, rischio ipotetico e rischio apparente: nel primo caso il valore che si ottiene è un valore oggettivo che deriva da analisi dei dati; nel secondo caso esso deriva da stime tecnico-scientifiche ed infine nel caso di rischio apparente si ottiene un valore soggettivo in quanto dipendente dalla percezione umana del rischio.

5.4.1 La modellizzazione del rischio: le catene di Markov

La ricerca di modelli di rischio in letteratura spesso richiama l'utilizzo delle teorie markoviane. Modelli di Markov sono utili in diverse situazioni: quando un problema di decisione comporta dei rischi, quando il modello che si rappresenta è continuo, e quando gli eventi da valutare possono verificarsi più di una volta.

Rappresentare situazioni cliniche con alberi decisionali tradizionali è difficile e può richiedere ipotesi semplificative irrealistiche. I modelli di Markov presuppongono che il paziente durante il suo percorso nei reparti ospedalieri si trovi in uno degli stati del processo, chiamati stati di Markov. Con le teorie Markoviane gli eventi sono rappresentati come transizioni tra i diversi stati e di essi se ne può realizzare una rappresentazione matriciale. In quest'ottica il processo di Pronto Soccorso può essere modellizzato tramite catene Markoviane in cui gli stati rappresentano le diverse fasi del processo e le transizioni tra essi sono definite in base alle probabilità di “passare” ad una fase successiva; lo studio della catena può portare a valutare la probabilità di accadimento di un determinato evento e nel caso in considerazione la probabilità di un evento di rischio.

CAPITOLO 6

Le logiche Lean e Agile in ambito sanitario

6.1 – Definizione e principali caratteristiche del servizio

E' bene chiarire il concetto di servizio e specificarne le differenze rispetto ai prodotti fisici.

Il professore Christian Grönroos definisce un servizio come *“un’attività o una serie di attività di natura più o meno intangibile che normalmente, ma non necessariamente, ha luogo nell’interazione tra cliente ed impiegato e/o risorse fisiche o prodotti e/o sistemi di fornitura del servizio, il quale viene fornito come soluzione ai problemi del cliente”*.

La letteratura riconosce unanimemente ad un servizio le seguenti peculiarità:

- un servizio è un processo e, in quanto tale, è costituito da una sequenza di *task* che generano un *output* finale (standard o personalizzato) rivolto al cliente (interno od esterno);
- un servizio è fisicamente intangibile.

È, a tal punto, opportuno andare ancor più a fondo ed esaminare le caratteristiche di un servizio. In particolare, il prof. Sergio Cherubini, docente di Marketing, nel testo *“Il Marketing nei Servizi. Per lo Sviluppo Competitivo e la Customer Satisfaction”*, ne individua quattro:

1. **Intangibilità:** un servizio è propriamente una pratica ed in quanto tale non può essere tangibile.

2. **Deperibilità:** un servizio non può essere né immagazzinato, né trasportato.
3. **Regolamentazione:** tale caratteristica si riferisce in particolare alla categoria dei servizi pubblici che, per loro stessa natura, sono definiti attraverso interventi legislativi, ai quali, dunque, devono necessariamente attenersi.
4. **Interazione Produttore/Consumatore:** tale punto assume una rilevanza molto spiccata in quasi tutti i servizi, i quali costituiscono il frutto di un'interazione diretta tra produttore e consumatore, nella quale si esauriscono contemporaneamente i momenti di produzione del servizio e di erogazione dello stesso.

Ulteriori differenze sono riportate in tabella:

Tabella 2: Principali differenze tra beni e servizi

Fonte: Grönroos, 1994

PRODOTTI FISICI	SERVIZI
Tangibili	Intangibili
Omogenei	Eterogenei
Produzione e distribuzione separata dal consumo	Produzione, distribuzione e consumo sono processi simultanei
Una cosa, un oggetto	Un'attività o un processo
Il valore essenziale viene prodotto in fabbrica	Il valore essenziale viene prodotto nelle interazioni venditore/acquirente
I clienti (normalmente) non partecipano al processo di produzione	I clienti partecipano alla produzione
Possono essere tenuti in magazzino	Non possono essere tenuti in magazzino
Trasferimento di proprietà	Non c'è trasferimento di proprietà

Risulta evidente, quindi, la caratteristica del servizio di essere *people-intensive* per cui la clientela diviene parte integrante del processo di fornitura.

L'erogazione di un servizio ha gli stessi obiettivi della produzione industriale: l'implementazione dei processi a valore aggiunto e l'utilizzo efficiente delle risorse a disposizione.

Una peculiarità di un sistema che eroga servizi, è rappresentata da un alto grado di *volatilità della domanda*, significativamente superiore rispetto ai sistemi di produzione manifatturiera. Ciò accade innanzitutto per l'impossibilità di stoccaggio suddetta, che non permette di accumulare scorte nei periodi di stanca per poi utilizzarle nei periodi di picco della domanda, il che comporta una certa difficoltà nel mantenere un livello di produzione più o meno stabile; in secondo luogo questo stato di variabilità è dettato anche dalla natura "personalizzata" del servizio, il cui processo di erogazione ha come destinatari clienti diversi che hanno spesso esigenze differenti, che nel corso del processo, sperimentano diversi livelli di esperienza e possono richiedere differenti quantitativi di transazioni, comportando in questo modo una certa variabilità nei tempi di erogazione del servizio stesso.

È indispensabile inoltre evidenziare il fatto che i servizi non sono necessariamente soggetti tutti alle medesime regole (quel che funziona bene nell'erogazione di un certo tipo di servizio può rivelarsi disastroso per un altro), che alla qualità del lavoro eseguito non corrisponde necessariamente la qualità del servizio, e che spesso per una gestione efficace, risultano molto utili conoscenze nell'ambito della gestione della produzione, del marketing e della gestione del personale; spesso infatti, la possibilità di buona riuscita del servizio, è funzione del grado di addestramento acquisito dal personale a volte ancor prima di entrare a far parte dell'azienda. Quindi la formazione del personale ed un attento monitoraggio della *customer satisfaction* e dei feedback può aiutare a mantenere standard elevati.

I requisiti salienti di un servizio possono essere classificati in base alla relazione tra la soddisfazione del cliente e il grado di adempimento delle specifiche.

In ciò ci viene in aiuto il Diagramma di Kano.

Si possono individuare tre tipi di requisiti¹⁸:

- *Requisiti di base o impliciti*: rappresentano quelle caratteristiche che il cliente si aspetta implicitamente dal servizio. Queste caratteristiche sono sottintese dal cliente; egli non tende ad esplicitarle in quanto le dà già per scontato. La loro presenza non genera soddisfazione, mentre la loro assenza genera un forte malcontento;
- *Requisiti espliciti o prestazionali*: rappresentano quelle caratteristiche che il cliente vuole nel servizio. Vengono espresse direttamente dal cliente, di conseguenza la loro presenza genera soddisfazione;
- *Requisiti “eccitanti”*: sono caratteristiche “accessorie” che il cliente non sa di volere e quindi non chiede; sono requisiti che generano molto più facilmente soddisfazione.

Dalla figura 11 è possibile notare che sull’asse delle ascisse si posiziona il grado di adempimento delle specifiche mentre sull’asse delle ordinate la soddisfazione del cliente:

¹⁸ Dispensa del Corso di Sistemi di Produzione di Beni e Servizi, A.A. 2008/2009

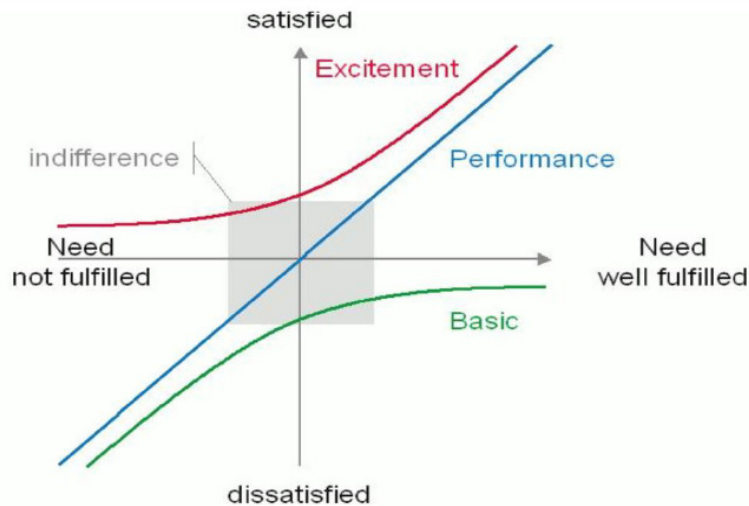


Figura 17: Diagramma di Kano

Secondo il modello di Kano esistono due tipi di qualità:

- Qualità *spoken*, quella che il cliente richiede espressamente, e che coincide con i requisiti espliciti o prestazionali. È caratterizzata da una relazione lineare tra il grado di adempimento delle specifiche e la soddisfazione del cliente, all'aumentare dell'una aumenta anche l'altra (curva blu);
- Qualità *unspoken*, quella di cui il cliente non parla, o perché non ne è a conoscenza (*qualità excitement*) o perché la ritiene implicita nel prodotto (*qualità basic*). La prima coincide con i requisiti "eccitanti", si tratta di una qualità latente (eccitante), perché quando il cliente la prova si entusiasma proprio perché non ne era a conoscenza (curva rossa). La seconda, invece, coincide con i requisiti base o impliciti, ed è rappresentata dalla curva verde.

Il futuro di un'azienda è tutto basato sulla *qualità excitement* (o latente): i clienti, infatti non si accontentano mai e desiderano sempre qualcosa di nuovo.

6.2 - Lean HealthCare

Quando si parla di settore sanitario si è in presenza di aziende di servizi e non più produttive, questo implica che i principi e le tecniche della Lean Manufacturing precedentemente esposti dovranno essere applicati in un contesto completamente diverso.

L'applicazione della Lean Manufacturing, all'industria dei servizi sanitari, consente di passare ad una visione orizzontale dell'erogazione del servizio, rendendo più agevole l'identificazione dei reali fabbisogni del paziente, e l'individuazione di tutte quelle attività che in realtà non creano valore aggiunto per il paziente e che dunque potranno essere ridotte o eliminate.

Tra le aree di forte criticità, comuni a tutte le strutture, quelle sulle quali la Lean Healthcare potrebbe intervenire con successo rintracciamo:

- Gestione delle liste di attesa chirurgiche;
- Riduzione dei costi su tutte le aree;
- Mancanza di posti letto;
- Aumento del tasso di infezioni acquisite in ospedale;
- Appropriatezza delle cure;
- Lunghezza della degenza.

Come è intuibile, la progettazione di un piano di azione, gestito da una task force adeguatamente formata, potrebbe introdurre i principi Lean e intervenire direttamente sulle problematiche più diffuse, riducendone l'impatto. In ogni caso, la riorganizzazione dei processi non avrà alcun effetto positivo se contemporaneamente non si proceda all'adeguamento dell'atteggiamento del personale nelle mansioni quotidiane, attraverso l'adozione comune della cultura Lean.

Inoltre, è errata l'assunzione che attraverso l'approccio Lean, si aumenti il carico di attività di un processo quando quest'ultimo è in ritardo rispetto alla TimeLine aziendale. Al contrario, per allineare quel processo ai tempi corretti, sarà necessario riorganizzarlo e nel farlo, emergeranno quasi sicuramente una serie di attività superflue da eliminare. Questo è un caso molto frequente in Sanità, dove riscontriamo che uno dei principali problemi è il tempo di attesa per una qualsiasi prestazione. A tal merito, il primo passo è censire quali attività lo compongono e identificare quelle a valore aggiunto e quelle senza valore.

Secondo Berwick il valore non deve essere visto come un concetto a livello individuale, ma come una proprietà di sistema, per cui il punto di leva sarà il sistema, e non l'individuo. Berwick dice *“every system is perfectly designed to achieve exactly the results it gets”*¹⁹ (“ogni sistema è perfettamente progettato per ottenere esattamente i risultati che ottiene”); in sostanza se il valore in un sistema è da migliorare, saranno da migliorare gli aspetti operativi e socio-tecnici di tale sistema.

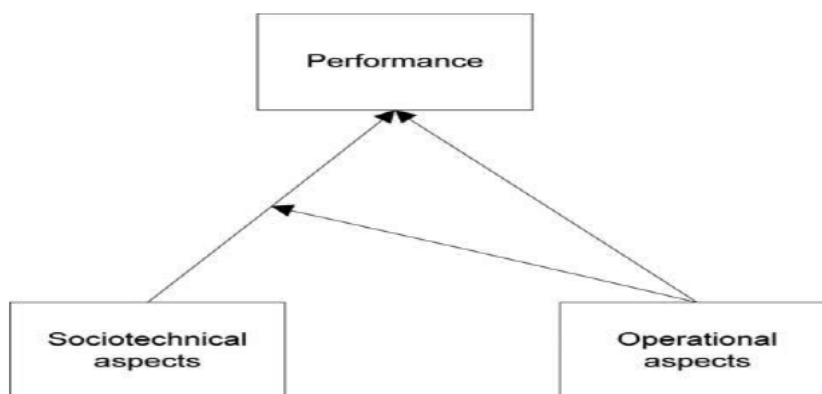


Figura 18: Quadro concettuale del Lean Thinking

Fonte: *D.M.Berwick, Improvement, trust, and and the healthcare workforce, Qual Saf Health Care 2003*

¹⁹ D.M.Berwick (2003), *“Improvement, trust, and and the healthcare workforce”*, Qual Saf Health Care.

Nicosia invece afferma che *“In sanità il valore ha caratteristiche peculiari, molto profonde.”* La gestione del paziente comporta una serie di responsabilità e difficoltà nelle scelte poiché non è in gioco un lotto di merce, bensì la vita e la salute di un essere umano. In questo senso la Lean Healthcare si propone di porre il paziente al centro del processo e degli sforzi dell’organizzazione, senza andare ad alterare in alcun modo il codice etico e deontologico del medico.

Detto ciò, è necessario rintracciare qual è il valore finale di quel processo (visita, analisi, intervento ecc.) per l’utente e focalizzare l’attenzione sulle attività che contribuiscono al raggiungimento di questo valore. Se l’identificazione del valore non è facilmente visibile, lo stesso vale per lo spreco. Si dovrà fare attenzione agli sprechi di materiali, di spazi, di tempo e di lavoro delle risorse, e trovare il modo per eliminarli via via.

Per aumentare la creazione di valore negli ospedali si proceda prima a comprendere quali sono i flussi primari di attività in cui è direttamente coinvolto il paziente e poi quali, altrettanto importanti, di supporto al funzionamento della struttura.

Una caratteristica degli ospedali e del sistema sanitario intero, che aggiunge complicazioni, è che il soggetto principale dei processi, il paziente, attraversa trasversalmente più funzioni di una struttura e più strutture.

La sfida della Lean Healthcare è la creazione di valore aggiunto attraverso la gestione del flusso del paziente nel suo insieme, coordinando il sistema sanitario nella sua interezza e gestendo il percorso del paziente dall’inizio alla fine.

Un’altra peculiarità delle strutture ospedaliere, risiede nell’alto grado di variabilità della domanda di prestazioni, sia in termini di volume sia in termini di tipologie di cure da erogare. La variabilità conduce senza dubbio all’aumento degli sprechi e in questo senso è necessario cercare di gestirla.

Lo snellimento dei processi di cura, il coordinamento tra le risorse appartenenti a flussi diversi ma in stretto contatto tra loro, la visione del sistema sanitario a tutto tondo, l'eliminazione degli sprechi e l'aumento di creazione del valore per il paziente sono soltanto alcuni tra i punti da inserire in un piano di progetto di ampie vedute che abbracci tutto il sistema sanitario orientandolo verso l'approccio Lean Healthcare.

Il processo che si realizza all'interno dell'azienda ospedaliera è la "trasformazione" da paziente malato ("prodotto" input del processo) a paziente sano (output del processo) in uscita dalla struttura ospedaliera. Il servizio sarà detto di qualità qualora soddisfi il paziente.

Lo spreco in ambito sanitario si potrà quindi definire come una qualunque azione che non incrementi la soddisfazione del paziente.

Si riportano di seguito alcuni esempi dei sette sprechi in sanità²⁰:

1. SOVRAPPRODUZIONE

- Fornire più informazioni di quelle necessarie;
- Fornire più informazioni in anticipo;
- Analisi anticipate per favorire le esigenze del laboratorio;
- Fornire copie di un rapporto a persone che non lo hanno richiesto e che non lo leggeranno.

2. TEMPO

- Attesa del paziente (dalla chiamata all'arrivo) per problemi di reparto;
- Attesa per tempo di cambio paziente (pulizia sala, cambio strumentario, preparazione chirurgo);
- Attesa di materiale sterilizzato per variazione lista;
- Attesa per materiale di impianto per variazioni intervento;

²⁰ A. Galgano, *IL SISTEMA TOYOTA NELLA SANITA': più Qualità meno Sprechi*, Roma 9 Maggio 2006

- Attesa per trasporto del paziente a causa di variazioni della sequenza operatoria.

3. TRASPORTO

- Spostamenti pazienti per trattamenti;
- Spostamenti pazienti da un posto all'altro;
- Spostamenti campioni;
- Recupero e trasporto sacche di sangue non programmate da Centro Trasfusionale (per le urgenze).

4. PERDITE DI PROCESSO

- Cambio di posizionamento del paziente;
- Apertura di più kit per mancanza di uno standard;
- Eccessiva produzione di documenti cartacei;
- Procedure non necessarie.

5. SCORTE

- Eccessivo immagazzinamento di materiali;
- Obsolescenza materiali e medicinali;
- Documenti in attesa di elaborazione;
- Provini in attesa di analisi.

6. MOVIMENTI

- Recupero strumenti non inclusi nel set chirurgico;
- Recupero strumenti da altra sala operatoria;
- Ricerca pazienti;
- Gestione documenti cartacei.

7. DIFETTI

- Errori nelle medicazioni;
- Informazioni mancanti;
- Ripetizione anestesia per problemi;
- Ripetizione interventi.

La persistenza di file di attesa, soprattutto, può aggiungere molti giorni di degenza ai pazienti già ricoverati in ospedale. Questo crea un ulteriore ciclo di potenziale errore e spreco per quei pazienti che stanno in ospedale più del necessario, a causa di un'infezione o più, che rischiano di entrare così in una spirale di aggravamento. Ed è proprio per evitare tutto ciò che in ambito sanitario si fa sempre più ricorso a tecniche e strumenti Lean Manufacturing.

Usando i principi Lean, i processi vengono continuamente migliorati o riprogettati per eliminare gli sprechi, richiedendo meno personale e meno rilavorazioni e quindi migliorando la qualità.

È possibile sviluppare un sistema-ospedale che preveda un flusso senza interruzioni inutili:

- Progettando in anticipo il miglior percorso dei pazienti ad esempio portando il servizio a loro, piuttosto che farli “rimbalzare” da una parte e dall'altra;
- Creando cellule aperte e flessibili, con strumentazioni e staff intercambiabili;
- Aggiustando sul posto il processo che si interrompe, senza rinviare il problema;
- Utilizzando per attività produttiva eventuali tempi di fermo macchina del processo;
- Mettendo il processo in modo tale che non si debbano ripetere più volte le stesse cose;
- Usando la tecnica del one piece flow.

Le 5 criticità con le quali tutti gli ospedali sono chiamati a fare i conti:

- trovare un'adeguata conformità fra le aree di Pronto Soccorso (PS) e di Emergenza e il resto dell'ospedale per accogliere, trattare, dimettere o ricoverare con la massima appropriatezza;

- gestire le liste di attesa chirurgiche;
- ridurre i costi globali;
- considerare la lunghezza delle degenze;
- l'aumento delle infezioni acquisite in ospedale.

Dall'interazione fra questi punti emerge che la lunghezza della degenza è la madre di tutte le altre criticità, avendo impatto sulle infezioni, sui costi, sulle liste di attesa, e sull'impossibilità di ricovero per mancanza di posti letto disponibili.

Da alcuni anni anche vari ospedali italiani hanno intrapreso un cammino di trasformazione "Lean" che gradatamente sta interessando un numero sempre crescente di aree e servizi ospedalieri. E' in corso una intensa co-progettazione dei nuovi percorsi assistenziali basati sul flusso dei pazienti e dei nuovi spazi fisici che saranno chiamati ad ospitarli.

Con la formazione sul modello Lean Healthcare è iniziata una continua ricerca dei momenti di spreco nelle linee di attività cliniche e amministrative.

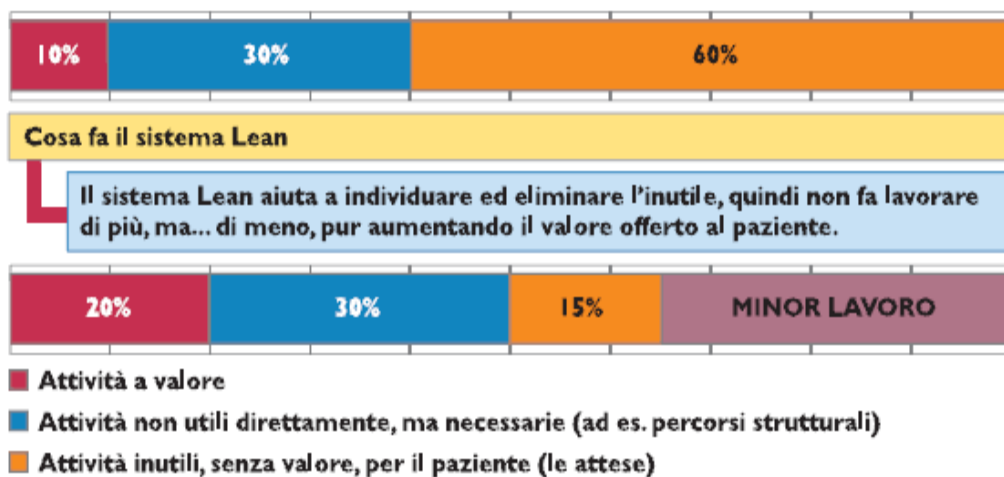


Figura 19: Gestione tipica delle attività in un processo reale

Fonte: Articolo, Così «Lean» taglia gli sprechi, *Il sole24ore*, 2011

Dalla figura 13 si evince che con l'applicazione della metodologia Lean Healthcare è possibile incrementare le attività a valore di circa il 10% e ridurre le attività inutili di circa il 45% ottenendo di conseguenza una riduzione pari al 35% del lavoro da eseguire.

Un ulteriore strumento Lean che può essere considerato, nell'ambito dell'erogazione dei servizi sanitari, è la standardizzazione. Consiste nello stabilire una sequenza di lavoro standard che sia ben eseguita ogni volta che deve essere fatto una particolare tipologia di attività. Il vantaggio di realizzare tale sequenza è quello di permettere ai lavoratori di riconoscere se c'è qualcosa di sbagliato in tempo reale e di intervenire per risolvere il problema immediatamente.

Tale strumento consente di migliorare le attività di cura nei confronti del paziente. Per raggiungere standardizzazioni ottimali occorre avere chiari i tre elementi che le compongono:

- ✓ Il tempo nel quale una funzione (ad esempio il lavaggio ferri, o la preparazione del carrello ferri chirurgici, oppure la sanificazione dell'ambiente) può essere svolta;
- ✓ La sequenza lavorativa, cioè l'ordine in cui le operazioni devono essere svolte;
- ✓ La quantità minima di beni strumentali perché quella funzione sia svolta regolarmente (numero e tipologia di ferri, numero e tipologia di strumenti per la sanificazione, posizionamento dei ferri per la migliore funzione).

La standardizzazione è un approccio praticabile in sanità, anche se ciò dipende dalla tipologia di variabilità presa in considerazione.

È possibile classificare la variabilità sanitaria in:

- ✓ Variabilità Clinica (ad esempio: differenti patologie; differenti livelli di gravità; differenti risposte alla terapia)
- ✓ Variabilità dei comportamenti clinico assistenziali (ad esempio: differenti livelli di abilità; differente training; differenti approcci e preferenze di trattamento; differenti comportamenti)

- ✓ Variabilità nei flussi (ad esempio: accessi programmati; accessi di emergenza/urgenza)

La natura della variabilità può essere:

- Naturale: variabilità ineliminabile, imprevedibile ed insita/connaturata allo stato delle cose.
- Artificiale: variabilità eliminabile attraverso interventi di natura organizzativa e spesso legata a comportamenti scorretti ed inadeguati. E' causata da disfunzioni nei processi.

Detto ciò è possibile identificare la natura attinente le variabilità nel sanitario, come riportato in figura 14, così da poter intervenire sulle variabilità di natura artificiale per cercare di ridurle e operare così la standardizzazione del lavoro.

	Naturale	Artificiale
Variabilità Clinica	X	
Variabilità dei comportamenti	X	X
Variabilità dei flussi	X	X

Figura 20: Natura delle variabilità in ambito sanitario

Fonte: *Giusepi I., (2012), "La gestione operativa in sanità - Operations management"*

6.2.1 - Modello per intensità di cure

Ormai da tempo si sta procedendo verso l'adozione di modelli di gestione dell'organizzazione e delle strutture sanitarie, alternativi rispetto al tradizionale impianto per reparto, che abbracciano la trasversalità dei processi di cura del paziente.

Si ricercano soluzioni ad un "vecchio" ospedale non più in grado di offrire efficienza, nonostante i suoi crescenti costi assistenziali. All'ospedale moderno si chiede di mettere al centro la persona e le sue necessità, di aprirsi al territorio e di integrarsi con la comunità sociale.

Una risposta tangibile è venuta dalla Regione Toscana che, prima in Italia, ha dato il via con la legge Regionale 40/2005 al piano di riordino ospedaliero prevedendo una grande innovazione non solo di natura edilizia, ma per l'appunto anche gestionale. La nuova organizzazione prevede una razionale valorizzazione delle risorse, ponendosi come definizione e scelta di appropriatezza organizzativa, il cui nuovo assetto è funzionale a garantire qualità ed efficienza di un Sistema centrato sul malato. Si tratta di una sorta di rivoluzione metodologica e organizzativa, la cui filosofia si fonda proprio sul moderno *lean thinking* (pensare snello), che prevede un sistema flessibile, competitivo, di lotta agli sprechi, e che auspica un utilizzo di risorse strettamente necessarie.

Il modello di intensità differenziata di cure ospedaliere può essere definito come il modello organizzativo strutturato per aree / settori dedicati a pazienti con esigenze omogenee di cura.

E' funzionale a garantire qualità ed efficienza di un sistema centrato sul malato.

Con il progetto "Ospedali per intensità di cura" si favorisce un approccio multidisciplinare, che supera una visione "d'organo", di reparto monodisciplinare e di dipartimento omogeneo. Attraverso la realizzazione di moduli/aree di ricovero "aperte e variabili", graduati per intensità di bisogno

assistenziale, si pensa che sarà possibile superare le criticità legate alla gestione del posto letto. Non dovendosi necessariamente “costituire uno specifico reparto”, si rende più facile il superamento del principale ostacolo alla costituzione di poli specialistici cui affidare, secondo criteri di appropriatezza clinica, la gestione del paziente. Dunque, il modello ruota sull’ineludibile centralità del malato, garantita da nuove figure professionali definite *tutor*, ma capovolge la tradizionale organizzazione ospedaliera fondata sulle specialità.

La sfida al nuovo paradigma del concetto di cura, quindi, parte dalla cancellazione dell’identità strutturale dei differenti reparti, accorpati funzionalmente in grandi aree più tecnologiche e multidisciplinari, in cui le componenti cliniche vengono disgiunte da quelle organizzative, proprie delle cosiddette piattaforme produttive (dipartimenti, blocchi, ambulatori).²¹

6.2.2- Lean: literature review

L’analisi del quadro teorico di riferimento ha permesso di individuare le linee di ricerca seguite e le posizioni concettuali ritenute più significative, sul tema dell’applicazione del Lean Thinking nei settori produttivi (area in cui ha avuto origine) e in ambito sanitario, mediante le quali è stato possibile ricostruire l’evoluzione dei modelli prevalenti e gli orientamenti teorici e operativi attuali.

Per quanto riguarda il settore sanitario, è emerso, un interesse crescente verso il ruolo delle «associazioni di attori» per l’efficace implementazione del Lean.

Nello specifico, attraverso la teoria Actor-Network Theory (ANT) (Papadopoulos, 2010) è stato dimostrato come il passaggio ai Lean Processes in sanità rimanga un problema critico a causa dell’eterogeneità degli attori

²¹ Francesco Maria Bovenzi, (2011),” *Ospedale per intensità di cura: scegliere leggero, un compito pesante*”

coinvolti, per il quale è necessario elaborare strategie fondate sull'interazione di elementi tangibili e intangibili, guidate dalla diffusione e applicazione della conoscenza a ogni livello, attraverso la partecipazione sinergica alle attività di miglioramento continuo e di ottimizzazione dei processi. La ANT è un modello nel quale interagiscono attori umani (stakeholder) e non umani (tecnologie, documenti, processi). È utilizzato per esplorare e studiare le relazioni all'interno di sistemi socio-tecnici molto complessi, al fine di identificare le dinamiche intercorrenti tra elementi umani e non umani all'interno di un network e tra network e promuovere la nascita di nuove identità e forme di organizzazione ed interazioni, sostenibili nel lungo termine (Papadopoulos, 2010).

Tale prospettiva è presente, pur se con impostazioni differenti, in un'elevata percentuale di studi nei quali si enfatizza da un lato, l'importanza dei principi lean, rigorosamente adattati al contesto, che concorrono a configurare un «Sistema Sanitario Snello» (Perrella e Leggeri, 2007, Nicosia, 2010), dall'altro i concetti di valore e qualità derivabili dal modello suddetto e dalla letteratura prevalente in tema di dimensioni del valore (Young et al., 2004) e della qualità (Damiani e Ricciardi, 2006). Lo studio e l'analisi di tali impostazioni, secondo un'ottica multidimensionale, ha consentito l'elaborazione di una possibile rappresentazione del valore «globale», in termini di piena soddisfazione del customer.

Ottimisticamente, l'Istituto per Healthcare Improvement (Miller, 2005) afferma che non c'è un motivo per cui non dovrebbe essere possibile applicare i principi lean in un contesto di assistenza sanitaria e che dovrebbero creare gli stessi guadagni, in termini di efficienza e qualità, visti in altri settori. Tuttavia, in pratica, tradurre i principi e gli strumenti lean nelle cure sanitarie risulta impegnativo. Una delle difficoltà che si riscontrano sta nel definire il “valore”.

Il concetto di valore sotteso alla dimensione clinica è riconducibile al raggiungimento del migliore risultato in termini di salute per il paziente che, in realtà, dovrebbe essere condiviso da pazienti e operatori medici e sanitari.

La dimensione operativa attiene alle modalità di erogazione delle prestazioni e dei servizi, la cui valutazione deve avvenire in termini di rapporto benefici/costi. Il valore, in questo contesto, implica la soddisfazione dei bisogni ai costi più bassi.

La prospettiva esperienziale è connotata dalla capacità delle strutture sanitarie di mantenere nel tempo i rapporti con i pazienti, migliorando contestualmente il grado di soddisfazione, la fiducia e la fedeltà degli stessi e razionalizzando le attività di erogazione e di comunicazione. Quest'ultima dimensione richiama aspetti estremamente soggettivi ed emozionali, legati alla percezione del paziente in merito all'esito della prestazione ricevuta, alle modalità attraverso cui essa viene erogata e alla qualità della relazione che si crea con gli erogatori della cura.

Il concetto di valore multidimensionale è comprensivo di altre due dimensioni ritenute significative: economica e quella sociale.

La dimensione economica, a cui ci si riferisce in questa sede, è legata al fenomeno del rischio clinico; il concetto di valore è riconducibile, da un lato alla potenziale riduzione dei costi (in termini di risarcimenti) sostenuti dalle aziende, come conseguenza della diminuzione del livello del rischio clinico, dall'altro all'aumento dei ricavi conseguiti in virtù della «vendita» di maggiori prestazioni (il paziente preferisce recarsi presso le strutture ritenute complessivamente più affidabili). La creazione di valore è direttamente connessa, quindi, alla capacità di ottenere la soddisfazione del customer che, a sua volta, rappresenta la percezione che il customer avverte su quanto e come le sue attese e aspettative siano state realizzate (Perrella e Leggeri, 2007).

La dimensione sociale del valore attiene alla capacità di mantenere e migliorare lo stato di salute globale della popolazione, un fattore ritenuto

fondamentale per favorire il progresso e la vitalità economica di un Paese. In questo caso, l'aspetto relazionale della qualità riguarda sia le relazioni intercorrenti tra i professionisti all'interno delle strutture e le strutture e gli utenti, sia la creazione di network finalizzati alla diffusione e condivisione delle conoscenze e allo scambio di esperienze e best practices, all'interno del sistema sanitario, considerato nella sua totalità.

Il valore «globale» fondato sulla sistematica interazione degli elementi descritti, dovrebbe contribuire a semplificare il processo di individuazione dell'obiettivo principale di un'organizzazione snella, ovvero l'innovazione, attraverso la riduzione degli sprechi, il risparmio delle risorse e il miglioramento complessivo della struttura (Womack et al, 2005).

Nella letteratura si è, infatti, riscontrato un notevole interesse, sia per la natura e le caratteristiche degli sprechi che ricorrono più frequentemente in sanità, al fine di procedere alla corretta identificazione ed eliminazione degli stessi (Zidel, 2006; Perrella e Delli Rocili, 2008; Graban, 2009), sia per i fattori contestuali (complessità delle strutture) e culturali (resistenza al cambiamento), che concorrono a rallentare o, addirittura, a impedire la diffusione dell'approccio Lean nelle strutture sanitarie (McAuliffe et al., 2005; Radnor e Walley, 2008; Burgess et al., 2008; Carr et al., 2012).

Si sottolinea, inoltre, l'importanza delle principali tecniche e strumenti lean utilizzate nelle aziende sanitarie (Value Stream Mapping, Visual Management, 5S) per la riprogettazione degli aspetti operativi del processo di fornitura della cura, diretti a rilevare in una prima fase, lo status quo dell'area/dipartimento/azienda e successivamente lo stato ideale futuro, attraverso procedure di lavoro standardizzate (Lodge e Bamford, 2008; Dickson et al., 2009).

Altri studi, infine, sono fondati prevalentemente sulla sperimentazione del modello Lean nell'intera struttura o in alcune aree (progetti pilota).

Essi hanno dimostrato, attraverso analisi empiriche, l'effettivo miglioramento delle performance aziendali in termini di: riduzione dei costi, tempi di attesa

e di risposta; aumento della qualità, soddisfazione dei pazienti, del personale e dei fornitori (Laursen et al., 2003; Mechi e Gemmi, 2008; Kim e Spahlinger, 2006; Ghosh e Sobek, 2006; Nicosia, 2008; Radnor, 2011; Brannmark et al., 2012).

La complessità delle relazioni che emerge dal modello di gestione impone l'adozione di modalità di coordinamento delle attività diverse rispetto a quelle funzionali tradizionali, nuove modalità di erogazione della cura (nuovi processi), l'elaborazione di nuove strategie e l'acquisizione di una nuova cultura, focalizzata sui network intra-organizzativi (team work, a livello micro) e inter-organizzativi (collaborazione con altre strutture, a livello macro) (Bifulco, 2009). Secondo tale prospettiva dinamica, l'approccio Lean, in un contesto di complessità dei processi di cura, di diagnosi e trattamento, può essere definito disruptive innovation (Womack et al., 2005), dal momento che prevede lo sviluppo graduale di nuovo sistema, un nuovo modello di gestione che va a scontrarsi e sostituirsi con l'esistente, richiedendo un processo di strategia separato, prima di concretizzare il processo definitivo di integrazione ed essendo finalizzato a fornire un percorso continuo di cura (care coordination), focalizzato sui singoli bisogni dell'individuo (patient-centered), a garanzia di maggiore qualità e sicurezza.

L'obiettivo di qualsiasi iniziativa Lean è quello di concentrarsi sulle esigenze dei clienti al massimo livello, individuando ed eliminando definitivamente gli sprechi. Le unità sanitarie riconoscono il paziente come un cliente primario e come un fattore critico da prendere in considerazione durante la progettazione del processo e nel fornire assistenza. Inoltre, gli ospedali dovrebbero fornire servizi di assistenza sanitaria di qualità più elevata e dare maggiore attenzione ai pazienti puntando sempre sulle loro esigenze in sede di attuazione dei principi lean in ospedale. Gli studi dimostrano che orientare l'assistenza sanitaria in tutte le preferenze e le esigenze dei pazienti ha il potenziale per migliorare la soddisfazione dei pazienti con la loro cura, così come i loro risultati clinici. Così, l'ospedale che fornisce assistenza centrata

sul paziente avrà alcuni dei benefici finanziari quali minor costo per caso, diminuzione degli eventi avversi, ridotti costi operativi e l'aumento della quota di mercato.

Il sistema sanitario si trova ad affrontare le sfide e le opportunità di un ambiente operativo in rapido cambiamento, tra cui l'aumento delle aspettative sulla qualità dell'assistenza sanitaria.

Rexhepi e Shrestha affermano che c'è bisogno dell'approccio Lean nel settore sanitario. Questo non è applicabile solo negli ospedali, ma anche nelle cliniche private o casa di cura. Hagg et al., afferma che la Lean è uno strumento efficace per l'individuazione e l'eliminazione dei rifiuti dal processo. Il vantaggio e l'obiettivo di applicare l'approccio Lean al settore sanitario è quello di un approccio migliore per ridurre i rifiuti, oltre a ridurre i tempi di attesa, e per la costruzione di qualità, velocità e flessibilità nell'organizzazione.

Secondo Rexhepi e Shrestha l'attuazione dell'approccio Lean non è possibile senza il supporto del vertice aziendale. Pertanto, la leadership è riconosciuta come aspetto critico per l'attuazione di sistemi di qualità in approccio Lean.

L'organizzazione non può avere successo nella Lean se non ha la cultura necessaria, lavoratori qualificati, il buy-in dal top management e una forte leadership. Angueloy et al., afferma che il livello di impegno della leadership è un fattore che ha colpito molto l'esito della pratica lean.

Abdullah et al., ha sottolineato che il coinvolgimento dei dipendenti è uno dei fattori più importanti per il miglioramento della qualità. Questo perché, il comportamento dei lavoratori svolge un ruolo importante sulla percezione del cliente della qualità del servizio.

Oltre a ciò, l'aumento nel coinvolgimento del personale è stato motivo di un aumento della soddisfazione sul lavoro, dell'impegno, della soddisfazione del paziente e dei risultati per il paziente.

Implementato nel contesto sanitario, quindi, l'approccio Lean è comunemente usato per ridurre i ritardi dei pazienti in pronto soccorso, eliminare gli errori terapeutici e medici, e prevenire le procedure inadeguate. Tale approccio nel settore sanitario è adattato a partire dal Toyota Production System, e si concentra sul miglioramento dell'efficienza dei processi attraverso l'eliminazione degli sprechi, o muda, definiti come qualsiasi attività che consuma risorse, ma che non generano alcun valore salvifico per il paziente (Womack & Jones, 1996).

A rafforzare questa definizione, Naylor et al. (1999) descrivono Lean come lo sviluppo di un flusso di valore per eliminare tutti i rifiuti, compreso il tempo. Anche per l'assistenza sanitaria, le applicazioni lean di solito includono elementi come la creazione di forti collaborazioni tra fornitori, riducendo il numero di fornitori, muovendosi verso la fornitura di servizi just-in-time e cercando continuamente di migliorare la qualità (Aronsson et al, 2011). Lean, in questo senso, è uno sforzo di riprogettazione rivolto principalmente ad aumentare l'efficienza delle operazioni.

Lean è attualmente la serie più comune di strumenti volti verso la definizione di nuove strategie di gestione per la fornitura di assistenza sanitaria (Shah et al, 2008).

La revisione della letteratura esistente aiuta ad evidenziare anche lo spazio operativo circostante la Lean Production. Secondo Monden (1983) ci sono tre sotto-obiettivi per raggiungere l'obiettivo primario della riduzione dei costi attraverso l'eliminazione degli sprechi: il controllo della quantità, la garanzia di qualità e il rispetto per l'umanità, raggiungibili attraverso quattro principali concetti: JIT, automazione, forza lavoro flessibile e capitalizzazione su suggerimento dei lavoratori.

Secondo Ohno (1988) il TPS può essere descritto come un tentativo di rendere le merci, per quanto possibile, in un flusso continuo. La base del TPS

è l'eliminazione assoluta degli sprechi. I due pilastri necessari per sostenere il TPS sono il JIT e l'autonotation (automazione con un 'tocco' umano, trasferendo l'intelligenza umana alle macchine).

Molti libri permettono di percepire gli strumenti e i metodi di Toyota Production System (TPS), quello di Liker (2004), invece, è unico nella sua spiegazione dei principi più ampi del lavoro nella cultura Toyota. Nel suo libro, Liker spiega i sistemi di management, il pensiero e la filosofia che costituiscono il fondamento del successo di Toyota.

Secondo l'autore il Toyota Way può essere brevemente riassunto attraverso i due pilastri che lo sostengono: "Miglioramento Continuo" e "Rispetto per le persone". Miglioramento continuo, spesso chiamato kaizen, definisce il vero valore del miglioramento continuo che risiede nella creazione di un clima di continuo apprendimento e un ambiente che non solo accetta, ma in realtà abbraccia il cambiamento. Un tale ambiente può essere creato solo dove c'è rispetto per le persone, da cui il secondo pilastro della Toyota Way. Ciò si realizza offrendo sicurezza sul lavoro e cercando di coinvolgere i membri del team attraverso una partecipazione attiva nel migliorare i loro posti di lavoro.

L'autore descrive un sistema progettato per fornire alle persone gli strumenti per migliorare continuamente il loro lavoro. Esso può essere riassunto in 14 principi organizzati in quattro sezioni: la filosofia a lungo termine, il processo giusto produrrà i risultati giusti, aggiungere valore all'organizzazione attraverso lo sviluppo del proprio popolo e risolvere i problemi alla radice spingendo in questo modo continuamente l'apprendimento organizzativo.

L'idea di base del TPS è produrre il tipo di unità necessarie, al tempo necessario e nella quantità necessaria tale che possono essere eliminate inutili scorte di prodotti intermedi e finiti. Quindi quando si parla di TPS ci si riferisce anche al JIT.

Come già detto in precedenza, storicamente l'origine del concetto di Lean lo si colloca nella produzione automobilistica giapponese nel 1960, ma fu reso popolare da Womack nel libro "the Machine that changed the world" ("la macchina che ha cambiato il mondo") (1990). "La macchina che ha cambiato il mondo" è essenzialmente la storia del mondo di fabbricazione di automobili Toyota.

Il libro identifica la Lean Production come una tecnologia che sta rimodellando la produzione di automobili. La forza trainante l'adozione della Lean è la necessità di fornire una maggiore varietà di prodotti a minor costo con cicli di sviluppo più brevi.

La produzione snella deve essere considerata come una strategia per raggiungere una leadership ottimale, al di là del taglio dei costi. In primo luogo, essa aumenta la soglia di qualità accettabile ad un livello tale che la produzione di massa non riesce facilmente a raggiungere. In secondo luogo, offre sia l'espansione della varietà dei prodotti che una rapida risposta ai mutevoli gusti dei consumatori; qualcosa che la produzione di massa difficilmente riesce a contrastare, tranne grazie a prezzi sempre più bassi. Inoltre, può utilizzare pienamente l'automazione in un modo in cui la produzione di massa non può.

Lewis (2000) con la sua pubblicazione mira a stabilire quale impatto la Lean Production ha avuto sulle posizioni competitive globali delle imprese che l'hanno adottata. Il documento asserisce che la produzione snella è in grado di sostenere il vantaggio competitivo se l'impresa è in grado di appropriarsi dei risparmi di produttività che crea e di creare risorse strategiche per mantenere un vantaggio competitivo sostenibile. Il documento suggerisce che implementare la pratica di LM, impostata come un sistema integrato, è fondamentale e questo dovrebbe essere fatto in tutta l'organizzazione, inclusi sia i fornitori che i clienti.

S. Bhasin (2008) propone un sistema robusto che si concentra non solo sulle attività immateriali e intellettuali, ma comprende anche diversi orizzonti temporali e gli interessi dei diversi stakeholders. Il sistema proposto è un dynamic multi-dimensional performance (DMP) che permette di misurare con successo le performances della Lean. Questo abbraccia cinque dimensioni e dimostra di essere più robusto rispetto ai suoi predecessori e sottolinea la necessità di utilizzare un insieme ridotto di metriche multidimensionali che sono strettamente allineate alle strategie di un'organizzazione. Quest'ultimo documento giunge quindi alla conclusione che non è sufficiente concentrarsi esclusivamente sulla realizzazione di alcune pratiche LM e strumenti/tecniche a livello dello stabilimento/ fabbrica, ma è necessario partecipare ad aspetti come il cambiamento culturale e quello del Lean Thinking organizzativo.

C'è anche un ampio consenso sul fatto che il successo di una trasformazione Lean non dipende solo dall'applicazione di strumenti e tecniche, ma che per raggiungere i benefici sostenibili di questa, è necessario prestare attenzione al fattore umano e alla creazione di una cultura che sostiene la trasformazione Lean. J.K. Liker and M. Hoseus (2010) ritengono che la gestione delle risorse umane sia di fondamentale importanza per un'ottima implementazione della Lean. C'è quindi bisogno di un eccezionale sviluppo delle risorse umane diretto ad un programma Lean. I membri del team devono saper riconoscere le deviazioni da alcuni standard stabiliti come problemi; devono essere ben addestrati a capire le norme; non possono temere che esporre un problema possa causare loro delle difficoltà; devono essere addestrati nei metodi di problem-solving ed infine devono essere motivati per aiutare l'azienda a raggiungere i suoi obiettivi.

Anche Hines (2008) afferma che è necessario che le tecnologie, gli strumenti e le tecniche di Lean Manufacturing siano affrontate congiuntamente alla

strategia, alla leadership del management, al comportamento e all'impegno di tutti i membri della società e alla gestione dei processi.

Questi ultimi concetti di coinvolgimento dei membri dell'organizzazione per l'attuazione della Lean erano già stati discussi da C.L. Comm e D.F.X. Mathaisel (2005). Essi hanno concluso che un atteggiamento proattivo per affrontare il cambiamento in Lean, la leadership e l'impegno del top management e il cambiamento della cultura organizzativa si concentrano sul ruolo delle persone, sulla formazione, sulla comunicazione, su un adeguato sistema di indicatori, sul controllo e sorveglianza continui.

Nel documento di U. Dombrowski and T. Mielke (2013) è data una nuova definizione della leadership Lean e sono identificati i principi fondamentali. Al fine di raggiungere una migliore cultura di miglioramento, il leader all'interno di un'organizzazione Lean deve essere un modello per i suoi dipendenti. L'implementazione sostenibile di sistemi di produzione snella richiede un cambiamento giornaliero nella cooperazione tra lavoratori e dirigenti.

La letteratura che troviamo sulla Lean è spesso ricca di descrizioni di valori e filosofia associati alla produzione snella che sono meno facilmente misurabili rispetto alle pratiche. Ad esempio, secondo Schonberger, (1986) la filosofia della Lean Production si concentra su alcuni aspetti: evitare i sette sprechi cardinali e rispettare i clienti, i dipendenti ed i fornitori. Sicuramente questi sono principi che si riflettono nella realizzazione delle pratiche Lean.

Concetti ripresi anche da Bhim (2010), per il quale gli obiettivi di produzione snella sono ridurre i rifiuti, lo sforzo umano e le scorte, raggiungere il mercato in tempo e riuscire a gestire le scorte di produzione che sono altamente reattive alle richieste del cliente, producendo prodotti di qualità nel modo più efficiente ed economico possibile.

In particolare, uno studio sull'importanza dei clienti ed anche dei fornitori rispetto all'implementazione della Lean è stato fatto da J. Moyano-Fuentes (2012). Maggiori livelli di cooperazione con i fornitori non influiscono sull'intensità di adozione della Lean production, mentre maggiori livelli di cooperazione con i clienti hanno un effetto significativo. Inoltre, maggiore è la cooperazione con i clienti e le informazioni integrate con essi, maggiore è l'intensità di adozione della Lean Production.

La Lean Production è un sistema di gestione focalizzato sull'eliminazione delle tipologie di rifiuti individuate dal Ohno (1988) e di altri rifiuti connessi con la variabilità interna e la variabilità esterna prodotta da rapporti nella catena di approvvigionamento. Come tale, può essere definita come "un sistema socio-tecnico integrato il cui principale obiettivo è quello di eliminare gli sprechi e, contemporaneamente, ridurre il numero dei fornitori, clienti, e la variabilità interna" Shah e Ward (2007). L'obiettivo principale di questo sistema di gestione è, quindi, per ottenere la massima efficienza, effettuare operazioni ad un costo minimo e con zero scarti.

Questo principio è trattato anche da Antony J. (2011), secondo il quale il concetto di Lean Thinking originato da Toyota Production System (TPS), determina il valore di qualsiasi processo, distinguendo le attività a valore aggiunto da quelle a non valore aggiunto e tende all'eliminazione degli scarti in modo che ogni passaggio aggiunga valore al processo.

Scherrer-Rathje M. (2009) sostengono che non è facile essere Lean. La Lean è una filosofia di gestione focalizzata sull'identificazione e l'eliminazione degli sprechi lungo l'intero flusso di valore di un prodotto, che si estende non solo all'interno dell'organizzazione, ma anche lungo la rete di Supply Chain aziendale. Essa promette benefici significativi in termini di riduzione degli sprechi e maggiore comunicazione ed integrazione della Supply Chain. Tuttavia l'implementazione della Lean è un compito arduo, poiché bisogna raggiungere alti livelli di impegno organizzativo, autonomia dei dipendenti e

trasparenza per le informazioni necessarie per garantire il suo successo. Questo articolo descrive in dettaglio due progetti di implementazione Lean all'interno della stessa azienda. Il primo progetto fu un fallimento, mentre il secondo un successo. Sono giunti ad alcuni fondamenti necessari per un'implementazione Lean di successo: la Lean non avrà successo senza impegno nella gestione visibile, sviluppo dei meccanismi formali per consentire l'autonomia, rivelazione degli obiettivi Lean a medio e lungo termine, certezza che i meccanismi in atto siano per la sostenibilità della Lean a lungo termine, comunicazione delle vittorie Lean fin dall'inizio, valutazione continua durante l'implementazione della Lean.

Come si è visto molti sono stati gli autori che hanno studiato i diversi aspetti del concetto di Lean. Alcuni si sono soffermati maggiormente sui suoi principi fondamentali mentre altri hanno dato maggiore importanza alle tecniche ed agli strumenti per seguirli ed altri ancora hanno visto quali sono gli obiettivi raggiungibili.

6.3 - Agile Healthcare

L'Agile Manufacturing è un insieme di strategie organizzative, che ottimizzano l'erogazione dei servizi in ambienti con domanda volatile e con esigenze dei clienti altamente variabili, concentrandosi sul rafforzamento del sistema di risposta ai clienti attraverso il coordinamento delle risorse e una migliore flessibilità, ridisegnando strutture organizzative, sistemi informativi, processi logistici, ed euristiche decisionali di gestione.

L'assistenza sanitaria si divide in molteplici attività, per alcune delle quali la domanda è estremamente variabile e imprevedibile.

Proprio per questo motivo la metodologia Agile Manufacturing è stata recentemente proposta come un mezzo per migliorare l'erogazione dei servizi di assistenza sanitaria, ma le specifiche politiche gestionali e pratiche operative, non sono ancora state sviluppate per catene di servizi, tra cui l'assistenza sanitaria. Inoltre, l'efficacia comparativa delle singole pratiche Agile è sconosciuta, mentre concetti teorici Agile Manufacturing sembrano perfettamente adatti per migliorare la gestione delle organizzazioni sanitarie complesse.

Spinta dalla necessità di mantenere un vantaggio competitivo e ridurre al minimo i costi, la pratica comune nel settore sanitario è quella di definire i livelli di personale pari alla domanda media del servizio, in contrasto con il concetto di soddisfazione della domanda di picco.

Sebbene tali strategie di gestione del personale contribuiscono a ridurre i costi legati al lavoro, questa tendenza porta alla conseguenza indesiderata di unità di cura sempre più carenti durante i periodi di picco della domanda. Ciò comporta l'uso di lavoro straordinario eccessivo come soluzione di gestione per ottenere la flessibilità necessaria.

Il lavoro straordinario eccessivo è un problema diffuso nel settore sanitario. Straordinari eccessivi comportano maggiori tassi di fatica e di errore, rafforzando ulteriormente la tesi che la gestione sanitaria deve essere migliorata, e contribuendo all'aumento dei rischi per la sicurezza del paziente e al deterioramento della qualità delle cure.

Il settore della sanità è molto indietro, rispetto ad altri settori, per quanto riguarda il successo della gestione della supply chain nell'erogazione dei servizi. I manager sanitari hanno bisogno di nuove strategie di gestione del servizio per essere in grado di rispondere efficacemente ai cambiamenti nella domanda dei pazienti e per mitigare gli effetti negativi della variabilità della domanda sulla cura del paziente.

La supply chain nel contesto sanitario, è caratterizzata da alcune peculiarità uniche, che rendono difficile la diretta applicazione delle conoscenze raccolte, nei settori industriali, al settore sanitario. L'imprevedibilità e la domanda stocastica, guidano verso la necessità di servizi personalizzati. L'incapacità di mantenere in un buffer un prodotto finito, l'incertezza inerente la durata dei processi di cura, e altre caratteristiche specifiche delle attività del servizio sanitario hanno finora impedito l'applicazione operativa della metodologia Agile Manufacturing nel settore in esame.

La supply chain Agile deve identificare e sviluppare un nuovo paradigma di gestione del servizio che ospita l'incertezza e la variabilità intrinseca alle cure sanitarie, in particolare deve condurre operazioni di ricerca per migliorare la progettazione e la gestione della catena di servizi sanitari.

Stabilire la metodologia migliore da applicare nel settore healthcare non è una scelta facile poiché la struttura ospedaliera rappresenta una grande realtà contenente un numero elevatissimo di reparti e servizi ospedalieri dalle caratteristiche più diverse.

In questo elaborato si intende confrontare l'applicazione delle metodologie di gestione finora esposte, ad parte di una struttura nosocomiale, per poter valutare quale risulta essere la più opportuna, in termini di efficacia ed efficienza.

Si è pensato di valutare l'applicazione della metodologia Agile Manufacturing all'ambito sanitario, anche se finora non è stata mai resa operativa, poiché in tale contesto la domanda è spesso imprevedibile ed, essendo molteplici le patologie che possono colpire il paziente, è necessaria una marcata flessibilità da parte del personale specialistico così da poter prontamente trovare la cura più adatta da somministrare ad ogni paziente, dove tale cura dovrà essere personalizzata.

Per minimizzare i tempi di attesa dei pazienti, è necessario che il personale medico e infermieristico lavori in sincronia, così da rendere il servizio fortemente reattivo.

6.3.1 - Agile: literature review

Nel paper “*Adapting Agile Strategies to Healthcare Service Delivery*” di Rust, Saeed, Isa Bar-On e Pavlov, viene definita l’Agility come “a fundamental characteristic of successful supply chains faced with volatile and unpredictable demand, and has been suggested as a promising new paradigm for improving healthcare delivery” ovvero come paradigma che consente di migliorare l’ambiente Healthcare, essendo l’Agility una caratteristica adatta a supply chain aventi domanda imprevedibile e volatile. Nell’articolo appena citato inoltre si definisce l’Healthcare come uno dei sistemi più complessi da gestire. Per migliorare tale gestione viene proposta la clinical network ovvero la creazione di una rete di medici, abolendo l’idea di medico come singolo individuo, che si occupa di prestare assistenza sanitaria. Secondo gli autori, i manager sanitari devono trovare nuove strategie di gestione dei servizi, per poter rispondere in modo efficace alle variazioni della domanda dei pazienti e per mitigare gli effetti negativi della variabilità della domanda sulla cura del paziente.

Sia dal punto di vista teorico sia da quello pratico, le operazioni di servizio sanitario sono in procinto di cambiare significativamente.

Tuttavia, la variabilità e imprevedibilità inerente alla domanda di assistenza sanitaria e le operazioni interne, rendono questo settore difficile da gestire (Li et al, 2002). Nuovi sviluppi nella condizione del paziente, i risultati diagnostici inaspettati o reazioni sorprendenti ai farmaci possono richiedere

cambiamenti improvvisi nei processi pianificati con effetti a catena in tutta la supply chain del servizio.

Kane et al (2007) trova che la mancata corrispondenza tra le risorse e picchi della domanda è la principale fonte di fatica e ridotta qualità delle cure nella maggior parte dei servizi sanitari. Con i sistemi di gestione attualmente in vigore, questa variante porta a errori di erogazione delle cure e un aumento dei rischi per la sicurezza del paziente. In particolare, le sollecitazioni da variabilità poste sul sistema sanitario portano ad un aumento degli errori di medicazione, infezioni nosocomiali, pazienti più gravi, e sono una delle principali cause di risultati avversi per il paziente (Needleman et al, 2002; Berens, 2000; Pronovost et al, 1999).

Garantire la corretta sincronizzazione delle risorse per soddisfare i picchi e le depressioni della domanda, è chiaramente di importanza cruciale per fornire un'assistenza sanitaria di alta qualità.

La pratica di gestione comune nel settore sanitario è quello di soddisfare le fluttuazioni della domanda tramite lavoro straordinario poiché la necessità di mantenere un vantaggio competitivo e ridurre i costi, spinge a definire i livelli di personale pari al fabbisogno medio di servizi (Litvak et al, 2005).

I comuni strumenti Lean, come Total Quality Management, o Six Sigma non funzionano in modo efficace in sistemi con alti livelli di variabilità del processo intrinseco (Lee, 2004).

Gran parte della ricerca sulla supply chain nei servizi è ancora teorica e poco focalizzata sull'operatività (Sampson e Froelhe, 2006); attualmente non fornisce un grande aiuto per i manager impegnati a ridisegnare i sistemi sanitari e l'integrazione dei processi di cura.

I manager sanitari, infatti, affrontano un divario significativo nella conoscenza intorno alla progettazione ottimale e alla gestione di sistemi di cura complessi che garantiscono una cura efficace del paziente.

Al momento in letteratura non è stato possibile trovare ulteriori autori che hanno trattato l'argomento dell'Agile Healthcare, mentre invece sono state rintracciati molteplici autori che hanno definito l'Agile in ambito manifatturiero.

Agility, come concetto nel settore manifatturiero, è stato coniato da un gruppo di ricercatori dell'Iacocca Institute, Lehigh University, nel 1991 per descrivere le pratiche osservate e considerate come aspetti importanti della produzione durante la loro indagine.

I sostenitori dell'Agility di Iacocca Institute of Lehigh University (USA) l'hanno definita come: "...A manufacturing system with extraordinary capabilities (Internal capabilities: hard and soft technologies, human resources, educated management, information) to meet the rapidly changing needs of the marketplace (speed, flexibility, customers, competitors, suppliers, infrastructure, responsiveness). A system that shifts quickly (speed, and responsiveness) among product models or between product lines (flexibility), ideally in real-time response to customer demand (customer needs and wants)..." ("... Un sistema di produzione con capacità straordinarie (capacità interne: tecnologie hard e soft, risorse umane, management educato, informazioni) per soddisfare le rapide mutevoli esigenze del mercato (velocità, flessibilità, clienti, concorrenti, fornitori, infrastrutture, reattività). Un sistema che si sposta velocemente (velocità e reattività) tra i modelli dei prodotti o tra le linee di prodotto (flessibilità), idealmente in risposta in tempo reale alle richieste del cliente (bisogni e voleri dei clienti)").

Agility, successivamente, da Goldman et al. (1995) è stata definita, in termini di risultati, come "dynamic, context specific, aggressively change embracing and growth oriented...succeeding...winning profits, market share and customers... Agility means delivering value to customers, being ready for change, valuing human knowledge and skills, and forming virtual partnership" ("dinamico, specifico contesto, che abbraccia in modo

aggressivo il cambiamento ed è orientato alla crescita...riuscendo...in profitti vincenti, quote di mercato e clienti...Agilità significa fornire valore ai clienti, essere pronti per il cambiamento, valorizzando conoscenze e competenze umane, e formando un partenariato virtuale”).

Qualche anno dopo anche Jain (2008) nelle sue ricerche ha cercato di valutare l'agilità di un'azienda riconoscendo i criteri di flessibilità, redditività, qualità, innovazione, proattività, velocità di risposta, costi e affidabilità. In altre parole, l'agilità è la capacità di un'azienda di crescere in un mercato competitivo e volubile, di rispondere repentinamente ai rapidi cambiamenti dei mercati guidati dalla valorizzazione dei prodotti e dei servizi basati sull'esigenze dei clienti.

Concetto ripreso e ampliato successivamente da Gosling (2010) che individua l'Agile nell'integrare efficacemente la Supply Chain e creare relazioni strette a lungo termine con clienti e fornitori.

Focalizzando l'attenzione sulla produzione, Gehani (1995) e Gligor e Holcomb (2012) concordano che un'organizzazione agile, in grado di soddisfare rapidamente gli ordini dei clienti, può introdurre nuovi prodotti frequentemente in modo tempestivo e può anche entrare e uscire rapidamente dalle sue alleanze strategiche.

L'agilità, però, non dovrebbe essere equiparata solo con la velocità di fare le cose, perché va oltre la velocità e richiede enormi cambiamenti strutturali e infrastrutturali. Equiparare la produzione agile con velocità di risposta o di flessibilità è una visione ristretta di ciò che costituisce agilità.

Ulteriori approfondimenti sull'agilità potrebbero essere acquisiti guardando i temi specifici e operativi. Ad esempio, secondo Kidd (1994), per rendere operativa l'agilità, essa può essere definita come “a combination of a number of enterprises such that each has some core skills or competencies that they contribute to a joint business operation.

This enables the cooperative enterprises to adapt and respond quickly to changing customer requirements” (“la combinazione di un numero di imprese ciascuna delle quali ha alcune abilità di base o competenze che portano ad una operazione di avventurarsi in comune consentendo in tal modo alle imprese cooperative di adattarsi e rispondere rapidamente alle mutevoli esigenze dei clienti”). Anni dopo anche Gligor e Holcomb, (2013) definiscono l’agilità in termini operativi come “a firm's ability to quickly adjust tactics and operations within its Supply Chain to respond or adapt to changes, opportunities, or threats in its environment” (“la capacità di un'azienda di regolare rapidamente le tattiche e le operazioni all'interno della sua catena di fornitura per rispondere o adattarsi ai cambiamenti, opportunità o minacce nel suo ambiente”).

Una definizione abbastanza specifica e succinta di agilità è stata proposta da Kumar e Motwani: “...ability to accelerate the activities on critical path and...time-based competitiveness” (“... la capacità di accelerare le attività sul percorso critico e ... la competitività basata sul tempo”). In altre parole le organizzazioni agili sono in grado di competere sulla base della ‘compressione’ del tempo.

Vastag (1994) definisce l’agilità come “Intra-enterprise and inter-enterprise integration for flexibility and speed to market, as enabled by technologies for advanced manufacturing, communication and transportation” (“Integrazione internamente all'azienda e con le altre aziende per la flessibilità e la velocità di mercato, come consentito dalle tecnologie per la fabbricazione avanzata, dalla comunicazione e dal trasporto”), guardando all’agilità come quella caratteristica di cooperazione per raggiungere la flessibilità; pensiero presente anche nella definizione di Yusuf (1999) che asserisce che “Agility is successful exploration of competitive bases through the integration of reconfigurable resources and best practices in a knowledge-rich environment to provide customer-driven products and services in a fast changing market

environment” (“ L'agilità è l'esplorazione di successo delle basi competitive, attraverso l'integrazione delle risorse riconfigurabili e delle migliori pratiche in un ambiente ricco di conoscenze per fornire prodotti e servizi orientati al cliente in un contesto di mercato in rapida evoluzione”) ed in quella di Mason-Jones et al. (2000) che affermano che “Agility means using market knowledge and virtual corporation to exploit profitable opportunities in a volatile market place” (“Agilità significa usare la conoscenza del mercato e la società virtuale per sfruttare le opportunità di profitto in un mercato volatile”).

Questa cooperazione è di fondamentale importanza per la competitività come sottolinea Yusuf (2004). Egli afferma che la catena Agile ha un forte impatto sulla competitività, perché consente la mobilitazione delle risorse globali per monitorare le modifiche dell'evoluzione tecnologica e lo sviluppo dei materiali, nonché i cambiamenti del mercato e le aspettative dei clienti.

Come si è visto caratteristica dell'agilità che spesso viene sottolineata è la flessibilità così come si può evincere anche dalle ricerche di Aitken (2002), nelle quali l'agilità è considerata come la capacità di avere visibilità della domanda, risposta flessibile e rapida e operazioni sincronizzate.

Riuscire tramite cooperazione a saper gestire i mercati volatili, comporterà benefici anche economici per l'azienda come descritto da Mason-Jones and Towill, (1999) che asseriscono che l'Agile consiste nell'utilizzo di conoscenza del mercato e società virtuale per sfruttare le opportunità redditizie in ambienti di business volatili.

S.L Goldman e R.N Nagel (1993) definiscono la produzione Agile come una produzione che “assimilates the full range of flexible production technologies, along with the lessons learned from total quality management, 'just-in-time' production and 'lean' production” (“assimila l'intera gamma di tecnologie di produzione flessibile, insieme con le lezioni apprese dalla total

quality management, produzione jit e lean production”); quindi la produzione Agile è, in questo caso, vista come una forma evolutiva del sistema produttivo, perché sintetizza e incorpora approcci precedenti del Lean Thinking.

Questo concetto è ripreso da Van Hoek, Harrison, & Christopher, (2001) i quali arrivano alla conclusione che “Agility is all about customer responsiveness and market turbulence and requires specific capabilities that can be achieved using ‘lean thinking’” (“l’agilità è tutto ciò che riguarda la risposta ai clienti e le turbolenze di mercato e richiede capacità specifiche che possono essere raggiunte usando ‘Lean Thinking’”).

Sempre sulla linea del pensiero lean e della riduzione degli sprechi, ma con un occhio anche alla flessibilità e alla riduzione dei tempi di consegna tipici del pensiero Agile, Romano (2000), e successivamente Agarwal (2006), Prajogo and Sohal (2006), Naim and Gosling (2011), Kisperska-Moron and de Haan (2011) con diverse sfumature sostengono che l’Agile implica flessibilità efficace e gestione della qualità per ridurre gli sprechi ed evitare insoddisfazione del cliente. Sono richieste anche strategie di prodotto e differenziazione del servizio, nonché le misure di performance di qualità del prodotto, innovazione di prodotto e di processo; tutti orientati verso la flessibilità e riduzione dei tempi di consegna.

Ma agilità non è vista solo come semplice flessibilità e cooperazione, essa tende alla soddisfazione del cliente come definita da Flidner & Vokurka, (1997) “Agility is an ability to produce a broad range of low-cost, high quality products with short lead times in varying lot sizes, built to individual customer specification” (“L’agilità è la capacità di produrre una vasta gamma di prodotti a basso costo, di alta qualità in tempi brevi in varie dimensioni di lotto, costruiti su specifica del singolo cliente”).

Pensiero ripreso successivamente anche da Christopher (2000) che definisce l'agilità come "...the ability of an organization to respond rapidly to changes in demand, both in terms of volume and variety" ("...la capacità di un'organizzazione di rispondere rapidamente alle variazioni della domanda, sia in termini di volume che di varietà").

Definita similmente anche da Braunscheidel e Suresh, (2009) che misurano l'agilità sulla base di risposta alla domanda, programmazione congiunta, risposta ai clienti e visibilità.

I drivers e le caratteristiche che rendono un'impresa agile sono definiti da molti autori, tra questi ricordiamo Bohdana Sherehiy (2007) che li individuano in flessibilità, reattività, velocità, cultura del cambiamento, integrazione e bassa complessità, prodotti personalizzati e di alta qualità e la mobilitazione di competenze di base. Charlene A. Yauch (2007) nel suo articolo analizza gli attributi che un team aziendale dovrebbe avere per agevolare la produzione Agile descrivendone quattro: le squadre agili sono multifunzionali, al fine di coniugare le conoscenze e le competenze necessarie per arricchire il cliente; dinamiche e in rapida riconfigurazione, dando la flessibilità dell'organizzazione e consentendo una rapida riconfigurazione; cooperative, sia all'interno che tra le imprese, per consentire una cooperazione internamente ed esternamente all'azienda necessaria per migliorare la competitività; virtuali, che permettono all'azienda di combinare le risorse (persone e informazioni), se necessario, al fine di perseguire obiettivi imprenditoriali.

Agli inizi degli anni 2000 si rivela di fondamentale importanza l'informazione e la comunicazione tramite l'IT così come descritto da Swafford (2008) che vede un diretto collegamento tra l'utilizzo dell'IT con la flessibilità che a sua volta, come più volte detto, è una caratteristica principale delle Supply Chain Agile; importanza già presente nell'articolo di Ashish Agarwal (2007) il quale, utilizzando una modellazione interpretativa

strutturale, hanno ottenuto le interrelazioni delle variabili che influenzano l'agilità della Supply Chain. Queste variabili sono state classificate in base al loro potere di driver e di dipendenza. Tutte e tre le variabili, cioè l'uso di strumenti informatici, la pianificazione centralizzata e collaborativa e l'integrazione di processo, sono drivers importanti, quindi, devono essere la priorità assoluta per una Supply Chain Agile.

Ovviamente, diversi aspetti di agilità sono stati sottolineati da vari autori, ciò ha portato a varie opinioni riportate nella letteratura. Queste definizioni dovrebbero essere considerate contemporaneamente al fine di ottenere una migliore comprensione di ciò che costituisce agilità.

Sebbene ciascuna delle definizioni tratte da ricerche/pubblicazioni evidenzia questioni distinte, ci sono temi comuni a tutte. I più presenti possono essere riassunti come sensibilità del cliente, sfruttamento dell'impatto delle persone e delle informazioni, cooperazione e buona gestione del cambiamento e dell'incertezza.

CAPITOLO 7

Elementi innovativi di modellazione Agile

7.1- Matrice delle caratteristiche: Lean vs Agile

La Lean Production è una metodologia che si pone come obiettivo, strategico e gestionale, la massimizzazione del valore attraverso la minimizzazione dello spreco.

Si definisce Agile Manufacturing, invece, l'insieme delle strategie organizzative che si concentra sul rafforzamento del sistema di risposta ai clienti attraverso il coordinamento delle risorse e una migliore flessibilità, ridisegnando strutture organizzative, sistemi informativi, processi logistici, ed euristiche decisionali di gestione.

E' possibile riepilogare le principali caratteristiche delle metodologie esposte finora attraverso la tabella 3, così da rendere chiaramente visibili le differenze tra la Lean Production e l'Agile Manufacturing.

Tabella 3: Caratteristiche Lean Production vs Agile Manufacturing

		LEAN	AGILE
AMBIENTE	DI	Make to order	Make to order
PRODUZIONE			
LOGICA		Pull / Just in time	Pull / Just in time
OBIETTIVO		One Piece Flow	Reactivity Flow
STRATEGICO			
GRANDEZZA	DI	$Takt = \frac{T_{disp}}{D}$	
RIFERIMENTO		$I = \frac{Lt}{Tv.a.}$	

CARATTERISTICHE STRUTTURALI	<ul style="list-style-type: none"> Layout Manutenzione Setup Qualità Movimentazione Logistica Personale 	Cell Design Total Productive Maintenance (TPM) Single Minute Exchange of Dies (SMED) Quality Function Deployment (QFD) KANBAN 5S	Cell Design Single Minute Exchange of Dies (SMED) Quality Function Deployment (QFD) Agile Information Technology Knowledge Culture
CARATTERISTICHE FUNZIONALI	<ul style="list-style-type: none"> Processo Produzione Rilevazione Anomalie Miglioramento Fornitura 	Standard Hejiunka Jidoka Kaizen Supply Chain Network	Modular Product Design Flexible Jidoka Kaizen Corporate Partners
STRUMENTI	<ul style="list-style-type: none"> Analitici Pianificazione e Controllo 	Value Stream Map (VSM) $T_{pitch} = Takt * n$ $EPEI = \frac{Tic}{Tdisp}$	Value Stream Map (VSM)

In ambiente Lean Production è possibile assumere una serie di grandezze di riferimento, tra cui il Takt Time, mentre, al momento, in letteratura, per l'Agile Manufacturing, non è stata esplicitata analiticamente alcuna caratteristica.

Nei paragrafi seguenti si cercherà di ricavare delle grandezze di riferimento valide in ambiente Agile Manufacturing.

7.1.1 - Lean Manufacturing: calcolo analitico del Takt Time

E' bene distinguere tra domanda aggregata e domanda puntuale.

Per domanda puntuale si intende la domanda in un determinato istante di tempo.

In ambiente Lean Manufacturing l'andamento della domanda puntuale assume valore costante poiché la quantità richiesta è sempre la stessa in ogni istante di tempo. Per cui, tale domanda, si definisce livellata.

Di conseguenza, la domanda aggregata, che altro non è che la sommatoria delle domande puntuali in un certo orizzonte di tempo, avrà andamento lineare e crescente.

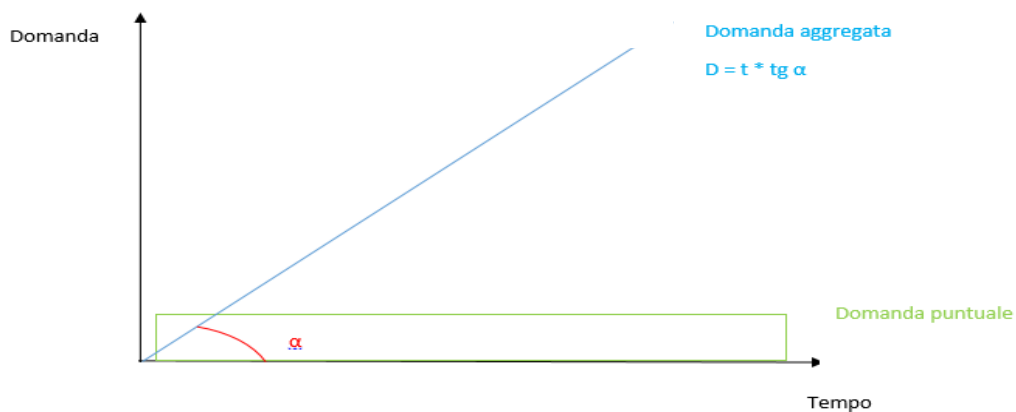


Figura 21: Esempio dell'andamento della domanda in ambiente Lean Manufacturing

La domanda aggregata sarà, quindi, pari a:

$$D(t) = t * \text{tg } \alpha$$

dove t rappresenta il tempo totale, cioè l'orizzonte temporale di riferimento.

L'incremento puntuale di domanda aggregata è pari alla derivata della domanda aggregata rispetto al tempo t, per cui sarà pari a $\text{tg } \alpha$ assumendo quindi valore costante.

$$D'(t) = \text{tg } \alpha$$

Il tasso di incremento puntuale della domanda aggregata è pari alla derivata rispetto al tempo dell'incremento puntuale della domanda aggregata ed assumerà valore nullo.

$$D''(t) = 0$$

È possibile definire, a questo punto, il Takt Time che è pari a:

$$\text{Takt Time} = \frac{\text{Tempo Disponibile}}{\text{Domanda Aggregata}}$$

Sostituendo nella formula la domanda aggregata si ottiene:

$$\text{Takt Time} = \frac{1}{\text{tg } \alpha}$$

che non è altro che l'inverso dell'incremento puntuale di domanda aggregata.

Tale quantità assume valore costante e rappresenta il tempo che intercorre tra due uscite successive di prodotto appartenente alla domanda aggregata.

7.1.2 - Scelta della grandezza Agile d riferimento

Per la determinazione della caratteristica d'interesse, in ambiente Agile, è possibile ricorrere a due differenti approcci, quello deterministico e quello stocastico.

È bene sottolineare che, per ricavare una grandezza d'interesse per l'Agile, si fa riferimento all'ambito manifatturiero, ma ciò solo per ragioni semplificative. In generale, la caratteristica di riferimento dovrà essere

adattabile anche ad altri settori di applicazione, come ad esempio quello dell'erogazione delle prestazioni sanitari.

Con l'approccio stocastico è possibile ricavare la probabilità che il sistema risponda alle variazioni della domanda, quindi la probabilità che il sistema sia reattivo e agile. Tale probabilità rappresenta il livello di servizio.

Per cui, con il seguente approccio, nel settore sanitario, ad esempio è possibile determinare la probabilità che il paziente si presenti al tempo t e che il paziente ha di essere servito in un determinato orizzonte temporale.

Con l'approccio deterministico, invece, è possibile ricavare il dimensionamento dell'impianto tale da risultare agile alle variazioni della domanda entro un limite massimo di agilità consentita dall'impianto stesso. Quindi permette di ricavare il dimensionamento che consente di ottenere un prestabilito livello di servizio. L'approccio stocastico, invece, consente di ricavare il livello di servizio a partire da un definito dimensionamento.

Poiché l'obiettivo del seguente elaborato di tesi è la riconfigurazione ottimale dell'impianto, in termini di risorse presenti, tale da massimizzare l'efficienza nell'erogazione dei servizi sanitari, in termini di tempi di esecuzione delle attività, nel rispetto dei principi stabiliti per la tutela della salute degli individui, nonché la massimizzazione dell'efficienza delle prestazioni offerte, si ritiene opportuno utilizzare l'approccio deterministico.

Inoltre, tale scelta è condizionata dal fatto che il sistema produttivo in esame ha dei vincoli a garanzia degli stakeholders. Questi ultimi sono rappresentati dai pazienti, ai quali non è possibile parlare in termini di probabilità di reazione del sistema ma è necessario assicurare loro la reattività e l'agilità del sistema entro un valore limite di domanda, essendo il vincolo preso in esame di natura sanitaria.

In particolare, nell'ambiente sanitario, il livello di servizio da assicurare al paziente deve essere pari al 100%.

Si ritiene necessario, quindi, utilizzare un approccio deterministico poiché è indispensabile avere un netto controllo sia sul sistema di erogazione dei servizi sanitari sia sull'agilità di reazione di quest'ultimo.

Per lo sviluppo analitico della caratteristica dei sistemi Agile, si suppone di avere una serie storica, che si compone di orizzonti temporali frazionabili in intervalli statisticamente significativi.

Per serie storica si intende la classificazione di diverse osservazioni di un fenomeno rispetto al tempo, il quale rappresenta un carattere qualitativo.

Nel seguente elaborato considero serie storiche deterministiche assumendo che i valori della variabile considerata possono essere esattamente determinati sulla base dei valori precedenti.

La suddivisione dell'orizzonte temporale dovrà essere fatta in intervalli statisticamente significativi, ovvero in intervalli che contengono almeno un'osservazione del fenomeno in esame. Inoltre tale intervallo sarà il minimo possibile, ovvero non sarà possibile considerare intervalli di ampiezza minore conservando la proprietà di significatività.

7.2 – Agile Manufacturing: formulazione della caratteristica di riferimento

In ambiente Agile Manufacturing il sistema deve garantire una reazione, ad una variazione di domanda, fino ad una massima variazione di produttività consentita dal sistema stesso.

In questo caso non è possibile parlare di domanda puntuale livellata e di conseguenza non si avrà un andamento linearmente crescente della domanda

aggregata. La domanda puntuale è variabile, poiché influenzata da un mercato in continua evoluzione.

Si immagina di avere un profilo di domanda nel tempo con evidenza a pettine come in Figura 16.

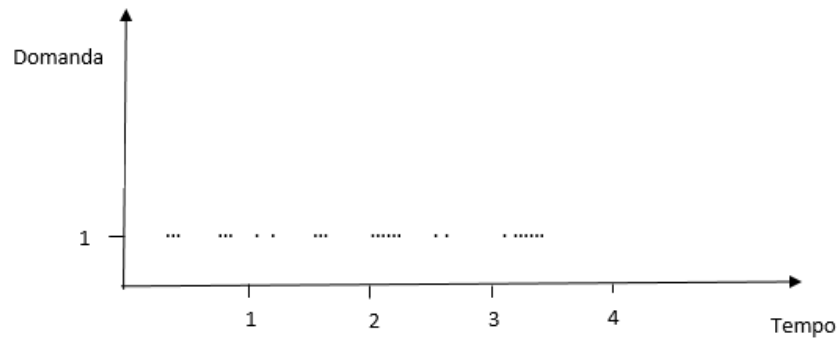


Figura 22: Profilo di domanda nel tempo

Tale profilo di domanda è costruito in ipotesi di non contemporaneità di due fenomeni, quindi nell'ipotesi in cui non è possibile registrare due richieste nello stesso istante di tempo.

Raggruppando detto profilo su intervalli di tempo successivi, statisticamente significativi, di identica durata, si ottiene un istogramma.

Unendo poi i punti medi di ogni barra dell'istogramma si ottiene un profilo che chiameremo *caratteristica di sollecitazione puntuale del sistema*.

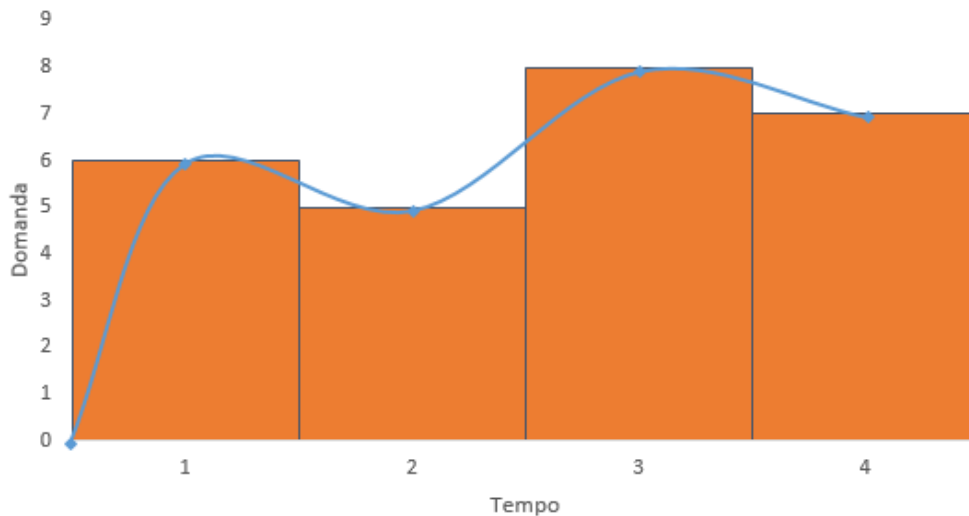


Figura 23: *Caratteristica di sollecitazione puntuale del sistema*

Il profilo in figura 17 altro non è che la curva di distribuzione delle frequenze di eventi domanda.

In regime di conoscenza deterministica del profilo di sollecitazione puntuale del sistema, è possibile conoscere “le curve” che compongono i successivi archi della caratteristica.

È perciò possibile conoscere:

- la frequenza F di sollecitazione per ogni intervallo, definito sulla condizione di significatività anzi detta;
- la massima accelerazione e decelerazione Δ imposta al sistema dalla caratteristica.

Un Sistema Agile è, quindi, un sistema produttivo per il quale valga l’ipotesi di caratteristiche di sollecitazione per le quali, a loro volta, valga il coefficiente Ω definito come di seguito:

$$\Omega = \frac{(F_{\max} - F_{\min})}{F_{\max}}$$

Tale quantità deve essere tale che:

$$\Omega = \frac{(F_{\max} - F_{\min})}{F_{\max}} \gg 0 \quad (1)$$

Analogamente per un Sistema Lean deve valere la condizione secondo cui:

$$\Omega = \frac{(F_{\max} - F_{\min})}{F_{\max}} = 0 \quad (2)$$

Se è vera la condizione (1), un sistema Agile è caratterizzato da 3 elementi che sono:

- ✓ F_{\max} ;
- ✓ F_{\min} ;
- ✓ $\Delta = \text{MAX} (F_{\max} - F_{\min})$.

Se è vera, come vera è, la (2), per un Sistema Lean si avrà che:

- ✓ $F_{\max} = F_{\min}$;
- ✓ $\Delta = 0$.

Nell'ipotesi di mancata frazionabilità dell'intervallo di significatività temporale che, peraltro, determina il carattere di minima durata dello stesso (cioè se viene scelto un intervallo più piccolo esiste almeno un orizzonte temporale nella serie storica considerata, per la quale si perde di significatività, nel senso che non ci sono registrazioni di fenomeni), non è possibile caratterizzare la distribuzione di frequenza di eventi all'interno di quell'intervallo, cioè non è possibile trovare una funzione di distribuzione rappresentativa dell'intervallo stesso.

In tale condizione, notiamo però che valgono le ipotesi della distribuzione di Poisson che sono di seguito riassunte:

- ❖ Si riceve una richiesta alla volta, ciò implica che non c'è contemporaneità nelle richieste per cui non si avranno lotti di richieste ma domande singole in ogni istante di tempo.
- ❖ $N(t+s) - N(t)$, ovvero il numero di richieste ricevute nell'intervallo $(t, t+s)$, è indipendente da $N(u)$ con $0 \leq u \leq t$, quindi è indipendente dalle

richieste ricevute negli istanti di tempo precedenti. Ciò implica l'indipendenza delle domande.

- ❖ La distribuzione di $N(t+s) - N(t)$ è indipendente da t , per ogni t , e $s \geq 0$.

Si dimostra che, sotto alcune condizioni, dalla distribuzione di Poisson è possibile arrivare alla distribuzione esponenziale e quindi è possibile sfruttare la funzione densità di probabilità, detta pdf, e la funzione di distribuzione cumulata, in breve Cdf, della distribuzione esponenziale. Per cui la distribuzione cumulata assumerà la seguente espressione:

$$D(t) = 1 - e^{-kt}$$

dove k rappresenta il tempo intercorrente tra due eventi successivi. È possibile notare che k non è altro che la frequenza prima definita.

La distribuzione cumulata rappresenta la domanda aggregata mentre la densità di probabilità rappresenta la domanda puntuale, in termini di probabilità. La pdf rappresenta, quindi, la probabilità che in un determinato istante di tempo si registri una richiesta.

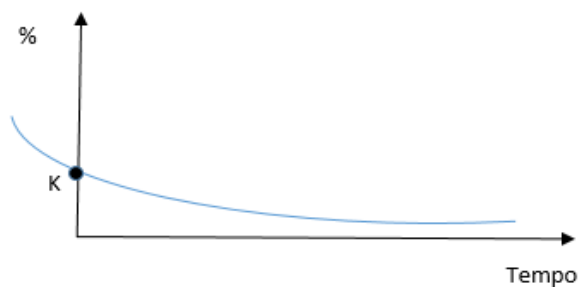


Figura 24: *Andamento pdf distribuzione esponenziale*

7.2.1 - Definizione di α e ρ

Se un Sistema Agile è definito secondo la condizione (1), cioè deve essere in grado di reagire a variazioni significative Ω della distribuzione di frequenza della domanda, è possibile definire due caratteristiche:

- ❖ ρ che rappresenta il ritmo di sistema. Per ritmo produttivo del sistema si intende l'inverso del Takt Time, quindi il rapporto tra la domanda e il tempo di disponibilità dell'impianto ed avrà come unità di misura il pezzo sul tempo;
- ❖ α che rappresenta l'agilità, anche detta reattività, intesa come capacità di reagire alle modificazioni di distribuzione della domanda.

Pertanto, sulla base di quanto detto finora, si ha che per i Sistemi in logica Lean Production valgono le seguenti equazioni:

$$F_1 = F_2 = t_{\text{TAKT}}^{-1} \quad (4)$$

$$\Delta = 0 \quad (5)$$

Per sistemi in logica Agile Manufacturing, invece, valgono le seguenti equazioni:

$$\rho_{\max} = k_{\max} \quad (6)$$

$$\rho_{\min} = k_{\min} \quad (7)$$

$$\alpha = \text{MAX} (k_{\max} - k_{\min}) \quad (8)$$

E' però necessario fare delle osservazioni su quanto finora detto.

In ambiente Agile Manufacturing, a rigore di logica basterebbero le equazioni (6) e (8). Nella realtà invece la condizione (7) risulta indispensabile, poiché il sistema deve essere progettato in modo tale da assicurare buone performance sia per l'efficacia che per l'efficienza. Il sistema deve assicurare buone performance di efficacia in corrispondenza del k_{\max} e quindi della massima domanda ma deve al tempo stesso assicurare

l'efficienza in corrispondenza del k_{\min} e quindi di bassi valori di domanda. Ad esempio, nel settore sanitario il sistema agile deve risultare efficace laddove si registrano picchi di domanda, ma allo stesso tempo efficiente nel caso in cui si registra l'arrivo di un solo paziente presso le strutture ospedaliere. Per rendersi reattivo ai picchi il sistema dovrebbe prevedere, ad esempio, un elevato numero di personale medico, ma nel caso in cui si registrano pochi accessi alle strutture, avere delle risorse inutilizzate genera una forte inefficienza. Quindi il sistema dovrà rendersi il più possibile agile ma nel rispetto dei vincoli di efficienza.

Un'altra osservazione, che si ritiene opportuno fare, riguarda i tempi di soddisfazione della domanda. Per sistemi Agile, a differenza di quelli Lean, si possono prevedere dei tempi di soddisfazione della domanda, denominati *delivery time* nella Teoria di Wortmman, differenziati da domanda a domanda.

Per *delivery time* si intende, quindi, la quantità di tempo necessario ad ottenere i prodotti pronti per la consegna.

Quanto detto finora vale nell'ipotesi in cui si registrano eventi domanda a *delivery time* costanti, ovvero indirizzabili su parti dell'impianto che realizza prodotti con *delivery time* simili tra loro. Ad esempio, nel caso dei sistemi sanitari, i *delivery time* dei pazienti risultano diversi da paziente a paziente essendo questi affetti da patologie differenti e quindi per ognuno di loro sarà necessario un *production time* diverso. Ma è possibile suddividere i pazienti in reparti o unità operative che trattano patologie pressoché simili quindi reindirizzare i pazienti in modo tale da avere in ogni reparto dei *delivery time* piuttosto omogenei.

7.2.2 – Osservazione sulla validità delle condizioni

Le condizioni ricavate, nei precedenti sottoparagrafi, hanno validità per sistemi impianto-mercato-ambiente monoprodotta e monoprocesso, denominati sistemi puntuali.

Nel caso in cui si considerano sistemi multiprodotta e multiprocesso, è necessario ragionare in termini matriciali.

Le matrici considerate saranno costituite da tante colonne quanti sono i prodotti e da tante righe quanti sono i processi eseguiti.

Tali matrici possono assumere dimensioni elevate nell'ambito sanitario essendo innumerevoli le possibili patologie, di cui i pazienti possono essere affetti, e i possibili percorsi di cura, a cui i pazienti possono essere sottoposti.

Le colonne della matrice sono rappresentate dalle attività costituenti il percorso di cura del paziente, come ad esempio il laboratorio di analisi chimico-cliniche, la radiologia, l'intervento, la degenza in terapia intensiva ecc. Le righe, invece, sono rappresentate dalle possibili cure da patologia.

Nel caso di Sistemi in logica Lean Production, l'obiettivo strategico è il one piece flow ed i processi che costituiscono il sistema risultano essere perfettamente accoppiati, ovvero tali processi usufruiscono del medesimo tempo di disponibilità. La matrice dei ritmi avrà, quindi, la seguente configurazione:

$$\begin{array}{c|cccc}
 & P_1 & P_2 & \dots & P_m \\
 P_a & \frac{D1}{T} & \frac{D2}{T} & & \frac{Dm}{T} \\
 P_b & \frac{D1}{T} & \frac{D2}{T} & & \\
 \cdot & \cdot & \cdot & & \cdot \\
 \cdot & \cdot & \cdot & & \cdot \\
 \cdot & \cdot & \cdot & & \cdot \\
 P_n & \frac{D1}{T} & \frac{D2}{T} & & \frac{Dm}{T}
 \end{array}$$

dove T rappresenta il tempo di disponibilità dell'impianto, ed è uguale per tutti i processi, mentre D_i è la domanda.

Tale matrice può essere riscritta mettendo in evidenza T e si ottiene:

$$T^{-1} \begin{vmatrix} D_1 & D_2 & \dots & D_m \\ D_1 & D_2 & \dots & D_m \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ D_1 & D_2 & \dots & D_m \end{vmatrix}$$

In ipotesi di perfetto accoppiamento dei processi e quindi di perfetto one piece flow è possibile affermare che:

$$\begin{vmatrix} F_1 & F_2 & \dots & F_m \\ F_1 & F_2 & \dots & F_m \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ F_1 & F_2 & \dots & F_m \end{vmatrix} = T^{-1} \begin{vmatrix} D_1 & D_2 & \dots & D_m \\ D_1 & D_2 & \dots & D_m \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ D_1 & D_2 & \dots & D_m \end{vmatrix}$$

In forma compatta si ha la seguente equazione:

$$\underline{F} = \underline{\rho} = T^{-1} * \underline{D}$$

ovvero la matrice delle frequenze naturali della domanda è uguale alla matrice dei ritmi del sistema, dove la matrice dei ritmi non è altro che il prodotto della matrice delle domande per l'inverso del tempo di disponibilità del sistema.

Nel caso di Sistemi in logica Agile bisogna distinguere tra due casi:

- ✓ Perfetto accoppiamento dei processi;

✓ Disaccoppiamento dei processi.

Nel caso in cui i processi godono della proprietà di perfetto accoppiamento, la matrice di riferimento assume la seguente configurazione:

$$\begin{array}{c}
 \\
 P_a \\
 P_b \\
 \cdot \\
 \cdot \\
 \cdot \\
 P_n
 \end{array}
 \begin{array}{cccc}
 P_1 & P_2 & \cdot & \cdot & P_m \\
 F_1(t) & F_2(t) & & & F_m(t) \\
 F_1(t) & F_2(t) & & & F_m(t) \\
 \cdot & \cdot & & & \cdot \\
 \cdot & \cdot & & & \cdot \\
 \cdot & \cdot & & & \cdot \\
 F_1(t) & F_2(t) & & & F_m(t)
 \end{array}
 = \underline{F}$$

In questo caso la matrice di riferimento è la matrice delle frequenze naturali della domanda. I pedici sono riferiti ai prodotti poiché le frequenze naturali sono indipendenti dal processo.

In ambiente Agile la frequenza è variabile nel tempo e quindi potrà assumere valori compresi tra un valore massimo ed un valore minimo in corrispondenza di k_{\max} e k_{\min} . I valori k rappresentano i limiti delle frequenze. Sono valori legati alla sollecitazione puntuale del sistema.

Sotto l'ipotesi di perfetto accoppiamento dei processi è possibile affermare che le frequenze assumono valore pari a k_{\max} e tali valori dipenderanno dai prodotti, ma per ogni singolo prodotto saranno uguali in tutti i processi necessari alla realizzazione del prodotto stesso. Per cui:

$$\text{Max } \underline{F}(t) = \underline{K}_{\max}$$

dove la matrice \underline{K}_{\max} è costruita come di seguito:

	P_1	P_2	. . .	P_m	
P_a	k_{max1}	k_{max2}		k_{maxm}	
P_b	k_{max1}	k_{max2}		k_{maxm}	
.	.	.		.	
.	.	.		.	
.	.	.		.	
P_n	k_{max1}	k_{max2}		$k_{max,m}$	

$$= \underline{K_{max}}$$

Il ritmo produttivo ρ dovrà essere pari a:

$$\rho \geq \sum k_{max j} \quad \text{con } j = 1..n$$

Da ciò è possibile fare una considerazione, secondo cui, in funzione del livello di servizio desiderato, è possibile calcolare il ritmo produttivo necessario alla realizzazione dello specifico livello di servizio.

Tale concetto è visibile in la Figura 19, che segue.

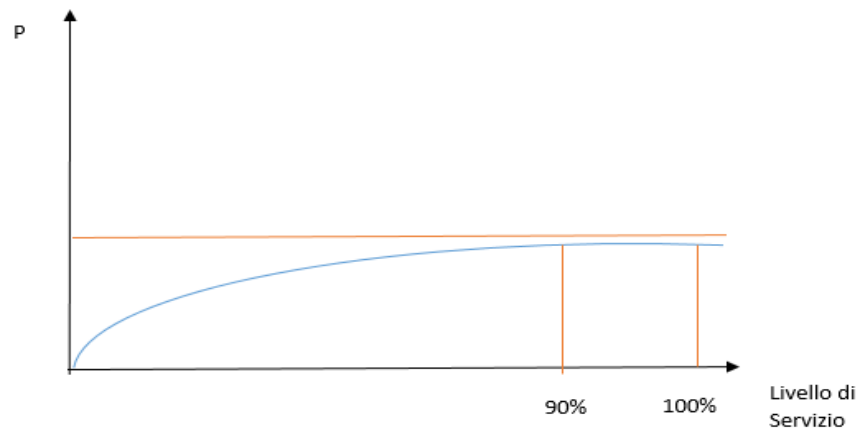


Figura 25: Grafico tramite cui è possibile calcolare il ritmo del sistema in funzione del livello di servizio

È possibile concludere che nel caso di perfetto accoppiamento dei processi, tutti i prodotti saranno a richiesta massima e quindi le frequenze naturali

saranno pari al k_{\max} e il ritmo produttivo del sistema, in logica Agile, sarà pari a:

$$\rho_A(t) = \rho_L(t) + \Delta\rho(t) \geq \sum k_{\max j} \quad \text{con } j = 1..n$$

dove la ρ_L coinciderà con la ρ_{\min} , ovvero il ritmo produttivo associato ad un sistema Lean coincide con il minimo ritmo.

Nel caso in cui, il Sistema in logica Agile non gode della proprietà di perfetto accoppiamento dei processi, non sarà più possibile affermare che la frequenza naturale assume valore pari a k_{\max} e che i k_{\max} sono indipendenti dai processi. In caso di disaccoppiamento, le frequenze, e quindi i k , sono dipendenti sia dai prodotti che dai processi, per cui le F assumo valore pari a k_{\max} ma tali k_{\max} non saranno uguali per tutti i processi relativi allo stesso prodotto, bensì saranno diversi.

Per quanto riguarda l'agilità α , invece, essa è pari alla derivata, rispetto al tempo, del ritmo del sistema:

$$\underline{\alpha} = \underline{\rho}'(t)$$

$$\text{dove } \underline{\rho}'(t) = \frac{\Delta\rho}{\Delta t}$$

Ma la derivata rispetto al tempo del ritmo produttivo, rappresenta la derivata rispetto al tempo di k , per cui si ottiene:

$$\underline{\rho}'(t) = \underline{K}'(t)$$

La derivata, rispetto al tempo, di k , a sua volta, rappresenta la massima variazione di k nell'unità di tempo, ovvero la reattività del sistema.

Per cui, è possibile affermare che un sistema in logica Agile deve avere un'agilità maggiore o uguale alla massima reattività dell'impianto e quindi:

$$\underline{\alpha} \geq \underline{\Delta K_{\max}}$$

In particolare guardando la matrice delle k , che segue:

$$\begin{array}{c}
 P_1 \quad P_2 \quad \cdot \quad \cdot \quad \cdot \quad P_m \\
 \left| \begin{array}{ccc}
 P_a & k_{\max 1} & k_{\max 2} & \dots & k_{\max m} \\
 P_b & k_{\max 1} & k_{\max 2} & \dots & k_{\max m} \\
 \cdot & \cdot & \cdot & \dots & \cdot \\
 \cdot & \cdot & \cdot & \dots & \cdot \\
 \cdot & \cdot & \cdot & \dots & \cdot \\
 P_n & k_{\max 1} & k_{\max 2} & \dots & k_{\max m}
 \end{array} \right| = \underline{K_{\max}}
 \end{array}$$

è possibile concludere che, in ipotesi di perfetto accoppiamento dei processi costituenti il sistema, l'agilità o reattività di processo deve rispettare la seguente condizione:

$$\alpha \geq \frac{\sum \Delta K_{\max j}}{\Delta t} \quad \text{con } j=1..m$$

mentre la reattività di prodotto deve rispettare la condizione secondo cui:

$$\alpha \geq \frac{\Delta K_{\max}}{\Delta t}$$

CAPITOLO 8

Pronto Soccorso: problema dell'abbandono e catene di Markov

8.1 Il reparto di pronto soccorso

Alla luce dell'importanza della gestione dei servizi sanitari resi ai pazienti in caso di emergenza, in questo capitolo si affronterà uno studio più approfondito del reparto di pronto soccorso, rappresentando e descrivendo nello specifico tutte le fasi caratterizzanti il processo di erogazione del servizio sanitario in questo particolare reparto.

Il pronto soccorso gioca un ruolo vitale nel fornire le prime cure ai pazienti ed è inoltre riconosciuto per il contributo che dà all'intera società. Le statistiche²² disponibili dicono che questo tipo di servizio sanitario è indispensabile per ogni paese che si affida a questo reparto per fornire servizi medici ai pazienti ventiquattro ore su ventiquattro. In USA ci sono state circa 107,5 milioni di visite all'Emergency Department solo nel 2001, più del 10% in più rispetto al 1997, mentre il numero di ospedali che offrivano questo servizio è drasticamente diminuito.

Per PRONTO SOCCORSO OSPEDALIERO si intende la Struttura complessa dedicata all'attività diagnostica e terapeutica d'emergenza e d'urgenza, funzionante in ospedale 24 ore su 24 ore. Compito primario del Pronto Soccorso è quello di gestire le emergenze e le urgenze mediche,

²² Dall'articolo: *Value Stream Mapping the Emergency Department*, reperibile on-line al sito: www.iienet2.org

stabilizzando i pazienti con alterazioni delle funzioni vitali per poterli poi affidare al reparto di degenza di competenza. Il Pronto Soccorso svolge anche attività ambulatoriale per le urgenze minori, di accettazione medica per le persone che necessitano di ricovero urgente (non già programmato – quindi – dai reparti di degenza) e di osservazione breve intensiva (OBI)²³.

In particolare un ospedale per essere sede di Pronto Soccorso, deve rispondere ai seguenti requisiti:

- gli interventi diagnostico - terapeutici di urgenza compatibili con le specialità di cui è dotata la struttura;
- l'esecuzione di un primo accertamento diagnostico clinico strumentale e di laboratorio;
- gli interventi necessari alla stabilizzazione dell'utente;
- il trasporto protetto.

8.1.1 Requisiti minimi strutturali²⁴

I locali e gli spazi devono essere correlati alla tipologia e al volume delle attività erogate.

L'unità minima dovrà prevedere:

- camera calda (area coperta e riscaldata di accesso diretto per mezzi e pedoni);
- locale per la gestione dell'emergenza;
- locale visita;
- locale osservazione;
- locale attesa utenti deambulanti e accompagnatori;

²³ Standard organizzativi delle Strutture di Emergenza-Urgenza, SIMEU-FIMEUC, Ottobre 2011

²⁴ http://www.sito.regione.campania.it/sanit%E0/delibere/requisiti_minimi/SEZIONEB.pdf

- locale attesa utenti barellati;
- locale lavoro infermieri;
- servizi igienici del personale;
- servizi igienici per gli utenti con vasca/doccia;
- locale/spazio per barelle e sedie a rotelle;
- deposito pulito;
- deposito sporco;
- spazio registrazione – segreteria – archivio;
- spazio/armadio per deposito materiale d'uso, attrezzature e strumentazioni;
- spazio/armadio per deposito attrezzature igiene ambientale;
- le superfici devono risultare resistenti al lavaggio e alla disinfezione, lisce con raccordo arrotondato al pavimento. Quest'ultimo deve essere non inquinante, del tipo monolitico, resistente agli agenti chimici e fisici, antisdrucchiolo.

8.1.2 Requisiti minimi impiantistici²⁵

Ogni unità deputata al pronto soccorso deve possedere i seguenti requisiti:

- impianto elettrico di emergenza, con gruppo di continuità per le tecnologie indispensabili di mantenimento dei parametri vitali;
- impianto di gas medicali.

²⁵ http://www.sito.regione.campania.it/sanit%E0/delibere/requisiti_minimi/SEZIONEB.pdf

8.1.3 Requisiti minimi tecnologici²⁶

La dotazione minima strumentale deve prevedere:

- elettrocardiografo;
- cardiomonitor e defibrillatore;
- attrezzature per rianimazione cardiopolmonare caratterizzate come minimo da:
 - un letto da rianimazione;
 - un ventilatore;
 - sistema monitoraggio respiratorio ed emodinamico;
 - lampada scialitica;
 - diafanoscopio a parete.

Le strutture deputate all'emergenza-urgenza si articolano su più livelli operativi e devono possedere requisiti tecnologici e dotazione strumentale adeguati alla tipologia e complessità delle prestazioni così come indicato nelle Leggi Regionali nn. 2/94 e 2/98.

8.1.4 Requisiti minimi organizzativi²⁷

Il Pronto Soccorso non è dotato di organico autonomo; l'assistenza è assicurata dal personale delle unità operative proprie del presidio. Deve essere garantita:

- la presenza di guardia medica attiva in anestesia/rianimazione, medicina, chirurgia, ostetricia - ginecologia, pediatria;

²⁶ http://www.sito.regione.campania.it/sanit%E0/delibere/requisiti_minimi/SEZIONEB.pdf

²⁷ http://www.sito.regione.campania.it/sanit%E0/delibere/requisiti_minimi/SEZIONEB.pdf

- la presenza di guardia medica o reperibilità in cardiologia, ortotraumatologia, laboratorio di analisi con banca del sangue, radiologia;
- la presenza di almeno due infermieri per turno.

Per ogni turno di presenza deve essere individuato il responsabile delle attività di Pronto Soccorso.

Nell'ambito dell'accettazione ospedaliera deve essere garantita la diversificazione organizzativa dell'attività di accettazione dei ricoveri programmati dall'attività di pronto soccorso.

Devono essere predisposti piani di emergenza interna (accettazione contemporanea di un elevato numero di pazienti).

Devono essere definite le modalità organizzative in riferimento alle situazioni di emergenza/urgenza psichiatrica.

8.2 Descrizione delle principali aree di pronto soccorso

Per meglio comprendere la successiva presentazione dell'intero processo che subisce un paziente del Pronto Soccorso, si riportano di seguito le principali aree di questo importante reparto:

8.2.1 Triage

Il Triage, quale primo momento di accoglienza delle persone che giungono in PS, è una funzione infermieristica volta alla definizione delle priorità assistenziali attraverso la valutazione della condizione clinica dei pazienti e del loro rischio evolutivo. Garantisce la presa in carico degli utenti e definisce l'ordine di accesso al trattamento. La funzione di Triage non riduce

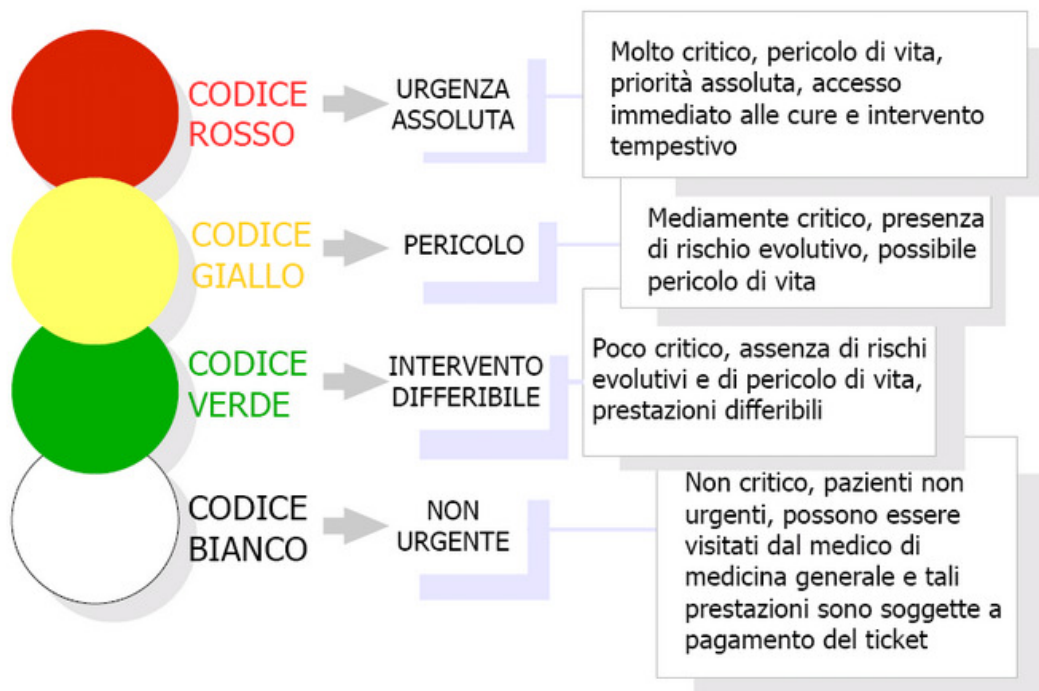
i tempi d'attesa di tutti gli utenti, ma li ridistribuisce a favore di chi ha necessità di interventi urgenti, assegnando ad ogni paziente un codice colore, come mostrato, anche, nella raffigurazione riportata in *Sch.3*:

- **CODICE ROSSO** (paziente in pericolo di vita): il Codice Rosso ha la priorità assoluta si tratta di pazienti con patologie molto gravi che portano all'alterazione o compromissione dei parametri vitali (respiro, circolo, coscienza). L'utente viene accolto e trattato immediatamente.
- **CODICE GIALLO** (paziente in potenziale pericolo di vita): sono pazienti con minaccia imminente di cedimento delle funzioni vitali (coscienza, respiro, circolo). Particolari aspetti contribuiscono all'assegnazione del codice:
 - intensità del sintomo;
 - parametri vitali alterati correlati al sintomo principale;
 - patologie tempo dipendenti.

Per i pazienti a cui è stato attribuito il “codice giallo”, l'accesso alla visita medica è, compatibilmente con altre urgenze/emergenze in atto, quasi immediato. Nel caso in cui, il paziente non sia trattato entro 20 min., deve essere rivalutato, controllando che non siano insorte condizioni per classificarlo come codice rosso.

- **CODICE VERDE:** paziente che necessita di una prestazione medica differibile (non necessita di una valutazione medica immediata e non ha segni/sintomi della severità descritta per i codici rossi e gialli); al momento della valutazione presenta le funzioni vitali integre e i parametri vitali nella norma. L'accesso agli ambulatori avviene dopo i codici rossi e gialli.
- **CODICE BIANCO:** i criteri per attribuire un codice bianco sono:

- non vi è alcuna alterazione delle funzioni vitali;
- non è presente alcuna sintomatologia critica o a rischio di aggravamento;
- la sintomatologia è minore o è presente da qualche giorno o è cronica;
- l'utente potrebbe trovare soluzione al problema utilizzando i percorsi sanitari alternativi al PS (MMG, Pediatri di libera scelta, Guardia Medica, Specialistica Ambulatoriale Esterna). L'accesso agli ambulatori avviene dopo i codici rossi, gialli e verdi.



Schema 3 - Codici Colore

8.2.2 Red Point

Area attrezzata per l'accesso dei codici rossi, in qualsiasi momento della giornata, con tutte le risorse necessarie per tutti gli interventi salvavita e pronta ad accogliere il team (anche multidisciplinare) che sarà chiamato ad intervenire. La commistione di tale area con l'area di gestione dei codici gialli (numericamente consistenti - fino al 20% in alcune realtà) rischia di ridurre l'efficienza (per consumo di risorse, affollamento, etc) del team nella situazione di emergenza assoluta. Ha un dimensionamento tale da prevedere il lavoro contemporaneo sul paziente di un team multidisciplinare (fino a 3 Medici e 4 infermieri).

8.2.3 Urgenza (Codici Gialli/Verdi)

Area con box/moduli multipli ed equivalenti e comunque attrezzati singolarmente per gli interventi salvavita, in cui si svolge la valutazione e il trattamento iniziale dei pazienti con codice giallo e verde. La caratteristica peculiare di quest'area è la FLESSIBILITA' del sistema nel far fronte anche a flussi supplementari di pazienti (crowding) con l'attivazione di nuovi punti visita/trattamento.

8.2.4 Codici Minori

Area destinata alla valutazione e al trattamento di pazienti con problematiche minori o specialistiche (laddove sono presenti percorsi di fast-track specialistico) che saranno pertanto non barellati e si prevede che abbiano necessità di singole prestazioni (interventi mono-risorsa).

8.2.5 Attesa Assistita

Area di stazionamento dei pazienti che devono completare il percorso diagnostico e/o terapeutico e quindi sono ancora in fase di "processo" (4-6 ore) e ovviamente necessitano di assistenza infermieristica e di assistenza alla persona. Inoltre in quest'area potrebbero trovare collocazione i "boarders", cioè quei pazienti che pur avendo completato il percorso

diagnostico/terapeutico e pur avendo già ottenuto la decisione di ricovero in regime d'urgenza, non riescono a trovare posto letto a causa del sovraffollamento dei reparti ospedalieri (Access block e Overcrowding).

I pazienti “destinati” al ricovero, che sono costretti a rimanere in Pronto Soccorso (per tempi talora molto lunghi), a causa della mancanza di posti letto nell'Ospedale, proprio perché presentano problematiche acute tali da richiedere il ricovero, assorbono una notevole quantità di risorse (umane e di tempo) che vengono sottratte alla loro funzione istituzionale (assistenza dei pazienti che si presentano al PS/h24).

8.2.6 Osservazione Breve (OB)

L'Osservazione Breve è riservata a quei pazienti per i quali non è possibile prendere una decisione di esito (dimissione o ricovero) nell'arco delle prime 4-6 ore dall'arrivo in PS e che necessitano pertanto di un periodo più lungo di valutazione per il raggiungimento della massima appropriatezza possibile nei ricoveri e nelle dimissioni. Tale area è collocata in uno spazio attiguo al PS cui è funzionalmente collegato. Il paziente può rimanere in regime di osservazione fino ad un massimo di 36 h.

Le funzioni :

- Osservazione longitudinale clinico-strumentale
- Approfondimento diagnostico
- Completamento e verifica di efficacia della terapia di urgenza

Criteri di ammissione e gestione:

- Il paziente deve presentare un solo problema clinico preminente da definire o risolvere.

8.2.7 Osservazione Breve Intensiva (OBI)

L'attività di quest'area comprende la gestione, in regime di degenza che non superi le 72 ore di pazienti con problemi clinici diversi e di diversa complessità, anche traumatologici e tossicologici, il cui iter diagnostico-terapeutico d'urgenza non è esauribile nelle poche ore a disposizione della gestione di Pronto Soccorso. La collocazione di tale articolazione organizzativa in area attigua ai locali del Pronto Soccorso garantisce la continuità diagnostico-terapeutica in urgenza.

8.3 Rappresentazione del processo sanitario di Pronto Soccorso

Per garantire la valutazione del processo produttivo sanitario, e per consentire una corretta analisi ai fini del miglioramento incrementale, occorre sviluppare una descrizione adeguata dei processi sanitari per permettere di identificare come ogni attività contribuisce al successo dell'azienda sanitaria incoraggiandola verso la logica del miglioramento continuo. Nella descrizione dei processi sanitari, la prima fase coinvolge l'individuazione delle attività che compongono il complessivo processo sanitario del paziente/utente.

Conoscendo cosa è effettivamente svolto durante un processo sanitario, è possibile giungere a porsi domande circa il come ed il perché determinate attività sono erogate. Si può osservare, così, come le attività sono combinate tra loro, come l'organizzazione le rende disponibili, con l'integrazione delle differenti unità partecipanti al complessivo processo, in quali tempi ed in quali luoghi. Nello specifico, il fattore tempo ha assunto una straordinaria importanza nella gestione aziendale e, in particolare modo, il *lead time*, il tempo effettivo d'attraversamento del processo. Misurando i tempi dei

differenti processi aziendali, si ottengono i *lead time* delle diverse fasi componenti il complessivo processo sanitario.



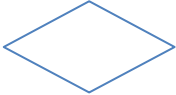

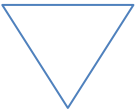

8.3.1 Il diagramma di flusso: descrizione e simbologia

La rappresentazione di un processo può avvenire tramite una mappatura: una descrizione sequenziale e dettagliata delle fasi del processo.

E' possibile utilizzare un diagramma di flusso, il quale consente di individuare le attività operative all'interno di ogni processo e le sequenze delle risorse umane, dei luoghi organizzativi, della tecnologia e del materiale di consumo.

Un *Process Flow Chart* (o diagramma di flusso) è uno strumento grafico che mostra i vari step presenti in un processo. In esso ciascuno step è rappresentato con un simbolo diverso a seconda del tipo di azione cui si riferisce ed è corredato da un breve testo al fine di chiarire ancor meglio il significato.

La simbologia tipicamente utilizzata in un process flow chart è quella indicata nel prospetto che segue.

Simbolo	Nome	Descrizione
	Terminator	Indica l'inizio o la fine del processo e funge da elemento di collegamento con altri processi.
	Process Step	Descrive una procedura o un'altra attività del processo.
	Decision	Simboleggia un nodo decisionale e gli steps successivi del processo.
	Data	Identifica la necessità di dati in input per lo svolgimento dell'attività che segue.
	Storage	Rappresenta un punto di immagazzinamento di un bene.
	Flow Line	Indica la direzione secondo cui fluisce il processo.

Prosp.3 - Simbologia diagramma di flusso

Il Process Flow Diagram è uno strumento modulare ed, in virtù di ciò, a seconda dell'obiettivo dell'analisi può essere applicato a tre livelli:

- a livello macro
- a livello medio

– a livello micro

A livello macro (ovviamente con riferimento all'ambito organizzativo) il diagramma illustra le varie fasi che le risorse in input ad un'azienda (in un contesto manifatturiero si tratta tipicamente di materie prime, componenti e/o semilavorati) devono attraversare per giungere alla produzione o erogazione dell'output dell'azienda stessa. Dunque, nel caso di prodotti materiali si potrà seguire l'iter relativo alla loro trasformazione fisica, concreta e visibile, mentre nel caso di prodotti immateriali od intangibili, quali sono ad esempio le informazioni, si potrà seguire il loro percorso di elaborazione.

Ad un livello medio di analisi è possibile descrivere i vari steps di trasformazione che si susseguono durante lo svolgimento di un sottoprocesso, inteso come una fase parziale di un processo di produzione aziendale. Ancora a tal livello, è possibile impiegare un process flow chart per dettagliare un processo svolto all'interno di una delle macrostrutture in cui risulta divisa l'impresa.

Infine, è possibile operare a livello micro. Ogni fase del processo produttivo è, infatti, a sua volta scomponibile in più operazioni elementari. In tal modo, si ha la possibilità di esaminare in maniera puntuale ciascuna di queste ultime.

Pur nella sua semplicità simbolica, un *process flow chart*, grazie alla sua duttilità ed alla sua modularità d'impiego, consente la rappresentazione grafica di realtà anche molto complesse. Tuttavia, come ogni strumento rappresentativo, esso possiede dei limiti; ad esempio, non potendo rappresentare la tridimensionalità, rende con difficoltà la presenza contemporanea di più processi che interagiscono, così come non è in grado di rappresentare la variabile tempo e le relazioni tra soggetti.

8.3.2 Il processo di attraversamento del reparto di P.S.

Descritte le principali aree di un P.S. e mostrata la simbologia utilizzata per descrivere il processo studiato, si posseggono gli strumenti per affrontare, nello specifico, il percorso effettivamente seguito dai pazienti all'interno del P.S., dall'ingresso all'uscita dal reparto.

L'infermiere del triage prende in carico tutti gli utenti in ingresso e decide se devono accedere alla sala visita oppure aspettare in sala di attesa, assegnando loro un codice colore che, come precedentemente specificato, dipende dalla severità delle condizioni del paziente. L'utente codice rosso (priorità assoluta) riceve immediatamente tutte le cure di cui ha bisogno nella zona del pronto soccorso ad egli dedicata; dopo aver ricevuto le prestazioni di primo intervento il paziente, se non è deceduto, viene ricoverato in un reparto specifico per il suo problema di salute.

Gli altri codici colore, condividendo le risorse del P.S., attendono in sala d'attesa la disponibilità di un letto, se necessario, o del medico. Priorità maggiore è del codice giallo che dovrà per primo ricevere il trattamento necessario; si comprende dunque che non saranno visitati prima i pazienti arrivati per primi, ma quelli più gravi. L'utente in attesa, può essere sottoposto a visita medica quando arriva il suo turno, ma superato un certo tempo può essere sottoposto a rivalutazione per verificare che le sue condizioni non si siano aggravate, nel qual caso cambia il codice colore ed il paziente avrà una priorità maggiore.

La "presa in carico" da parte dell'infermiere di triage, inizia con l'assegnazione del codice colore e termina con l'acquisizione informatica dell'utente da parte del collega in sala visita o del medico, quindi il triagista è responsabile del cliente, per tutto il tempo dell'attesa. Esiste, poi, la possibilità che un paziente, le cui condizioni di salute non siano gravi (codici bianchi e verdi), decida, superato un tempo massimo di attesa pari a circa 3 ore, di abbandonare il P.S. senza essere visitato.

I pazienti a cui è stato assegnato un codice verde e giallo condividono le risorse del reparto e si può assumere che subiscano un simile trattamento, perciò il percorso di questi utenti può essere descritto allo stesso modo.

Non appena si rende disponibile un letto, l'utente (prima il codice giallo e poi quello verde) lo occupa per essere visitato da un medico, il quale stabilisce la necessità di effettuare analisi radiologiche o di laboratorio.

Il medico effettua la sua diagnosi in base alla quale sceglie la terapia più opportuna da somministrare al paziente che viene effettuata, per un tempo di massimo 6 ore, in una zona detta *attesa assistita*²⁸ dove il paziente è controllato da un infermiere.

Al termine della terapia il medico visita nuovamente il paziente per stabilire se può essere dimesso, se deve essere ricoverato in reparto o se invece necessita di ulteriori controlli e la terapia non può considerarsi conclusa. In quest'ultimo caso l'utente può essere sottoposto ad un ricovero breve in P.S. nelle aree di O.B.I. oppure O.B., a seconda delle sue condizioni. Al termine del periodo di tempo (fino ad un massimo di 72 ore in OBI e 36 in OB) il paziente potrà essere ricoverato in reparto oppure dimesso.

Il diagramma di flusso di un generico pronto soccorso è presentato nel prossimo capitolo in Figura 5.1

8.3.3 Area per i codici bianchi

Il Pronto soccorso deve essere una struttura per pazienti in condizioni gravi o comunque con problemi che si possono curare solo in ospedale.

Oggi le strutture accolgono spesso persone che lamentano lievi malanni, non solo in Italia ma in tutto il mondo. Il risultato è il sovraffollamento e, di conseguenza, lunghe attese prima di poter essere visitati. Per di più si disperdono risorse ed energie in quanto il Pronto soccorso finisce per

²⁸ vedi sottoparagrafo 4.2.5

rappresentare una sorta di calderone, dovendosi occupare di problemi che vanno dalla caviglia slogata fino all'infarto. Negli ospedali più affollati si possono avere anche 300 accessi in un giorno e tre quarti dei casi, come avviene in tutte le strutture, non sono gravi. Almeno il 5%, poi, è rappresentato addirittura da accessi impropri (avrebbero dovuto rivolgersi al medico di famiglia).

Per risolvere questi problemi una possibile soluzione per gli ospedali, è di creare ambulatori al di fuori del Pronto soccorso da riservare ai codici bianchi. Per i pazienti non gravi potrebbero essere realizzati persino ingressi e uscite autonomi rispetto al Pronto soccorso. Questa, infatti, è la nuova proposta (maggio 2012) della regione Lombardia.²⁹ «Se una persona non ha un'alternativa, va a intasare una struttura che dovrebbe essere riservata alle emergenze - sottolinea Brazzoli³⁰ - Allora abbiamo pensato di creare un percorso fisicamente separato rispetto a quello del Pronto soccorso, per governare meglio la richiesta. Sono due modi diversi di fare ospedale e uno non deve intralciare l'altro. L'obiettivo è quello di utilizzare al meglio le risorse e di evitare disagi sia agli operatori dell'ospedale sia agli utenti».

Il personale in servizio al Pronto soccorso viene in sostanza suddiviso tra la struttura “tradizionale” e i nuovi ambulatori, infatti, alcune patologie si affrontano e si curano con calma in ambulatorio, mentre il Pronto soccorso non ha questa funzione, non serve a ottenere esami diagnostici evitando la trafila della prenotazione e l'attesa.

Molto spesso, però, i codici bianchi assegnati in ingresso rappresentano una piccola percentuale di tutti i pazienti che si rivolgono al pronto soccorso, che non giustificerebbe l'investimento di risorse per la realizzazione di un ambulatorio riservato ai codici meno gravi. Esiste, allora, per poter gestire al

²⁹ http://www.laprovinciadico.it/stories/cant%C3%B9%20%20mariano/287983_pronto_soccorso_separato_per_i_casi_non_urgenti/

³⁰ Giuseppe Brazzoli, direttore sanitario dell'azienda ospedaliera Sant'Anna di Como.

meglio il flusso dei pazienti, la possibilità di introdurre un quinto codice colore da poter assegnare al triage. A riguardo non esiste ancora una normativa specifica ma rappresenta il prossimo futuro degli ospedali. L'ulteriore codice colore, che si potrebbe indicare con l'azzurro, permetterà agli infermieri del triage di superare l'ostacolo di assegnare il codice bianco, rappresentando un livello di gravità intermedio, tra il codice bianco e quello verde, (rappresenta in sostanza i codici verdi non gravi). I triagisti, infatti, molto spesso assegnano "prudenzialmente" un codice verde che nella realtà dei fatti si rivela essere un codice bianco.

In questo modo, dunque, esiste il reale vantaggio a realizzare una zona dedicata a pazienti con problemi di salute non gravi, in quanto si potrebbe destinare al fast-track i codici sia bianchi che "azzurri", ovvero circa il 20% dei pazienti in ingresso al P.S.

Gli utenti con codice bianco, come già detto, non necessitano effettivamente del pronto soccorso. Per poter snellire il flusso dei pazienti e ridurre i tempi d'attesa, dunque, i codici bianchi sono inviati ad un'area differenziata a loro dedicata. Qui il paziente viene visitato, si effettua la relativa diagnosi, e dopo aver stabilito le cure necessarie il medico decide se è necessario il ricovero nel reparto di degenza specifico, o come più spesso accade dimette il paziente dalla struttura verso il proprio domicilio.

8.4 Diagramma di flusso del processo di Pronto Soccorso

La sequenza delle attività che avvengono in pronto soccorso formano un processo; come precedentemente esposto nel capitolo 4 tale processo viene comunemente descritto tramite un diagramma di flusso (vedi Fig 5.1).

La mappatura del processo si rende necessaria al fine di una corretta comprensione dello stesso: grazie ad essa è possibile evidenziare le potenziali criticità che si manifestano durante il percorso di attraversamento del paziente in PS e successivamente implementare e realizzare misure correttive. La mappatura segue il percorso del paziente a partire dalla fase iniziale di identificazione e triage fino a giungere alla dimissione o ricovero dell'ammalato passando per le attività di visita e analisi.

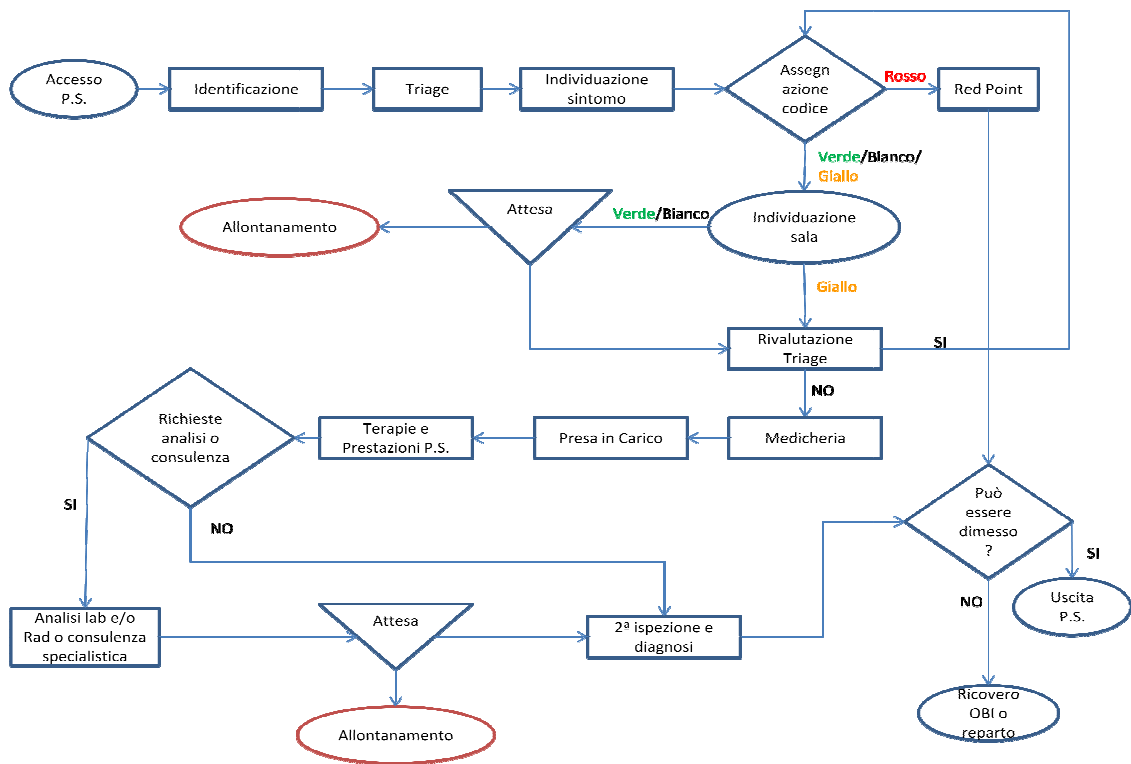
L'analisi del processo ha permesso di individuare un punto di particolare criticità: l'allontanamento non tracciato del paziente. Questo evento normalmente può insorgere durante le seguenti fasi del processo:

- Prima della valutazione del triage;
- Durante la permanenza in attesa in PS;
- Dopo la dimissione da PS in OBI o altro reparto dell'ospedale.

In questa sede viene analizzato unicamente il caso di allontanamento non tracciato del paziente durante la permanenza in attesa, in quanto la mancanza di dati relativi alle fasi iniziali e finali dell'intero processo di P.S. non permette di poterne valutare la probabilità di abbandono relativa.

8.4.1 Diagramma di flusso del processo di Pronto Soccorso

Di seguito è riportato il diagramma di flusso relativo al processo di pronto soccorso (*Diag.1*) in cui vengono evidenziati gli eventi di allontanamento volontario del paziente dal reparto di emergenza.



Diag.1 - Diagramma di flusso del Pronto Soccorso

Il diagramma richiama le seguenti fasi del processo:

- All'arrivo in P.S. del paziente avviene l'identificazione dello stesso da parte degli infermieri di triage.
- Gli infermieri in base all'analisi dei parametri vitali (coscienza, respirazione e circolazione del sangue) assegnano ad ogni paziente un codice di priorità ed una risorsa medica (medica, chirurgica, ortopedica) a seconda dell'urgenza e della tipologia di malessere che egli presenta.
- Se il paziente presenta uno dei tre parametri vitali alterato gli viene assegnato codice rosso e necessiterà un intervento immediato; viene quindi direttamente condotto in un'area apposita (detta Red Point) esercitando diritto di prelazione, ovvero interruzione di servizio, su pazienti di qualsiasi altro codice ogni volta che tutte le risorse sono impegnate.

- Se invece il paziente presenta sintomi per cui l'aiuto medico può essere ritardato gli verrà assegnato un codice prioritario di colore giallo, verde o bianco e verrà condotto in una sala d'attesa. Se necessario, i pazienti ricevono una prima assistenza (ad esempio una fasciatura) o possono essere compiuti dei primi controlli prima di entrare "in coda" e attendere in una sala sorvegliata dagli infermieri di triage, in modo che qualsiasi aumento di gravità possa essere immediatamente rilevato.
- Non appena le risorse sono disponibili e secondo la priorità, i pazienti ricevono la prima visita medica, seguendo una regola di tipo FIFO per stesso codice colore. Durante questa prima ispezione vengono stabilizzate le condizioni del paziente e viene assegnata una terapia.
- In seguito si decide tra la richiesta di ulteriori analisi (esami di laboratorio come l'ematochimico o l'esame delle urine, o esami strumentali come radiografie ed ecografie) e la dimissione o il ricovero.
- Si prosegue con una seconda ispezione dopo l'attesa per i risultati delle analisi ed infine si decide per dimissione o ricovero in un reparto ospedaliero o in OBI.

Si può notare come tra ogni fase del processo esista un momento di "attesa" : è proprio in queste fasi che si rischia l'abbandono volontario non tracciato del paziente dal pronto soccorso. A causa di sovraffollamento del reparto, congestione o indisponibilità delle risorse necessarie all'avanzamento del processo, i tempi di attesa possono risultare elevati, generando di conseguenza insoddisfazione nel paziente che è spinto ad abbandonare il reparto di emergenza. Normalmente la maggior causa di attesa all'interno del pronto soccorso è l'ottenimento dei risultati delle analisi, ed in particolare quelle di laboratorio. I tempi di attesa variano anche in base alla priorità

assegnata e al tipo di risorsa impegnata: codici di priorità più bassa tenderanno ad attendere di più poichè giudicati meno “urgenti”.

Si avranno quindi tempi di attesa differenti per i 12 casi: codice bianco chirurgico, medico o ortopedico; codice verde chirurgico, medico o ortopedico; codice giallo chirurgico, medico o ortopedico; codice rosso chirurgico, medico o ortopedico.

Il codice di priorità maggiore (rosso) non dovrebbe subire tempi di attesa elevati poichè, come già esposto, gode del diritto di prelazione sugli altri pazienti; per questo motivo per questo tipo di codice in seguito non verrà effettuata alcuna analisi sul rischio di allontanamento volontario dal reparto di urgenza.

8.4.2 Modello del processo tramite l’approccio Markoviano

Nel precedente paragrafo si è evidenziato tramite un diagramma di flusso in quali fasi del processo può sorgere l’evento “abbandono”. Il diagramma di flusso non è l’unico modo per rappresentare un processo, e soprattutto non è il modo più efficace per rappresentare un evento di “rischio”; in seguito, quindi, si farà riferimento ad un approccio di tipo Markoviano. Si vuole quindi mostrare come questo processo possa essere rappresentato da una Catena di Markov.

Le catene di Markov sono una classe speciale di sistemi dinamici che evolvono in modo probabilistico; esse si distinguono dagli Automi a Stati Finiti in quanto per questi ultimi le transizioni tra gli stati avvengono con certezza (i.e. con probabilità unitaria). Si può quindi affermare che Automi a Stati Finiti caratterizzati da probabilità di transizione minori dell’unità rappresentano Catene di Markov. Per le C.M. il risultato di ogni “esperimento” (transizione) è influenzato da quello precedente; gli elementi peculiari che caratterizzano la catena sono:

- Un insieme di configurazioni ammissibili per il sistema;

- Le probabilità di transizione da uno stato al successivo del sistema considerato;
- Rappresentazione del sistema in base ad un grafo orientato in cui i nodi rappresentano gli stati e gli archi tra i nodi le probabilità di transizione tra gli stati che collegano.

Una volta definite le principali caratteristiche di questo tipo di rappresentazione, si può convertire la rappresentazione del processo precedentemente descritto tramite il diagramma di flusso in una rappresentazione Markoviana. L'approccio seguito è quello di evidenziare da quali stati si può incorrere nell'evento abbandono.

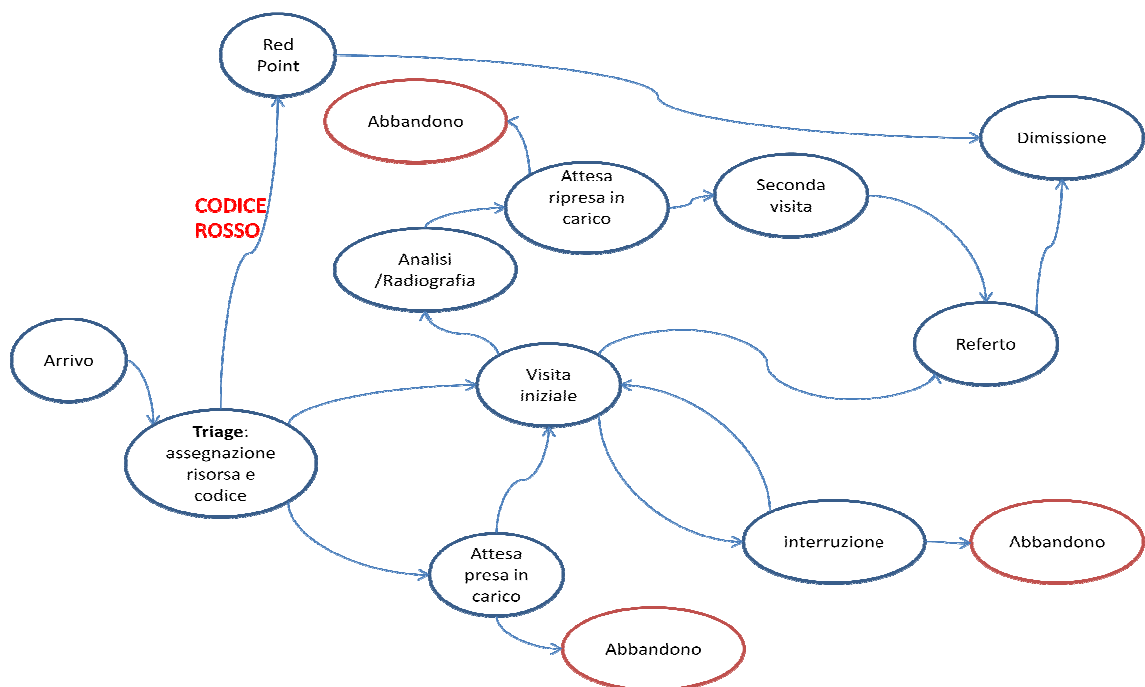
La catena graficata in *Mod.5* mostra unicamente gli stati in cui si può trovare il paziente all'interno del processo di pronto soccorso, e non sono quindi presenti le probabilità di transizione che possono essere dedotte dai dati di un generico pronto soccorso in termini di tassi di arrivo per codice e risorsa medica e numero di pazienti che effettuano le analisi. Questa catena, quindi, serve unicamente a definire gli stati che interessano la dinamica del "sistema-paziente" considerata; il modello si focalizza sul rischio di allontanamento del paziente dal P.S. per questo motivo sono evidenziati gli stati correlati a questo evento.

Il modello mostra la struttura e l'architettura della catena; è' plausibile, poichè questo tipo di modellizzazione può essere utilizzato indipendentemente da qualsiasi Pronto Soccorso ed ha quindi valore universale: le differenze per le diverse aziende sanitarie ricadono quindi nelle probabilità di transizione tra gli stati. Lo stato dell'abbandono è rappresentato con colore diverso, in modo da evidenziare il focus dell'analisi presentata in seguito.

La catena si inizializza con l'evento "arrivo" ed può valere per uno qualunque dei dodici casi precedentemente elencati, che si distinguono per codice di priorità e per risorsa medica. L'evento di uscita dalla catena può

essere invece rappresentato sia dall'abbandono che dalla dimissione dall'ospedale. Le ipotesi sottostanti al modello sono:

- 1) E' un modello unico che cambia da ospedale a ospedale in base alle probabilità di transizione da porre sugli archi;
- 2) Le quote sugli archi si ottengono dalle dinamiche di ingresso e dai tassi di utilizzo delle risorse;
- 3) Il modello è valido per singolo codice di priorità.



Mod.5 - Catena di Markov del processo di Pronto Soccorso

8.4.3 La Catena di Markov “comportamentale”

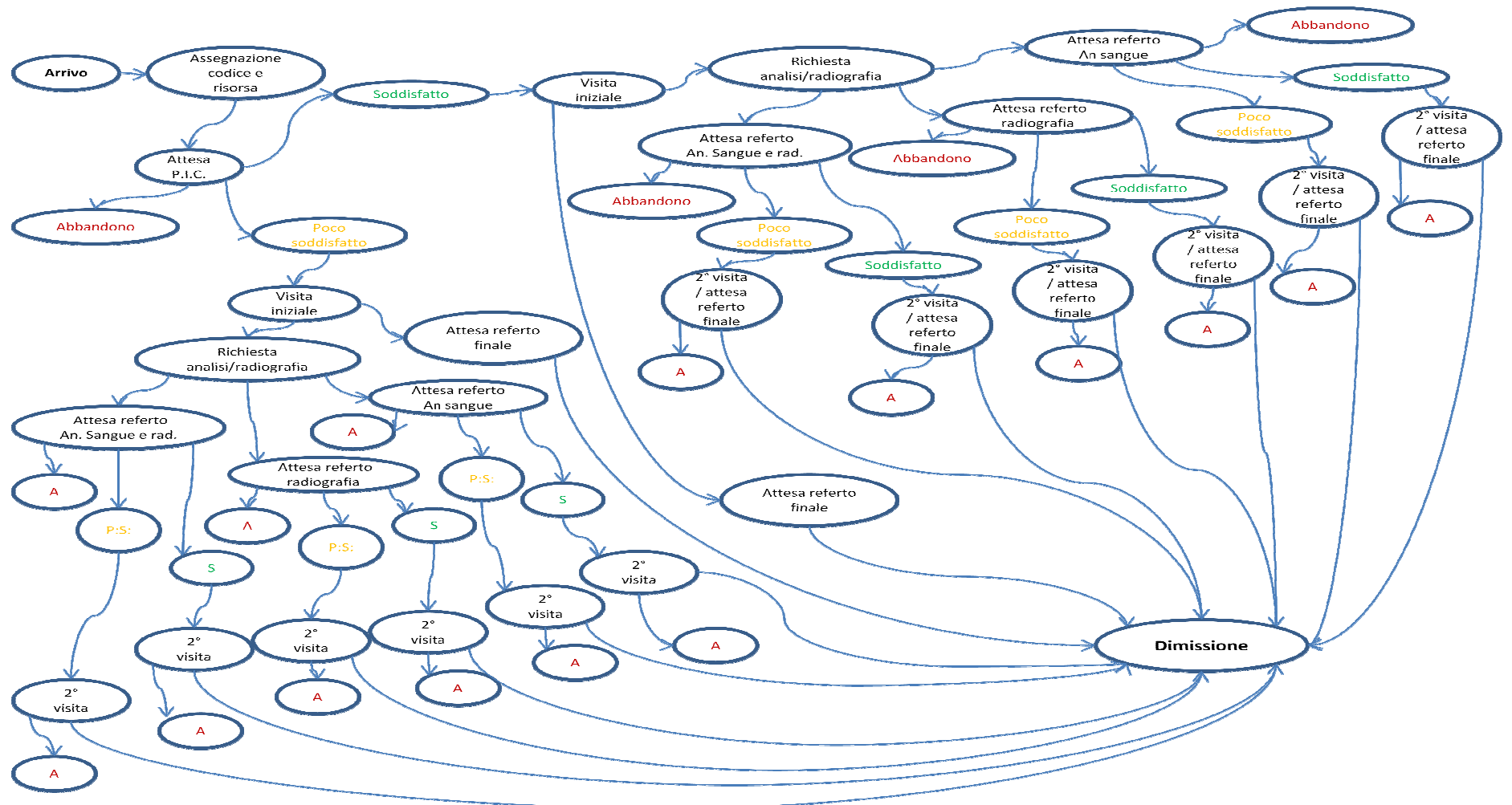
Una volta mostrato il modello generico dell'Automa a Stati Finiti bisogna includere le probabilità di transizione sugli archi congiungenti gli stati. Prima di descrivere in che modo possono essere derivati i valori da apporre sugli archi si vuole mostrare un ulteriore modello (d'ora in poi Catena di Markov comportamentale) in cui si evidenzia come la soddisfazione del paziente può influire sull'allontanamento volontario dal reparto di pronto soccorso.

L'analisi seguente poggia su tale assunto: l'abbandono da parte del paziente è conseguenza del raggiungimento di un elevato livello di insoddisfazione e l'insoddisfazione è dovuta principalmente a lunghe attese. Nel passaggio tra i diversi stati il paziente può definirsi soddisfatto, poco soddisfatto o insoddisfatto: in quest'ultimo caso si scatena l'evento dell'allontanamento; nel primo caso (paziente soddisfatto) egli proseguirà verso lo stato successivo con probabilità di abbandono nulla; il caso di scarsa soddisfazione genera una maggiore probabilità di insoddisfazione (e quindi di transizione verso lo stato di abbandono) nelle fasi successive. La catena di Markov sopra descritta, viene quindi resa più complessa dalla presenza di stati che rappresentano il grado di soddisfazione.

La descrizione della catena riportata di seguito (*Mod.6*) richiama quella precedentemente esposta. Si possono però distinguere degli ulteriori stati di "soddisfazione": ad esempio in seguito all'assegnazione del codice di priorità e della risorsa medica, il paziente, a seconda del tempo trascorso in attesa, può cambiare il proprio "stato di soddisfazione" nel passare allo stato successivo (attesa Presa In Carico). Ciò vuol dire che potrà raggiungere tre stati differenti: Attesa P.I.C. Soddisfatto, attesa P.I.C. Poco Soddisfatto, Abbandono (ovvero Insoddisfazione); a partire dai primi due stati si "apriranno" due catene identiche ma con probabilità di transizione differenti in termini di soddisfazione. Questa rappresentazione deriva dalla considerazione che se, nonostante le lunghe attese, il paziente è disposto ad aspettare ulteriormente, la sua propensione all'abbandono aumenterà nelle fasi successive: ciò implica che le probabilità di transizione verso lo stato di abbandono saranno più elevate nel caso la parte di catena considerata ha origine in uno stato di poca soddisfazione.

Per semplicità di rappresentazione nella parte finale della catena lo stato di abbandono è indicato con una A rossa e lo stato di poca soddisfazione con P.S. in giallo. Inoltre, si evidenzia che a partire dallo stato 2° visita/attesa

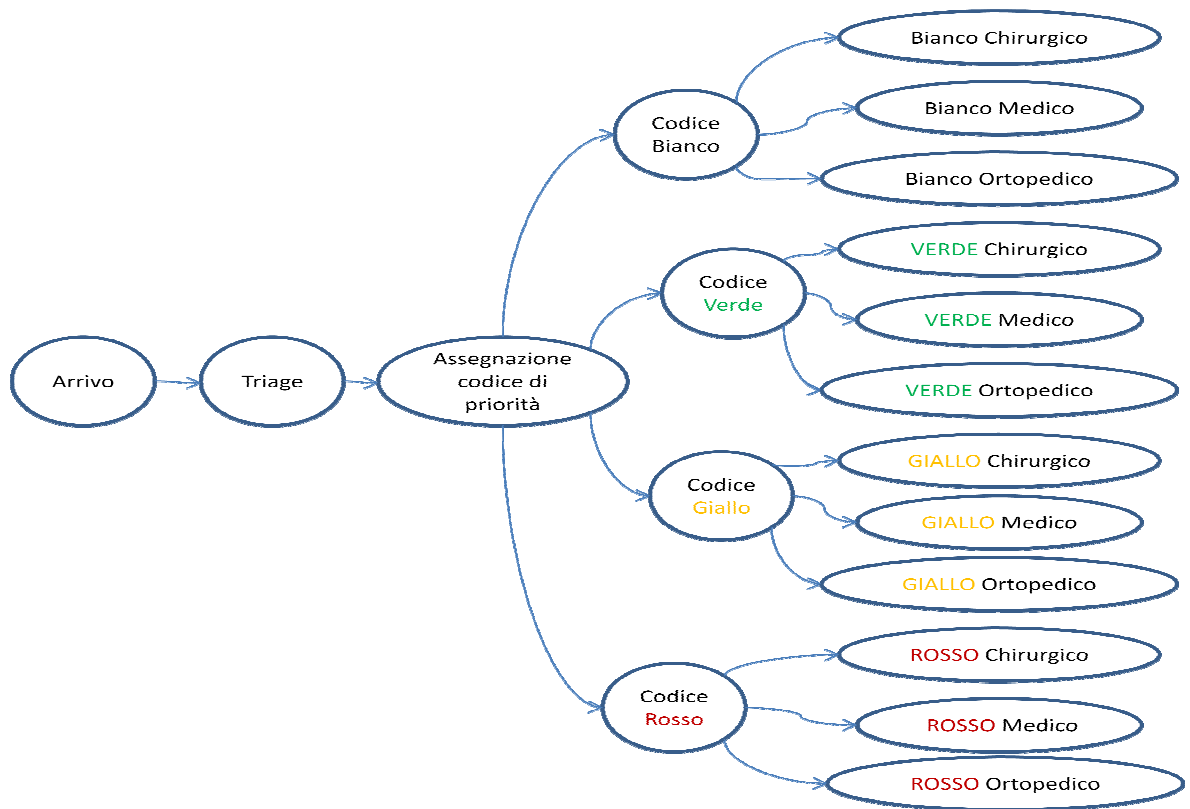
referto non compare lo stato Poco Soddisfatto poichè si suppone che il paziente nella parte terminale del processo attende fino alla dimissione o abbandona.



Mod.6 - Catena di Markov Comportamentale

Come si può dedurre dalla rappresentazione della catena comportamentale, essa vale per solo uno dei 12 casi elencati nel paragrafo precedente (per un unico codice prioritario a cui è assegnata una particolare risorsa medica): per ognuno dei dodici casi cambieranno le probabilità di transizione, poichè ogni caso sarà caratterizzato da priorità, disponibilità di risorse e conseguentemente tempi di attesa differenti.

Si può pensare di considerare una catena “precedente” a quella appena descritta (*Mod.7*), in cui si evidenziano le transizioni tra l’arrivo e l’assegnazione del codice e successivamente l’assegnazione della risorsa; a partire da ognuno dei 12 stati partirà poi la catena comportamentale in fig 5.3. .



Mod.7 - Catena Assegnazione risorsa e codice

E’ ora possibile descrivere le probabilità di transizione. Per ottenere tali probabilità è necessario ottenere i dati del processo di pronto soccorso

relativi ai tassi di arrivo, ai tempi di attesa ed al numero medio di pazienti del reparto di emergenza che necessitano analisi di tipo radiografico e/o del sangue.

Si distingue tra due tipologie di probabilità di transizione:

- le probabilità che discendono dai tassi di arrivo e dal numero medio di pazienti che necessita di ulteriori analisi;
- le probabilità che discendono dai tempi di attesa è che conseguentemente definiscono la transizione verso uno stato di soddisfazione.

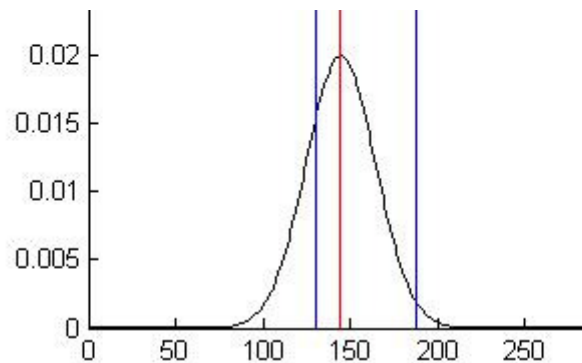
Per la prima tipologia si è ipotizzato di far corrispondere le percentuali di arrivo alla probabilità di transizione: ad esempio nel caso della fig 5.4 per la probabilità di transizione dallo stato assegnazione codice allo stato codice di priorità si è considerata la percentuale di pazienti a cui (nel periodo considerato) è stato assegnato quel particolare codice prioritario. La stessa ipotesi è stata fatta per l'assegnazione della risorsa medica e per le transizioni da richiesta analisi allo stato attesa risultato analisi/ radiografia (si è considerata la percentuale di pazienti di un determinato codice e di una determinata risorsa medica per cui sono state richieste le analisi specifiche).

La stima del secondo tipo di probabilità si presenta invece più complessa, poichè si tratta di trasformare un indice di soddisfazione in una probabilità di transizione. L'approccio scelto deriva da due considerazioni:

- i dati a disposizione sono i tempi in attesa del paziente nel reparto di emergenza;
- l'insoddisfazione è generata da tempi attesa troppo lunghi.

E' stata quindi considerata una distribuzione normale con media corrispondente alla media campionaria e varianza alla varianza campionaria. Su questa distribuzione sono stati poi considerati dei limiti (in termini orari) di soddisfazione: un limite inferiore che rappresenti l'attesa massima perchè

un paziente risulti soddisfatto e un limite superiore che rappresenti l'attesa massima oltre la quale il paziente si ritiene insoddisfatto ed è portato all'abbandono; tra i due limiti vi è un'area che corrisponde alla "poca soddisfazione", in cui il paziente permane in attesa ma con una probabilità di abbandono maggiore nelle fasi successive. Definiti questi limiti, le probabilità di transizione verso gli stati "comportamentali" vengono calcolate valutando l'area sottesa alla curva e delimitata dai valori limite per la soddisfazione: si osservi il *Diag.2*, in cui sono tracciati sulla normale la media campionaria (in rosso), un limite sinistro ed uno destro (in blu).



Diag.2 - Distribuzione normale tempo di attesa referto analisi di laboratorio (codice verde medico)

Una volta definite tutte le caratteristiche di questo tipo di rappresentazione (stati, probabilità e grafo) si vuole descrivere come quanto mostrato sino ad ora può essere applicato ad un caso concreto; nei prossimi paragrafi verrà quindi particolarizzata la catena di Markov, comportamentale e non, con le probabilità di transizione derivate dai dati forniti dalla struttura ospedaliera A.O.R.N. "A.Cardarelli" di Napoli. In seguito verrà poi calcolato, con l'utilizzo della teoria Markoviana, il tasso di abbandono volontario dal reparto di Pronto Soccorso.

8.5 Validazione dello studio presso A.O.R.N. “A . Cardarelli”

8.5.1. Descrizione A.O.R.N. “A. Cardarelli”³¹

L'Ospedale Antonio Cardarelli di [Napoli](#) è il maggior nosocomio della [Campania](#) ed il principale dell'intero [Meridione](#) e a livello nazionale per la cura dei grandi ustionati.

L'Azienda Ospedaliera di Rilievo Nazionale “Antonio Cardarelli” è una grande azienda erogatrice di servizi sanitari, cui competono funzioni e responsabilità rilevanti, nell'esercizio di assicurare l'assistenza sanitaria su un ampio territorio, il quale abbraccia un bacino di utenza esteso, oltre alla città, anche alla provincia ed alla Campania, e rappresenta un punto di riferimento per altre regioni d'Italia, in particolar modo del Meridione. L'Azienda Ospedaliera di Rilievo Nazionale e di Alta Specializzazione “A. Cardarelli” ha acquisito il ruolo di rilevanza nazionale, in base ai riconoscimenti di funzione attribuiti con il DPCM 8-4-93 e con il Decreto n° 12255 del 22-12-1994 della Regione Campania. L'Ospedale “Cardarelli” svolge un ruolo di primo piano per quanto riguarda l'assistenza sanitaria di urgenza: è infatti sede di Dipartimento di Emergenza-Accettazione di secondo livello, assicurando prestazioni di pronto soccorso in molteplici specialità. E' fra l'altro sede del Centro Grandi Ustionati, del Centro Antiveleni e del Centro per i Trapianti Epatici, (Centri di Emergenza Regionali); è presente altresì il Centro di Terapia Iperbarica, di recente istituzione. Presso tale dipartimento è ubicata inoltre la Centrale Operativa “118”. L'azienda “A. Cardarelli” si distingue anche per le attività specialistiche di elezione di area medica e chirurgica, rappresentando quindi un riferimento per la rete sanitaria della regione. L'attività assistenziale produce un elevato numero di ricoveri annui ordinari ed in day-hospital, mediamente superiori a 90.000, nonché di prestazioni erogate in regime

³¹ <http://www.ospedalecardarelli.it/ospedale/>

ambulatoriale nell'ambito delle diverse specialità. Le opere di ristrutturazione edilizia e di ammodernamento tecnologico dell'azienda, attualmente in corso, consentono di mettere a disposizione degli operatori e degli utenti una struttura sempre al passo con i progressi tecnologici e progressivamente adeguata al meglio con le esigenze di comfort dei pazienti. L'ospedale "A. Cardarelli", situato nel cuore della zona ospedaliera, possiede una struttura "a padiglioni" che occupa nel complesso una superficie di 250.000 metri quadrati. Di questi, 50.000 metri quadri sono rappresentati da edifici, ed i restanti 200.000 da viali alberati e pinete che di fatto costituiscono un vero e proprio "polmone verde". Dei 21 padiglioni esistenti, costruiti in diverse epoche a partire dal 1927 fino al 1990, quattordici sono destinati alle attività di diagnosi e cura, ed i restanti sette ai servizi tecnici. Oltre che dalla rete viaria interna, gli edifici dell'azienda sono aggregati da collegamenti sotterranei per le attività tecniche e di servizio, i quali si sviluppano in corrispondenza dell'ampia superficie dell'ospedale, costituendo una vera e propria "rete viaria" per gli operatori. Nell'ambito della superficie dell'Azienda è situato un eliporto che costituisce il fulcro dei trasferimenti rapidi nell'ambito dell'emergenza intra ed extraregionale, ove sia richiesto l'intervento di eliambulanza. La struttura dell'eliporto, dotato dei più moderni strumenti tecnici di controllo e di sicurezza, consente l'atterraggio ed il decollo anche nelle ore notturne.



Mapa dell'azienda ospedaliera "A. Cardarelli"

Nel 1943 l'Ospedale fu intitolato ad Antonio Cardarelli (Civitanova 1832-Napoli 1927), clinico ricercatissimo, maestro venerato, senatore del Regno. Nel novembre del 1990 fu inaugurato, da Sua Santità Giovanni Paolo II, il Padiglione dell'Emergenza, sviluppato su cinque piani, che rappresenta tuttora una delle migliori strutture dipartimentali di Pronto soccorso in campo nazionale, per la modernità tecnologica, l'efficienza del personale e l'efficacia delle prestazioni erogate. L'assetto aziendale, tuttora vigente, con la nomenclatura di Azienda Ospedaliera di Rilievo Nazionale e di Alta Specializzazione "A. Cardarelli" è stato istituito con decorrenza dal 1° gennaio 1995, dalla Giunta della Regione Campania, con decreto n° 12255 del 22 dicembre 1994. L'Azienda Ospedaliera A. Cardarelli è un ospedale per acuti di Rilevanza Nazionale che eroga prestazioni di diagnosi, cura e riabilitazione per un ampio bacino di utenza. Esso è dotato di Unità Operative Complesse ad elevato grado di specializzazione, orientate a tre grandi campi di attività: assistenza, didattica e ricerca. La cooperazione e l'integrazione di tutte le professionalità presenti ed operanti nell'ospedale, permettono il

raggiungimento dell'obiettivo globale di carattere assistenziale, di rilevanza pubblica.

L'azienda afferma con forza la volontà di lavorare per produrre sviluppo in campo assistenziale e creare le condizioni per favorire un grande rilancio della sanità nella città di Napoli. A tal fine si propone di:

1. Porre i bisogni dell'assistito sempre al centro di ogni decisione di tipo sanitario;
2. Investire nel patrimonio di competenza professionale e nell'innovazione tecnologica e strutturale;
3. Aprirsi all'esterno rendendosi capace di collegarsi ed interagire con il contesto.

Le prestazioni erogate dall'Azienda Ospedaliera comprendono:

- Assistenza medica;
- Assistenza infermieristica;
- Ogni intervento e procedura diagnostica, terapeutica e riabilitativa necessari per risolvere i problemi di salute del paziente-utente, compatibilmente con il livello di dotazione strutturale e tecnologica delle singole strutture;
- Interventi di pronto soccorso nei confronti di malati o infortunati in situazioni di urgenza ed emergenza medica e/o chirurgica, incluso il trasporto in ospedale, ove necessario e in condizioni di emergenza;
- Interventi a supporto psicologico e sociale, per i pazienti ed i loro familiari;
- Collegamento con Enti e Strutture esterne, per assicurare la continuità dei trattamenti terapeutici;

- Interventi a tutela degli anziani e degli utenti con particolari problematiche sociali o con patologie invalidanti croniche (minori, stranieri, invalidi);
- Attività di ricerca;
- Banca dati del Centro Antiveneni;
- Attività di formazione continua diretta all'aggiornamento professionale del personale medico e non medico;
- Attività didattica universitaria, perseguita all'interno dell'Azienda, grazie alla presenza di un Polo didattico che espleta il corso di Laurea triennale: "Scienze Infermieristiche".

La Direzione Sanitaria Aziendale dirige sul piano strategico i servizi ospedalieri dell'Azienda Ospedaliera ai fini tecnico sanitari ed igienico-organizzativi. A tale scopo essa collabora, unitamente alla Direzione Amministrativa, con la Direzione Generale per la definizione delle linee di indirizzo e delle priorità al fine di realizzare la migliore composizione tra le caratteristiche di efficacia, efficienza, e qualità dei servizi che l'azienda offre al cittadino. Pertanto essa collabora al controllo di gestione, curando il sistema informativo sanitario e pianificando l'allocazione delle risorse umane, tecniche e strumentali nell'ambito della programmazione aziendale. La Direzione Sanitaria Aziendale, inoltre, coadiuva il Direttore Generale nel mantenimento dei rapporti con l'Università definendo i relativi protocolli d'intesa; esprime pareri sulle progettazioni edilizie in merito agli aspetti igienico-organizzativi e promuove le attività di aggiornamento e formazione del personale sanitario.

8.5.2 Il processo di Pronto Soccorso nell'A.O.R.N. "A. Cardarelli"

Si intende ora particolarizzare le rappresentazioni del processo di Pronto Soccorso mostrate nei paragrafi precedenti al caso particolare del nosocomio "A.Cardarelli".

Il processo di Pronto Soccorso è ovviamente identico a quello precedentemente descritto, bisogna però fare alcune precisazioni. In particolare il laboratorio di analisi non è dedicato unicamente al reparto di emergenza, ma condiviso con il resto della struttura ed è esterno al padiglione considerato; i campioni vengono inviati al laboratorio tramite un sistema di condotti pressurizzati ed i risultati inviati per via elettronica ai computer dei medici di turno. Questa caratteristica del P.S. fa sì che si generino rallentamenti sia in fase di presa in carico del campione da analizzare, sia in fase di ricezione del referto; inoltre, essendo il laboratorio una risorsa condivisa deve essere realizzata una efficiente gestione e scheduling delle attività da svolgere. La maggiore congestione all'interno del laboratorio si ha normalmente durante la mattina quando vengono svolte le analisi richieste da tutti i reparti.

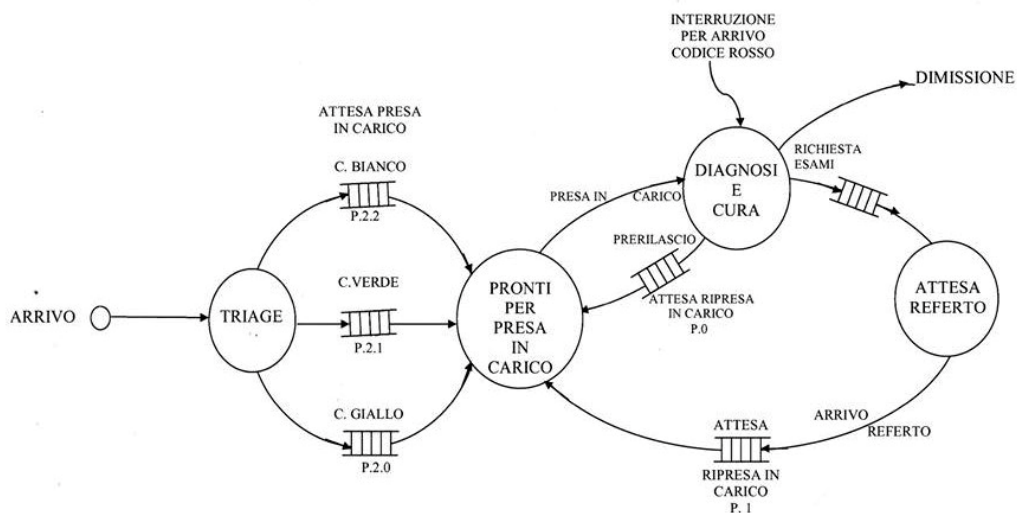
Il reparto di emergenza di questa struttura ospedaliera è inoltre caratterizzato da forte congestioni e sovraffollamenti: si è osservato un tasso di arrivo medio molto elevato (circa 7 pazienti all'ora). Attualmente la struttura si trova nella situazione critica di una sempre maggiore domanda che non sempre riesce a soddisfare a fronte di risorse materiali, economiche e personali talvolta insufficienti. Una delle maggiori cause di sovraffollamento e congestione lamentate dagli operatori sanitari all'interno del reparto di emergenza è l'arrivo di pazienti con codice prioritario bianco, ovvero di urgenza nulla: la maggior parte degli utenti di PS non ha effettivamente necessità di rivolgersi a una struttura di questo tipo che invece dovrebbe essere dedicata alle emergenze.

Un altro problema rilevato nell'analisi del processo di pronto soccorso è quello dell'allontanamento volontario non tracciato dal reparto di emergenza: trascorso un certo periodo di tempo in attesa il paziente decide volontariamente di abbandonare l'ospedale; la maggior parte dei casi di

allontanamento è collegata ai codici di priorità minore ed in particolare ai codici bianchi e verdi.

Lo studio effettuato, si propone di valutare il tasso di abbandono dei soli codici bianco, verde e giallo associati alla risorsa medica (si escluderanno quindi i casi ortopedici e chirurgici ed il codice rosso) poichè si presuppone soggetti ad una probabilità di allontanamento volontario più elevata rispetto ai casi esclusi.

Si riporta in seguito l'automa a stati finiti del processo fornito dall'azienda.



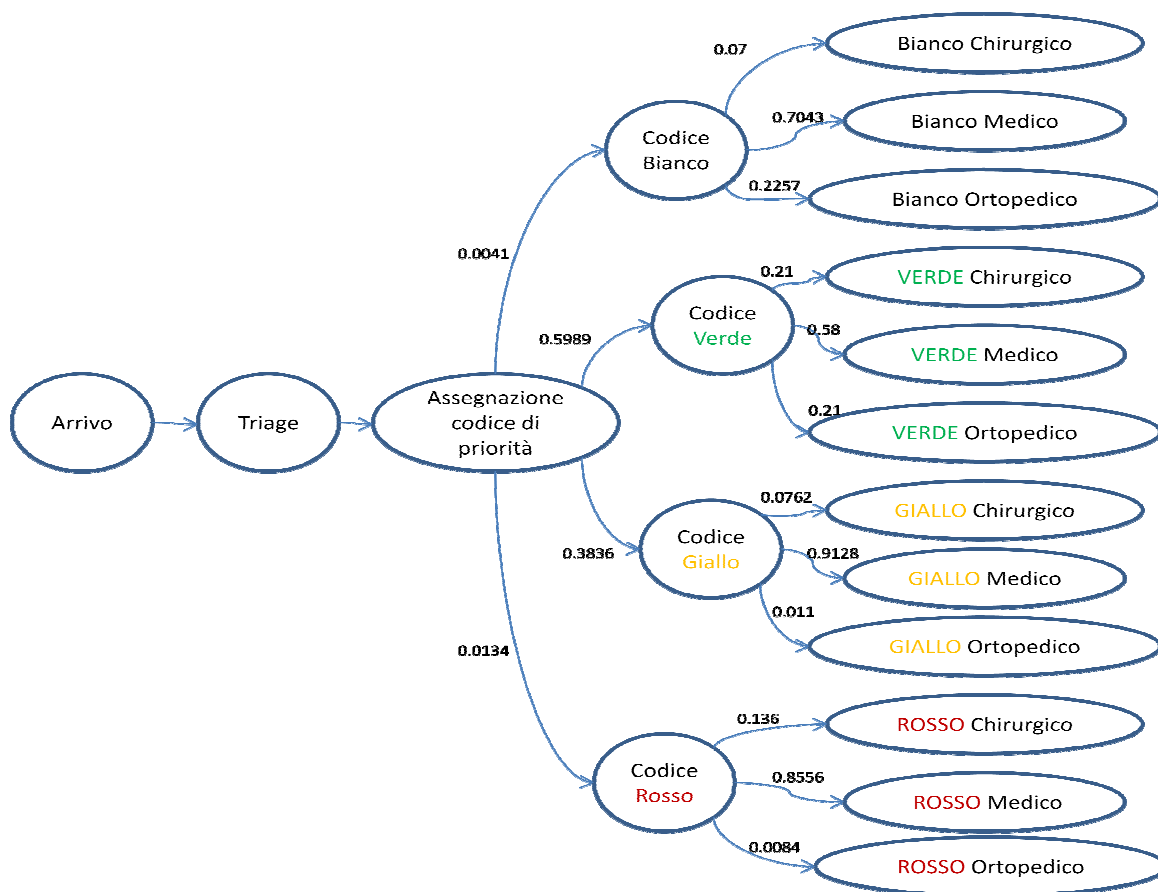
Mod.9 - Automa del Pronto Soccorso della A.O.R.N. "A. Cardarelli" di Napoli

Le rappresentazioni del diagramma di flusso, della catena di Markov e della catena comportamentale sono coincidenti con quanto già rappresentato nei precedenti *Diag.1, Mod.5 e Mod.6*. Non resta che inserire i valori delle probabilità di transizione corrispondenti. Si è fatto riferimento ai dati forniti dal C.E.D. dell'azienda in termini di tassi di arrivo e tempi di attesa; tali dati si riferiscono ad un arco temporale che va da Febbraio a Dicembre 2013. Per il tasso di arrivo e per il numero medio di pazienti sottoposti ad analisi si è considerato il numero totale di arrivi per codici e specialità e sono poi state calcolate le percentuali sul totale. Si sono ottenuti i seguenti dati:

Codice Prioritario	Risorsa medica	Risorsa Ortopedica	Risorsa Chirurgica	Totale
Codice Bianco	181	58	18	257
Codice Verde	21703	7909	7873	37485
Codice Giallo	21919	262	1832	24013
Codice Rosso	717	7	114	838
TOTALE	44520	8236	9837	62593

Prosp.4 - Tassi di arrivo Febbraio-Dicembre 2013

Trasformati questi dati in percentuali si sono ottenute le probabilità di transizione da inserire in *Mod.10*; si ottiene la seguente catena:



Mod.10 - Catena assegnazione risorsa e codice con probabilità di transizione

Per quanto riguarda le probabilità di transizione legate al numero di pazienti che sono stati sottoposti a analisi di laboratorio e/o radiografici si sono ottenuti i seguenti dati e le relative percentuali:

Codice Prioritario	Analisi di laboratorio	%	Diagnostica per immagini	%	Nessuna analisi	%	Arrivi
Codice Bianco	0	0	4	0.22	14	0.78	18
Codice Verde	180	0.02	4394	0.56	3299	0.42	7873
Codice Giallo	267	0.15	1292	0.70	273	0.15	1832
Codice Rosso	43	0.38	71	0.62	0	0	114
TOTALE	490	0.05	5761	0.59	3586	0.36	9837

Prosp.5 - Tassi pazienti con analisi ed esami caso chirurgico

Codice Prioritario	Analisi di laboratorio	%	Diagnostica per immagini	%	Nessuna analisi	%	Arrivi
Codice Bianco	0	0	18	0.10	163	0.90	181
Codice Verde	2236	0.10	4073	0.19	15394	0.71	21703
Codice Giallo	4648	0.21	3954	0.18	13317	0.61	21919
Codice Rosso	375	0.52	342	0.48	0	0	717
TOTALE	7284	0.16	8412	0.19	28874	0.65	44520

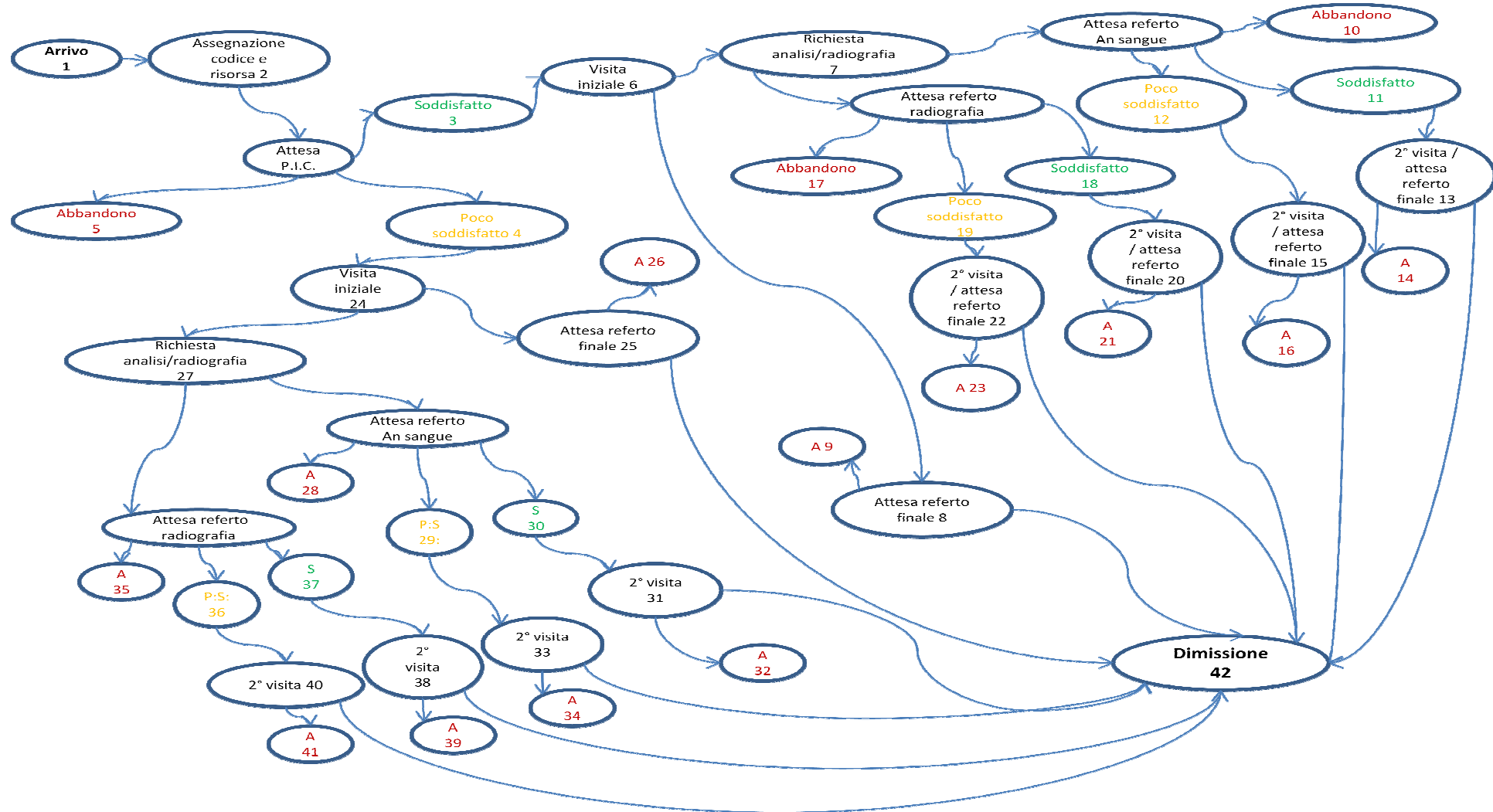
Prosp.6 - Tassi pazienti con analisi ed esami caso medico

Codice Prioritario	Analisi di laboratorio	%	Diagnostica per immagini	%	Nessuna analisi	%	Arrivi
Codice Bianco	0	0	9	0.16	49	0.84	58
Codice Verde	10	0.01	3100	0.39	4799	0.60	7909
Codice Giallo	7	0.03	133	0.51	122	0.46	262
Codice Rosso	1	0.14	1	0.14	5	0.72	7
TOTALE	18	0.01	3243	0.39	4975	0.60	8236

Prosp.7 - Tassi pazienti con analisi ed esami caso ortopedico

E' possibile osservare che non sono disponibili dati riguardo il numero di pazienti che sono stati sottoposti sia ad analisi di laboratorio sia ad analisi di diagnostica per immagini; per questo motivo la catena comportamentale del precedente *Mod.6* viene considerata in una forma ridotta, senza però perdere di generalità nel prossimo *Mod.11*; è stato inoltre inserito un numero per ognuno dei 45 stati.

Per le probabilità di transizione verso gli stati “comportamentali” si è seguito l’approccio descritto nei paragrafi precedenti: gli stati di soddisfazione vengono raggiunti con delle probabilità di transizione che dipendono dai tempi di attesa; per definire i limiti orari entro cui il paziente si ritiene soddisfatto, poco soddisfatto ed insoddisfatto si è deciso di sottoporre i pazienti in attesa in P.S. ad un questionario (riportato in seguito in *Prosp.8*).



Mod.11 - Catena comportamentale ridotta

Questionario sul grado di soddisfazione dei pazienti in attesa al P. S.	
1. Sesso:	M <input type="checkbox"/> F <input type="checkbox"/>
2. Età:
3. Codice di priorità:	Bianco <input type="checkbox"/> Verde <input type="checkbox"/> Giallo <input type="checkbox"/> Rosso <input type="checkbox"/>
4. Risorsa medica assegnata:	Chirurgica <input type="checkbox"/> Medica <input type="checkbox"/> Ortopedica <input type="checkbox"/>
5. E' soddisfatto della qualità del pronto soccorso?	Si <input type="checkbox"/> No <input type="checkbox"/>
6. Se no, indicare il motivo:	Tempi di attesa lunghi <input type="checkbox"/> Triage inadeguato <input type="checkbox"/> Scarsa attenzione dal personale medico <input type="checkbox"/> Ambiente poco confortevole <input type="checkbox"/> Altro.....
7. Quali tempi di attesa ritiene adeguati prima della Presa in Carico per potersi dire soddisfatto ?	
8. Quali sono i tempi di attesa massimi che è disposto ad accettare prima della Presa in Carico?.....	
9. Quali tempi di attesa ritiene adeguati per ottenere i risultati delle analisi del sangue per potersi dire soddisfatto (sia nel caso di piena soddisfazione che poca soddisfazione) ?	
10. Quali sono i tempi di attesa massimi che è disposto ad accettare per ottenere i risultati delle analisi del sangue (in caso soddisfatto e non)?	
11. Quali tempi di attesa ritiene adeguati per ottenere i risultati degli esami radiografici per potersi dire soddisfatto (in caso soddisfatto e non) ?	
12. Quali sono i tempi di attesa massimi che è disposto ad accettare per ottenere i risultati degli esami radiografici (in caso soddisfatto e non)?	
13. Quali tempi di attesa ritiene adeguati per ottenere i risultati sia delle analisi del sangue che degli esami radiografici per potersi dire soddisfatto (in caso soddisfatto e non)?	
14. Quali sono i tempi di attesa massimi che è disposto ad accettare per ottenere i risultati sia delle analisi del sangue che degli esami radiografici (in caso soddisfatto e non)?	
15. Quali tempi di attesa ritiene adeguati per ottenere la dimissione dall'ospedale, una volta ottenuti i risultati delle analisi (nei casi solo analisi, radiografia, nessuna analisi)? ...	
16. Da quanto tempo è in attesa (se è già avvenuta la dimissione indicare il tempo totale trascorso nel P.S.) ?	

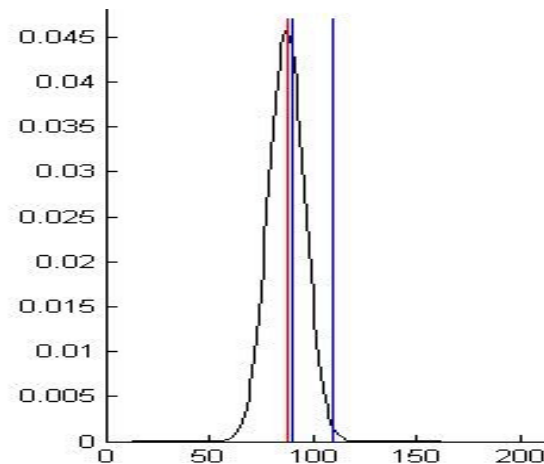
Prosp.8 - Questionario sul grado di soddisfazione del paziente in attesa in P.S.

Il questionario è stato sottoposto ai pazienti in attesa nella fase successiva al triage, talvolta è stato compilato riferendosi ad esperienze passate; purtroppo si sono raccolti unicamente 120 questionari, poichè non sempre si è incontrata la disponibilità del paziente alla compilazione. I questionari raccolti hanno evidenziato come il problema maggiormente riscontrato come indice di bassa qualità sia quello della lunga attesa: laddove il paziente si è dichiarato insoddisfatto il motivo dichiarato è stato nella quasi totalità dei casi “tempi di attesa lunghi”. Dei questionari raccolti sono stati considerati solo quelli che si riferiscono al caso medico e sono stati definiti poi i limiti orari di soddisfazione per ogni fase e per ogni codice. I questionari relativi al caso medico sono pari a 75 di cui 23 codici bianchi, 32 codici verdi e 20 codici gialli.

Viene mostrato in seguito come si sono calcolate le probabilità di transizione “comportamentali” facendo riferimento alla transizione dallo stato “assegnazione codice e risorse” allo stato “attesa Presa in Carico” per il caso codice verde medico. Viene considerata una distribuzione normale con media pari alla media campionaria e varianza pari alla varianza campionaria; su tale distribuzione sono poi stati evidenziati i limiti orari di soddisfazione: un limite inferiore prima del quale il paziente si ritiene soddisfatto ed un limite superiore dopo il quale il paziente si dichiara insoddisfatto e abbandona il Pronto Soccorso. Tali limiti sono stati dedotti dai questionari a cui sono stati sottoposti i pazienti in attesa attraverso una media aritmetica dei dati raccolti: laddove nel questionario si legge “tempi di attesa adeguati” si indicano i limiti inferiori, mentre i “tempi di attesa massimi per la soddisfazione” indicano i limiti superiori.

Nel caso considerato si è ottenuto: media campionaria 87.4 minuti (in media per la presa in carico di un codice verde medico si impiega quasi un’ora e mezza), varianza campionaria 76.2 (deviazione standard 8.73), limite inferiore 90 minuti (in media il paziente codice verde medico si ritiene

soddisfatto nell'attesa fino ad un'ora e 30 minuti) e limite superiore 110 minuti (oltre tale limite orario il paziente “minaccia” di abbandonare il reparto di emergenza). Si ottiene una distribuzione come in *Diag.3*, in cui le 3 linee verticali rappresentano in rosso la media ed in blu i limiti orari di soddisfazione del paziente.



Diag.3 - Distribuzione normale tempo attesa PIC codice verde medico

Per calcolare la probabilità di avere in questa transizione di stati un paziente soddisfatto, poco soddisfatto o insoddisfatto si è calcolato, tramite l'utilizzo di Matlab, le aree sottese alla curva e delimitate dai tempi di attesa minimi e massimi.

I codici utilizzati per il calcolo e la rappresentazione grafica sono esposti in seguito:

```
x=0:480;
% scelgo un intervallo di rappresentazione sull'asse delle x tra 0
minuti e
% 480 minuti
mi=87.4; %dichiaro la media dei tempi di attesa, precedentemente
calcolata
% con il comando mean (A) con A vettore dei tempi
sigma=8.73; % dichiaro la deviazione standard campionaria,
precedentemente
%calcolata con il comando std (A)
y=normpdf(x,mi,sigma); % definisco la distribuzione normale di
media x e dev
%standard sigma
%grafico della distribuzione
plot(x,y, 'k');
hold on
```

```

min=90; %definisco il limite inferiore
max=110; %definisco il limite superiore
% traccio sul grafico i limiti e la media
y= linspace(0,0.047);
xmin= min*ones(size(y));
plot(xmin,y, 'b')
hold on

y= linspace(0,0.047);
media= mi*ones(size(y));
plot(media,y, 'r')

xmax=max*ones(size(y));
plot(xmax,y, 'b')

% calcolo le aree sottese alla curva
PrSodd= normcdf(min,mi, sigma)
PrInsodd=1- normcdf(max,mi, sigma)
PrPSodd= normcdf(max,mi, sigma)-normcdf(min,mi, sigma)

```

Si è ottenuto:

-Pr soddisfatto= 0.62 (area a sinistra del limite inferiore)

-Pr poco soddisfatto= 0.375 (area compresa tra i due limiti)

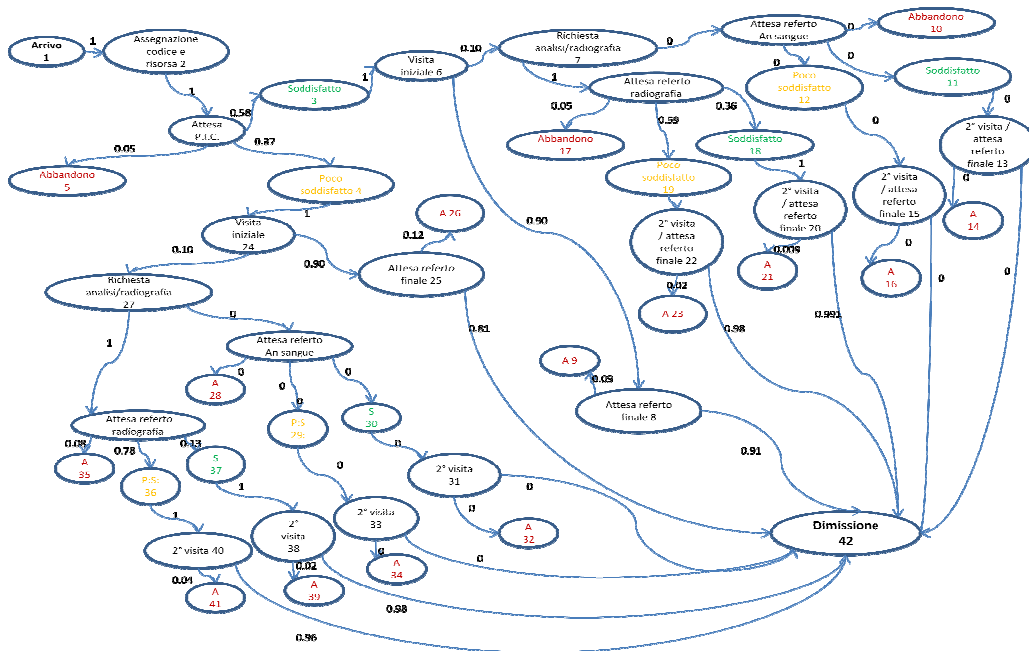
-Pr insoddisfatto=0.005 (area a destra del limite superiore).

Analogamente si sono ottenute le altre probabilità di transizione e sono state quindi apposte sulla catena di *Mod.11*. Sono state inoltre create delle tabelle con righe e colonne rappresentanti i 45 stati ed in corrispondenza delle caselle si è inserita la probabilità di transizione dallo stato riga verso lo stato colonna. Le matrici ottenute sono di notevoli dimensioni ma sparse: la maggior parte degli elementi è nullo, poichè la transizione è possibile solo tra alcuni di tutti gli stati. Queste tabelle non sono altro che le matrici di transizione per le singole catene: si avranno la matrice di transizione P_b per il codice bianco medico, P_v per il codice verde medico e P_g per il codice giallo medico. Le catene correlate di probabilità di transizione che rappresentano i tre casi (codice bianco medico (*Mod.13*), codice verde medico (*Mod.14*) e codice giallo medico (*Mod.15*)) sono rappresentate in seguito.

Si riporta successivamente a titolo di esempio la tabella con la matrice di transizione per il codice verde (Tab 5.5).

Si possono riscontrare alcune particolarità:

- il codice bianco non effettua analisi di laboratorio, per cui tutti le transizioni che seguono da “attesa referto analisi del sangue” hanno probabilità nulla;
- per il codice giallo si è rilevata una probabilità di insoddisfazione tale da portare all’abbandono particolarmente bassa: questo poichè il paziente di gravità più elevata pur di ottenere una cura è disposto ad attendere anche tempi elevati. Ciò nonostante i pazienti codice giallo si sono dichiarati quasi sempre poco soddisfatti dei tempi di attesa dalla presa in carico fino alla dimissione.



Mod.13- Catena di Markov comportamentale caso codice bianco medico

8.5.3 Validazione del modello

Le tre catene di Markov che si ottengono considerando la figura 5.9 con gli archi contrassegnati con le probabilità di transizione precedentemente calcolate e mostrate nelle figure 5.12, 5.13, 5.14 devono essere ora validate.

Per farlo, si è deciso di fare utilizzo di un software di calcolo: sia la validazione che il calcolo della distribuzione limite della catena sono stati realizzati con l'utilizzo di MATLAB. MATLAB sta per Matrix Laboratory, ma non si limita al solo calcolo matriciale e numerico; oltre ad essere un ambiente interattivo per il calcolo numerico, è un linguaggio di alto livello che permette l'analisi e la visualizzazione dei dati oltre che la programmazione. Il linguaggio, gli strumenti e le funzioni matematiche incorporate permettono di esplorare diversi ambiti e di raggiungere una soluzione più veloce rispetto all'utilizzo di fogli di calcolo o linguaggi di programmazione tradizionali come C / C++ o Java. MATLAB è utilizzato in una vasta gamma di applicazioni, tra cui analisi dei segnali e sistemi di telecomunicazioni oltre che per l'elaborazione di immagini e sistemi di monitoraggio video, test e misura, finanza computazionale e biologia computazionale.

Per la validazione delle catene di Markov, si è valutato l'equilibrio della soluzione attraverso il concetto di regolarità della catena e il teorema della distribuzione limite.

Si consideri la matrice di transizione P di dimensioni pari al numero di stati ammissibili (45x45). Sia n il numero di configurazioni ammissibili e sia S lo spazio di stati $S=\{O_1, \dots, O_n\}$, le proprietà di cui gode la matrice sono:

1. $P_{ij} \geq 0 \forall i, j \in S$
2. $\sum_{j=1}^n P_{ij} = 1$

Se queste proprietà sono verificate, la matrice P è detta stocastica; nel caso in cui anche la somma degli elementi delle singole colonne sia pari ad 1, la matrice P è detta doppiamente stocastica.

Si definisce anzitutto il concetto di Regolarità della catena di Markov:

Una catena di Markov è detta regolare se $\exists m > 0 : P^m$ ha solo elementi positivi.

Le catene di Markov relative ai tre casi codice bianco, verde e giallo medico raggiungeranno la regolarità per $m=14$.

Si definisce ora il Teorema di Distribuzione Limite:

Se P è la matrice di transizione di una catena di Markov regolare allora:

1. $\exists!$ vettore di probabilità $p^T > 0 : p^T P = p^T$
2. $\forall x_0^T = [0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0]$ (\forall distribuzione iniziale i – sima)

$$\lim_{n \rightarrow \infty} x_0^T P^n = v^T$$

$$v^T \equiv P^T;$$

3. $\lim_{n \rightarrow \infty} P^n = \underline{P}$

$$\underline{P} = \begin{bmatrix} P \\ P \\ P \\ \dots \\ P \end{bmatrix}; \underline{P} = P^T$$

Inizializzando la catena in un condizione iniziale x_0^T dopo un “elevato” numero di passi (asintoticamente) la probabilità di occupare o essere in una dato stato è data dalla i-esima componente del vettore di probabilità P.

$P^T = [P_1\ P_2\ \dots\ P_n]$ dove P^T è la distribuzione limite e corrisponde all’autovettore associato a $\lambda = 1$.

In Matlab si sono utilizzati i seguenti codici:

```
clc, clear all
```

```

%% Stato iniziale
x_passo_0 = zeros(27,1);
x_passo_0(1,1)=1;
dim_stato = length(x_passo_0)
% dichiaro un vettore che mi tiene conto del valore al passo
successivo
% inizialmente messo a tutti uno, anche se cambia valore alla prima
% iterazione
x_passo_successivo = x_passo_0*0 + 1;
%% Matrice P
load Pb.mat ;
P=Pb; %dove Pb rappresenta la matrice di transizione per il codice
bianco

%% Ciclo For

n = length(P)

for k=1:500
    %definisco x(k+1) = P x(k) ! inizialmente x_passo_successivo =
P* x(0)
    x_passo_successivo = P*x_passo_0;
    %ridefinisco la condizione iniziale in modo tale che x_passo_0
= x(k)
    x_passo_0 = x_passo_successivo;
end

```

8.5.4 Comportamento della catena nelle condizioni attuali

Nel caso in esame il comportamento delle catene in condizioni nominali è dato dalle distribuzioni limite calcolate in Matlab imponendo come condizione iniziale “l’arrivo” del paziente. Si è deciso di inserire degli ulteriori archi con probabilità unitaria a partire da ogni stato finale (dimissione o abbandono) verso lo stato iniziale di arrivo: in tal modo si può simulare la catena in continuo, come se ad ogni uscita di un paziente dal reparto di P.S. corrispondesse l’entrata di un altro paziente.

Le distribuzioni limite sono state ottenute iterando la matrice di transizione 650 volte; in questo modo si è ottenuta una matrice stazionaria in cui ogni riga rappresenta il vettore di probabilità limite: le componenti del vettore rappresentano le probabilità del paziente di trovarsi in ognuno degli stati della catena dopo un elevato numero di iterazioni.

Lo scopo del presente lavoro era il calcolo della probabilità di abbandono (allontanamento volontario dal reparto di emergenza) del paziente: nelle figure sono evidenziati in rosso gli stati di “abbandono” e le relative probabilità di trovarsi in dati stati. La somma di queste probabilità ci fornisce il tasso di allontanamento dal pronto soccorso lungo tutto il processo analizzato.

Per il codice bianco si è ottenuto:

Assegnazione codice e risorsa 2	Abbandono P.I.C. 5	Attesa P.I.C. Sodd 3	Visita iniziale 6	attesa referto finale 8	Abbandono attesa referto 9	Richiesta analisi o radiografia 7
0,198879	0,009944	0,115346	0,115348	0,103813	0,009343	0,011535
attesa referto rad. Sodd 18	2a visita /referto finale 20	Abbandono referto finale 21	attesa referto rad. Poco Sod 19	2a visita /referto finale sod 22	Abbandono referto finale 23	Abbandono attesa rad 17
0,004152	0,004153	0,000037	0,006805	0,006806	0,000136	0,000577
Attesa P.I.C. poco Sodd 4	Visita iniziale 24	Abbandono attesa referto 26	attesa referto finale sodd 25	Richiesta analisi o radiografia 727	attesa referto rad. Sodd 37	2a visita /referto finale sod 38
0,073583	0,073584	0,007947	0,066226	0,007358	0,001030	0,001030
Abbandono referto finale 39	attesa referto rad. Poco Sod 36	2a visita /referto finale sod 40	Abbandono referto finale 41	Abbandono attesa rad 35	Dimissione 42	
0,00002	0,00574	0,00574	0,00023	0,00059	0,17005	

-Tasso allontanamento volontario codice bianco: 2.88%.

Per il codice verde:

Assegnazione codice e risorsa 2	Abbandono P.I.C. 5	Attesa P.I.C. Sodd 3	Visita iniziale 6	attesa referto finale 8	Abbandono attesa referto 9	Richiesta analisi o radiografia 7
0,17166	0,00086	0,10643	0,10643	0,07556	0,00567	0,03086
Attesa referto analisi sangue 7s	Abbandono attesa analisi 10	attesa referto analisi sodd 11	2avisa/referto finale sodd 13	Abbandono referto finale 14	attesa referto analisi poco sod 12	2avisa/refertofinale sod 15
0,02006	0,00060	0,00582	0,00582	0,00017	0,01364	0,01364
Abbandono referto finale 16	Attesa referto radiografico 7r	attesa referto rad. Sodd 18	2a visita /referto finale 20	Abbandono referto finale 21	attesa referto rad. Poco Sod 19	2a visita /referto finale sod 22
0,00061	0,01080	0,00597	0,00597	0,00006	0,00465	0,00465
Abbandono referto finale 23	Abbandono attesa rad 17	Attesa P.I.C. poco Sodd 4	Visita iniziale 24	attesa referto finale sodd 25	Abbandono attesa referto 26	Richiesta analisi o radiografia 27
0,00009	0,00018	0,06437	0,06437	0,04570	0,00411	0,01867

Attesa referto analisi 27 s	Abbandono attesa analisi 28	attesa referto analisi sodd 30	2avisa/refertofinale sodd 31	Abbandono referto finale 32	attesa referto analisi poco sod 29	2avisa/refertofinale sod 33
0,00653	0,00052	0,00046	0,00046	0,00002	0,00555	0,00555
Abbandono referto finale 34	Attesa referto radiografico 27r	attesa referto rad. Sodd 37	2a visita /referto finale sod 38	Abbandono referto finale 39	attesa referto rad. Poco Sod 36	2a visita /referto finale sod 40
0,00033	0,01213	0,00364	0,00364	0,00015	0,00764	0,00764
Abbandono referto finale 41	Abbandono attesa rad 35	Dimissione 42				
0,00023	0,00085	0,15719				

-Tasso allontanamento volontario codice verde: 1.35%.

Ed infine per il codice giallo:

Assegnazione codice e risorsa 2	Abbandono P.I.C. 5	Attesa P.I.C. Sodd 3	Visita iniziale 6	attesa referto finale 8	Abbandono attesa referto 9	Richiesta analisi o radiografia 7
0,1621225	0,0000648	0,1107945	0,1107949	0,0675845	0,0000068	0,0432098
Attesa referto analisi sangue 7s	Abbandono attesa analisi 10	attesa referto analisi sodd 11	2avisa/referto finale sodd 13	Abbandono referto finale 14	attesa referto analisi poco sod 12	2avisa/refertofinale sod 15
0,0233334	0,0000070	0,0004760	0,0004760	0,0000000	0,0228504	0,0228503
Abbandono referto finale 16	Attesa referto radiografico 7r	attesa referto rad. Sodd 18	2a visita /referto finale 20	Abbandono referto finale 21	attesa referto rad. Poco Sod 19	2a visita /referto finale sod 22
0,0000002	0,0198766	0,0047684	0,0047684	0,0000000	0,0151062	0,0151062
Abbandono referto finale 23	Abbandono attesa rad 17	Attesa P.I.C. poco Sodd 4	Visita iniziale 24	attesa referto finale sodd 25	Abbandono attesa referto 26	Richiesta analisi o radiografia 27
0,0000002	0,0000020	0,0512631	0,0512633	0,0312705	0,0000003	0,0199926
Attesa referto analisi 27 s	Abbandono attesa analisi 28	attesa referto analisi sodd 30	2avisa/refertofinale sodd 31	Abbandono referto finale 32	attesa referto analisi poco sod 29	2avisa/refertofinale sod 33
0,0107960	0,0000032	0,0010764	0,0010764	0,0000001	0,0097164	0,0097164
Abbandono referto finale 34	Attesa referto radiografico 27r	attesa referto rad. Sodd 37	2a visita /referto finale sod 38	Abbandono referto finale 39	attesa referto rad. Poco Sod 36	2a visita /referto finale sod 40
0,0000010	0,0091966	0,0011027	0,0011027	0,0000001	0,0080930	0,0080930
Abbandono referto finale 41	Abbandono attesa rad 35	Dimissione 42				
0,0000008	0,0000009	0,1620355				

-Tasso allontanamento volontario codice giallo: 0.0009%.

Come atteso, la probabilità di abbandono è più elevata per codici di priorità più bassa e per il codice giallo è praticamente nulla: questo poichè i codici gialli pur di ottenere una cura sono disposti ad aspettare anche tempi molto lunghi. Traducendo queste probabilità in numero di pazienti che nel periodo considerato erano a rischio di abbandono si ottiene: su un totale di 31 codici bianchi un paziente a rischio di abbandono, mentre per il codice verde su un totale di 4073 accessi circa 55 pazienti a rischio abbandono.

Un notevole impatto sul tasso di allontanamento è generato dalla probabilità di abbandono in seguito ad attesa del referto finale: in generale il paziente intervistato si aspetta dei tempi per la dimissione più bassi rispetto alla media dei tempi rilevati; per il codice giallo è invece elevata l'insoddisfazione causata dall'attesa del referto dell'analisi di laboratorio.

8.6 Analisi di sensitività

Una volta validato il modello ed osservato il suo funzionamento nelle condizioni attuali, si vuole osservare in che modo possano cambiare le percentuali di abbandono nel caso in cui diminuissero determinati tempi di refertazione: si cerca ovvero di valutare come alcuni cambiamenti all'interno del processo, da intendersi come incremento della qualità del pronto soccorso, si possano ripercuotere sul tasso di allontanamento volontario dal reparto di emergenza. Questo paragrafo affronta, quindi, la cosiddetta analisi di sensitività: la valutazione degli effetti che si hanno sui risultati forniti dal modello a seguito di una variazione delle probabilità in gioco.

Dei tre casi presentati, il caso del codice giallo medico presenta una probabilità di abbandono notevolmente bassa, si preferisce quindi valutare i cambiamenti positivi che si possono ottenere nei casi dei codici di priorità più bassa. In particolare si vuole analizzare l'impatto sulla probabilità di

abbandono di una diminuzione dei tempi medi di dimissione e dei tempi medi di ottenimento del referto delle analisi.

Codice Bianco medico: diminuzione tempi di dimissione

Per il codice bianco la media campionaria dei tempi di dimissione fornisce un valore di 40,4 minuti. Si supponga un decremento nei tempi del 10%: si ottiene un tempo di attesa medio di 36.4 minuti. Considerando che, nel caso la catena parta da “Attesa P.I.C” soddisfatto, i limiti dichiarati sono stati di 45 min per la soddisfazione e 60 minuti per l’insoddisfazione, si ottiene a fronte delle precedenti probabilità di transizione (0.09 per l’abbandono e 0.81 per la dimissione):

$$\text{PrSodd} = 0.7849;$$

$$\text{PrInsodd} = 0.0152 ;$$

$$\text{PrPSodd} = 0.1999.$$

Nel caso la catena parta “Attesa P.I.C. poco soddisfatto” i limiti dichiarati sono di 35 min per la soddisfazione e 55 per l’abbandono. Si ottiene:

$$\text{PrSodd} = 0.4489;$$

$$\text{PrInsodd} = 0.0440;$$

$$\text{PrPSodd} = 0.5071.$$

Simulando nuovamente la catena con le nuove probabilità di transizione per 650 passi, si ottiene una nuova distribuzione limite in cui la somma delle probabilità delle transizioni verso l’abbandono fornisce il valore 1.6%. Si ottiene quindi una buona diminuzione del tasso di allontanamento volontario, a dimostrazione del forte peso che aveva la transizione verso l’abbandono da attesa referto finale.

Codice Bianco medico: diminuzione tempi refertazione radiografici

Per il codice bianco la media campionaria dei tempi di refertazione delle analisi radiografiche fornisce un valore di 97.2 minuti. Si supponga un decremento nei tempi del 10%: si ottiene un tempo di attesa medio di 87.5 minuti. Considerando che, nel caso la catena parta da “Attesa P.I.C” soddisfatto, i limiti dichiarati sono stati di 90 min per la soddisfazione e 130 minuti per l’insoddisfazione, si ottiene:

$$\text{PrSodd} = 0.5497;$$

$$\text{PrInsodd} = 0.0168;$$

$$\text{PrPSodd} = 0.4335.$$

Nel caso la catena parta “Attesa P.I.C. poco soddisfatto” i limiti dichiarati sono di 75 min per la soddisfazione e 125 per l’abbandono. Si ottiene:

$$\text{PrSodd} = 0.2660;$$

$$\text{PrInsodd} = 0.0304;$$

$$\text{PrPSodd} = 0.7036.$$

Simulando nuovamente la catena con le nuove probabilità di transizione per 650 passi, si ottiene una nuova distribuzione limite in cui la somma delle probabilità delle transizioni verso l’abbandono fornisce il valore 2.7%. La diminuzione nel tasso di abbandono non è molto elevata: la motivazione è da ricercarsi nella volontà del paziente di aspettare un tempo maggiore pur di ottenere il risultato degli esami diagnostici effettuati.

Codice Verde medico: diminuzione tempi di dimissione

Per il codice verde la media campionaria dei tempi di dimissione fornisce un valore di 70.45 minuti. Si supponga un decremento nei tempi del 10%: si ottiene un tempo di attesa medio di 63.4 minuti. Considerando che, nel caso la catena parta da “Attesa P.I.C” soddisfatto, i limiti dichiarati sono stati di 60

min per la soddisfazione e 100 minuti per l'insoddisfazione, si ottiene a fronte delle precedenti probabilità di transizione (0.075 per l'abbandono e 0.925 per la dimissione):

$$\text{PrSodd} = 0.4338;$$

$$\text{PrInsodd} = 0.0364;$$

$$\text{PrPSodd} = 0.5298.$$

Nel caso la catena parta "Attesa P.I.C. poco soddisfatto" i limiti dichiarati sono di 50 min per la soddisfazione e 97.5 per l'abbandono. Si ottiene a fronte delle precedenti probabilità di transizione (0.09 per l'abbandono e 0.91 per la dimissione):

$$\text{PrSodd} = 0.2556;$$

$$\text{PrInsodd} = 0.0473;$$

$$\text{PrPSodd} = 0.6971.$$

Simulando nuovamente la catena con le nuove probabilità di transizione per 650 passi, si ottiene una nuova distribuzione limite in cui la somma delle probabilità delle transizioni verso l'abbandono fornisce il valore 0.999347%. Si ottiene quindi una diminuzione del tasso di allontanamento volontario, analogamente al codice prioritario bianco.

Codice Verde medico: diminuzione tempi refertazione analisi di laboratorio

Per il codice verde la media campionaria dei tempi di refertazione delle analisi di laboratorio fornisce un valore di 160.8 minuti. Si supponga un decremento nei tempi del 10%: si ottiene un tempo di attesa medio di 144.72 minuti. Considerando che, nel caso la catena parta da "Attesa P.I.C" soddisfatto, i limiti dichiarati sono stati di 150 min per la soddisfazione e 200 minuti per l'insoddisfazione, si ottiene:

$PrSodd = 0.6041;$

$PrInsodd = 0.0029;$

$PrPSodd = 0.3930.$

Nel caso la catena parta “Attesa P.I.C. poco soddisfatto” i limiti dichiarati sono di 130 min per la soddisfazione e 188 per l’abbandono. Si ottiene:

$PrSodd = 0.2309;$

$PrInsodd = 0.0152;$

$PrPSodd = 0.7539.$

Simulando nuovamente la catena con le nuove probabilità di transizione per 650 passi, si ottiene una nuova distribuzione limite in cui la somma delle probabilità delle transizioni verso l’abbandono fornisce il valore 1.34%. La diminuzione nel tasso di abbandono non è molto elevata: la motivazione è da ricercarsi, come nel caso del codice bianco, nella volontà del paziente di aspettare un tempo maggiore pur di ottenere il risultato degli esami diagnostici effettuati; ciò nonostante il livello di insoddisfazione risulta essere minore, indice di un’opportunità di miglioramento nella percezione della qualità del reparto da parte del paziente.

In conclusione si può affermare che la maggior parte del rischio di abbandono si ha in fase di attesa/presa in carico ed in fase di attesa del referto finale; generalmente una volta sottoposto a richiesta di esami, il paziente, è disposto ad aspettare un tempo maggiore per conoscerne i risultati.

Per operare attività di miglioramento continuo nell’ottica della qualità, occorre eliminare gli sprechi durante tutto il processo, in modo da ridurre i tempi necessari allo svolgimento delle singole attività e di conseguenza i tempi di attesa per il paziente; in tal modo, come evidenziato anche dall’analisi di sensibilità si potrà ridurre il livello di insoddisfazione e di conseguenza il tasso di allontanamento volontario dal reparto di emergenza.

CAPITOLO 9

La modellazione delle U.O. di emergenza dei D.E.A.

9.1 – L'analisi dei processi di una U.O. di Pronto soccorso di un DEA

Per quanto fin qui esposto è possibile, a tutti gli effetti, considerare un ospedale come un impianto per la erogazione di servizi sanitari mediante l'implementazione di processi produttivi, le cure prestate ai pazienti, che risultano presentare le seguenti caratteristiche:

1. estrema diversificazione della produzione/erogazione;
2. *lead time* non definito (o, meglio, definito solo nominalmente);
3. elevato contenuto tecnologico dei processi;
4. sollecitazione delle risorse secondo due modalità: elezione (cioè programmati) e emergenza (non programmabili, né rifiutabili);
5. logica di erogazione dei processi di tipo *pull*

Ai fini della dimensionamento di un impianto ospedaliero, le caratteristiche menzionate devono essere oggetto di attenta valutazione. Come visto nella *literature review*, infatti, l'orientamento dei sistemi sanitari avanzati tende a separare i flussi che sollecitano l'impianto ospedaliero, predisponendo due tipologie differenziate di Dipartimento per la gestione dei pazienti (rispettivamente - vedi punto 4 -, quelli elettivi e i DEA), il cui afflusso variando nel tempo costituisce l'onda di carico dell'intero sistema.

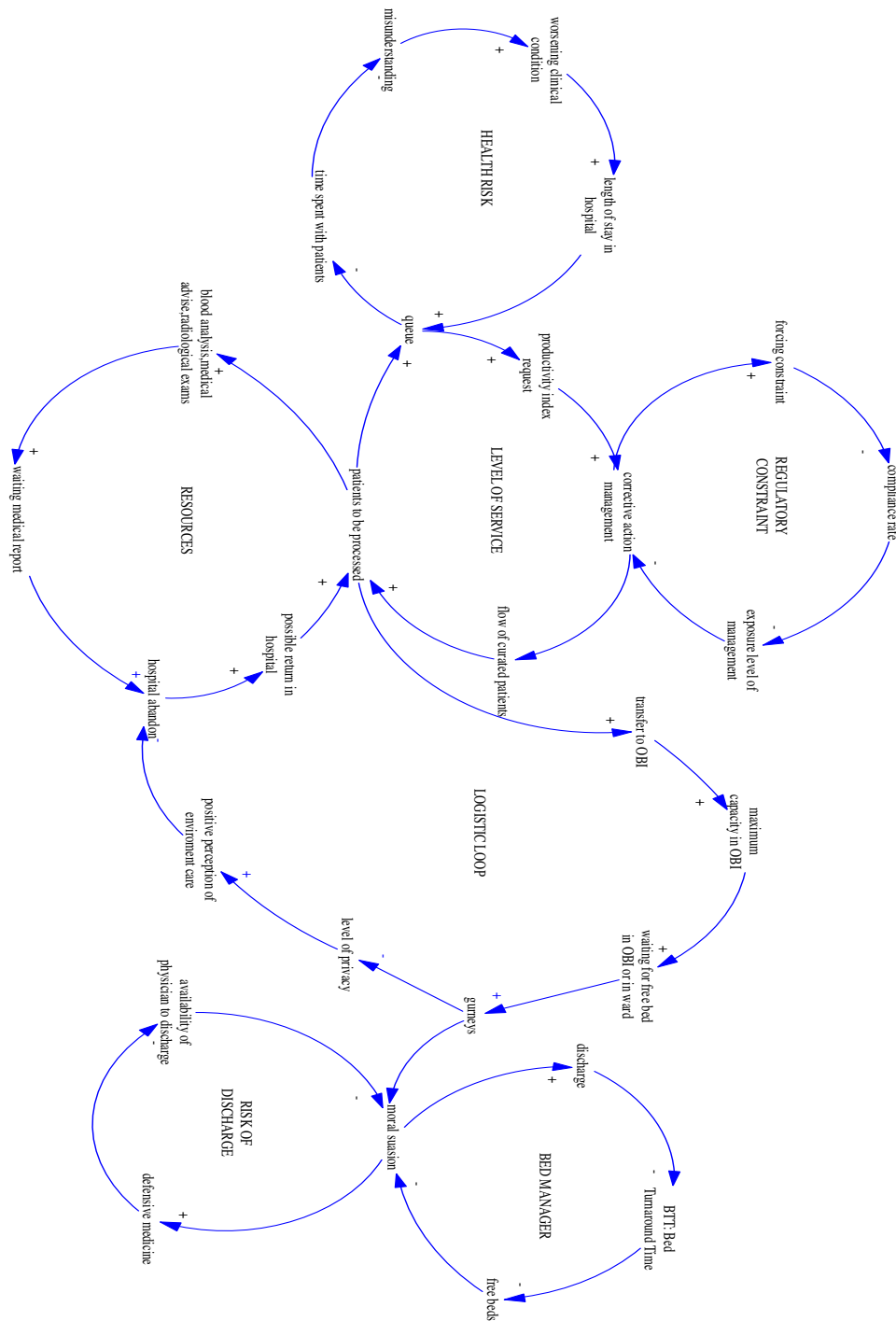
L'utilizzo del vocabolo onda va inteso con parziale abuso semantico del termine, essendo la sollecitazione di carico, immaginabile come una funzione periodica (in senso largo), pur non presentando le caratteristiche di regolarità tipica di un'onda (ad esempio sinusoidale).

Tanto premesso la parte di detta funzione imputabile agli arrivi programmati esula dai nostri obiettivi, ricadendo in una casistica del dimensionamento di impianto ampiamente trattata in letteratura e, soprattutto, ben consolidata.

La nostra attenzione si focalizzerà, invece, sui DEA ospedalieri (Dipartimento di Emergenza a Accettazione) che, articolati in U.O. di emergenza e U.O. di urgenza, devono essere dimensionati, nel rispetto delle caratteristiche sopra esposte, per rispondere alla variabilità tipologica e temporale dei processi da erogare.

A tal fine inizieremo, anche sfruttando le conclusioni a cui si è giunti nel precedente capitolo mediante l'approccio markoviano proposto, ad analizzare i processi di una U.O. di Pronto Soccorso, che riassume in sé l'erogazione emergenziale di un DEA, per poi passare nel Capitolo 10 all'analisi delle dinamiche dei reparti di urgenza.

Lo strumento che utilizzeremo sarà quello della simulazione in logica System Dynamics che, come già trattato, parte dalla valutazione dei loop fenomenologici che governano i sistemi.



Mod.16 – Causal Loop Diagram per le dinamiche di una U.O. di P.S.

Nel casual loop diagram riportato in *Mod.16* – sono state analizzate tutte le variabili che entrano in gioco nella gestione di un pronto soccorso e di un ospedale, in particolare sono stati individuati sette loop:

1. Livello di servizio;
2. Vincoli normativi;
3. Rischio sanitario;
4. Risorse disponibili;
5. Loop logistico;
6. Bed management;
7. Rischio dimissioni;

Il livello di servizio rappresenta il livello di performance di un ospedale in termini di qualità delle prestazioni offerte e qualità percepita dal paziente che vi si rivolge, all'aumentare del numero di pazienti, si vengono a creare delle colli di bottiglia in più punti del sistema, dal triage all'accesso in medicheria, dall'ingresso alle OBI al ricovero nei reparti del DEA o dell'intero ospedale. Il sovraffollamento di un ospedale aumenta l'indice di produttività richiesta e dunque gli interventi manageriali correttivi necessari. Qualora questi dovessero aumentare, aumenterà di conseguenza il flusso dei pazienti curati richiamando l'attenzione di nuovi pazienti.

Quest'ultimo loop (positivo) viene bilanciato dal loop negativo "vincoli normativi", infatti, all'aumentare dagli interventi manageriali collettivi, aumentano le forzature vincolari (ad esempio, barelle nei corridoi). Tale incremento induce a una riduzione della "tasso di conformità" e successivamente aumenta il livello di esposizione della direzione sanitaria comportando una diminuzione degli interventi correttivi.

È importante sottolineare il rischio sanitario dovuto al sovraffollamento negli ospedali (terzo loop). In un ospedale sovraffollato il tempo dedicato a ciascun

paziente si riduce, di conseguenza il rischio che i medici possano commettere degli errori aumenta. Quanto detto può far peggiorare le condizioni cliniche dei pazienti che sono costretti ad aumentare il loro tempo di permanenza all'interno della struttura e comportando inevitabilmente un aggravarsi dello stato di sovraffollamento.

Il quarto loop riguarda le risorse disponibili: all'aumentare del numero di pazienti, c'è una maggiore richiesta di esami di laboratorio, esami di radiologia, e dunque le attese per l'ottenimento di un referto aumentano. In generale nel pronto soccorso un numero notevole di pazienti attende i referti per un periodo di tempo superiore all'ora. Queste lunghe attese dei referti possono implicare, nei pazienti non affetti da gravi condizioni cliniche, l'allontanamento volontario dalla struttura ospedaliera.

Tale situazione può però, implicare un possibile ritorno del paziente che non ha concluso efficacemente il suo percorso di cura, rientrando nel flusso dei nuovi accessi in ospedale.

Sempre dalla variabile “pazienti da processare” parte un nuovo loop definito “loop logistico” poiché, i pazienti che fanno accesso al pronto soccorso possono non essere dimessi nell'arco della giornata, ma mentre alcuni necessitano di essere tenuti sotto osservazione per tempi brevi e quindi fanno accesso alle OBI, altri invece possono essere trasferiti nei reparti. All'aumentare del numero di pazienti nelle OBI o nei reparti si arriva alla loro saturazione e alla presenza di barelle lungo i corridoi della struttura.

L'aumento di barelle provoca da un lato un abbassamento del livello della privacy e dunque una riduzione della percezione positiva della cura da parte del paziente che tenderà a lasciare l'ospedale in cerca di soluzioni migliori.

D'altro canto, un numero sempre più elevato di pazienti in barella aumenta la sollecitazione che viene fatta ai capi dipartimento a dimettere i pazienti. In questo modo i bed turn around time (BTT) (tempi di permanenza)

diminuiscono e aumentano i posti liberi questo fa sì che la “moral suasion” si riduca. Questo loop è negativo, e sarà bilanciato dal fatto che una eccessiva “moral suasion” porterà a un rischio di dimissioni elevato che a sua volta farà sì che i medici non saranno più disposti a dimettere. Questo loop (positivo) si chiuderà collegando la variabile “disponibilità del medico a dimettere” con la moral suasion con segno negativo in quanto all’aumentare della prima la seconda diminuisce.

9.2 Lo Stock & Flow Diagramm per la simulazione di una U.O. di P.S.

Per definire il modello simulativo di un qualsiasi sistema complesso da analizzare con il metodo SD, è necessario tradurre le informazioni immesse dall'utente in equazioni matematiche atte a formulare il modello di riferimento.

Come abbiamo visto nel precedente Capitolo 4, dopo aver definito tutte le variabili e individuato quali siano quelle di flusso, quali quelle di livello e quali quelle ausiliarie è possibile costruire il modello e studiare il comportamento nel tempo.

I simboli utilizzati sono identici a quelli presentati per la System Dynamics nel capitolo dedicato agli strumenti di modellazione simulativo, sia in termini di diagrammi stock, che per quanto riguarda i costrutti relativi ai controlli sui flussi.

Particolare rilievo assume, invece, l’approccio utilizzato per realizzare ed articolare il modello

9.2.1 Approccio matriciale

A ciascuna variabile di stock, è stato attribuito un valore iniziale, non scalare ma di natura matriciale. La matrice ha dimensioni fisse (\mathbf{nxm}) dove, \mathbf{n} è il numero di pazienti, che fanno accesso all'ospedale nel corso del periodo della simulazione ed \mathbf{m} è il numero di caratteristiche associate a ciascun paziente. Nel nostro caso $m=15$, le prime tre colonne riguardano data e ora d'arrivo, le successive tre se il paziente necessita di cure mediche, ortopediche o chirurgiche. Di seguito a queste sei colonne abbiamo il codice di triage, una colonna con 1 o 0 a seconda se il paziente ha effettuato analisi cliniche, test radiologici o se necessita di consulenze specialistiche e infine se è stato in OBI, OBI2 o se è stato ricoverato, se il paziente (codice rosso) è arrivato mediante ambulanza o con mezzi propri, perché come vedremo seguirà un percorso diverso, più una colonna delle presenze con tutti uno con scopo puramente "tecnico".

Da un punto di vista modellistico, il ricorso alle matrici risponde all'esigenza di associare il maggior numero d'informazioni all'evoluzione della "materia fluida" del sistema in via di simulazione. Ove l'evoluzione della continuità del sistema fosse rimandata ad uno scalare ovvero ad una visione semplicemente n - dimensionale (espressa quindi da un vettore colonna) si renderebbe necessario, passaggio per passaggio e nel corso di tutta l'evoluzione del sistema, associare a quest'ultima una serie di ipotesi di natura chiaramente semplificativa, che ne indebolirebbero la valenza e l'affidabilità dei risultati. In tal senso, è possibile concludere, in punto meramente modellistico, che la realtà simulativa in System Dynamics, può essere, nella sua forma di maggiore dettaglio, sintetizzata attraverso la determinazione di due entità, chiaramente distinte :

- **Una materia evolvente**, di dimensioni (\mathbf{nxm}).

- **Un'architettura evolutrice**, composta di opportuni costrutti.

Potremmo arrivare ulteriormente a concludere che: al degenerare della materia evolvente verso un'entità monodimensionale (soluzione da un punto di vista logico-matematico sempre possibile) deve aumentare la complessità dell'architettura evolutrice, determinandosi in tal senso l'inevitabile introduzione di una serie di ulteriori costrutti, che allo stesso tempo nascondono, implicitamente, altrettante se non maggiori ipotesi (si pensi che la mancata distinzione tra due oggetti monodimensionali determina per la loro evoluzione, quanto meno, l'inserimento di un tempo medio, se non l'adozione di un corpo di soluzioni statistiche più evolute). Ai fini dei costrutti, nella visione modellistica evoluta, qui proposta, se ne distinguono, in via generale, tre categorie:

Architetture evolutive principali. I costrutti di cui si compongono elaborano il flusso di avanzamento della materia evolvente e si distinguono, al di là dell'ovvia ma non obbligatoria complessità, poiché ai loro estremi, ad inizio simulazione, si hanno stock iniziale con tutte le righe diverse da zero e stock finale costituito dalla matrice nulla, mentre, ad esecuzione della simulazione, si leggerà la situazione inversa. Il costrutto fondamentale che caratterizza questa tipologia di architettura (al netto delle differenti formalizzazioni in cui esso si rappresenta) è costituito dalla valutazione di j proprietà, con $j = 1, \dots, m$, che, opportunamente ordinate, consentono il processamento di un'attività rispetto ad un'altra, modificando la struttura dell'entità evolvente.

Architetture evolutive secondarie. Tra queste rientrano tutti i costrutti di natura diversa da quella primaria, che elaborano fenomeni, raccogliendo informazioni interne al modello e restituendole internamente allo stesso. Ad esempio indichiamo anelli chiusi ed anelli aperti di successione sottoprocessi, in cui una o più risorse, modellizzate con dimensioni (hxk) (potenzialmente diverse, quindi, da n ed m), evolvono per loro conto, tornando disponibili o

meno per processi appartenenti alle architetture principali ed assumendo , *timestep* per *timestep*, valori diversi.

Architetture di supporto e/o complemento. Si caratterizzano per essere strutturate da costrutti in grado di integrare la logica System Dynamics con logiche differenti, che meglio si adattano alla modellazione di fenomeni complementari alla fenomenologia *core* trattata nel modello.

Il modello, di seguito presentato, si è avvalso della logica di modellazione sopra presentata. In concreto, il passaggio di un paziente da una variabile di stock a quella successiva, comporta il trasferimento dell'intera riga associata a quel paziente. Questo consente di visualizzare, in qualsiasi istante, il numero di pazienti che si trovano in ciascuna variabile di stock del modello, visualizzando, per ogni paziente, tutte le specifiche ad esso associate.

L'utilizzo delle matrici all'interno del modello impedisce la perdita di informazioni, associate a ciascun paziente, consentendo di visualizzare, istante per istante, tutte le righe della matrice che attraversa l'intera struttura. In ogni variabile di stock del modello, tale matrice, di dimensione fissa, presenterà tutte righe nulle tranne quelle corrispondenti al paziente presente, in quell'istante, all'interno della variabile considerata.

Di seguito si riporta il modello realizzato con l'ausilio del software Powersim.

L'orizzonte temporale scelto per la simulazione è pari a 30 giorni ed il *timestep* pari ad un minuto.

È stato scelto tale intervallo in modo da prendere il massimo orizzonte temporale su cui la simulazione ha senso statistico, anche se scegliere un orizzonte temporale piccolo dà vantaggio in termini di validazione dato che per quest'ultima si rende necessario raccogliere 30 set di dati con ampiezza

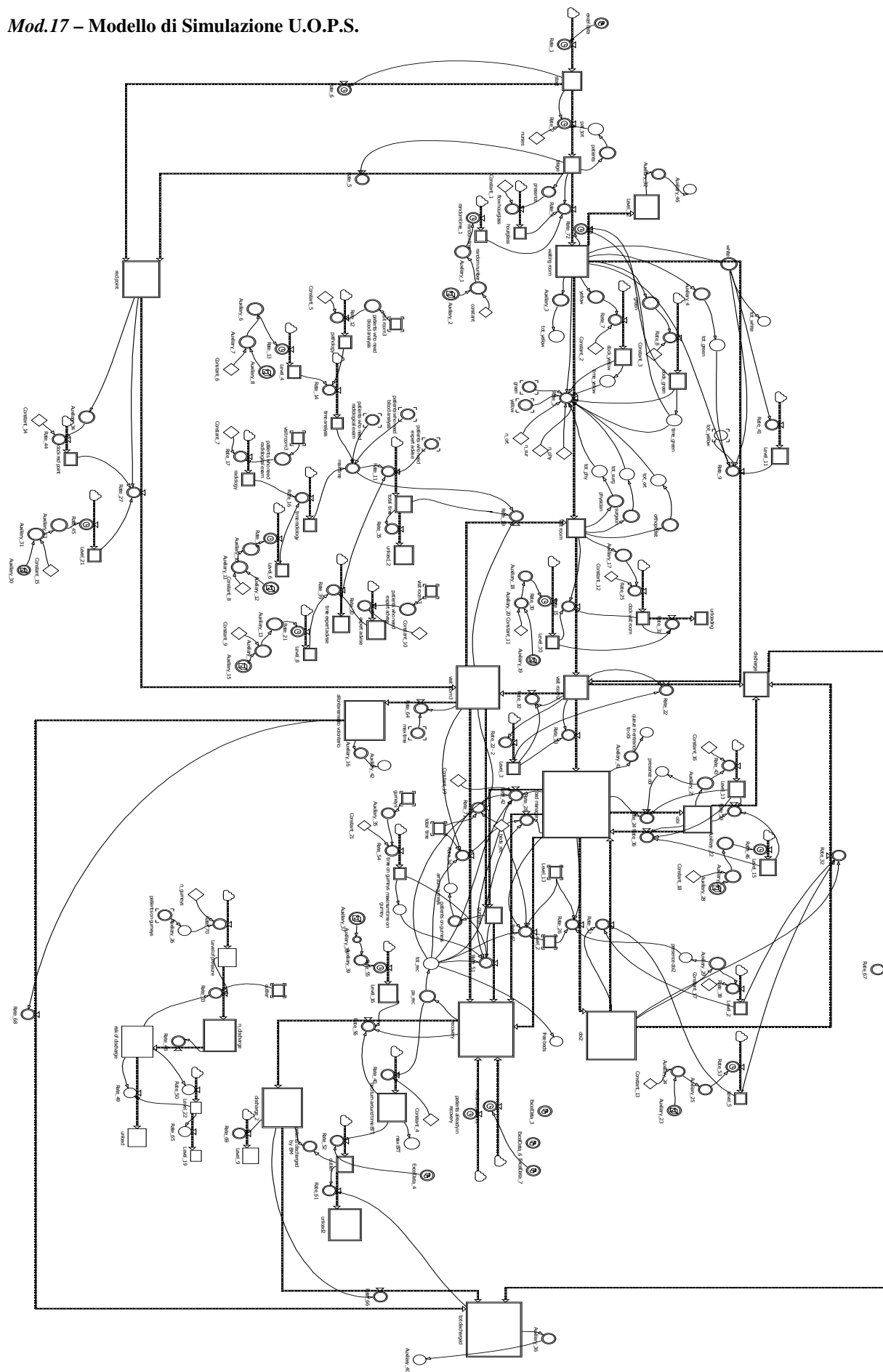
pari all' orizzonte temporale di riferimento ed inoltre è opportuno che i dati siano tutti rappresentativi del fenomeno in esame.

9.3 Il modello di Simulazione

In questo paragrafo descriveremo come è stato modellato il flusso di pazienti all' interno del Pronto soccorso.

Nella *Mod.17*, è rappresentato il modello per intero, in seguito sarà analizzato nel dettaglio.

Mod.17 – Modello di Simulazione U.O.P.S.



Il flusso in entrata passa prima per il Triage, dove gli infermieri eseguono una valutazione preliminare dello stato di salute del paziente e assegnano un codice di triage:

- I codici rossi vengono inviati direttamente al Red Point, dove un team di medici e infermieri esegue una prima operazione sul paziente.

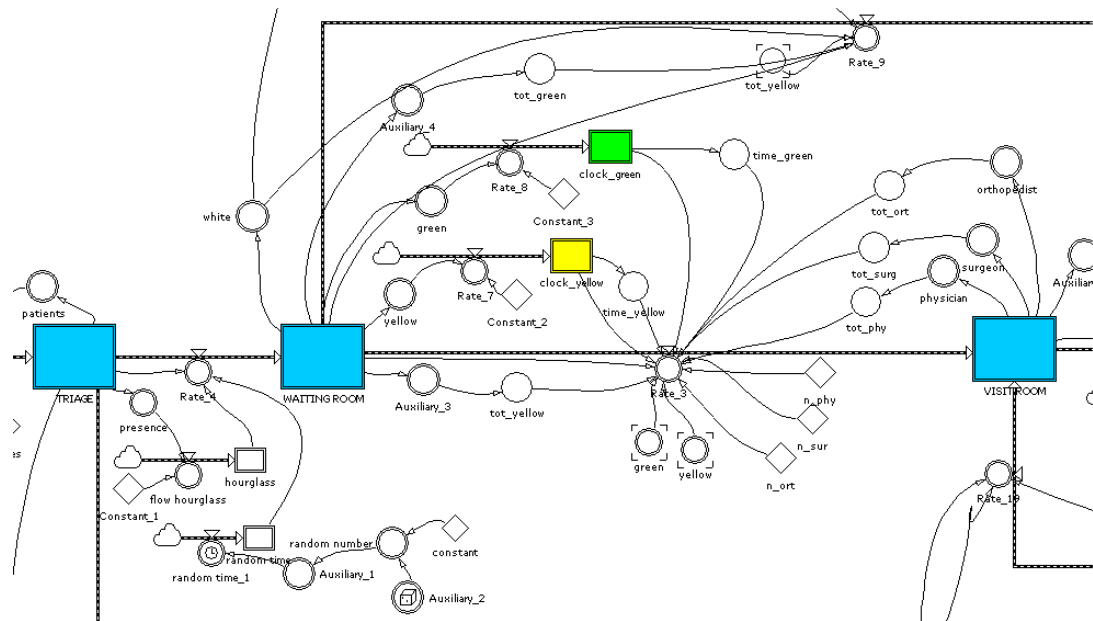
È importante notare che sono stati modellati anche i codici rossi che arrivano con l'ambulanza invece che con i propri mezzi di trasporto. Tali pazienti non effettuano la fase di triage, ma vanno direttamente al red point.

- I codici gialli, verdi e bianchi dopo il triage entrano nella Waiting room e attendono il loro turno. I pazienti entrano nella medicheria secondo uno schema di priorità, in particolare, è data priorità ai codici di colore giallo, poi i verdi e infine i bianchi.

I codici bianchi entrano in medicheria solo se in quell'istante i codici gialli o verdi sono assenti. Per questo motivo, nel modello abbiamo definito una condizione in base al tempo di permanenza del paziente (codice bianco) in sala d'attesa. Se infatti, un codice bianco rimane più di 24 ore volontariamente lascia l'ospedale.

Non appena un medico, un chirurgo o un ortopedico è libero, entra il paziente in relazione alla sua necessità (cioè se necessita di un medico, di un ortopedico o di un chirurgo). La priorità per l'ingresso è del paziente con il codice di gravità maggiore e a parità di codice entra colui che sta attendendo da maggior tempo.

Nella *Mod.18* è mostrato quanto detto:



Mod.18 – Sottosistema di accesso

La condizione scritta nella variabile di flusso “rate_3” ha lo scopo di gestire gli ingressi alla medicheria ed è la seguente:

```
FOR (i=1..11300; j=1..15 \ IF ('waiting room'[i;7]=1; IF ('waiting room'[i;4]=1 AND n_phy > tot_phy OR 'waiting room'[i;5]=1 AND n_sur > tot_surg OR 'waiting room'[i;6]=1 AND n_ort > tot_ort; IF (clock_yellow[i]=time_yellow; yellow[i;j]; IF ('waiting room'[i;7]=1; IF ('waiting room'[i;4]=1 AND n_phy > tot_phy OR 'waiting room'[i;5]=1 AND n_sur > tot_surg OR 'waiting room'[i;6]=1 AND n_ort > tot_ort; IF (clock_green[i]=time_green; IF (tot_yellow=0; green[i;j]; IF('waiting room'[i;7]=1; IF('waiting room'[i;4]=1 AND n_phy > tot_phy OR 'waiting room'[i;5]=1 AND n_sur > tot_surg OR 'waiting room'[i;6]=1 AND n_ort > tot_ort; white[i;j]; 0))))))))))
```

Le variabili di livello utilizzati per gestire l'ingresso al Pronto Soccorso sono:

- Triage: rappresenta il numero di pazienti che hanno bisogno di entrare in pronto soccorso.

- Sala d'attesa: rappresenta il numero di pazienti che hanno bisogno di entrare in medicheria e sono in attesa di un medico, un chirurgo o un ortopedico in base alle loro esigenze;
- Medicheria: numero di pazienti sotto osservazione;

Le variabili ausiliarie: “*giallo*”, “*verde*” e “*bianco*” contano il numero di persone con, rispettivamente codice bianco verde e bianco che sono presenti in sala d'attesa, mentre le costanti n_{phy} , n_{sur} e n_{ort} rappresentano rispettivamente il numero di medici, chirurghi e ortopedico che sono presenti nella medicheria.

I medici lavorano su tre turni: 08:00-14:00, 14:00-20:00 e 20:00-08:00, e ogni turno prevede quattro medici, un chirurgo e due ortopedici.

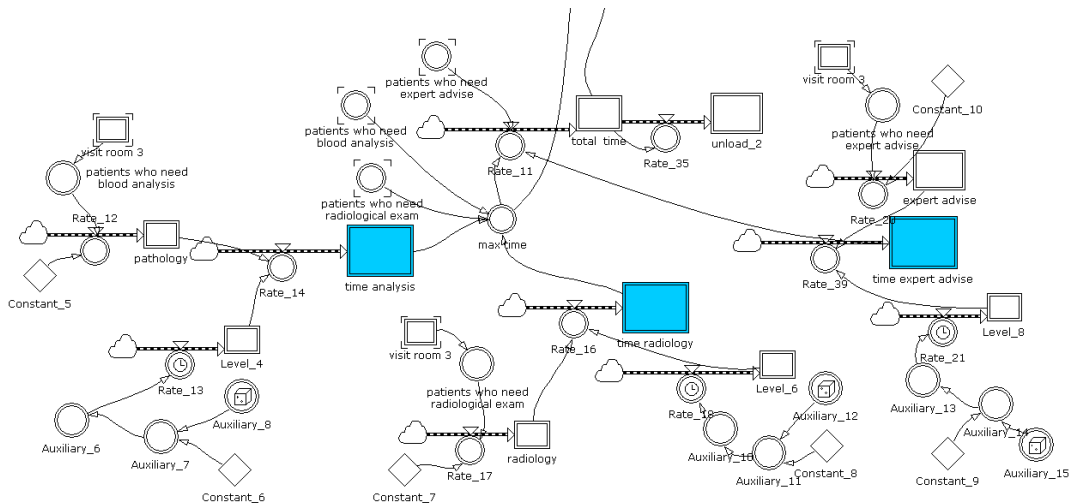
Il tempo di attesa dei pazienti durante la fase di triage è calcolata mediante una funzione random limitata in un intervallo temporale che ha un valore minimo di 6 minuti e un valore massimo di 10 minuti.

Ovviamente, il tempo di attesa in sala d'attesa dipende da quando i medici nella medicheria diventano disponibili per i pazienti. Per modellare questa situazione abbiamo introdotto un'altra funzione random per valutare il tempo necessario per il medico per effettuare una visita . I valori forniti dalla funzione casuale sono compresi tra 1 e 7 minuti.

Se il paziente ha bisogno di esami di laboratorio, esami radiologici o di un consulto da parte di specialisti allora seguirà il percorso che lo condurrà in “medicheria 3”, questa variabile livello è utilizzato solo per questi pazienti.

La medicheria 3 è una "medicheria fittizia" che è legata solo ad alcuni contatori. Questi contatori sono utilizzati nel modello per valutare i tempi di attesa necessari per ottenere le refertazioni di eventuali esami del sangue, esami radiologici o consulenze. Anche in questo caso è stata utilizzata una funzione random per valutare i tempi di attesa.

Nella Mod.19 sono mostrati i contatori.



Mod.19 – Costruito calcolo tempi in visit room

In particolare, i tempi di attesa del paziente durante la fase del “paziente presente in visit room 3” sono valutati nel seguente modo:

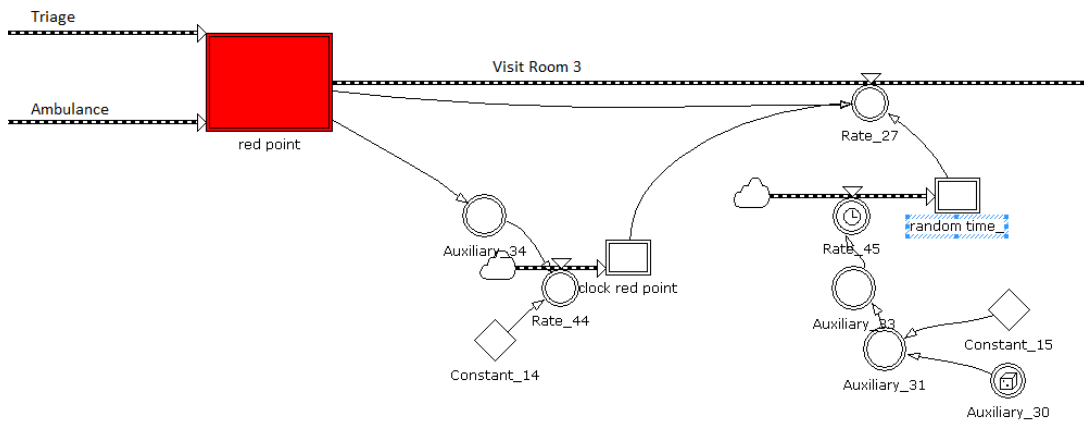
FOR (i=1..11300 \ IF ('max time'[i] > 0 AND 'expertadvise'[i] > 0; 'max_time'[i] + 'expert advise'[i]; IF ('patients who need expert advise'[i]=0 AND 'max time'[i] > 0; 'max time'[i]; IF ('patients who need expert advise'[i] > 0 AND 'max time'[i] = 0; 'patients who need expert advise'[i]; 0))))

Dove il tempo max è calcolato nel seguente modo:

FOR (i=1..11300 \ IF ('time analysis'[i] > 0 AND 'time radiology'[i] > 0 OR 'patients who need blood analysis'[i] = 0 AND 'time radiology'[i] > 0 OR 'patients who need radiological exam'[i] = 0 AND 'time analysis'[i] > 0; MAX('time analysis'[i]; 'time radiology'[i]) ;0))

Tale valore è calcolato prendendo il massimo tra il tempo per ottenere la refertazione delle analisi di laboratorio e il tempo per la radiologia (ovviamente se il paziente effettua solo una di queste due attività si prende solo quel tempo) a tale tempo massimo valutiamo anche il tempo necessario al medico per leggere le analisi e definire una diagnosi per il paziente.

I pazienti con codice rosso, che si trovano nel red point vanno tutti nella “medicheria 3”, dal momento che, data la gravità, saranno sottoposti ad esami diagnostici.



Mod.20 – Costrutto di gestione “Red Point”

successivamente il paziente va direttamente al ricovero solo se viene confermato il codice rosso, altrimenti seguirà lo stesso percorso dei codici gialli o verdi.

Tutti i pazienti che non necessitano di esami diagnostici o del ricovero lasciano l’ospedale passando per la variabile di livello “medicheria 2” attraverso la seguente condizione:

FOR (i=1..11300; j=1..15 \ IF (Level_3[i] > 2; IF ('visit room 2'[i;12] = 0 AND 'visit room 2'[i;13] = 0 AND 'visit room 2'[i;14] = 0 OR 'visit room 2'[i;9] = 0 AND 'visit room 2'[i;10] = 0 AND 'visit room 2'[i;11] = 0; 'visit room 2'[i;j]; 0)))

Le variabili di livello sono:

- Medicheria 2: rappresenta il numero di pazienti che non hanno bisogno di ulteriori esami o ricovero.
- Discharge: rappresenta il numero di pazienti dimessi.

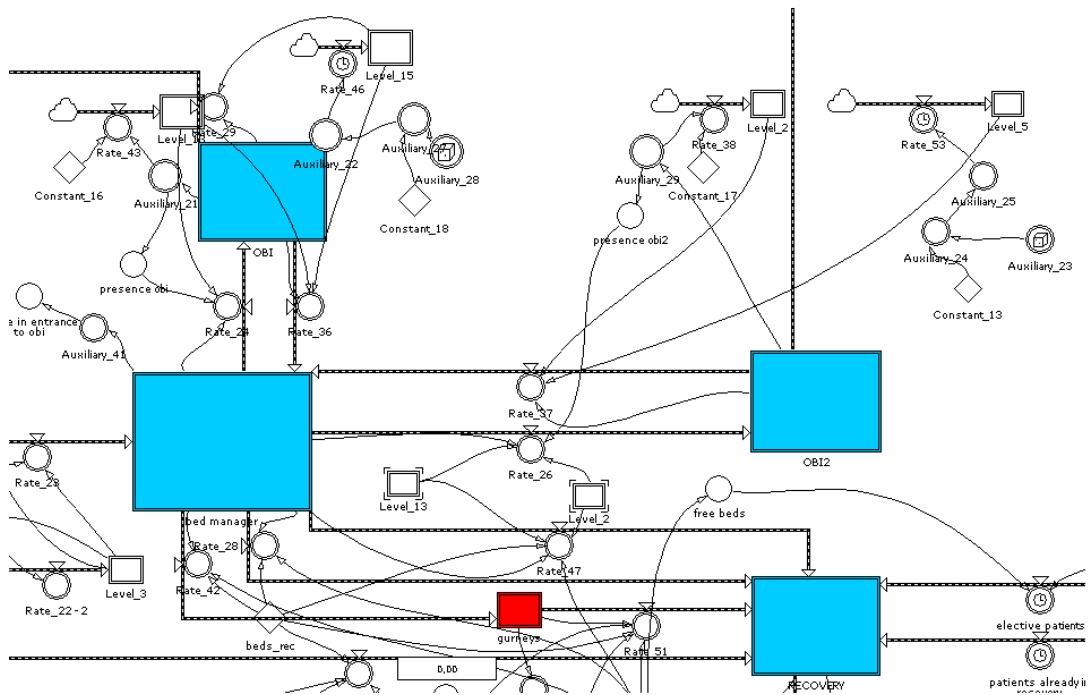
I pazienti per i quali non è stata definita una diagnosi (e, pertanto, devono comunque essere tenuti sotto osservazione) vanno nelle OBI (OBI oppure OBI 2).

Se è necessario un ricovero, il modello fornisce due possibilità. Nel caso in cui ci siano letti disponibili, il paziente sarà spostato direttamente al reparto di competenza. In caso contrario, il paziente sarà posizionato in barella fino a quando non si libererà un posto letto.

```
FOR(i=1..11300;j=1..15\IF('bed manager'[i;14]=1;IF(beds_rec <= tot_rec;'bed manager'[i;j];0)))
```

Nella configurazione “as-is” tutti i pazienti passano attraverso la variabile di livello “Bed Manager” che in questo momento si limita solo a far entrare i pazienti nelle OBI o nei reparti.

Se non sono presenti posti liberi, i pazienti sono posizionati in barella (variabile di livello gurneys) , come mostrato in *Mod.21*.



Mod.21 – Costrutto gestione barelle

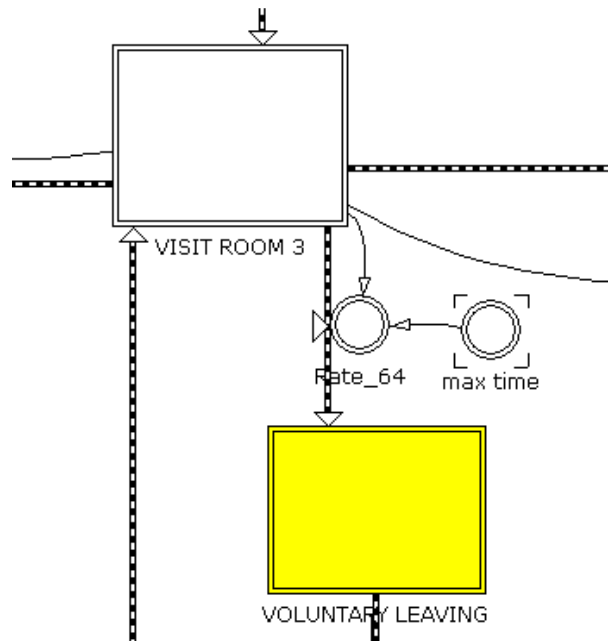
Le variabili di livello rappresentati in figura sono:

- Ricovero: comprende tutti i pazienti in fase di ricovero;
- Obi e Obi 2: i pazienti presenti nelle OBI.
- Gurneys: i pazienti posizionati in barella.

Inoltre, nel modello di simulazione abbiamo anche valutato altro aspetto critico del Pronto Soccorso ossia l' allontanamento volontario del paziente.

Le lunghe attese per la refertazione o per la visita medica spesso portano i pazienti a lasciare la struttura ospedaliera, ovviamente solo nel caso in cui le condizioni cliniche lo consentono. Nel modello è stato valutato l'allontanamento volontario del paziente per eccessive attese della refertazione, secondo le risultanze emerse nel precedente Cap.8. Quando il "Waiting Time" supera un valore limite ("max time") il paziente abbandona l'ospedale.(nei limiti stabiliti mediante il richiamato approccio markoviano).

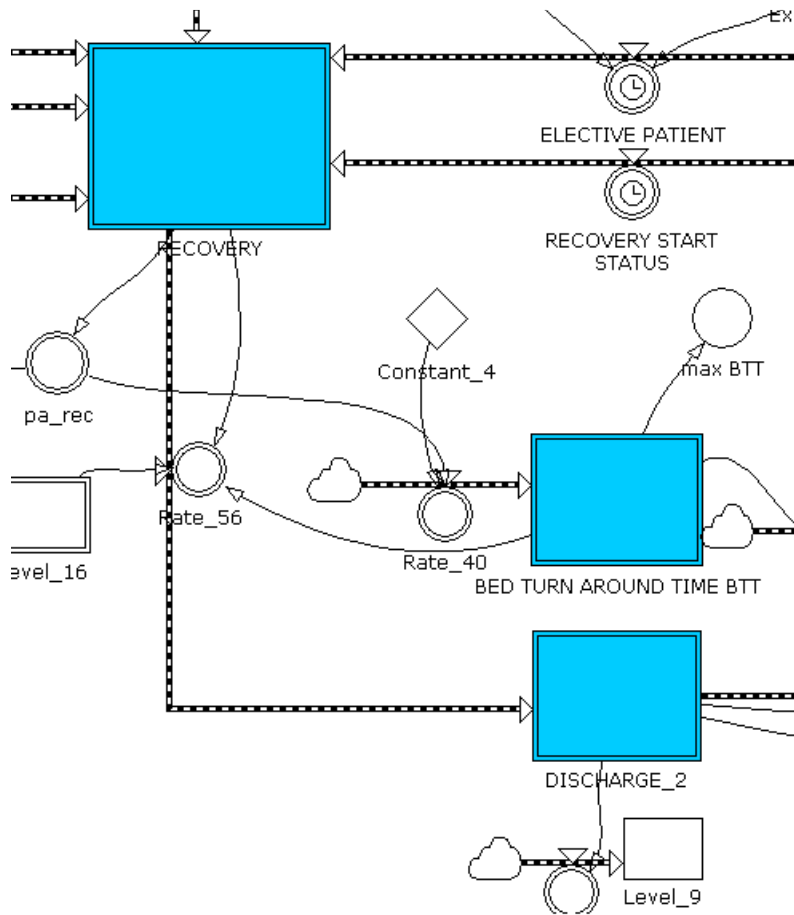
In figura è rappresentato quanto detto.



Mod.22 – Costrutto abbandono volontario

Finora, sono state descritte le principali attività che un paziente può svolgere all'interno del pronto soccorso. Tuttavia, l'argomento centrale di tale lavoro è la gestione di letti ospedalieri. Questo perché, come è stato detto più volte, quotidianamente bisogna far fronte ad un ingresso ad emergenza e uno programmato.

La variabile di flusso "elective patient" mostrata in *Mod.23* alimenta quotidianamente il ricovero di un certo numero di pazienti a seconda della disponibilità di letti. La variabile di flusso "Recovery Start Status" inserisce all'istante iniziale della simulazione, 800 pazienti.



Mod.23 – Costrutto di ingrasso elettivo

Le variabili di livello di questa figura sono:

- Bed Turnaround Time: è una variabile che misura la durata della degenza del i-esimo paziente.
- Discharge: dove troviamo tutti i pazienti dimessi.

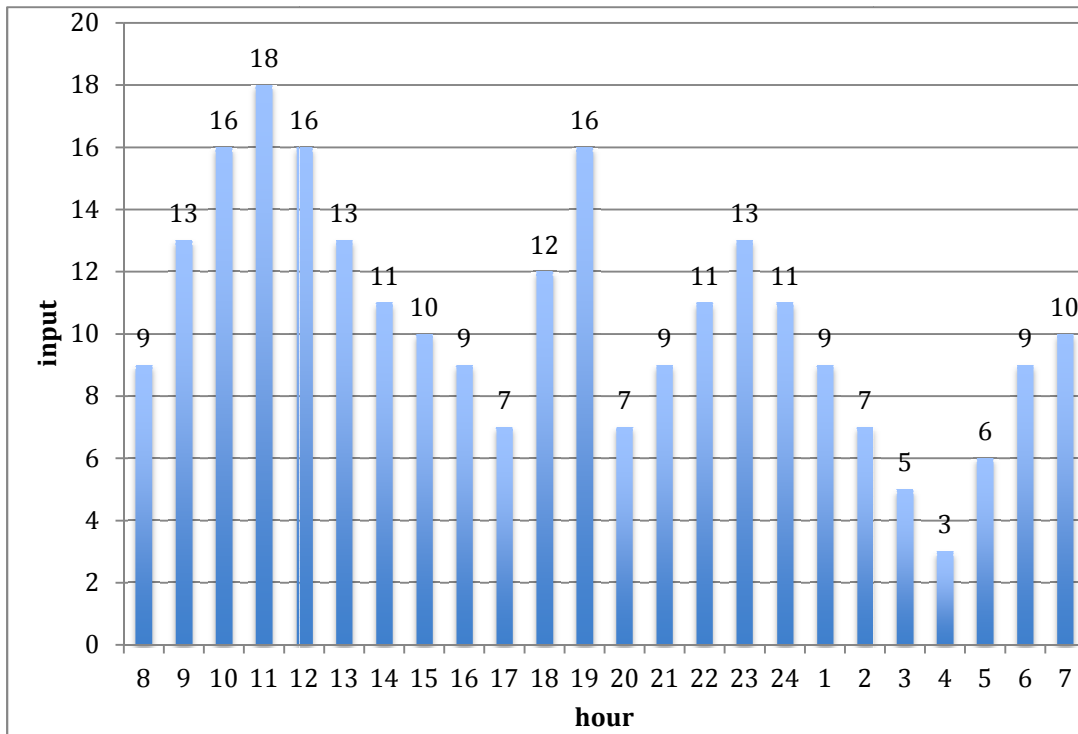
9.4 La modellazione degli arrivi per la determinazione dell'onda di carico

Al fine di condurre gli opportuni esperimenti simulativi sul modello elaborato nel precedente paragrafo, si rende necessario, anzitutto, effettuare alcuni rilievi presso un sistema reale, nell'ottica. Durante il periodo di analisi condotta presso l'Azienda Ospedaliera, è stato possibile analizzare tutte le attività che un generico paziente svolge all'interno del pronto soccorso, individuando punti critici e anomalie che causano situazioni di sovraffollamento all'interno del suddetto reparto.

Inoltre è stato svolto un lavoro di raccolta dei dati per un intervallo temporale di tre anni (2011-2013). Raccogliere i dati relativi ad un periodo ancora precedente sarebbe risultato inutile perché le innovazioni tecnologiche, i cambiamenti sociali e comportamentali hanno modificato le caratteristiche del sistema.

In riferimento a tale intervallo è stata poi effettuata una stratificazione dei dati, sono stati infatti eliminati i mesi singolari, ad esempio il mese di agosto oppure periodi in cui per la presenza di epidemie si sono registrati un numero di accessi al nosocomio superiore alla media.

Per modellare il flusso di pazienti in ingresso al Pronto soccorso, è stata analizzata la distribuzione degli accessi nell'arco della giornata. Mediamente circa 250 persone si rivolgono al dipartimento di emergenza di tale struttura secondo la seguente distribuzione:



Diagr.4 – Andamento arrivi in Pronto Soccorso

come notiamo dalla figura gli inputs al sistema sono altamente variabili durante il giorno, il picco si registra in tarda mattinata tra le ore 10:00 e le 12:00 e intorno alle ore 19:00.

Un processo di arrivi è un flusso di eventi casuali che avviene in un determinato sistema di riferimento. Se il sistema di riferimento è temporale, il flusso è costituito da una successione di eventi casuali associati a particolari istanti di tempo.

È noto che, sotto ipotesi abbastanza semplici, una distribuzione di arrivi casuali in un certo intervallo di tempo segue la distribuzione di Poisson e da qui il nome: Processi di arrivi di Poisson. Nel seguito vediamo brevemente come si definisce un processo di Poisson.

9.4.1 I processi di poisson

Un processo stocastico $\{N(t), t \geq 0\}$ è detto di Poisson se:

- 1.. I client arrivano uno alla volta.
- 2.. $N(t + s) - N(t)$ (il numero di arrivi nell' intervallo di tempo $(t, t + s]$) è dipendente da $\{N(u), 0 \leq u \leq t\}$.
- 3.. La distribuzione $N(t + s) - N(t)$ è indipendente da t , per ogni $t, s \geq 0$.

Le proprietà 1 e 2 sono caratteristiche di molti processi di arrivo. La proprietà 1 non è verificata se I clienti arrivano in lotti; la proprietà 2 afferma che il numero di arrivi nell' intervallo $(t, t + s]$ è indipendente dal numero di arrivi nell' intervallo $[0, t]$ e anche dai tempi ai quali questi arrivi si riferiscono. Questa proprietà potrebbe essere violata se, per esempio, un gran numero di arrivi in $[0, t]$ comporta che gli arrivi di clienti in $(t, t + s]$ rinuncino cioè vanno via senza essere serviti perché trovano il sistema già altamente congestionato.

La terza proprietà, d'altro canto, sarà violata in molti processi di arrivo reali dal momento che implica che il tasso di arrivi dei client non dipende dall' ora del giorno.

Se $\{N(t), t \geq 0\}$ è un processo di Poisson, allora il numero di arrivi in un intervallo di tempo di lunghezza s è una variabile aleatoria di Poisson con parametro λs (dove λ è un numero reale positivo). E dunque:

$$P[N(t + s) - N(t) = k] = \frac{e^{-\lambda s} (\lambda s)^k}{k!} \quad \text{per } k = 0, 1, 2, \dots \text{ e } t, s \geq 0$$

inoltre, $E[N(s)] = \lambda s$ e, in particolare, $E[N(1)] = \lambda$. Perciò, λ è il numero atteso di arrivi in ogni intervallo di lunghezza unitario. Definiamo λ il tasso di tale processo.

Le condizioni cui deve sottostare il processo sono definite dalle seguenti tre ipotesi:

- 1.. La probabilità di avere x arrivi nell'intervallo di tempo di ampiezza t dipende solamente dall'ampiezza dell'intervallo t e non dalla sua posizione sull'asse dei tempi;
- 2.. Gli arrivi si ripartiscono indipendentemente gli uni dagli altri (assenza di memoria);
- 3.. In un intervallo di tempo Δt sufficientemente piccolo è ammesso al massimo un arrivo.

Inoltre un arrivo è un evento aleatorio e dunque il tempo d'attesa di un arrivo aleatorio si distribuisce come una variabile casuale esponenziale negativa, è stato dimostrato che se il numero di arrivi in un intervallo di tempo t è una variabile casuale di Poisson di parametro λ , allora il tempo T tra due arrivi consecutivi è una variabile casuale esponenziale negativa di valor medio $\frac{1}{\lambda}$.

Dunque ho calcolato i λ e cioè il numero di pazienti in ingresso al minuto per le 24 ore.

Nel Prosp.9 sono mostrati i valori dei tassi d'arrivo λ valutati dai dati reali. Tali rates rappresentano il numero di accessi al pronto soccorso per ogni ora del giorno, di seguito sono mostrati come accessi al minuto.

λ_1	0.15	λ_9	0.15	λ_{17}	0.18
λ_2	0.22	λ_{10}	0.12	λ_{18}	0.15
λ_3	0.27	λ_{11}	0.2	λ_{19}	0.12
λ_4	0.3	λ_{12}	0.27	λ_{20}	0.08
λ_5	0.27	λ_{13}	0.12	λ_{21}	0.05
λ_6	0.22	λ_{14}	0.15	λ_{22}	0.1
λ_7	0.18	λ_{15}	0.18	λ_{23}	0.15
λ_8	0.17	λ_{16}	0.22	λ_{24}	0.17

Prosp.9 - Valori dei tassi d'arrivo λ

Attraverso questi rates è stata effettuata una generazione di numeri casuali (utilizzando il foglio di calcolo Excel). Il risultato di tale generazione è stato dato come input al sistema per il modello di simulazione.

9.5 Validazione del modello

La validazione consiste nel valutare se il modello realizzato fornisce risultati paragonabili al comportamento del sistema reale. Nello specifico, si deve verificare se le misure di prestazione del sistema reale sono bene approssimate dalle misure generate dal modello di simulazione. Se è verificata la seguente condizione, allora si potrà dire che le due popolazioni (“reale” e “simulata”) sono sovrapponibili. Nel caso in cui, dal confronto, avvenga un disallineamento, occorre intervenire per correggere il modello e studiare se le cause sono semplici errori imputabili ai parametri del processo oppure se sono errori più consistenti relativi ad un flusso logico/causale non coerente.

Come parametro di prestazione del sistema reale da confrontare con i corrispettivi output del modello, si è scelto il numero medio di barelle presenti nell'intero ospedale durante la giornata. Il motivo di tale scelta è da ricondursi al fatto che la carenza di posti letto a fronte di un così elevato numero di accessi, sia da pronto soccorso che da elezione, fa sì che il numero di pazienti che giacciono in barella è un fattore rilevante per comprendere lo stato di emergenza dell'ospedale in termini di sovraffollamento.

Gli input del modello di simulazione seguono delle distribuzioni statistiche e sono affetti da una certa variabilità che, ovviamente, si trasferisce anche sugli output. Non essendo nota la funzione di distribuzione della v.a. “numero di persone in barella” si ricorre al Teorema del Limite Centrale per applicare le tecniche di stima intervallare, a condizione che si utilizzino grandi campioni ($n \geq 30$) casuali.

Per la validazione del modello si effettua un test d'ipotesi per il confronto tra le medie di due popolazioni.

Se abbiamo motivi sufficienti per ritenere che le due variabili aleatorie siano s-indipendenti e abbiano Cdf gaussiane, di parametri rispettivamente μ_1, σ_1 e μ_2, σ_2 , per effettuare il test di uguaglianza delle medie delle due popolazioni, possiamo utilizzare il fatto che la differenza delle rispettive medie campionarie $Y = \bar{X}_1 - \bar{X}_2$ ha Cdf ancora gaussiana di parametri:

$$\mu_Y = \mu_1 - \mu_2 \quad (9.5.-1)$$

$$\sigma_Y = \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} \quad (9.5 - 2)$$

Molto frequentemente, è lecito supporre che $\sigma_1 = \sigma_2 = \sigma$, per cui la varianza se non è nota può essere stimata sull'insieme dei due campioni come somma di tutte le differenze quadratiche disponibili divisa per i corrispondenti gradi di libertà.

$$t = \frac{\bar{X}_2 - \bar{X}_1}{s \sqrt{1/n + 1/m}} \quad (9.5 - 3)$$

dove S^2 rappresenta la Varianza Pooled così definita:

$$S^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{(n+m-2)} \quad (9.5 - 4)$$

ed equivale alla media delle due varianze campionarie “pesate” con i rispettivi gradi di libertà.

A questo punto, per effettuare il test si procederà così come segue.

Si definiscono l'ipotesi nulla H_0 e l'ipotesi da rigettare H_1 :

$$H_0 = [\bar{x}_1 = \bar{x}_2]; H_1 = [\bar{x}_1 \neq \bar{x}_2] \quad (9.5 - 5)$$

Si calcolano i gradi di libertà ν , dati dalla differenza tra il n° di osservazioni ed il n° di gruppi messi a confronto, si definisce il livello di significatività α del test e si estrapolano dalle relative tabelle i valori critici $t_{0.025}$ e $-t_{0.025}$ che definiscono la regione di rifiuto di H_0 :

$$\nu = n + m - 2;$$

$$\alpha = 0.05;$$

$$t_{0.025} = 2,00039$$

È stato infine calcolato il numero medio di barelle alle ore 15:00 del giorno ed è stata fatta poi la media su 30 giorni. Mettendo i dati d'ingresso anche nel simulatore, otterremo gli output di 30 run di simulazione, dove l'intervallo di simulazione di ogni run è pari proprio a 30 giorni.

I parametri principali dei nostri campioni sono:

$$\bar{x}_1 = 230.8 \quad ; S_1^2 = 111.9$$

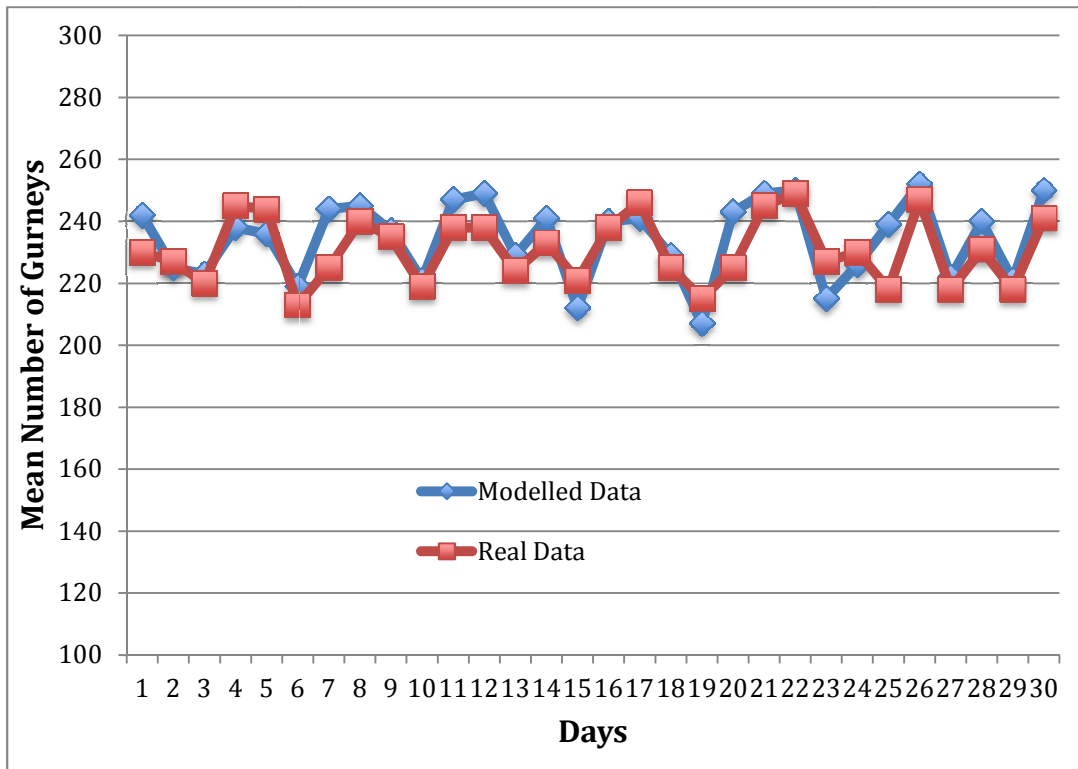
$$\bar{x}_2 = 234.4 \quad ; S_2^2 = 154.6$$

$$S^2 = 133.24 \quad ; \nu = 29 \quad ;$$

$$t = 1.2$$

Dal momento che $t_{0.05} = 1.69$ nel nostro caso $t < t_{0.05}$, ne consegue che il test di Student è superato, posto un intervallo di confidenza pari a $1 - \alpha = 0.95$.

In *Diag.5* è confrontato l'andamento dei dati reali e quelli ottenuti dalla simulazione. Già visivamente si nota che lo scarto tra le due serie di dati è minimo. Con il test d'ipotesi si è data validazione scientifica al risultato ottenuto.



Diagr.5 – Test di Validazione

Nelle stesse ipotesi di prima, ovvero considerando le realizzazioni come v.a. gaussiane indipendenti con varianza pari ad $\sigma_1^2 = \sigma_2^2 = \sigma$, è possibile eseguire un test d'ipotesi sulle varianze delle due popolazioni, infatti, indicando con S_1^2 e S_2^2 due stime campionarie delle varianze σ_1^2 e σ_2^2 otteniamo la v.a. Z di Fisher:

$$Z = \frac{S_1^2 \sigma_2^2}{S_2^2 \sigma_1^2} \quad (9.5 - 6)$$

con $n - 1$ e $m - 1$ gradi di libertà, dove n ed m le dimensioni dei due campioni. Fissate le ipotesi $H_0 = \{\sigma_1^2 = \sigma_2^2\}$ e $H_1 = \{\sigma_2^2 < \sigma_1^2\}$, assumiamo che la funzione (6), in cui poniamo $\sigma_1^2 = \sigma_2^2$ diventa:

$$Z = \frac{s_1^2}{s_2^2} = \frac{111.9}{154.6} = 0.98 < z_{0.95} = 1.86 \quad \text{con } \nu_1 = \nu_2 = 29$$

Secondo quanto dimostrato il modello di simulazione risulta essere validato.

9.5.1 Analisi dei risultati ottenuti in configurazione “as is”

La validazione del modello è stata effettuata limitandosi al confronto di un solo parametro di prestazione, vale a dire il numero di pazienti in barella alle ore 15:00 del giorno. I risultati ottenuti con gli scenari “to-be” invece fanno riferimento a più indicatori quali: numero di persone nella sala d’attesa, numero di persone nella medicheria, numero di persone nella “visit room 3” (cioè i pazienti che sono attesa dei risultati delle analisi) e il numero dei pazienti in barella.

Come abbiamo visto nel capitolo precedente la forza dell’utilizzo delle matrici all’interno del modello consente di visualizzare, istante per istante, tutte le righe della matrice che attraversa l’intera struttura. In ogni variabile di stock del modello, tale matrice, di dimensione fissa, presenterà tutte righe nulle tranne quelle corrispondenti al paziente presente, in quell’istante, all’interno della variabile considerata. Ed è proprio grazie a questo che siamo riusciti a tirare fuori i grafici rappresentativi del numero di persone che giorno per giorno affollavano tale reparto.

Il modello di simulazione consente all'utente di osservare i risultati raggiungibili utilizzando gli stessi dati in ingresso impiegati per l'analisi della situazione "as-is".

Attualmente la gestione del nosocomio è basata su un'ottica prettamente Lean, dunque il dimensionamento delle risorse necessarie è stato valutato su un valore di λ costante per le 24 ore, ciò è esemplificato dal fatto che nella medicheria abbiamo tre turni (ore 08-14, 14-20, 20-08) ognuno dei quali presenta lo stesso numero di risorse ovvero 4 medici, 2 ortopedici e 1 chirurgo.

Con un flusso d'ingresso medio di 250 persone al giorno, nel pronto soccorso molto spesso si vengono a creare situazioni di sovraffollamento in più punti del processo, primo tra tutti l'accesso alla medicheria.

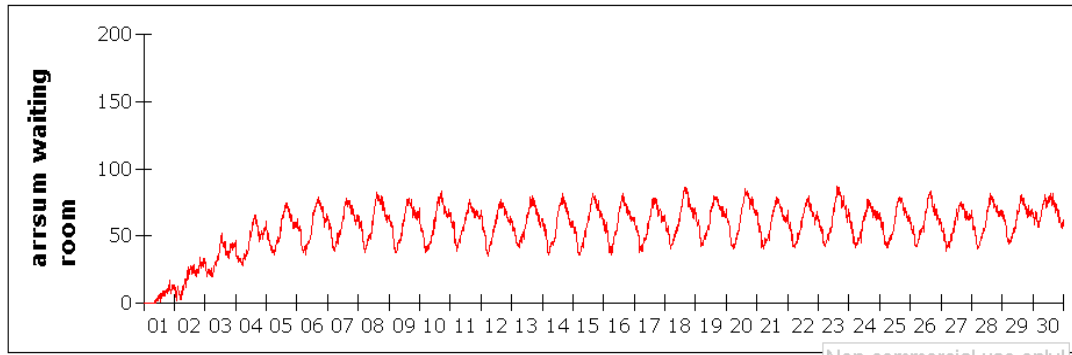
Una prima soluzione "to-be" è stata quella di considerare un λ variabile durante l'arco della giornata e di conseguenza dimensionare il numero di risorse impiegate. In particolare si è pensato di considerare una risorsa medica aggiuntiva all'interno della medicheria solo nelle ore in cui si stimano dei picchi di accessi al sistema (tarda mattinata).

Ragioniamo dunque in ottica *agile* che, come abbiamo visto nel capitolo II si adatta a tutti quei sistemi in cui fare una previsione della domanda risulta complicato data la natura aleatoria di quest'ultima, tale logica inoltre mi consente di spostare risorse all'interno della struttura ospedaliera solo per determinati periodi temporali senza assumere personale aggiuntivo, ma utilizzando in maniera efficiente quello già presente.

Di seguito riportiamo i grafici ricavati dal modello di simulazione in riferimento agli indicatori scelti, tutti i grafici hanno un trend crescente nei primi 2-3 giorni dovuto al fatto che si parte da una situazione iniziale nulla cioè del pronto soccorso vuoto, per cui tale fase può essere considerata come una fase di riempimento del sistema.

Analizziamo dunque la configurazione “as-is”:

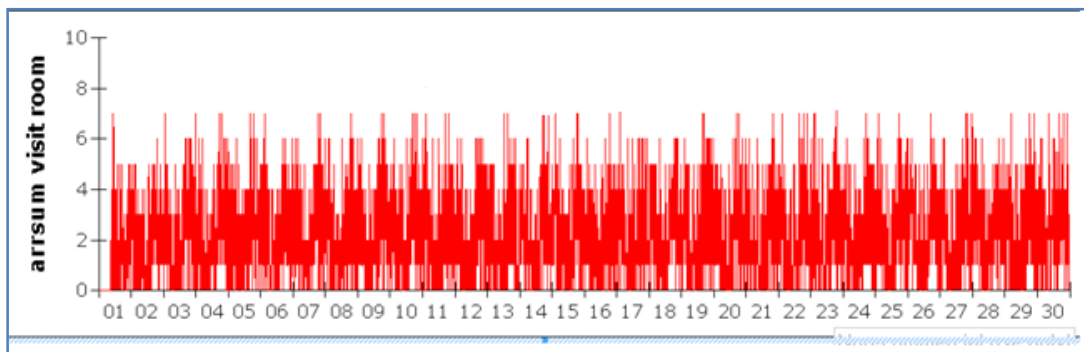
In figura – vediamo il numero di pazienti in sala d’attesa nei 30 giorni di simulazione.



Diagr.6 – “As Is”: Code pazienti in Sala Attesa

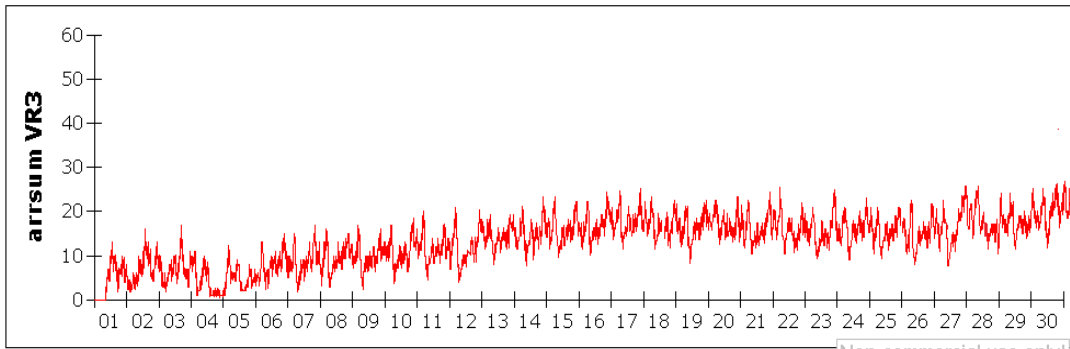
Tale valore varia da 30 a 80 persone, notiamo un picco che si ripete sempre in tarda mattinata in accordo ai dati inseriti nel sistema.

Successivamente è possibile mostrare l’andamento delle persone che si trovano nella medicheria. In questo caso le condizioni scritte nelle variabili di flusso fanno sì che al massimo 7 pazienti possono accedervi in accordo al numero delle risorse mediche disponibili.



Diagr.7 – “As Is”: Andamento pazienti trattati in “Medicheria”

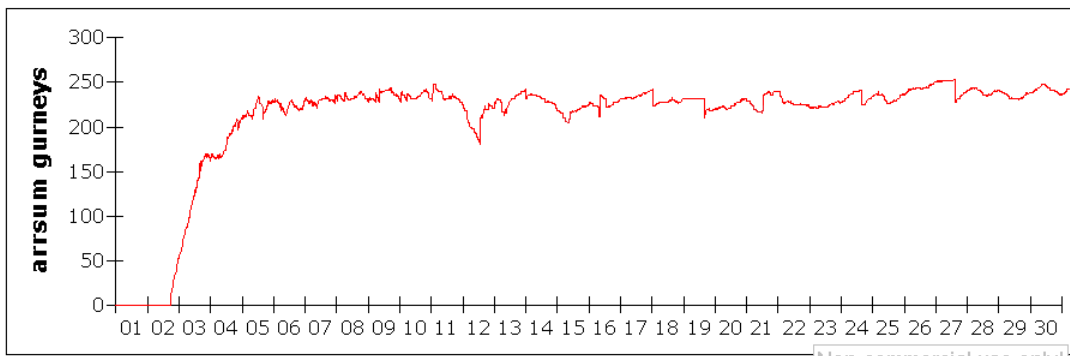
Il grafico successivo invece mostra il numero dei pazienti nella “visit room 3”, cioè tutti i pazienti che sono in attesa della refertazione.



Diagr.8 – “As Is”: Andamento pazienti in “Visit Room”

Come abbiamo precedentemente detto, la refertazione avviene all’ esterno del pronto soccorso, infatti le provette vengono inviate tramite posta pneumatica nel dipartimento di patologia clinica che è collocato in un padiglione diverso. In determinati istanti di tempo si arriva a quasi 23 persone che sono in attesa dei referti, queste persone contribuiscono ad uno stato di sovraffollamento del pronto soccorso.

Infine vediamo, sempre in configurazione “as-is” il numero medio di pazienti in barella.

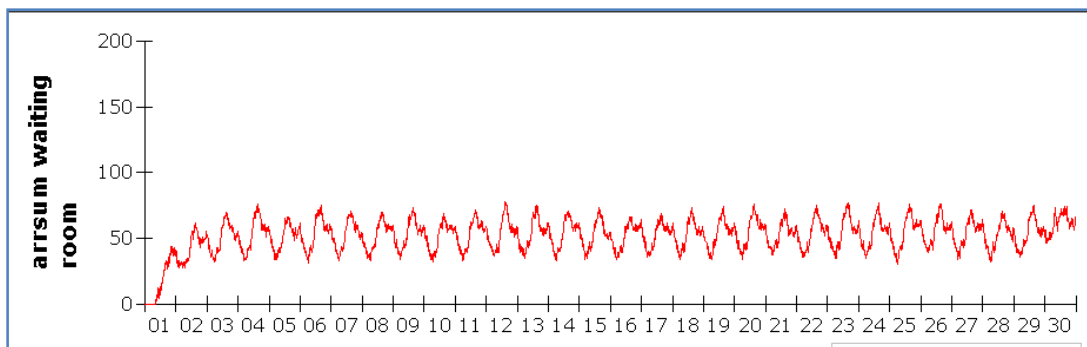


Diagr.9 – “As Is”: Andamento pazienti ricoverati in barella

Trascurando sempre i primi due giorni di riempimento, il numero medio di pazienti che quotidianamente è posizionato in barella è all’ incirca pari a 230.

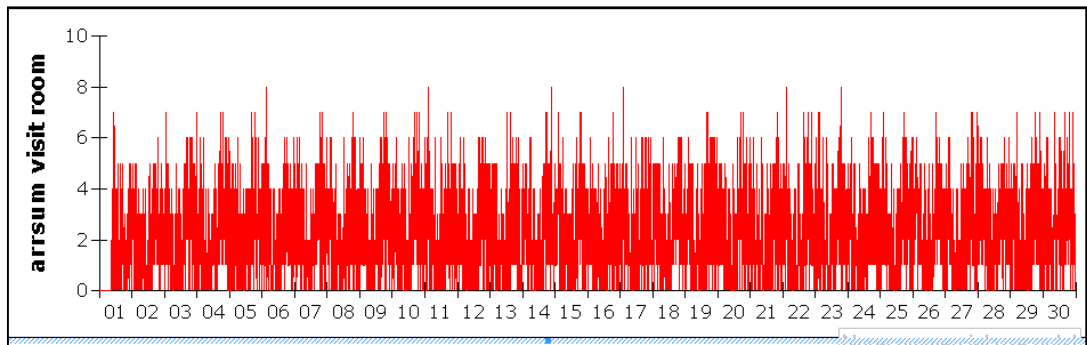
9.6 – Il modello in configurazione Agile

Al fine di valutare l'andamento del sistema in configurazione Agile, rispetto alla configurazione Lean "as is" (in cui esso si presenta attualmente, in vero in senso largo), di seguito riportiamo tutti i grafici visti precedentemente, aggiungendo, inizialmente però, una risorsa nel reparto di medicheria.



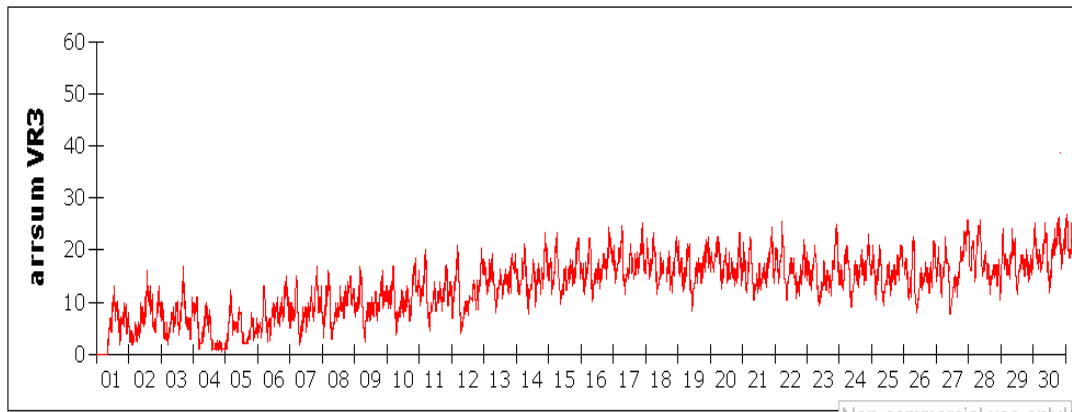
Diagr.10 – Primo "To Be": Code pazienti in Sala Attesa

Le presenze nella sala d'attesa variano da un valore minimo di 27 ad un massimo di 75 , dunque, evidenziano un miglioramento poco significativo.



Diagr.11 – Primo "To Be": Andamento pazienti trattati in "Medicheria"

Nella medicheria notiamo che in certi istanti di tempo il numero delle persone presenti raggiunge il valore di 8, dal momento che le risorse mediche disponibili sono aumentate di una unità.



Diagr.12 – Primo “To Be”: Andamento code pazienti in “Visit Room”

Anche le presenze nella visit room 3 sono piuttosto simili, con un valore medio pari a 18.

Dunque arriviamo alla conclusione del primo “To-Be” proposto secondo cui il reparto di pronto soccorso dimensionato in logica Lean, non dipende significativamente da rettifiche Agile dei sottosistemi operativi (medicheria).

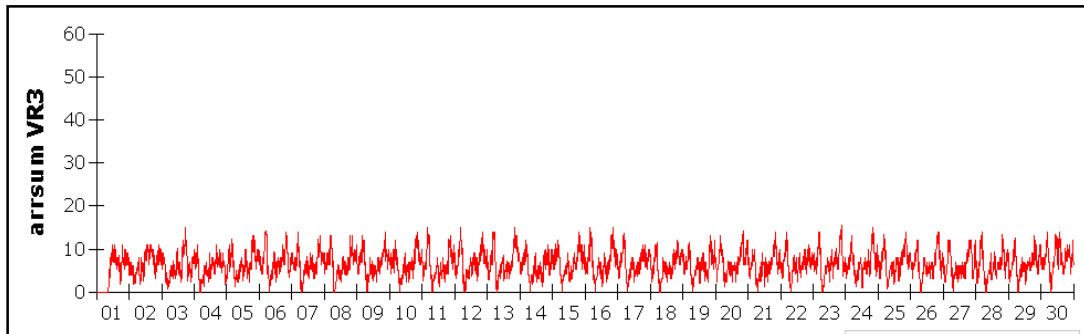
Immaginiamo di voler modificare anche i sottosistemi diagnostici, abbiamo già detto che tutti i pazienti che hanno bisogno di analisi del sangue, esami radiologici o consulenze specialistiche sono idealmente inseriti in una medicheria fittizia (“visit room 3”) che serve solo per contare il numero di persone che in ogni istante è in attesa di un referto.

La logica *Agile*, com’è stata definita in precedenza, ci permette di spostare risorse all’interno del nosocomio in alcuni periodi temporali dove effettivamente, dato l’afflusso di persone si creano situazioni di sovraffollamento.

Se, com’è stato visto dall’analisi dei dati relativi agli ingressi al pronto soccorso, durante l’arco di una giornata si verificano dei picchi in tarda mattinata è ipotizzabile introdurre una risorsa medica che esegua refertazioni con un certo ritardo rispetto ai fenomeni di picco. Tale scostamento temporale è dovuto al tempo che il paziente impiega dall’ingresso al pronto

soccorso fino all'istante in cui ha eseguito gli esami e sta attendendo la refertazione da parte di un medico.

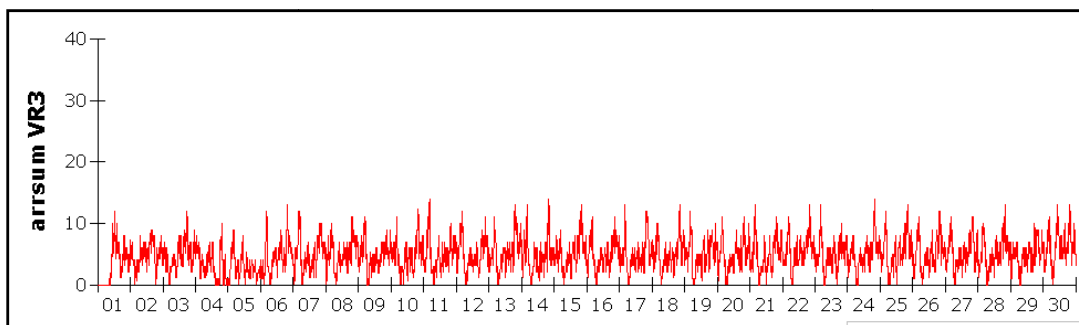
Di seguito mostriamo i risultati riguardanti il numero di persone in attesa post mediceria (visit room 3) in seguito all'introduzione di una risorsa medica in più nell'area dedicata alla refertazione degli esami obiettivi e radiologici.



Diagr.13 – Secondo “To Be”: Andamento code pazienti in “Visit Room” – Logica Agile

Il numero massimo di persone che è in attesa della refertazione è pari a 13, mentre il valore medio è pari 7, dunque il risultato ottenuto con il secondo “to-be” comporta un notevole miglioramento nelle performance del sistema abbassando notevolmente il numero di persone che in ogni istante è in attesa di esami o refertazioni mediche.

Dai risultati raggiunti potremmo pensare di confrontare tale soluzione con quella che si otterrebbe riportando il sistema in logica *Lean*, ossia livellando la produzione e dunque il numero di risorse impiegate per tutte le 24 ore, ovviamente aumentate di una unità. Il risultato che ne viene fuori dal modello di simulazione è rappresentato dal seguente *Diag.14*

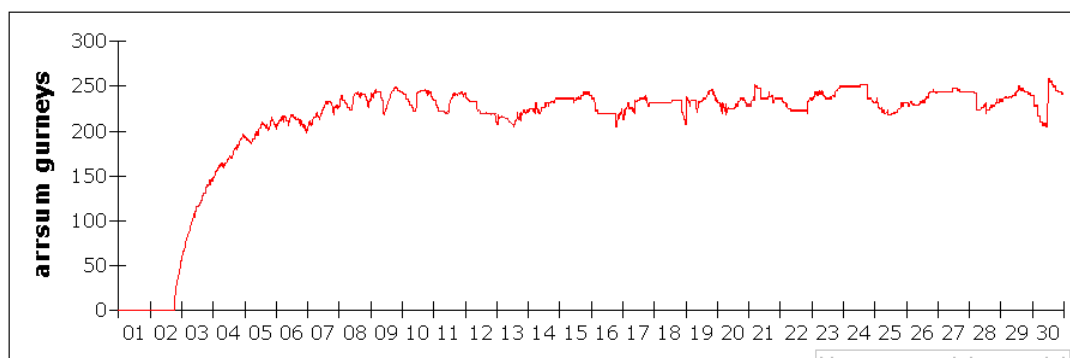


Diagr.14 – Secondo “To Be”: Andamento code pazienti in “Visit Room” – Logica Lean

Se a fronte del maggiore costo sostenuto per avere una risorsa medica aggiuntiva tutto il giorno diminuisse il valor medio delle code, allora avrebbe senso ritornare all’ottica lean, ma poiché il valore medio anche in questo caso è pari a 8 dunque la differenza è poco significativa.

Arriviamo alla conclusione (terzo To-Be) che ragionare in termini agile è la migliore soluzione in quanto a parità di risultati risparmio in termini di costi.

Essendo giunti alla conclusione che la logica *Agile* migliora il livello di servizio del pronto soccorso almeno nella gestione dei sistemi di supporto, vediamo cosa succede con queste ipotesi al numero di pazienti in barella.



Diagr.15 – Secondo “To Be”: Andamento pazienti ricoverati in barella

Il numero medio è invariato, tale conclusione dunque, non andrà ad influire sul problema dei posti letto e quindi sulla presenza di barelle nel pronto soccorso, nei reparti del DEA e lungo i corridoi dell’intero ospedale.

La presenza di queste ultime, infatti, costituisce un altro elemento critico non solo del pronto soccorso, ma dell’intero ospedale, soprattutto nei reparti di

emergenza poiché pur di tenere i pazienti in cura questi vengono tenuti in barella lungo i corridoi dell' ospedale, cosa che invece non accade per tutti coloro che sono ricoverati mediante elezione, ma che essendo "patologie non urgenti" il loro ricovero sarà ritardato. È evidente l'esigenza di migliorare la gestione dell'occupazione dei letti ospedalieri cercando di razionalizzare e ottimizzare l'utilizzo di tale risorsa.

La soluzione di seguito proposta riprende il concetto di gestione dei letti ospedalieri visto nel primo capitolo.

L'elevato numero di barelle che affolla tale reparto deriva da una insufficiente capacità del nosocomio di ospitare un numero così elevato di persone che richiedono tale servizio, ma anche da una cattiva gestione delle degenze.

La difficoltà che si riscontra nel gestire l'"emergenza" e il "programmato" comporta una serie d'inefficienze quali cancellazioni e/o slittamenti dei ricoveri ad elezione e pazienti (da pronto soccorso) che affollano i corridoi dell'ospedale sulle barelle.

Nella descrizione del modello abbiamo visto come la variabile di livello "bed manager" si limitasse a far fluire i pazienti verso le OBI o i reparti e solo nel caso di saturazione di questi ultimi, accumulava i pazienti nella variabile di livello "gurneys" .

Dalla letteratura presente sulla gestione dei letti ospedalieri l'apporto che viene dato in tale lavoro è dotare la figura professionale del Bed Manager di uno strumento attraverso il quale tale figura potesse esortare i primari, laddove fosse possibile a dimettere i pazienti. Questo strumento, già in parte presentato nel primo capitolo, è una valutazione statistica sui tempi medi di degenza per ogni patologia.

Senza dubbio ci possono essere casi in cui, peggioramenti delle condizioni cliniche dei pazienti, guasti a macchinari, ritardi di disponibilità nelle sale

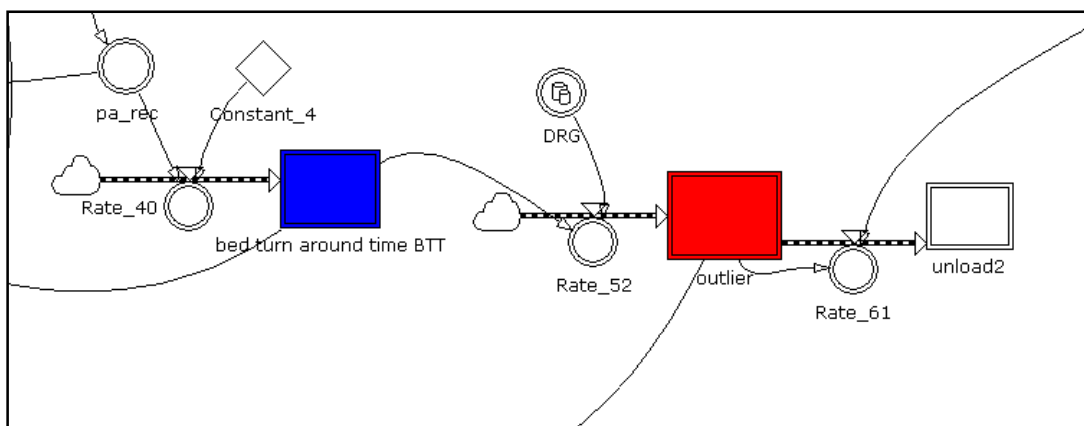
operatorie possono far oltrepassare il valore soglia per i DRG senza che queste degenze risultino anomale. In altri casi invece è opportuno effettuare una pressione sui capi dipartimento dell'ospedale a dimettere pazienti migliorando di conseguenza l'indice di rotazione dei pazienti ricoverati.

Ovviamente i benefici che si trarrebbero sono anche di trasparenza nella gestione dei posti letto ospedalieri, ma anche un risparmio di tempo per i medici del pronto soccorso nella ricerca di un posto letto libero per una particolare patologia, in quanto tale operazione sarebbe interamente gestita dal Bed Manager. Attualmente, infatti, viene utilizzato un sistema informativo che quotidianamente mostra il numero di letti disponibili per ogni reparto ma, nonostante ciò, nel caso in cui tale sistema non fosse aggiornato ora per ora, si procede con telefonate ai reparti e interventi persona. Attività che tolgono tempo prezioso ai medici di pronto soccorso.

La struttura proposta per l'introduzione del Bed Manager all'interno del modello è descritta nel seguito:

è stato introdotto nel modello un vettore contenente i tempi dei DRG per ciascun paziente.

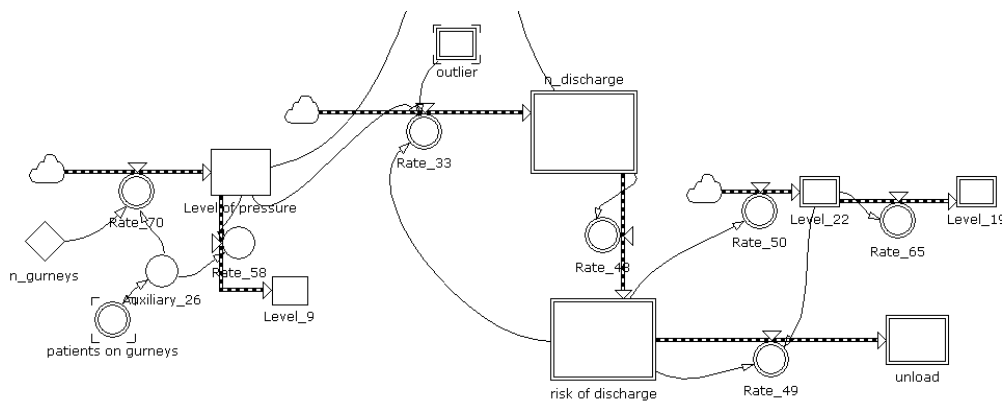
La variabile di livello "bed turn around time" conta il tempo di permanenza del paziente all'interno del reparto.



Mod.24 - Costrutto di calcolo per il "bed turn around time"

Non appena il paziente ha un BTT maggiore del suo tempo definito dal DRG, la variabile di livello “outlier” inizia a contare.

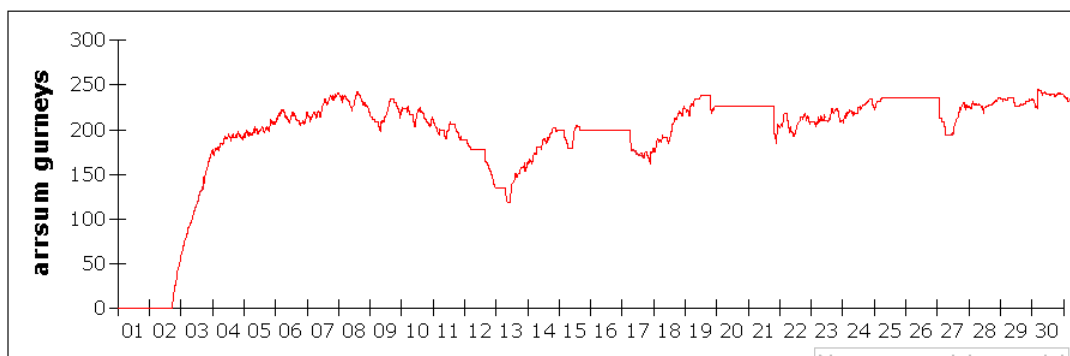
Nella successiva struttura invece sono confrontate il numero di persone che stanno in barella con un valore ritenuto fisiologico di presenze di barelle all’ interno dell’ ospedale, quando il valore reale supera quello fisiologico aumenta la variabile di livello “level of pressure” cioè siamo in una situazione di emergenza con un numero elevato di pazienti in barella.



Mod.25 – Costrutto della meccanica operativa del Bed Manager

La variabile di livello “n_discharge” si incrementa solo se “outlier” è maggiore di zero, cioè ci sono persone da dimettere, e “level of pressure” è anch’essa positiva, cioè stiamo in situazione di emergenza dal punto di vista delle barelle. Non appena questa variabile sarà maggiore di zero, il sistema dimetterà il paziente che sta da più tempo tra quelli che hanno superato il valore soglia. La variabile di livello “Risk of Discharge” si incrementa man mano che vengono effettuate dimissioni da parte del Bed Manager e rappresenta il rischio che si ha nel dimettere anticipatamente un paziente. Trascorse 24 ore questo livello viene azzerato attraverso lo scarico "unload".

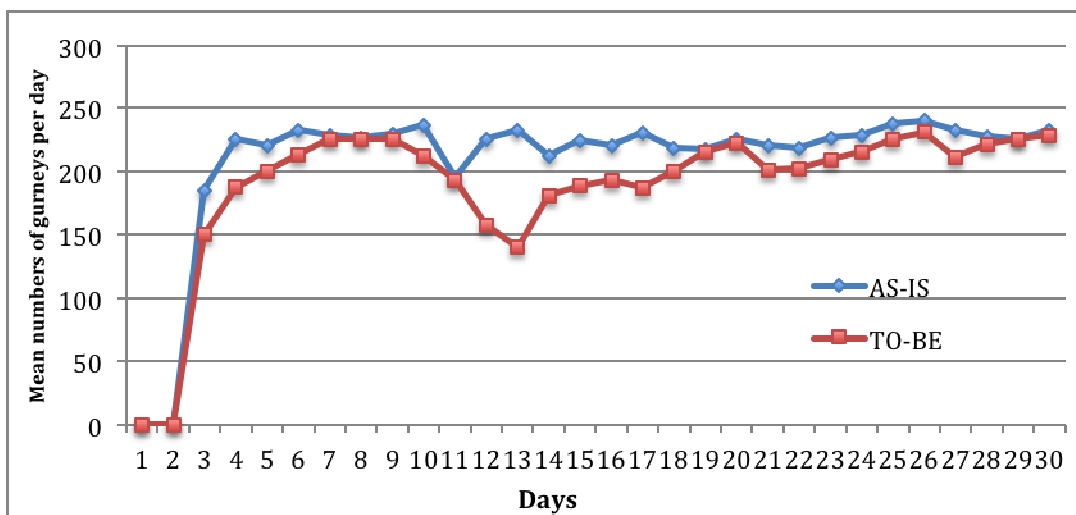
Analizziamo dunque i risultati dopo l’applicazione del “Bed Manager”:



Diagr.16 – Terzo scenario “To Be”: Andamento pazienti ricoverati in barella

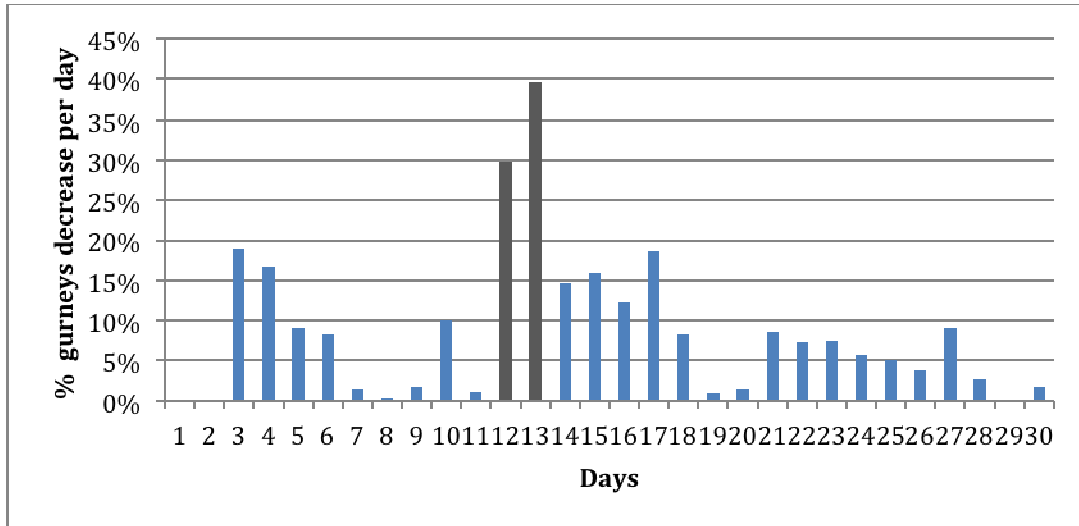
Dal grafico *Diagr.16* possiamo osservare qualitativamente che il numero delle barelle, da un valore medio pari a 230 è diminuito in alcuni giorni della simulazione.

Valutando la differenza tra il numero medio di barelle per ogni giorno tra le due configurazioni “as-is” e “to-be” notiamo che mediamente otteniamo una decremento del numero di pazienti in barella di circa l’8%. Tale valore è stato calcolato senza tenere in considerazione alcuni giorni della simulazione, in particolare il 1° e il 2° giorno poiché rientrano nella fase transitoria di riempimento dell’ospedale e il 12° e il 13° giorno caratterizzati invece da un’eccessiva dimissione da parte dell’ algoritmo di simulazione.



Diagr.17 – Comparazione “As Is” - Terzo scenario “To Be” degli andamenti relativi ai pazienti ricoverati in barella

Il risultato del terzo scenario To-Be espresso intermini di riduzione percentuale delle barelle occupate che siamo riusciti ad ottenere per i 30 giorni di simulazione è rappresentata nel seguente grafico:



Diagr.18 – Terzo scenario “To Be”: riduzione percentuale barelle

9.7 - Osservazioni sulle risultanze

Obiettivo principale di questo capitolo del lavoro è stata la realizzazione di un modello dinamico di accesso casuale in ospedale; abbiamo infatti creato un modello di simulazione di tipo matriciale, che rappresenta la dinamica del sistema complesso, pronto soccorso.

Anche se l'obiettivo finale è quello di razionalizzazione e gestione di posti letto negli ospedali, abbiamo scelto di modellare prima questa U.O. . Infatti, il buon funzionamento di quest'ultimo è un elemento critico per ogni ospedale e, al contempo, risulta una condizione sufficiente per l'efficienza dell'intero sistema ospedaliero.

Ecco perché si è pensato che un primo miglioramento doveva essere eseguita in questo particolare reparto. Quando una persona entra al pronto soccorso, il più delle volte, è tenuto in attesa per diverse ore per arrivare nella stanza visita a causa del sovraffollamento soprattutto in alcune ore della giornata; queste attese possano comportare un peggioramento delle condizioni cliniche del paziente, con il conseguente abbassamento dei livelli di performance dell'ospedale.

Le configurazioni reperite attualmente in letteratura ed effettivamente applicate nella realtà (almeno in via di principio), si rifanno ad una logica Lean Service che, a sua volta, fondandosi sulla livellabilità degli accessi, impone una configurazione One Piece Flow al sistema produttivo, non realizzabile da quest'ultimo, non per incapacità o insufficienza delle risorse ma, piuttosto, per la mancata sovrapposibilità della curva reale dei carichi (id est gli effettivi accessi), in combinato disposto con le altre quattro caratteristiche dei sistemi nosocomiali (vedi par. 9.1).

Una prima soluzione in logica Agile, quindi, potrebbe consistere nell'attivazione di una risorsa medica aggiuntiva nel reparto produttivo diretto del processo di cura (la medicheria), temporalmente in opportuno ritardo di fase rispetto ai quotidiani periodi di alta afflusso.

Grazie alla simulazione, realizzata in System Dynamics, è stato possibile rilevare l'inefficacia della soluzione: i risultati raggiunti, infatti, misurati in termini di coda dei pazienti in attesa ante-medicheria, evidenziano uno scarso miglioramento della situazione.

Ciò implica una prima conclusione controdeduttiva: i sistemi di P.S. dei DEA (dimensionati in logica lean) non presentano nei sottosistemi primari di cura l'elemento di maggiore criticità.

Un secondo scenario Agile ha configurato l'attivazione di una risorsa medica aggiuntiva nel reparto di supporto di refertazione diagnostica al processo di

cura (radiologia e esami di laboratorio), anche in questo caso temporalmente in opportuno ritardo di fase rispetto ai quotidiani periodi di alta afflusso.

In questo secondo scenario i risultati, sempre misurati in termini di coda dei pazienti in attesa ante-medicheria, evidenziano un miglioramento ben più confortante. Ciò prefigura una seconda conclusione controdeduttiva: i sistemi di P.S. dei DEA (dimensionati in logica lean) presentano nei sottosistemi di supporto diagnostico alla cura l'elemento di maggiore criticità.

In ogni caso la configurazione Agile migliora la situazione, potendo meglio rispettare i vincoli posti dal profilo di carico del sistema.

Ciononostante, pur migliorando l'unità di Pronto Soccorso, non migliora la performance dell'intero sistema degli accessi all'ospedale, misurata in termini di coda di attesa (numero barelle occupate) per il ricovero a reparto, una volta finito il percorso in Pronto Soccorso.

A tal fine ci siamo concentrati su una strategia che possa permettere di conciliare un elevato numero di accessi da pronto soccorso, con un altrettanto elevato numero di pazienti prenotati per ospedalizzazione (ricovero per elezione) all'interno dello stesso ospedale.

Poiché il numero di letti è una costante per l'ospedale la soluzione proposta e simulata è consistita nell'espansione e integrazione delle *best practices* reperite in letteratura sull'argomento *bed management*.

Sempre considerando il numero di pazienti in barella, come parametro di valutazione, abbiamo verificato mediante appositi lanci simulativi, poi opportunamente mediati, che la performance del sistema fa registrare una riduzione significativa delle code, quando si impone un maggiore rispetto (sebbene ben lungi dall'essere forzato) dei tempi medi DRG di ospedalizzazione valutati per le diverse patologie. In tale scenario abbiamo registrato circa il 20% di riduzione totale delle barelle. In questo modo, la velocità di rotazione di pazienti ospedalizzati aumentata significativamente.

Quest'ultimo risultato ci fa comprendere che il sistema ospedaliero assume evidenti caratteristiche di elevata complessità. Infatti pur migliorando la performance di un suo sottosistema vitale, come l'unità di Emergenza dei DEA, non migliora la performance complessiva del nosocomio. Pertanto (ed è questa la seconda conclusione di rilievo) sebbene fortemente influenzato dalla curva degli accessi, un sistema ospedaliero fa registrare miglioramenti complessivi delle sue performance, solo se risolve problematiche di *lead time* a valle delle unità di accettazione in emergenza.

Sarà questo il tema del prossimo capitolo

CAPITOLO 10

Dualità Lean Agile nei U.O. di urgenza dei D.E.A.

10.1 – Sviluppo del modello di simulazione per l’Urgenza

Il modello di simulazione costruito nel seguente elaborato di tesi raffigura un DEA -Dipartimento di Emergenza e Accettazione- e le strutture a supporto di esso. Nel dettaglio sono stati rappresentati i seguenti elementi:

- Reparti di emergenza / urgenza;
- Blocco operatorio d'emergenza e reparto di rianimazione;
- Laboratorio di analisi chimico-cliniche;
- Radiologia;
- Pronto Soccorso;
- Gestione dei farmaci.

L'accesso a tali strutture non è programmabile poiché i pazienti vi accedono senza prenotazione; non vi è possibilità di prevedere né il volume né la tipologia di patologia.

Nel modello di simulazione si è rappresentato il percorso che segue il paziente dall'ingresso fino alla dimissione dalle strutture costituenti il DEA.

Per la definizione del modello è stato utile costruire il flow chart delle seguenti aree: farmacia, laboratorio di analisi, radiologia, pronto soccorso e reparti del DEA. Il flow chart, anche detto diagramma di flusso, è un linguaggio di modellazione grafico per rappresentare il flusso di controllo o processo logico di alcune procedure.

Esso consente di descrivere in modo schematico:

- le operazioni da compiere, rappresentate mediante sagome convenzionali (come rettangoli, rombi, esagoni, parallelogrammi, rettangoli smussati) all'interno delle quali un'indicazione testuale descrive l'attività da svolgere
- la sequenza nella quale devono essere compiute, rappresentate con frecce di collegamento.

L'idea alla base del seguente lavoro di tesi prende spunto dall'esperienza e dalle evidenze raccolte presso l'A.O.R.N. "A. Cardarelli" di Napoli. Durante l'attività di analisi ivi condotta, è stato possibile notare che le strutture di emergenze/urgenza, sono affette da numerose criticità tra cui, la prima anche se non la più importante, la mancanza di spazi poiché progettate in tempi in cui non si prevedeva un tale aumento di accessi. Ciò causa un sovraffollamento delle strutture stesse.

Il sovraffollamento dei servizi di emergenza interferisce pesantemente sui processi assistenziali dei pazienti in termini di tempestività delle terapie, di incremento della mortalità e degli errori di trattamento, di aumento dei tempi di degenza ospedalieri, generando costi aggiuntivi elevatissimi.

Lo scopo del seguente elaborato consiste nell'individuazione della metodologia di gestione più opportuna da applicare agli elementi rappresentati nel modello, nonché il dimensionamento ottimale delle risorse che consente di migliorare l'erogazione dei servizi sanitari.

10.1.1 – Costruzione Causal Loop Diagram

Prima di procedere con la costruzione del modello, si è reso necessario costruire il Causal Loop Diagram che si presenta come uno strumento utile ad una prima interpretazione grafica del problema. È costituito da un insieme di variabili, tra loro interconnesse, mediante frecce, che segnalano l'influenza casuale tra le variabili stesse.

Il CLD fornisce una rappresentazione semplice e schematica delle ipotesi elaborate in relazione alle cause responsabili degli andamenti problematici illustrando chiaramente i meccanismi di feedback.

In figura 27 è possibile vedere che il CLD è costituito da 7 loop collegati uno all'altro mediante delle variabili in comune.

Nel dettaglio i loop sono rappresentativi delle seguenti problematiche:

- ❖ Gestione del Pronto Soccorso;
- ❖ Gestione dei Ricoveri;
- ❖ Gestione del Blocco Operatorio;
- ❖ Gestione dei Supporti;
- ❖ Gestione Replenishment;
- ❖ Sostenibilità Gestionale;
- ❖ Estendibilità universale delle cure.

Per la Gestione del Pronto Soccorso si è ritenuto opportuno mettere in evidenza la problematica legata ai lunghi di tempi di attesa, che spesso provocano l'abbandono dei pazienti dalla struttura stessa o comunque l'insoddisfazione dei pazienti, facendo così diminuire l'efficienza e l'efficacia delle performance del Pronto Soccorso.

Nel dettaglio all'aumentare degli arrivi presso il Pronto Soccorso aumentano le richieste di supporto e replenishment poiché generalmente al paziente vengono somministrati dei farmaci ed egli viene sottoposto ad esami di laboratorio e ove necessario ad esami radiologici. Ma l'aumentare delle richieste provoca la congestione dei processi con un conseguente aumento dei tempi di attesa. Ciò genera insofferenza nei pazienti che, nel caso non siano colpiti da patologie gravi, decidono di abbandonare la struttura. L'allontanamento del paziente dalla struttura, prima che egli sia servito, provoca non solo danni di efficacia alla struttura ma anche danni in termini di efficienza, nel caso in cui il paziente abbia iniziato il percorso di visita, venendo sottoposto ad esami, ma va via prima della chiusura della cartella clinica. In tal caso la struttura si espone a spese che si traducono in sprechi. I lunghi tempi di attesa sono legati a vari fenomeni, uno dei quali dipende dai

tempi necessari all'esecuzione e alla refertazione degli esami di laboratorio e di quelli radiologici.

Per la Gestione dei Ricoveri la problematica riguarda analogamente i tempi di esecuzione e di refertazione degli esami e la conseguenza a cui questi portano, intesa come allungamento dei tempi di degenza con diretta conseguenza della diminuzione delle dimissioni. Avere dei letti occupati più del necessario provoca uno spreco in termini finanziari, e quindi un'inefficienza, nonché un calo dell'efficacia poiché diventa difficile dedicare cure ad altri pazienti con un maggiore bisogno. Spesso, per sopperire all'insufficienza dei letti disponibili, si ricorre all'utilizzo di barelle, nonostante ciò non sia consentito dalla legge. La barella non è stata concepita come posto letto per degenti. Ciò comporta anche un aumento del rischio sanitario a cui i pazienti sono esposti. L'aumento dei ricoveri genera, quindi, un aumento delle richieste dei servizi di supporto, provocando un aumento del tempo di processamento del paziente e quindi un aumento della permanenza del paziente presso la struttura ospedaliera con conseguente diminuzione del numero di dimissioni. All'aumentare del numero di dimissioni si ha un aumento del numero di pazienti stabilizzati e ciò comporta un miglioramento delle performance della gestione dei ricoveri.

La Gestione del Blocco Operatorio evidenzia una reale difficoltà nello smaltimento degli interventi e delle degenze in rianimazione nel caso in cui gli accessi saturano la capacità della struttura ospedaliera in termini di risorse umane e materiali. L'aumento del numero di pazienti in arrivo presso la struttura ospedaliera comporta un aumento del numero di interventi da eseguire. Ciò comporta un aumento del numero di degenti presso la rianimazione con conseguente diminuzione dei letti liberi. Al diminuire del coefficiente di occupazione dei letti corrisponde un aumento dei pazienti stabilizzati che genera a sua volta un miglioramento delle performance del blocco operatorio.

Nel loop dedicato alla Gestione dei Supporti si è cercato di scomporre il tempo di esecuzione e refertazione dell'esame, clinico-chimico e radiologico, per capire a quale parte del processo è imputabile una maggiore criticità. Con l'aumento del numero di richieste di servizi di supporto si genera una congestione sia nel processo di esecuzione di analisi sia nella refertazione con un conseguente aumento dei relativi tempi di esecuzione. Tali tempi si ripercuotono sui tempi di processamento del paziente andando a condizionare sia le performance di tali attività sia le performance dell'intera struttura.

Nella Gestione Replenishment si è messa in evidenza una problematica comune alle varie unità operative costituenti la struttura ospedaliera. Tale criticità riguarda la gestione dei farmaci poiché non sempre risulta facile la scelta dell'opportuna politica di gestione degli ordini. Gli scenari percorribili sono due:

- Si ordina più del necessario e quindi si va incontro ad uno spreco di merce per inutilizzo e scadenza;
- Si ordina meno del necessario rischiando di rimanere sprovvisti di farmaci e di non poter dedicare le giuste cure al paziente.

All'aumentare del numero di pazienti da servire, aumenta l'utilizzo di farmaci comportando quindi una diminuzione di disponibilità dei farmaci stessi presso il Pronto Soccorso e/o le unità operative. Avere una maggiore disponibilità di farmaci consente si migliori performance in termini di minor tempo necessario alla somministrazione del farmaco, ma d'altra parte se i farmaci sono in eccesso rispetto a quelli che poi vengono effettivamente utilizzati, si va incontro ad uno spreco per scadenza generando un peggioramento delle performance.

Infine si evidenzia la netta opposizione tra due loop che sono la Sostenibilità Gestionale e l'Estendibilità Universale di Cure. Quest'ultimo tende a salvaguardare il diritto alla salute del cittadino spingendo verso un aumento

dell'efficacia del servizio sanitario con una conseguente diminuzione del rischio sanitario. In particolare con l'aumento del numero di pazienti curati nel tempo si ottiene una diminuzione della permanenza del paziente in ospedale. Ciò consente la diminuzione del rischio clinico inteso come rischio di infezioni o rischi legati alla cattiva sanità ad esempio per mancata lucidità del medico a causa di turni di lavoro troppo lunghi o per sovraffollamento delle strutture. La sostenibilità gestionale, invece, tende a salvaguardare l'efficienza gestionale mirando alla minimizzazione degli sprechi, attraverso opportune azioni correttive, e della probabilità di un cattivo uso delle risorse disponibili. Nel dettaglio l'aumento degli sprechi spinge verso un aumento delle azioni correttive che consentono un aumento degli investimenti verso nuove attività. D'altra parte però la diminuzione degli sprechi consente un miglioramento in termini di efficienza gestionale la quale, a sua volta, comporta un aumento dell'indice di sostenibilità, il quale rappresenta il rapporto tra i ricavi unitari e i costi unitari.

10.1.2 – L'architettura del modello

Il modello di simulazione è stato realizzato con l'ausilio del software Powersim Studio, che consente una rappresentazione grafica del modello elaborato.

Ogni simbolo grafico rappresenta una tipologia di variabile, di conseguenza, la connessione tra i vari simboli, rappresenta l'interazione logica tra le diverse variabili del modello.

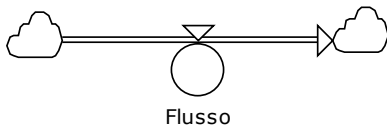
Powersim, come già detto, consente di sviluppare un modello di simulazione attraverso un linguaggio grafico, senza utilizzare un vero e proprio linguaggio di programmazione.

La simbologia utilizzata in Powersim è la seguente:

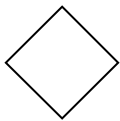


Livello

La variabile di livello anche detta stock è rappresentata da una casella. Indica un'accumulazione che può essere modificata nel tempo solo attraverso i flussi.

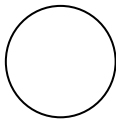


La variabile di flusso influenza le variabili di stock, collegandole tra loro.



Costante

La costante è definita da un valore iniziale e mantiene questo valore durante tutto il periodo di tempo della simulazione, a meno che l'utente non cambi manualmente il suo valore.



Variabile ausiliaria

E' una variabile al cui interno si svolgono dei calcoli, sulla base di altre variabili. Va collegata al flusso ma mai direttamente alle variabili di stock, che possono essere modificate solo attraverso le variabili di flusso. Tale variabile, a differenza della variabile di stock, non ha alcuna memoria.



La freccia, anche detta link, indica un collegamento di informazioni che può avvenire tra i diversi tipi di variabili. E' lo strumento con cui si rappresenta l'influenza che una variabile ha su un'altra variabile.

Come già detto in precedenza, il modello in esame rappresenta il percorso seguito dal paziente dall'entrata fino alla dimissione dalle strutture del DEA.

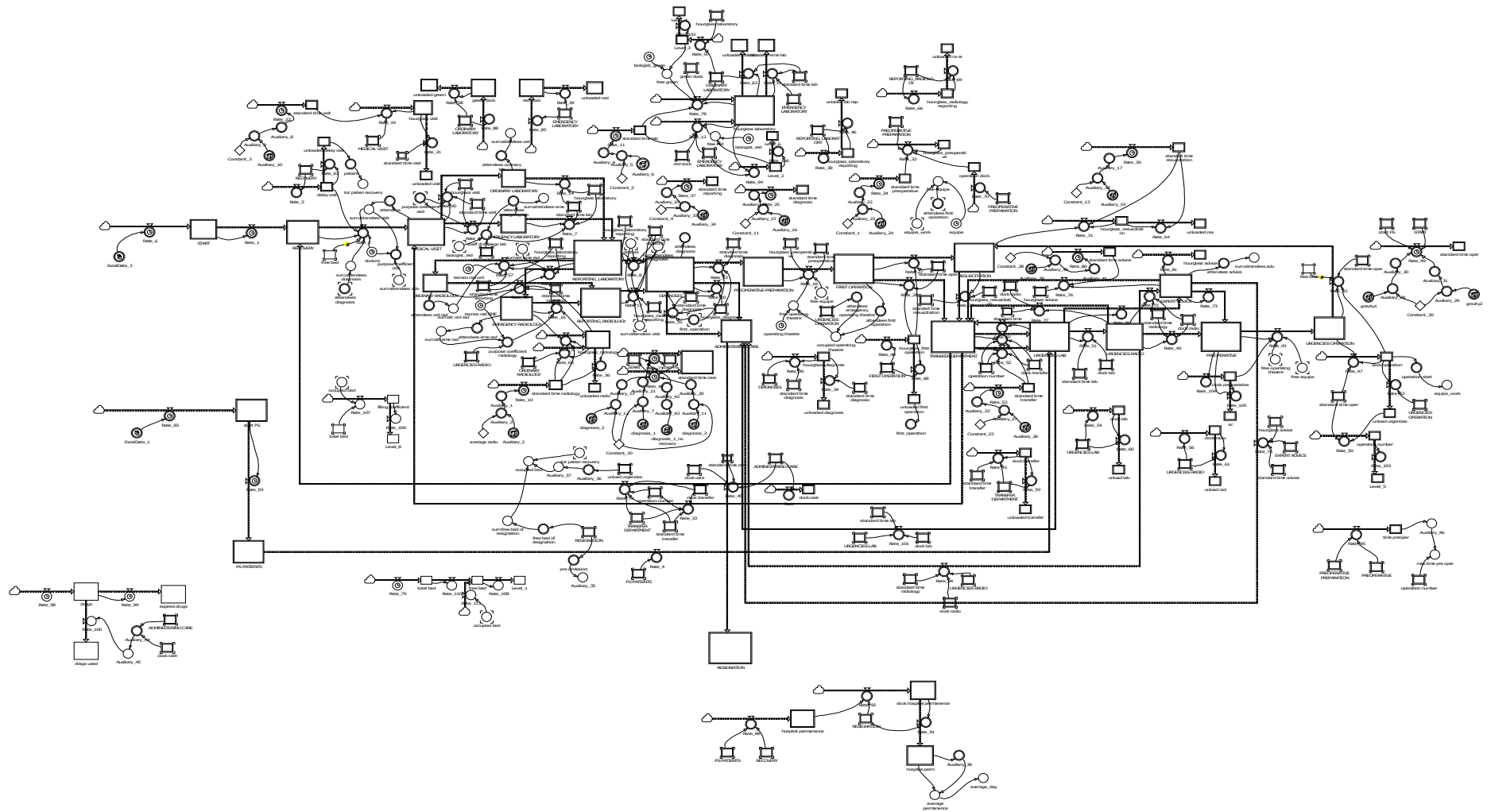


Figura 27: Modello di simulazione

Sono stati immaginati due possibili ingressi presso il Dipartimento d’Emergenza e Accettazione: pazienti provenienti dal Pronto Soccorso e pazienti provenienti da altre unità operative o da altre strutture ospedaliere.

Come è possibile notare in figura 27, le variabili di livello rappresentano le attività a cui è sottoposto il paziente nella sua permanenza presso i locali ospedalieri. Tali attività sono:

- ✓ il ricovero presso una delle unità operative del DEA;
- ✓ la visita medica;
- ✓ il laboratorio, ordinario o d’emergenza;
- ✓ la refertazione di laboratorio;
- ✓ la radiologia, ordinaria o d’emergenza;
- ✓ la refertazione di radiologia;
- ✓ la diagnosi;
- ✓ la preparazione preoperatoria, ove necessario l’intervento;
- ✓ l’intervento operatorio;
- ✓ la degenza in rianimazione;
- ✓ il trasferimento in reparto;
- ✓ la consulenza;
- ✓ la somministrazione di cure;
- ✓ la dimissione.

Inoltre sono state utilizzate delle altre variabili di livello per rappresentare il passaggio dei pazienti provenienti dal Pronto Soccorso presso il laboratorio di analisi chimico-cliniche, la radiologia, la preparazione preoperatoria e l’intervento operatorio. Tali variabili in realtà non rappresentano il passaggio del paziente in strutture fisicamente separate da quelle elencate in precedenza, ma sono servite solo per una più facile rappresentazione del percorso del paziente.

Tali variabili sono:

- ✓ il laboratorio d'urgenza;
- ✓ la radiologia d'urgenza;
- ✓ la preparazione preoperatoria d'urgenza;
- ✓ l'intervento d'urgenza.

È bene notare che usufruiscono delle attività d'urgenza anche i degenti sottoposti a più di un intervento.

Inoltre anche la differenziazione tra il laboratorio di emergenza e quello ordinario e la radiologia di emergenza e quella ordinaria, in realtà non rappresenta una separazione fisica dell'attività in esame, ma solo una separazione logica. Tale separazione si è resa necessaria poiché, presso il laboratorio di analisi e la radiologia, si hanno delle risorse dedicate solo ai pazienti soggetti ad emergenza, quindi, provenienti dal Pronto Soccorso e delle risorse dedicate esclusivamente ai pazienti degenti presso le unità operative costituenti il DEA. Mentre il laboratorio e la radiologia di emergenza lavorano h24, i servizi ordinari lavorano solo in alcune fasce orarie. Per tale motivo, nel caso in cui nei reparti di degenza si presenta la necessità di usufruire dei servizi su detti, in orari in cui i servizi ordinari non sono attivi, è possibile appoggiarsi a quelli di emergenza.

Le variabili di flusso presenti nel modello si sono rese utili alla rappresentazione del passaggio del paziente da un'attività all'altra sotto opportune condizioni.

Nella creazione di tali condizioni sono stati sfruttati due elementi, che sono il *ciclo for* e le relazioni *if...;then*.

Sono, poi, state aggiunte le cosiddette SnapShot. Queste sono una sorta di "alias" di una variabile di origine. Sono utili per il collegamento di

variabili situate in diverse parti del modello e per rendere più chiara possibile la lettura del modello.

Per evidenziare il fatto che i possibili accessi alle strutture sono due, si è ritenuto opportuno utilizzare due variabili di livello, denominate *START* e *START PS*. Nel modello, gli accessi e le indicazioni relative all'individuazione del percorso di cura per ogni paziente, sono rappresentati da 2 fogli Excel, uno per ogni tipologia di accesso, contenenti matrici di dati. Le variabili di livello *START* e *START PS* vengono inizializzate a zero e si riempiranno solo quando sono verificate le condizioni presenti rispettivamente nei *Rate_6* e *Rate_83*. Tali condizioni impongono il caricamento dei dati nel primo giorno della simulazione. Una volta caricati i dati, attraverso le condizioni presenti nei *Rate_1* e *Rate_93* si ottiene il ricovero o l'arrivo del paziente al Pronto Soccorso. I Rate appena citati consentono il passaggio al livello successivo solo quando il giorno e l'orario della simulazione coincidono con quello di arrivo del paziente presso il DEA.

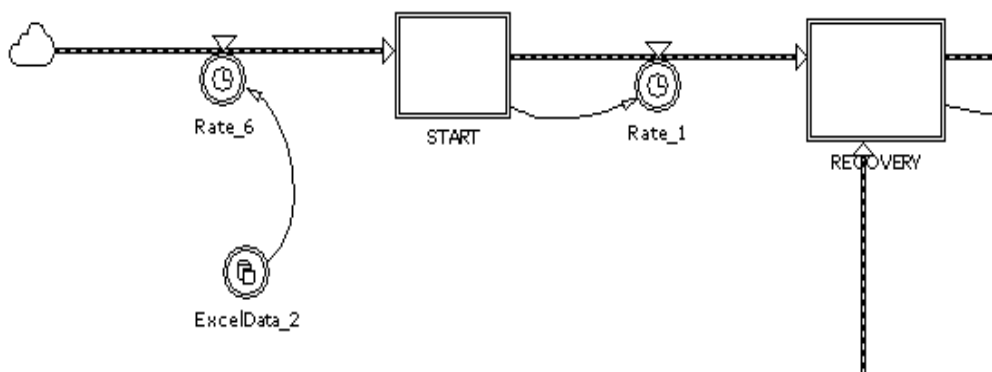


Figura 28: Costrutto che consente il caricamento dei dati relativi ai pazienti provenienti da altre unità operative o altre strutture ospedaliere

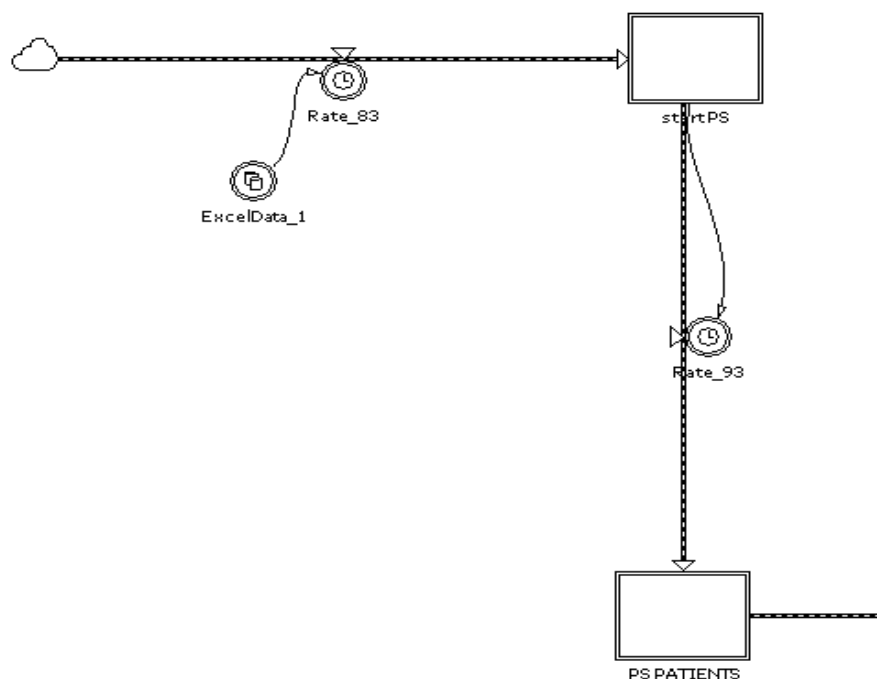


Figura 29: Costrutto che consente il caricamento dei dati relativi agli accessi presso il Pronto Soccorso

I pazienti degenti presso le unità operative costituenti il DEA possono essere sottoposti alle seguenti attività: visita medica; laboratorio di analisi; radiologia; refertazione; diagnosi; preparazione preoperatoria; intervento; degenza in rianimazione; trasferimento in reparto; somministrazione di cura e dimissione.

I pazienti provenienti dal Pronto Soccorso possono usufruire delle seguenti attività: laboratorio di analisi, radiologia, consulenza, somministrazione di cura e dimissione. Nel caso in cui per il paziente in esame è previsto il ricovero presso la struttura ospedaliera, egli usufruirà anche delle attività previste per i pazienti in degenza.

Chiaramente per ogni paziente è previsto un percorso di cura personalizzato e il passaggio o meno presso alcune attività, a seconda della patologia di cui è affetto.

È prevista la possibilità di molteplici passaggi presso le attività sopra elencate nel caso tali passaggi risultino necessari per la stabilizzazione del paziente.

La permanenza del paziente nelle varie attività è determinata attraverso un tempo random generato sulla base della patologia del paziente in esame e del tempo medio necessario allo svolgimento di tale attività. Il passaggio del paziente all'attività successiva sarà autorizzato solo quando saranno verificate tutte le condizioni presenti nel flusso che collega le due attività considerate.

Le condizioni costruite tengono conto sia della disponibilità delle risorse umane e materiali, necessarie allo svolgimento dell'attività, sia del tempo effettivamente trascorso. In particolare, è possibile effettuare l'attività seguente, se c'è disponibilità delle risorse necessarie e se il tempo trascorso è proprio pari al tempo random assegnato ad un determinato paziente per quella specifica attività.

Per verificare l'uguaglianza in termini temporali è possibile costruire una clessidra che inizia a contare gli istanti di tempo trascorsi da quando il paziente è effettivamente arrivato presso l'attività in esame.

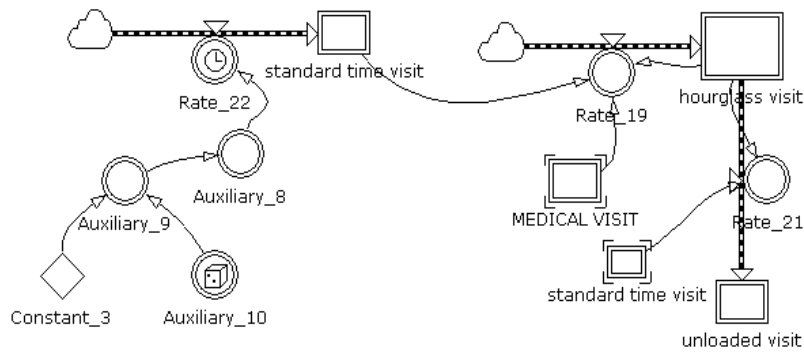


Figura 30: Costrutto clessidra-tempo random per la visita medica

In figura 30 si può vedere il costrutto realizzato per verificare quando il paziente ha effettivamente terminato la visita medica. La variabile *standard time_visit* rappresenta un vettore di tempi dove in ogni riga è possibile leggere il tempo random assegnato al paziente *i*-esimo per terminare la visita. La costante *constant_3* contiene il tempo medio necessario all'esecuzione dell'attività. La variabile ausiliaria *Auxiliary_10* rappresenta il vettore contenente i tempi random. Le altre variabili fungono da ausilio nella definizione del costrutto. In figura 30 si nota la presenza della variabile di livello *hourglass_visit* che funge da contatore. È un vettore che conta il tempo di permanenza del paziente alla visita medica dall'istante in cui egli arriva all'attività in esame. Viene effettuato lo scarico della clessidra, e quindi si carica lo stock *unloaded visit*, nel momento in cui tale clessidra assume valore pari a quello del tempo random generato per il paziente *i*-esimo.

In generale, lì dove sono state costruite le variabili di livello raffiguranti le clessidre, intese come contatori del tempo di permanenza del paziente in una determinata attività, sono state previste anche delle variabili di livello che fungono da scarico di tali clessidre. Si sono resi necessari tali scarichi per rendere possibile molteplici passaggi del paziente per una stessa attività.

Un esempio di condizione contenuta nel flusso è mostrato di seguito:

```
FOR(i=1..7720;j=1..21\IF(hourglass_preoperative[i]='standardtimepreoperative'[i];IF('freeoperatingtheatre'>0;IF('freeequipe'>0;'PREOPERATIVE PREPARATION'[i;j];0))))
```

Tale condizione significa che è possibile passare dall'attività di preparazione preoperatoria all'intervento quando le risorse necessarie all'intervento sono disponibili, quindi se c'è un equipe medica disponibile e se c'è una sala operatoria libera, e il tempo necessario alla preparazione preoperatoria è trascorso. In questa condizione è stato usato il ciclo for, così da sottoporre a tale condizione tutti i pazienti per i quali è necessario l'intervento, e l'approccio if-then secondo cui se e solo se sono rispettate le condizioni il paziente può procedere verso l'attività successiva.

È bene notare che il passaggio ad alcune attività, come ad esempio all'attività di radiologia, sia essa di emergenza o ordinaria, è consentito solo nel momento in cui le risorse necessarie all'espletamento dell'esame radiologico sono effettivamente disponibili. Invece il passaggio ad altre attività, come ad esempio il laboratorio di analisi, è consentito nel momento in cui l'attività precedente è effettivamente finita, al di là della disponibilità delle risorse. Ciò implica che può accadere che la provetta, contenete il sangue prelevato al paziente, attenda all'arrivo nel laboratorio prima di essere sottoposta alle attività svolte nel laboratorio stesso. Ciò accade per quelle attività per cui non c'è lo spostamento fisico del paziente; al laboratorio, ad esempio, verrà inviata solo la provetta. Di conseguenza per tale attività la clessidra comincerà a contare il tempo di permanenza, da confrontare col tempo generato in maniera randomica, solo quando c'è l'effettiva disponibilità delle risorse necessarie all'espletamento dell'attività. È comunque prevista una variabile di livello che conta il tempo di permanenza della provetta nel laboratorio dal suo arrivo, denominata *green clock o red clock*, che viene utilizzata per stabilire l'ordine di processamento delle provette nel caso in cui queste si accumulano all'interno del laboratorio. Sono stati considerati due

clock, uno verde ed uno rosso, come è possibile vedere in figura 31, per distinguere le provette accompagnate da scheda verde da quelle con scheda rossa. I colori delle schede indicano la priorità, intesa in termini di urgenza, assegnata al paziente. Le provette accompagnate dalla scheda rossa saranno lavorate nel laboratorio di emergenza e quelle con scheda verde nel laboratorio ordinario. La precedenza sarà data, a parità di colore della scheda di accompagnamento, alla provetta che è arrivata prima nel laboratorio, quindi alla provetta che registrerà il massimo tempo sull'orologio.

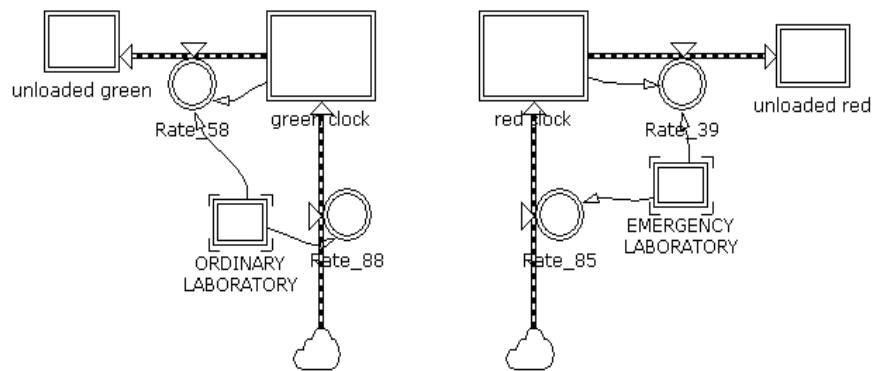


Figura 31: *Costrutto orologi verde/rosso associati ai pazienti*

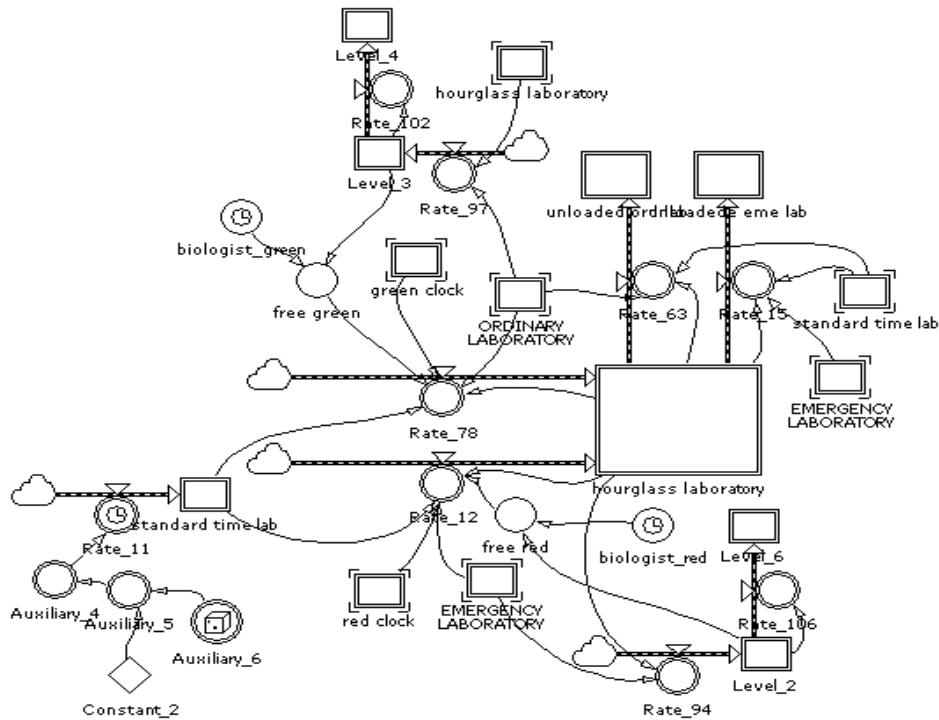


Figura 32: Costrutto clessidra per il laboratorio di analisi

$FOR(i=1..7720 \mid IF('ORDINARYLABORATORY'[i;6]=1; IF('hourglasslaboratory'[i] < 'standardtimelab'[i]; IF('freegreen' > 0; IF('greenclock'[i]=ARRMAX('green\ clock'); 1; 0))))))$

Tale condizione, appunto, significa che, se c'è disponibilità delle risorse e alla provetta i -esima corrisponde il massimo tempo registrato sul vettore orologio verde, allora tale provetta può essere lavorata e il vettore *hourglass_laboratory* può iniziare a contare il tempo di esecuzione dell'attività nell' i -esima posizione.

Nel modello è stato previsto un loop nel caso in cui un paziente debba essere sottoposto a più di un intervento chirurgico. Dal seguente loop il paziente uscirà solo nel momento in cui ha terminato tutti gli interventi.

farmaci, da parte delle unità operative costituenti il DEA, e conseguente scarico per utilizzo o per scadenza.

Lo scarico per utilizzo è stato realizzato tenendo conto delle patologie di cui sono affetti i pazienti e quindi della quantità di farmaci necessari per curare quella determinata patologia, mentre lo scarico per scadenza è stato realizzato su base percentuale.

La ricezione degli ordini presso le unità operative viene fatta con cadenza settimanale. Si suppone di ordinare ogni settimana una quantità fissa di farmaci e di riceverli in un giorno prestabilito della settimana.

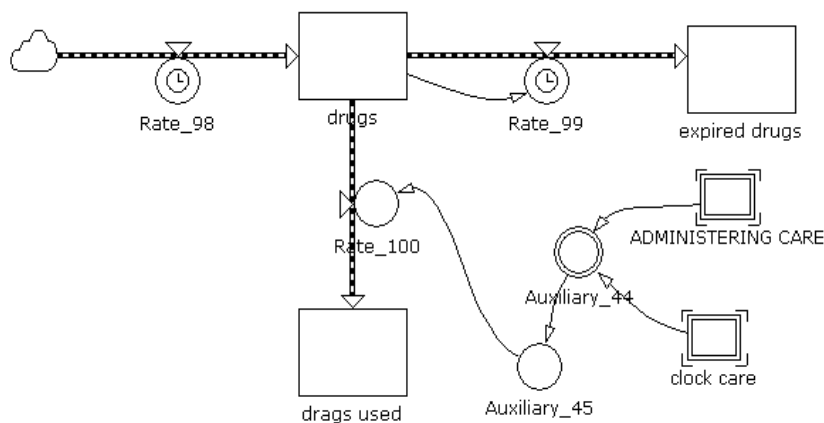


Figura 34: Costrutto della gestione dei farmaci

Nel modello è possibile trovare risorse che lavorano su più attività. In questo caso si è reso necessario tener conto dell'eventuale spostamento delle risorse mediante un coefficiente di impiego.

Ad esempio, le risorse che si occupano di effettuare la visita medica appena il paziente accede al ricovero, coincidono con quelle che effettuano la diagnosi dopo l'esecuzione e la refertazione degli esami di laboratorio e/o radiologici.

Inoltre, lì dove fosse necessaria una consulenza per un paziente presentatosi al Pronto Soccorso, la risorsa erogante la consulenza dovrà arrivare dal reparto lasciando quindi un posto vacante nell'unità operativa di provenienza. Tale risorsa coincide con quella operante alla visita medica.

Il coefficiente di impiego è rappresentato da una variabile ausiliaria al cui interno è possibile ritrovare la seguente condizione:

FOR(i=1..7720\IF(doctors<('sumattendeesvisit'+ 'sumattendeesadv');doctors/(doctors-'sum attendees adv');1))

Ovvero se il numero di medici risulta minore della somma dei pazienti che stanno effettuando la visita medica, di quelli sottoposti a diagnosi e di quelli sottoposti a consulenza, tale coefficiente assume valore pari al rapporto tra il numero di medici e la differenza tra i medici e i pazienti sottoposti a consulenza. In caso contrario il coefficiente assume valore unitario.

Per cui il flusso che consente di passare dalla visita medica al laboratorio conterrà la seguente condizione:

FOR(i=1..7720;j=1..21\IF('MEDICALVISIT'[i;6]=1;IF('standardtimevisit'[i;7]=('hourglassvisit'[i]/purposecoefficientvisit'[i]);IF('MEDICALVISIT'[i;7]=1;'MEDICAL VISIT'[i;j;0))))

In questo caso la condizione sul tempo di permanenza presso la visita medica assume una forma diversa, non si dovrà più verificare l'uguaglianza tra il tempo generato in maniera randomica e la clessidra, ma si dovrà avere l'uguaglianza tra il tempo random e il rapporto tra la clessidra e il coefficiente d'impiego. Questo perché se la risorsa medica è stata costretta a recarsi presso il Pronto Soccorso per erogare una consulenza lasciando inconclusa la visita medica, il tempo di permanenza presso l'attività in esame ne dovrà tener conto e quindi sarà certamente più lungo rispetto a quello preventivato.

10.2 – Descrizione della struttura di urgenza dell' A.O.R.N.

L'Azienda Ospedaliera di Rilievo Nazionale "Antonio Cardarelli" di Napoli è una grande azienda erogatrice di servizi sanitari, cui competono funzioni e responsabilità rilevanti, nell'esercizio di assicurare l'assistenza sanitaria su un ampio territorio. La struttura abbraccia un bacino di utenza esteso, oltre alla città, anche alla provincia ed alla Campania, e rappresenta un punto di riferimento per altre regioni d'Italia, in particolar modo del Meridione. L'Azienda Ospedaliera di Rilievo Nazionale e di Alta Specializzazione "A. Cardarelli" ha acquisito il ruolo di rilevanza nazionale, in base ai riconoscimenti di funzione attribuiti con il DPCM 08.04.93 e con il Decreto n°12255 del 22.12.1994 della Regione Campania. La struttura svolge un ruolo di primo piano per quanto riguarda l'assistenza sanitaria di urgenza: è infatti sede di Dipartimento di Emergenza-Accettazione di secondo livello, assicurando prestazioni di pronto soccorso in molteplici specialità. E' fra l'altro sede del Centro Grandi Ustionati, del Centro Antiveneni e del Centro per i Trapianti Epatici, (Centri di Emergenza Regionali); è presente altresì il Centro di Terapia Iperbarica. Presso tale dipartimento è ubicata inoltre la Centrale Operativa "118". L'azienda si distingue anche per le attività specialistiche di elezione d'area medica e chirurgica, rappresentando quindi un riferimento per la rete sanitaria della regione. L'attività assistenziale produce un elevato numero di ricoveri annui ordinari ed in day-hospital, mediamente superiori a 90.000, nonché di prestazioni erogate in regime ambulatoriale nell'ambito delle diverse specialità.

L'ospedale "A. Cardarelli", situato nel cuore della zona ospedaliera, possiede una struttura "a padiglioni" che occupa nel complesso una superficie di 250.000 metri quadrati. Di questi, 50.000 metri quadri sono rappresentati da edifici, ed i restanti 200.000 da viali alberati e pinete che di fatto costituiscono un vero e proprio "polmone verde". Dei 21 padiglioni esistenti, identificati con le lettere A, B, C, D, E, F, G, H, I, L, M, N, O, P, Q, R, S, T,

U, V e Y, costruiti in diverse epoche a partire dal 1927 fino al 1990, circa 14 sono destinati alle attività di diagnosi e cura, ed i restanti 7 ai servizi tecnici. Nell'ambito della superficie dell'Azienda è situato un eliporto che costituisce il fulcro dei trasferimenti rapidi nell'ambito dell'emergenza intra ed extraregionale, ove sia richiesto l'intervento di eliambulanza. La struttura dell'eliporto, dotato dei più moderni strumenti tecnici di controllo e di sicurezza, consente l'atterraggio ed il decollo anche nelle ore notturne.

La struttura è costituita dai seguenti 9 Dipartimenti e 2 Aree:

- Dipartimento di Emergenza e Accettazione;
- Dipartimento Medico Specialistico;
- Dipartimento Chirurgico Specialistico;
- Dipartimento dei Trapianti;
- Dipartimento Oncopneumoematologico;
- Dipartimento delle tecnologie avanzate diagnostico-terapeutiche;
- Dipartimento dei servizi;
- Dipartimento di Anestesia e Rianimazione;
- Dipartimento Amministrativo;
- Area Generale delle Attività Sanitarie;
- Area Generale della Direzione Strategica.

I Dipartimenti/UOC/UOS/reparti/Servizi sono le strutture organizzative all'interno delle quali sono distribuiti i 3800 operatori sanitari ai quali si aggiungono pazienti, utenti, partecipanti ad attività didattiche e di ricerca, visitatori, volontari, fornitori e dipendenti di imprese esterne.

Di seguito riportiamo una piantina estesa della struttura ospedaliera per capirne meglio la suddivisione territoriale dei vari padiglioni ed attività.



Figura 35: *Estensione territoriale e suddivisione in pianta della struttura ospedaliera “A. CARDARELLI”*

10.3 – Implementazione del modello

Lo scopo per il quale si è deciso di utilizzare un software di simulazione per modellare il comportamento di un sistema complesso, quale può essere il DEA di una struttura ospedaliera, è principalmente quello di verificare se è possibile ottenere effettivi miglioramenti apportando modifiche in ottica Agile, così come suggerito dallo stato dell’arte in materia.

Un software di simulazione permette, infatti, di verificare i risultati, positivi o negativi, di modifiche, eventualmente anche strutturali, prima di realizzarle praticamente, risparmiando, quindi investimenti inutili qualora i risultati dovessero essere negativi. Rappresenta pertanto uno strumento di supporto alle decisioni.

Lo scopo di questo lavoro è, dunque, quello di fotografare lo stato attuale del Dipartimento d’Emergenza e Accettazione e, dopo averne individuato le eventuali criticità, proporre dei miglioramenti.

I risultati di questi eventuali miglioramenti possono essere valutati con il modello di simulazione precedentemente descritto.

L’implementazione del modello è stata fatta inserendo una matrice di dati nel programma di simulazione. Il programma di modellizzazione/simulazione *Powersim* permette di importare i file *Excel*; quindi, i dati relativi ai pazienti in ingresso presso la struttura ospedaliera, sono stati organizzati in fogli di lavoro *Excel*, uno per ogni tipo di accesso, in cui ogni riga della matrice rappresenta un paziente e ogni colonna rappresenta una caratteristica del paziente.

La generazione dei dati di simulazione è stata eseguita a partire dalla serie storica degli accessi presso l’A.O.R.N “A. Cardarelli”. Essendo nota la serie storica degli arrivi dei pazienti, è possibile suddividere l’orizzonte temporale di riferimento, che è pari a 1 giorno, considerando un giorno ritenuto rappresentativo, dopo opportuna stratificazione dei dati, in intervalli statisticamente significativi. Sono stati considerati 24 intervalli di ampiezza pari a 1 ora. Per ogni intervallo di tempo è possibile affermare che la distribuzione degli accessi segue una distribuzione di Poisson, ognuna caratterizzata da un k diverso. Quindi, per ogni orizzonte temporale statisticamente significativo è stato possibile calcolare un k che rappresenta la frequenza con cui si registrano gli accessi. A questo punto è stata generata la matrice dei dati di simulazione, riferiti ad un periodo temporale pari a 30 giorni, tenendo conto dei k riscontrati nella serie storica dei dati reali.

La matrice di dati che è stata utilizzata ha dimensioni pari a 7720 righe con 21 colonne, ovvero sono stati considerati 7720 pazienti, essendo questo

mediamente il numero di accessi presso il DEA registrato in 30 giorni, e di ogni paziente sono state considerate 21 caratteristiche.

Alcune delle caratteristiche considerate sono: identificativo del paziente, data di accesso esplicitata in giorno, ora e minuto di ingresso, diagnosi di ingresso e specifica delle attività fatte, intese come analisi di laboratorio chimico-cliniche, esami radiologici, intervento, tipologia di intervento, numero degli interventi, consulenza e degenza in rianimazione.

Nella cella relativa ad uno specifico paziente ed ad una specifica attività, ci sarà un valore che, attraverso la legenda della matrice, rivelerà informazioni del tipo se la prestazione è prevista o meno per quel determinato utente.

L'orizzonte di simulazione scelto è pari a 30 giorni con un time step di un minuto. La scelta dell'orizzonte di simulazione risulta critica poiché è bene scegliere un orizzonte di riferimento sufficientemente grande così da soddisfare l'esigenza di significatività del periodo in esame, ma d'altro canto, scegliere un orizzonte piccolo dà un vantaggio in termini di validazione dato che per quest'ultima si rende necessario raccogliere 30 set di dati con ampiezza pari all'orizzonte temporale di riferimento ed inoltre è opportuno che i dati siano tutti rappresentativi del fenomeno in esame ovvero che risultino tali da non evidenziare innovazioni tecnologiche o cambiamenti radicali dei processi aziendali.

Dunque è opportuno scegliere il massimo orizzonte temporale tale per cui la serie storica di riferimento assuma senso statistico.

10.4 - Onda di carico e condizione di stazionarietà degli impianti

Data una serie storica di dati è possibile valutare la distribuzione della domanda e nel caso tale distribuzione presenti un andamento di Poisson è

possibile calcolare il valore di k , ovvero la frequenza con cui si registrano gli eventi del fenomeno in esame.

Se dalla distribuzione della domanda è possibile ricavare un solo valore di k allora la domanda risulta essere livellata. In generale, la domanda può essere livellata totalmente o parzialmente. Lo è totalmente se risulta livellata lungo tutto l'orizzonte di riferimento, invece lo è parzialmente se risulta livellata solo all'interno di alcuni tratti dell'orizzonte temporale, come ad esempio in alcune fasce orarie o all'interno di alcuni turni di lavoro oppure se lo è in tutti i turni ma in ognuno di questi si registra un differente livellamento caratterizzato da un differente valore di k .

Nel caso in cui la domanda risulta livellata è possibile affermare che tale caratteristica presenta una configurazione di tipo Lean.

Se, invece, all'interno di alcune fasce orarie, l'onda di carico non risulta avere configurazione Lean, non è possibile creare un turno di lavoro in cui operare Lean.

Si definisce stazionaria una configurazione in cui non si registrano variazioni significative.

La configurazione stazionaria delle risorse è una condizione necessaria ma non sufficiente affinché un sistema di produzione possa essere definito di tipo Lean. Per poterlo definire Lean, il sistema di produzione deve avere la più efficiente configurazione, stazionaria e livellata, rispetto all'onda di carico.

Nel caso in cui la domanda presenti un andamento non livellato, è possibile affermare che sistemi a configurazione stazionarie cambiano onde di carico non livellate, quindi a k differenziati, in profili di performance tendenzialmente livellati.

Inoltre, se la domanda presenta una configurazione non livellata è necessario reingegnerizzare l'impianto produttivo secondo una configurazione non stazionaria delle risorse e tale impianto deve essere in grado di compensare i k , e i relativi Δk , della domanda. Tale configurazione potrà essere Agile se

sono verificate le condizioni su dette ed inoltre gode della proprietà di massima efficienza.

Per cui è possibile affermare che:

Dato un impianto industriale, sollecitato da un'onda di carico non naturalmente livellata (e non industrialmente livellabile), il sistema di produzione associato all'impianto è di tipo Agile se e solo se:

1. le risorse dell'impianto sono in configurazione non stazionaria;
2. intervallo (di significatività statistica) per intervallo, vale che:

$$k(t) = \rho(t);$$

3. la matrice del ritmo produttivo è tale che:

$$\underline{\rho}_A(t) = \underline{\rho}_{\min}(t) + \underline{\Delta\rho}(t)$$

con $\underline{\rho}_{\min}(t) = \underline{\rho}_L(t) = \frac{1}{\text{Takt Time}}$, ovvero $\underline{\rho}_{\min}$ risulta uguale al ritmo

produttivo registrato in ambiente Lean e $\underline{\Delta\rho}(t) = \underline{\Delta K}(t)$

4. la matrice dell'agilità, anche detta della reattività, è pari a:

$$\underline{\alpha} = \frac{d}{dt} \underline{\rho}_A(t) = \frac{dK}{dt}$$

Nel caso in esame l'onda di carico risulta essere non livellata e le risorse hanno una configurazione parzialmente stazionaria, ovvero stazionaria all'interno del turno di lavoro ma non stazionaria nell'arco dell'intera giornata.

In particolare l'onda di carico presenta k differenziati per ogni intervallo statisticamente significativo.

È necessario, quindi, reingegnerizzare il sistema a “configurazione parzialmente stazionaria delle risorse” in una “configurazione agile delle risorse”.

La reingegnerizzazione può essere fatta per singolo processo ma il ragionamento può essere reiterato più volte così da osservare molteplici miglioramenti. In particolare se si considera un sistema costituito da n

processi, al più in n iterazioni è possibile ottenere una totale configurazione agile delle risorse per l'intero sistema.

Si riporta di seguito un'analisi dettagliata dell'onda di carico in esame.

A partire dalla serie storica degli accessi presso il DEA dell'A.O.R.N. "A. Cardarelli" di Napoli, è stato scelto un giorno ritenuto rappresentativo dell'andamento medio degli accessi e su questo, come già detto, è stato effettuato il calcolo dei k per ogni intervallo statisticamente significativo. Come intervallo statisticamente significativo è stata scelta l'ora ed è pari al minimo intervallo in cui è possibile registrare almeno un accesso.

Dopo di che sono state calcolate le relative variazioni e quindi i Δk .

La specifica dei valori ottenuti è riportata la Tabella 6, che segue.

Tabella 4: Calcolo di k e Δk negli intervalli di tempo statisticamente significativi

FASCIA ORARIA	k	INTERVALLO ORARIO	Δk
0-1	8		
1-2	9	0-2	1
2-3	7	1-3	2
3-4	3	2-4	4
4-5	6	3-5	3
5-6	10	4-6	4
6-7	10	5-7	0
7-8	10	6-8	0
8-9	10	7-9	0
9-10	14	8-10	4
10-11	15	9-11	1
11-12	17	10-12	2
12-13	17	11-13	0
13-14	18	12-14	1
14-15	16	13-15	2
15-16	12	14-16	4
16-17	7	15-17	5
17-18	8	16-18	1
18-19	10	17-19	2
19-20	10	18-20	0
20-21	9	19-21	1

21-22	11	20-22	2
22-23	9	21-23	2
23-24	8	22-24	1

Il massimo valore di k si registra nell'orizzonte temporale tra le 13 e le 14 ed è pari a 18. Il k_{\min} invece è pari a 3 e si registra tra le 3 e le 4. Dalla tabella 6, inoltre, è possibile notare che la massima decelerazione si registra nelle fascia oraria tra le 15 e le 17 ed è pari a 5 mentre, invece, la massima accelerazione è pari a 4 e si manifesta tra le 8 e le 10 e tra le 4 e le 6.

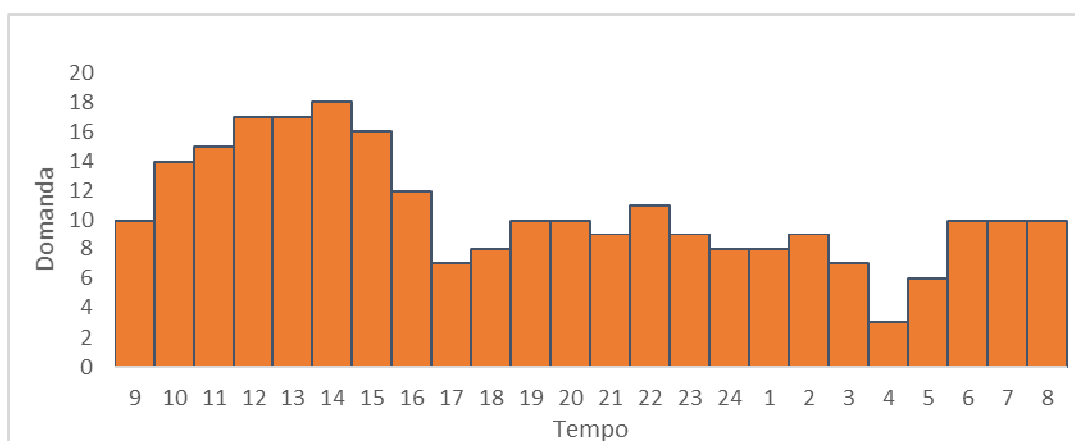


Figura 36: *Andamento della domanda in una giornata ritenuta significativa*

Di seguito sono riportati gli andamenti degli accessi divisi per turno di lavoro. In particolare nel turno che va dalle 8 alle 14 si registra il 34% degli accessi giornalieri, dalle 14 alle 20 si registra il 26% della domanda mentre invece nel turno notturno, dalle 20 alle 8, si avrà il 40% degli accessi giornalieri.

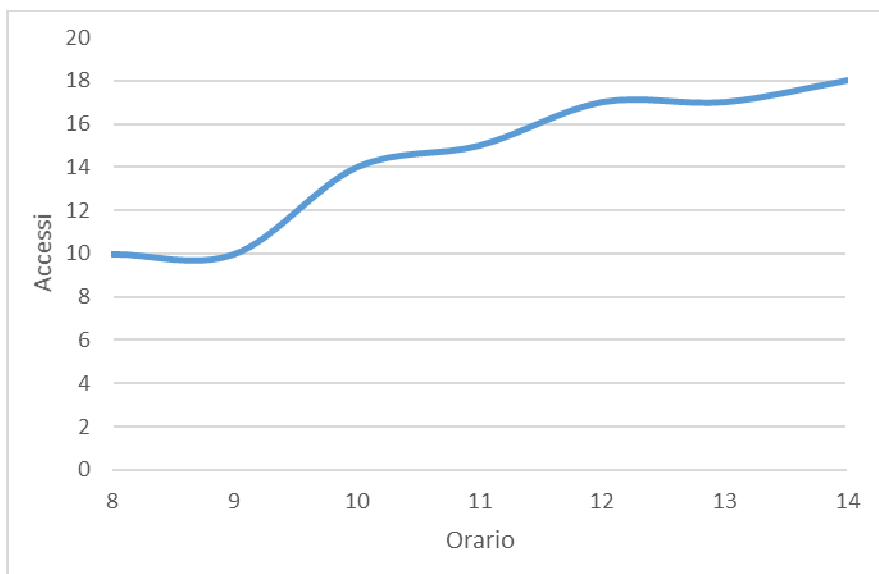


Figura 37: Andamento degli accessi nel turno di lavoro 8-14

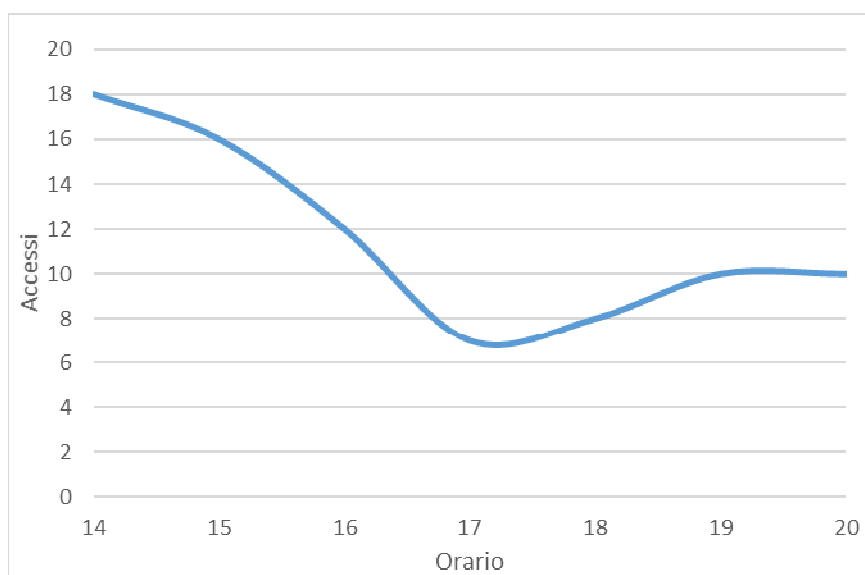


Figura 38: Andamento degli accessi nel turno di lavoro 14-20

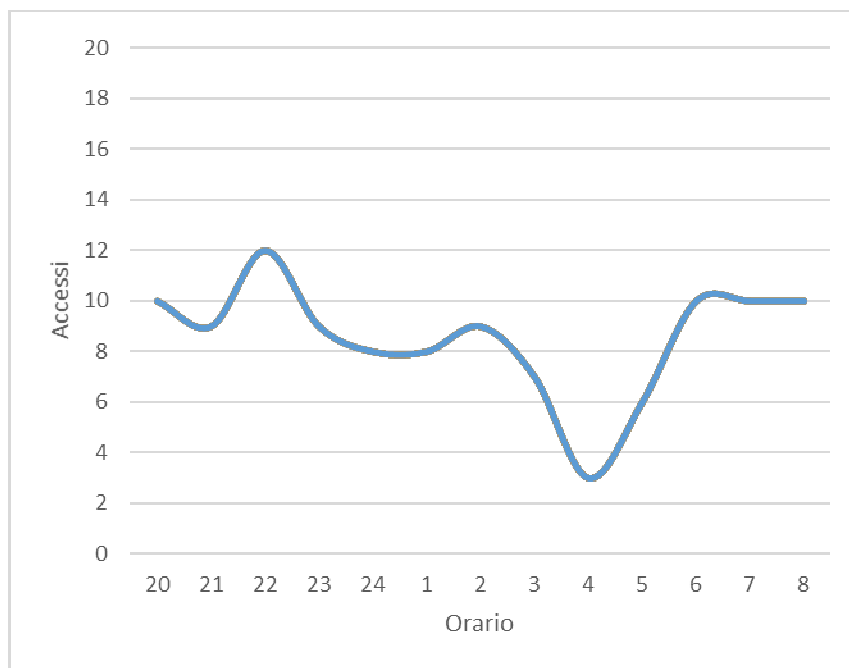


Figura 39: Andamento degli accessi nel turno di lavoro 20-8

10.5 – Validazione del modello

È importante verificare che il modello di simulazione elaborato, una volta terminato, riproduca il comportamento del sistema reale, oggetto di studio. Nello specifico, occorre verificare che le misure di prestazione del sistema reale siano bene approssimate dalle misure generate dal modello di simulazione. Se è verificata la seguente condizione allora si potrà dire che le due popolazioni, reale e simulata, sono sovrapponibili. È necessario, dunque, validare il modello. La validazione consiste nel confrontare i risultati ottenuti dalla simulazione con i dati reali dell'azienda in studio. Nel caso in cui, dal confronto, si verifichi un disallineamento, occorre intervenire per correggere il modello e studiare se le cause sono semplici errori imputabili ai parametri del processo oppure se sono errori più consistenti relativi ad un flusso logico/causale non coerente.

Dunque, se si riuscisse ad ottenere, nel periodo di simulazione, in configurazione AS-IS, un modello che si comporti come il processo vero, se pur non allo stesso modo ma ad un livello che si ritiene sufficientemente accettabile, allora presumibilmente gli effetti che si otterrebbero in simulazione si otterrebbero anche nel mondo reale.

Oltre alla validazione del modello è necessaria anche la verifica dello stesso, dove per verifica si intende quell'attività in cui si determina se il modello di simulazione funziona secondo le intenzioni di colui che lo ha creato. Dunque, durante la costruzione del modello stesso è necessario effettuare delle prove utilizzando dei risultati noti in modo da verificare che tutti gli elementi importanti siano stati inclusi nel modello.

La validazione del modello verrà fatto con l'ausilio dei due test statistici riportati di seguito:

- test di Student;
- test di Fisher.

Come parametro di prestazione del sistema reale, da confrontare con i corrispettivi output del modello, si è scelto l'andamento del coefficiente di riempimento dei posti letto. Tale coefficiente è calcolato mediante il costrutto definito nella figura 40.

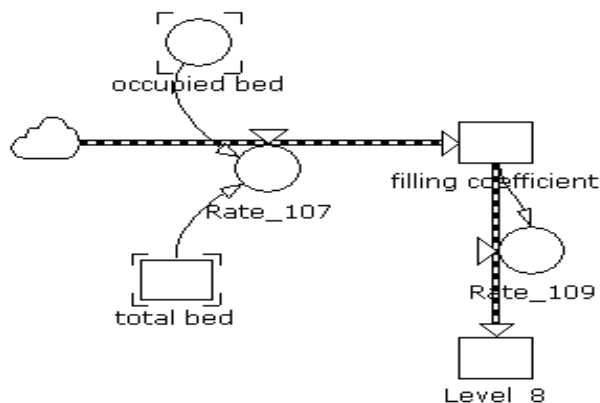


Figura 40: *Costrutto che consente il calcolo dell'andamento del coefficiente di riempimento dei posti letto*

Il *filling coefficient* è definito dal rapporto tra il numero di letti occupati, inteso come somma dei posti letto più le barelle, e il numero di letti totali, moltiplicato per 100.

Di seguito si riporta la specifica dei posti letto presenti per ogni unità operativa considerata.

Tabella 5: Posti letto presenti nelle unità operative costituenti il DEA

Unità operativa	Numero letti
Chirurgia d'urgenza	33
Rianimazione	20
Gastro d'urgenza	13
Ostetricia	14
Terapia intensiva neonatale (TIN)	16
Ginecologia	13
Nido	19
Neurologia	20
Neurochirurgia	30
Unità terapia intensiva cardiologica (UTIC)	8
Cardiologia con UTIC	12
Medicina d'urgenza	30
TOTALE	228

Come noto, gli input del modello di simulazione seguono delle distribuzioni statistiche e sono affetti da una certa variabilità che, ovviamente, si trasferisce anche sugli output.

Si ricorre al Teorema del Limite Centrale, per applicare le tecniche di stima intervallare, a condizione che si utilizzino grandi campioni ($n \geq 30$) casuali. Il Teorema del Limite Centrale afferma che la somma di una successione di n variabili aleatorie, s-indipendenti ed equidistribuite, con media μ e varianza σ^2 , converge in distribuzione alla variabile aleatoria Gaussiana con media $n\mu$ e varianza $n\sigma^2$, qualunque sia l'originaria Cdf delle n variabili aleatorie.

Per la validazione del modello è necessario effettuare un test d'ipotesi, per il confronto tra le medie di due popolazioni con varianza incognita (ma supposta uguale), avendo a disposizione le statistiche del primo (X_1 e X_2) e del secondo ordine (S_1^2 e S_2^2) di due grandi campioni ($n = m = 30$) casuali.

Nell'ipotesi in cui le due popolazioni sono s-indipendenti ed abbiano Cdf Gaussiane, di parametri rispettivamente μ_1, σ_1 e μ_2, σ_2 , per effettuare il test di uguaglianza delle medie delle due popolazioni, è possibile affermare che la differenza delle rispettive medie campionarie $Y=X_1-X_2$ ha Cdf ancora Gaussiana, di parametri:

$$\mu_Y = \mu_1 - \mu_2$$

$$\sigma_Y = \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}$$

E' lecito supporre che $\sigma = \sigma_1 = \sigma_2$, per cui la varianza (se non è nota) può essere stimata, sull'insieme dei due campioni, come somma di tutte le differenze quadratiche disponibili divisa per i corrispondenti gradi di libertà.

Tale varianza è nota come *Varianza Pooled* ed equivale alla media delle due varianze campionarie pesate con i rispettivi gradi di libertà:

$$S^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}$$

Per effettuare il test di uguaglianza delle medie delle due popolazioni, dunque, è possibile utilizzare la seguente statistica t di Student:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S\sqrt{(1/n + 1/m)}}$$

A questo punto, per effettuare il test si procederà così come segue.

Si definiscono l'ipotesi nulla H_0 , secondo cui i dati simulati approssimano bene i dati reali, e l'ipotesi da rigettare H_1 :

$$H_0 = [x_1 = x_2]$$

$$H_1 = [x_1 \neq x_2]$$

Si calcolano i gradi di libertà v , dati dalla differenza tra il n° di osservazioni ed il n° di gruppi messi a confronto, si definisce il livello di significatività $1-\alpha$ del test e si estrapolano, dalle relative tabelle, i valori critici $t_{0,025}$ e $-t_{0,025}$ che definiscono la regione di rifiuto di H_0 :

$$v = n+m-2;$$

$$\alpha = 0,05;$$

$$t_{0,025} = 2,00039$$

Nello specifico caso in esame, a partire da una serie storica di dati raccolti lungo un orizzonte temporale pari a 3 anni, effettuando un'opportuna stratificazione dei dati, ovvero eliminando eventuali valori anomali, sono stati considerati, come dati reali, gli andamenti medi del coefficiente di riempimento dei posti letto di 30 orizzonti temporali ciascuno di ampiezza pari a 30 giorni. Come dati simulati, invece, sono stati considerati gli andamenti medi del coefficiente per 30 run di simulazione. Nella tabella 5, che segue, sono riassunti gli output reali e quelli simulati.

Tabella 6: Confronto tra dati reali e dati ottenuti dalla simulazione relativamente al Coefficiente di riempimento dei posti letto

PERIODO	OUTPUT REALE	OUTPUT SIMULAZIONE
1	118	116
2	121	118
3	123	121
4	119	118
5	119	120
6	116	116
7	118	115
8	117	118
9	117	117
10	121	118
11	117	116
12	120	121
13	118	117
14	123	118
15	122	119
16	119	118
17	120	119
18	123	120
19	121	119
20	118	117
21	118	116
22	122	118
23	117	115
24	119	116
25	121	117
26	122	119
27	116	117

28	121	118
29	120	119
30	122	120

Dai dati a disposizione è possibile calcolare:

- La media dei dati reali pari a 119.633;
- La varianza dei dati reali pari a 4.566;
- La media dei dati simulati pari a 117.867;
- La varianza dei dati simulati pari a 2.582.

A questo punto, sostituendo i valori su detti nella formula della t di Student, si ottiene che $t=1.84$.

Essendo $|t| < t_{0.025}$, si deduce che non può essere rigettata l'ipotesi H_0 , ed è quindi possibile affermare che il test d'ipotesi è superato posto un intervallo di confidenza pari a $1 - \alpha = 0,95$.

Nelle ipotesi su dette, è possibile effettuare anche il test di uguaglianza delle varianze delle due popolazioni. Indicando con S_1^2 e S_2^2 due stime campionarie delle varianze σ_1^2 e σ_2^2 , è possibile utilizzare la seguente funzione ancillare:

$$Z = \frac{S_1^2 \sigma_2^2}{S_2^2 \sigma_1^2}$$

Tale funzione è definita come Z di Fisher con n-1 e m-1 gradi di libertà (essendo n e m le dimensioni dei due campioni).

Si sceglie come livello di significatività, anche in questo caso, $1-\alpha = 0.95$ e si fissano le seguenti ipotesi:

$$H_0 = [\sigma_1^2 = \sigma_2^2]$$

$$H_1 = [\sigma_1^2 < \sigma_2^2]$$

$$v_1 = 29$$

$$v_2 = 29$$

$$Z_{0,95} = 1.86$$

Sostituendo i valori delle varianze, ottenute dai dati reali e da quelli simulati, e assumendo $\sigma_1^2 = \sigma_2^2$ si ottiene che la Z di Fisher è pari a:

$$Z=1.77$$

Tale valore è minore della $Z_{0,95}$, quindi risulta fuori dalla zona di rigetto ed induce a non rifiutare l'ipotesi H_0 .

Per cui il test d'ipotesi risulta verificato con un livello di significatività pari a $1-\alpha=0.95$.

In Figura 41 è confrontato l'andamento dei dati reali e quello ottenuti dalla simulazione. Già visivamente si nota che lo scarto tra le due serie di dati non è da ritenersi significativo. Con i test d'ipotesi si è data validazione scientifica al risultato ottenuto.

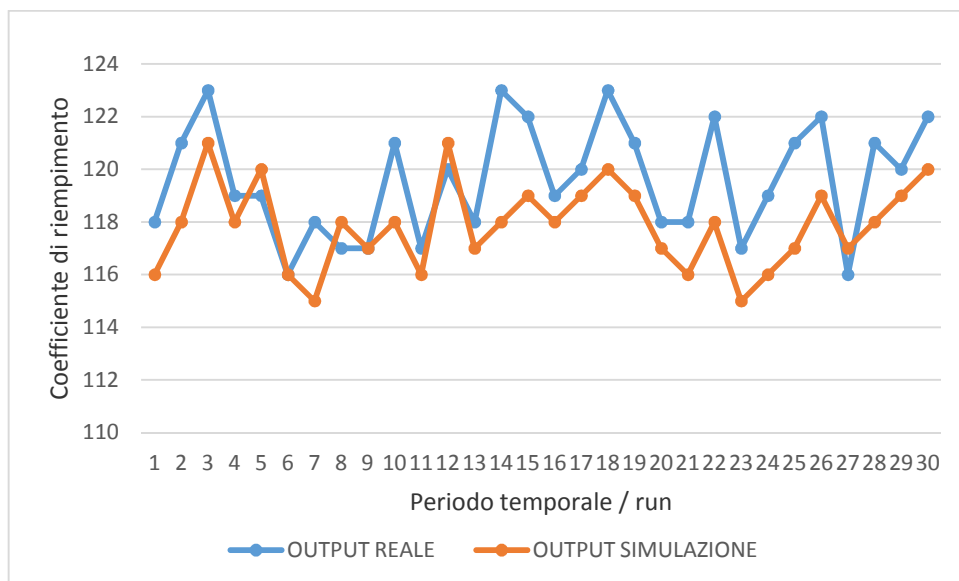


Figura 41: Confronto tra output reali e output simulati

10.6 - Soluzioni proposte: Scenario TO BE

Il miglioramento che si propone è la reingegnerizzazione del sistema, attraverso il dimensionamento delle risorse, che consente di passare da una “configurazione parzialmente stazionaria delle risorse” ad una “configurazione agile delle risorse”.

Per fare ciò è necessario bilanciare le accelerazioni che si registrano, ovvero i picchi di domanda presso la struttura ospedaliera. Si bilanciano solo le accelerazioni e non le decelerazioni poiché si considera di avere, come configurazione minima, una configurazione di tipo Lean ovvero una configurazione che consente di massimizzare l'efficienza anche se però a discapito dell'efficacia.

Le maggiori criticità si sono riscontrate nella gestione dei supporti, a causa dei lunghi tempi di refertazione, e nella gestione degli interventi, essendo il DEA un dipartimento d'emergenza e presentando quindi numerosi pazienti bisognosi di intervento chirurgico urgente.

Per bilanciare i Δk_{\max} si prevede di disporre di una risorsa in più, solo su parte del turno di lavoro, e tale risorsa sarà prevista nella fascia oraria che corrisponde alla presenza del picco di domanda più il lead time necessario affinché il paziente arrivi all'attività collo di bottiglia, ovvero all'attività in cui si registra la congestione delle richieste del servizio. Quindi, dato che i picchi di domanda si registrano tra le 4 e le 6 e le 8 e le 10, si prevede di disporre di una risorsa in più, per l'attività di refertazione, tra le 6 e le 8 e tra le 10 e le 12. Per tale attività si ritiene opportuno lo spostamento di una risorsa proveniente dal processo di visita medica poiché è stato possibile osservare che, nelle fasce orarie in cui si registra un picco di domanda di refertazione, si hanno invece poche richieste di visita medica quindi si opera lo spostamento della risorsa dall'attività più “scarica” all'attività più “carica”. Tale risorsa dovrà essere un medico con caratteristiche versatili ovvero con caratteristiche flessibili così da poter svolgere molteplici compiti con risultati

soddisfacenti. Dovrà essere quindi competente sia nella refertazione che nella visita medica, ovvero dovrà essere in grado di interpretare i risultati delle analisi di laboratorio e di leggere le lastre nonché di stabilire quale sia il miglior percorso di cura da intraprendere per il paziente.

Per quanto riguarda l'attività di intervento chirurgico invece si propone di poter disporre di una sala operatoria e di un'equipe operatoria in più, nella fascia oraria a cui corrisponde il picco di domanda più il lead time necessario affinché il paziente arrivi a tale attività, a partire dal momento in cui egli accede alla struttura ospedaliera. Quindi si aggiungeranno tali risorse nelle fasce orarie che vanno dalle 9 alle 11 e dalle 13 alle 15. Sia l'equipe operatoria che la sala verranno aggiunte alla gestione delle emergenze e sottratte dalla gestione degli interventi programmati.

Il miglioramento proposto è stato valutato attraverso il modello di simulazione, ed in particolare monitorando i valori assunti dal coefficiente di riempimento dei posti letto, che si configura come indicatore primario di riferimento, e da altri 3 indicatori durante l'esecuzione di 30 run con la nuova configurazione delle risorse.

I 4 indicatori considerati, quindi, sono:

- ❖ il coefficiente di riempimento dei posti letto, descritto in precedenza;
- ❖ il tempo di permanenza media nella struttura ospedaliera;
- ❖ il tempo massimo di attesa a cui il paziente può essere esposto prima di essere sottoposto all'intervento;
- ❖ il tempo medio di refertazione.

La riconfigurazione delle risorse è stata eseguita considerando separatamente i due processi critici. Sono stati individuati quindi due scenari TO BE, uno in cui si opera il dimensionamento dei fattori produttivi per l'attività di refertazione ed uno in cui si riconfigurano le risorse del blocco operatorio.

In entrambi i casi la struttura del modello di simulazione è rimasta invariata mentre invece è stata apportata una modifica nella condizione inserita

all'interno della variabile ausiliaria riferita alla tipologia di risorsa in questione.

Come primo TO BE si considera di apportare le modifiche su dette al processo di refertazione (TO BE 1).

In tabella 7 sono riportati i valori del coefficiente di riempimento dei posti letto, precedentemente descritto, ottenuti dai 30 run in configurazione Agile delle risorse.

Tabella 7: Valori Coefficiente di riempimento dei posti letto nei due scenari (TO BE 1)

RUN	OUTPUT AS IS	OUTPUT TO BE 1
1	116	116
2	118	117
3	121	120
4	118	119
5	120	119
6	116	116
7	115	116
8	118	117
9	117	117
10	118	118
11	116	115
12	121	120
13	117	119
14	118	118
15	119	118
16	118	119
17	119	120
18	120	119
19	119	118

20	117	118
21	116	117
22	118	117
23	115	116
24	116	116
25	117	116
26	119	118
27	117	118
28	118	117
29	119	120
30	120	120

La media ottenuta dai valori dei 30 run in configurazione AS IS delle risorse è pari a 117.867 mentre la media dei coefficienti di riempimento registrati in configurazione TO BE 1 è pari a 117.8.

Quindi con il cambiamento proposto non è stato possibile ridurre il valore medio del coefficiente di riempimento dei posti letto.

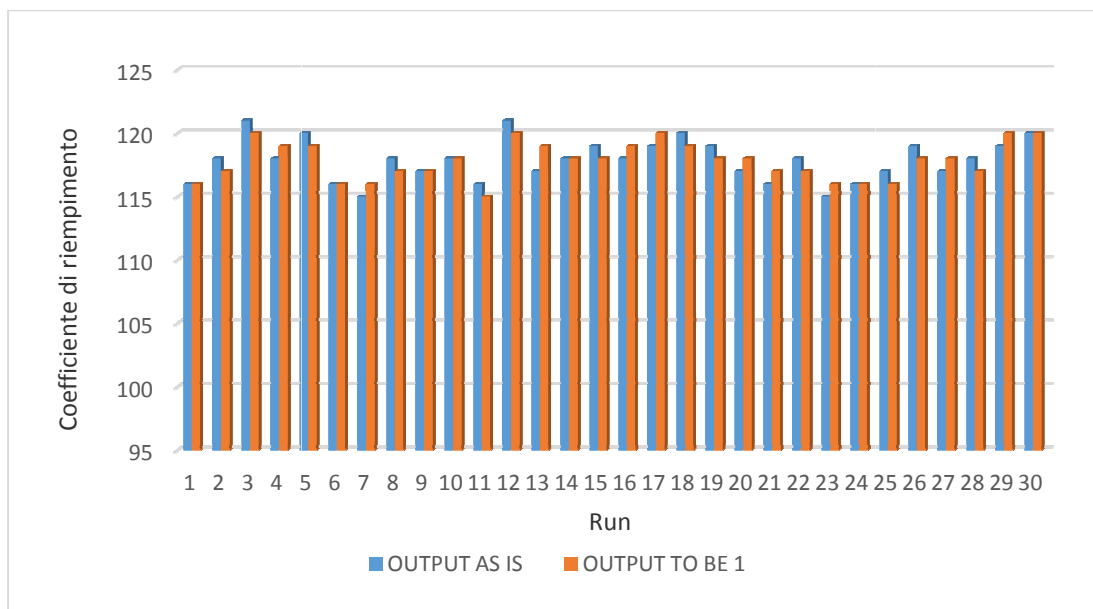


Figura 42: Confronto tra i valori del coefficiente di riempimento nei due scenari (TO BE 1)

Per valutare il tempo di permanenza media nella struttura ospedaliera è stato necessario progettare un apposito costrutto che si riporta di seguito.

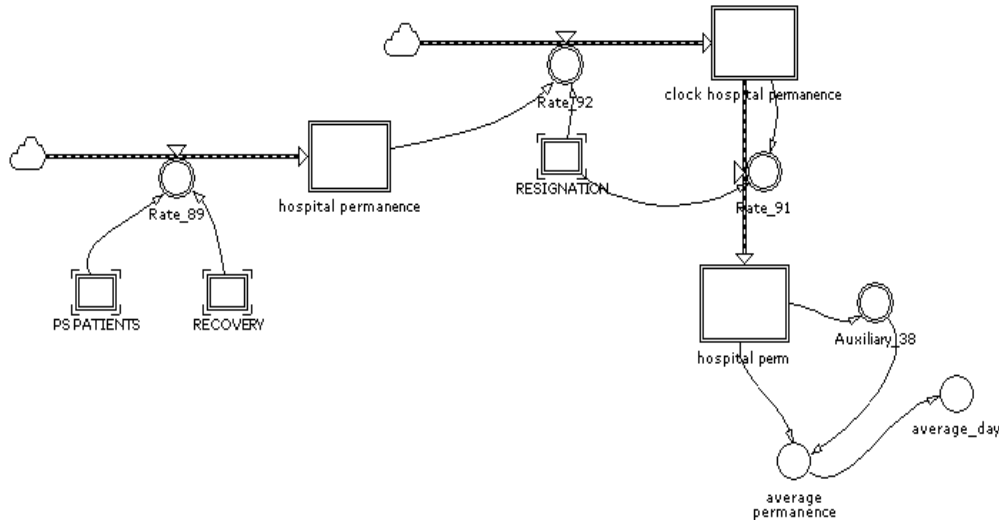


Figura 43: Costrutto per valutare la permanenza media presso il DEA

Mediante una clessidra, costruita con una variabile stock vettoriale, è stato possibile calcolare il tempo di permanenza nel DEA per ogni paziente e poi con l’ausilio di un’apposita variabile è stata fatta la media di tali valori. La clessidra inizia a contare il tempo di permanenza dal momento in cui il paziente viene ricoverato presso i reparti del Dipartimento d’Emergenza e Accettazione.

Tabella 8: Confronto tempi di permanenza media presso il DEA nei due scenari

(TO BE 1)

RUN	OUTPUT AS IS	OUTPUT TO BE 1
1	7	6
2	6	7

3	7	7
4	6	6
5	5	6
6	6	5
7	7	6
8	5	6
9	6	6
10	6	7
11	5	5
12	6	6
13	7	6
14	6	7
15	6	6
16	5	5
17	7	6
18	6	7
19	8	7
20	6	7
21	6	6
22	5	6
23	5	5
24	7	6
25	6	6
26	6	7
27	7	6
28	5	5
29	7	7
30	6	6

Dalla tabella 8 si evince che, attraverso la configurazione agile delle risorse, i tempi di permanenza medi presso il DEA restano pressoché invariati. Tale risultato è in accordo con quello registrato dalla valutazione del coefficiente di riempimento dei posti letto.

Il confronto tra i valori del tempo di permanenza dei pazienti presso la struttura ospedaliera nelle due configurazioni è visibile mediante il grafico che segue.

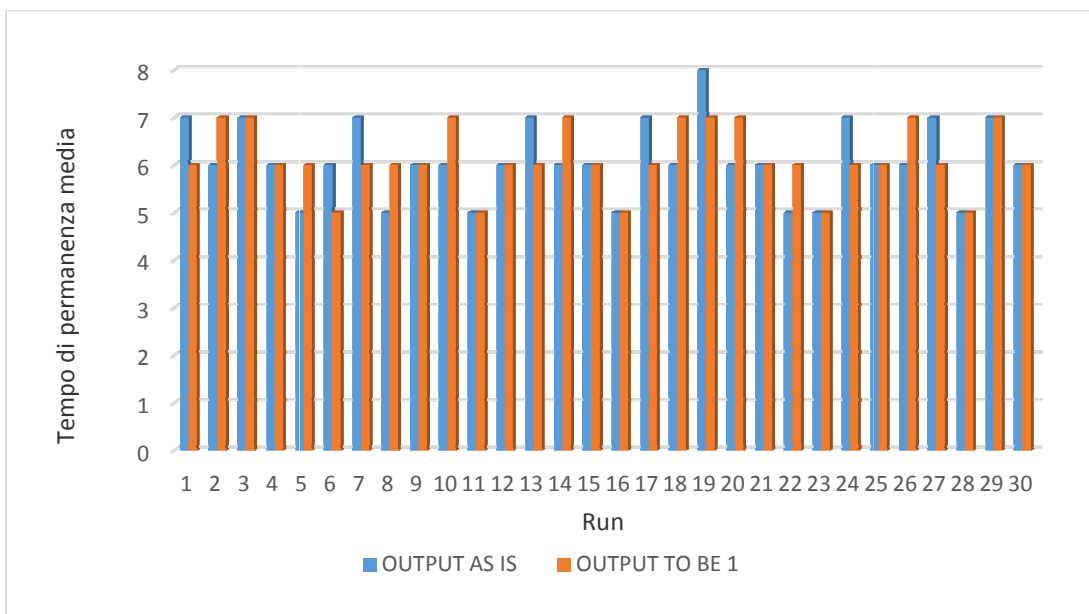


Figura 44: Grafico di confronto dei tempi di permanenza nei due scenari (TO BE 1)

La media dei valori registrati dal modello di simulazione in configurazione AS IS è circa 6 giorni così la media in configurazione TO BE 1. Quindi non è stato registrato alcun miglioramento.

Per calcolare invece il tempo massimo di attesa a cui il paziente può essere esposto prima di essere sottoposto all'intervento si è fatto riferimento al costruito che segue.

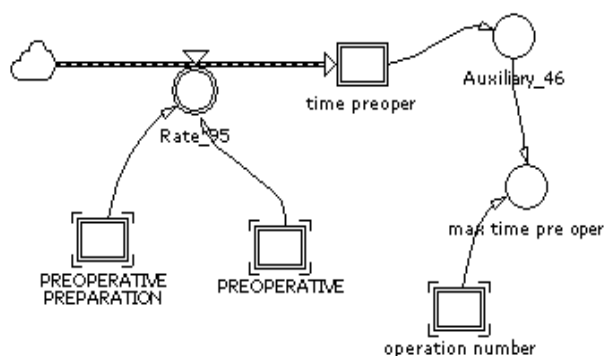


Figura 45: Costrutto del massimo tempo di attesa per essere sottoposti ad intervento chirurgico

Attraverso la clessidra *time preoper* sono stati registrati i tempi di attesa per essere sottoposti ad intervento. Poi con l'ausilio della variabile *max time pre oper* è stato calcolato il massimo valore registrato dalla clessidra.

In tabella 9 sono riportati i valori del massimo tempo di attesa registrati su 30 run nei due scenari.

Tabella 9: Confronto massimo tempo di attesa intervento nei due scenari (TO BE 1)

RUN	OUTPUT AS IS	OUTPUT TO BE 1
1	450	448
2	458	460
3	436	437
4	442	441
5	459	460
6	460	458
7	448	449
8	449	450
9	444	445

10	457	455
11	453	453
12	448	449
13	445	446
14	451	450
15	453	451
16	449	448
17	448	450
18	452	450
19	458	457
20	450	448
21	446	447
22	447	448
23	443	441
24	451	450
25	450	448
26	454	456
27	449	447
28	453	455
29	448	448
30	450	451

Dai valori riportati in tabella si evince che in configurazione TO BE 1 non si registrano variazioni del massimo tempo di attesa del paziente affinché possa essere sottoposto ad intervento chirurgico. Il valore medio che assume l'indicatore su 30 run è pari a 450 minuti per la configurazione parzialmente livellata delle risorse, mentre in configurazione agile tale valore è pari a circa 449.8 minuti.

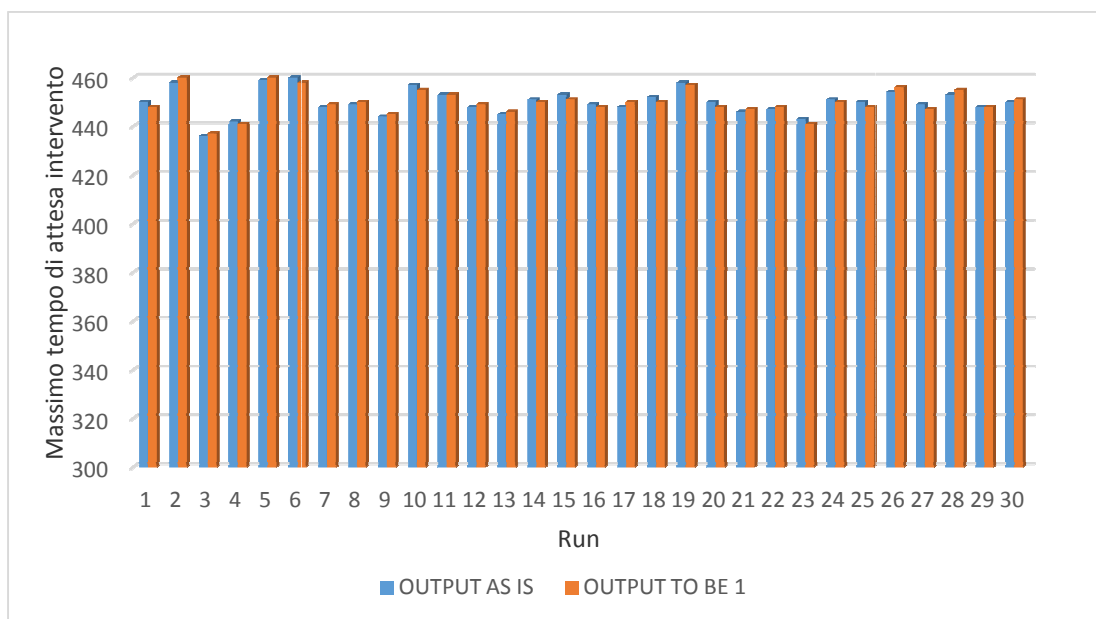


Figura 46: *Grafico dei max tempi di attesa per l'intervento registrati su 30 run nei 2 scenari (TO BE 1)*

Infine, per calcolare il tempo medio di refertazione, è stata utilizzata una variabile ausiliaria che ha consentito di fare la media dei valori registrati dalla clessidra, rappresentata da una variabile stock vettoriale, che conta i tempi di esecuzione dell'attività in esame.

I valori registrati nel corso dei 30 run nei due scenari sono riportati di seguito.

Tabella 10: *Confronto tra le medie dei tempi di refertazione nei due scenari (TO BE 1)*

RUN	OUTPUT AS IS	OUTPUT TO BE 1
1	36	29
2	38	29
3	35	27
4	37	29
5	38	28
6	34	26
7	39	28

8	41	33
9	37	30
10	38	27
11	36	29
12	37	26
13	38	26
14	37	27
15	40	29
16	39	28
17	38	28
18	39	28
19	36	27
20	37	30
21	38	28
22	37	29
23	36	27
24	37	30
25	38	28
26	35	28
27	37	29
28	36	28
29	35	28
30	37	30

Dalla tabella 10 è possibile vedere che in configurazione TO BE 1 si è registrata una diminuzione del tempo medio di refertazione.

Nel dettaglio, dai 30 run si evince che in configurazione AS IS il tempo medio di refertazione è pari a 37 minuti mentre invece in configurazione TO BE 1 assume valore pari a 28 minuti.

I valori registrati nei 30 run sono stati rappresentati mediante il grafico che segue per rendere facilmente visibile il miglioramento ottenuto.

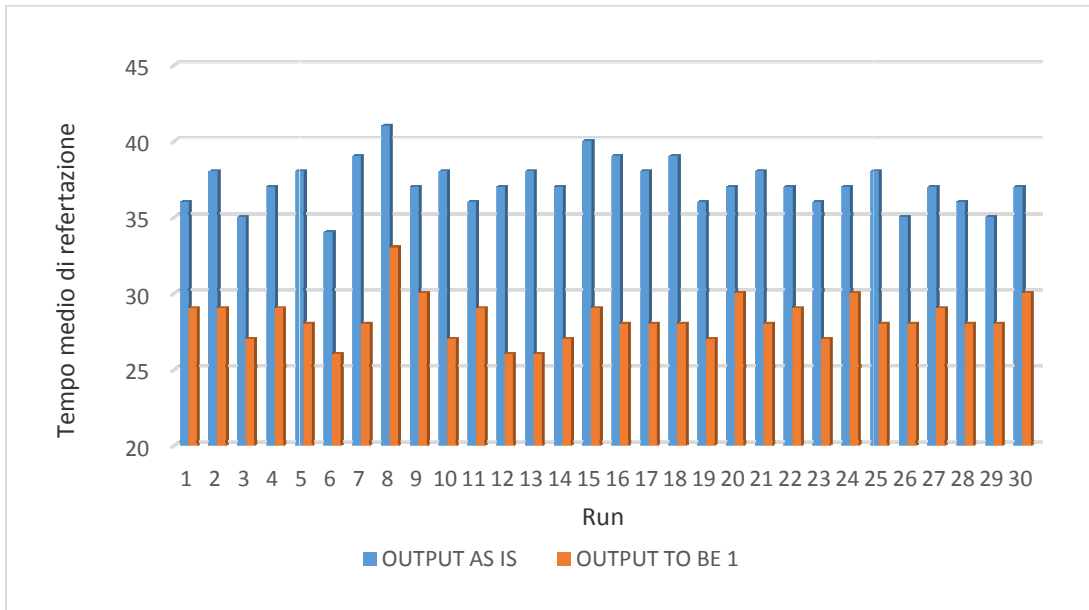


Figura 47: *Grafico dei tempi medi di refertazione registrati in 30 run nei due scenari (TO BE 1)*

Per cui con il primo scenario TO BE è stato possibile ridurre il tempo necessario al processo di refertazione ma non sono stati registrati miglioramenti sugli altri indicatori.

Come secondo TO BE si considera di apportare le modifiche su dette alla configurazione delle risorse del blocco operatorio.

In tabella 11 sono riportati i valori del coefficiente di riempimento dei posti letto ottenuti dai 30 run.

Tabella 11: *Valori del Coefficiente di riempimento dei posti letto nei due scenari (TO BE 2)*

RUN	OUTPUT AS IS	OUTPUT TO BE 2
1	116	107

2	118	108
3	121	106
4	118	106
5	120	105
6	116	107
7	115	108
8	118	107
9	117	106
10	118	109
11	116	108
12	121	106
13	117	106
14	118	107
15	119	105
16	118	106
17	119	107
18	120	105
19	119	108
20	117	107
21	116	108
22	118	106
23	115	107
24	116	107
25	117	109
26	119	107
27	117	106
28	118	107
29	119	105
30	120	105

La media ottenuta dai valori dei 30 run in configurazione AS IS delle risorse è pari a 117.867 mentre invece la media dei coefficienti di riempimento registrati in configurazione TO BE 2 è pari a 106.7.

Quindi con il miglioramento proposto è stato possibile ridurre il valore medio del coefficiente di riempimento di circa il 10.167 %.

È stato possibile dunque migliorare le prestazioni del sistema in termini di efficacia anche se però per massimizzare l'efficacia è necessario portare il coefficiente di riempimento al di sotto del 100%.

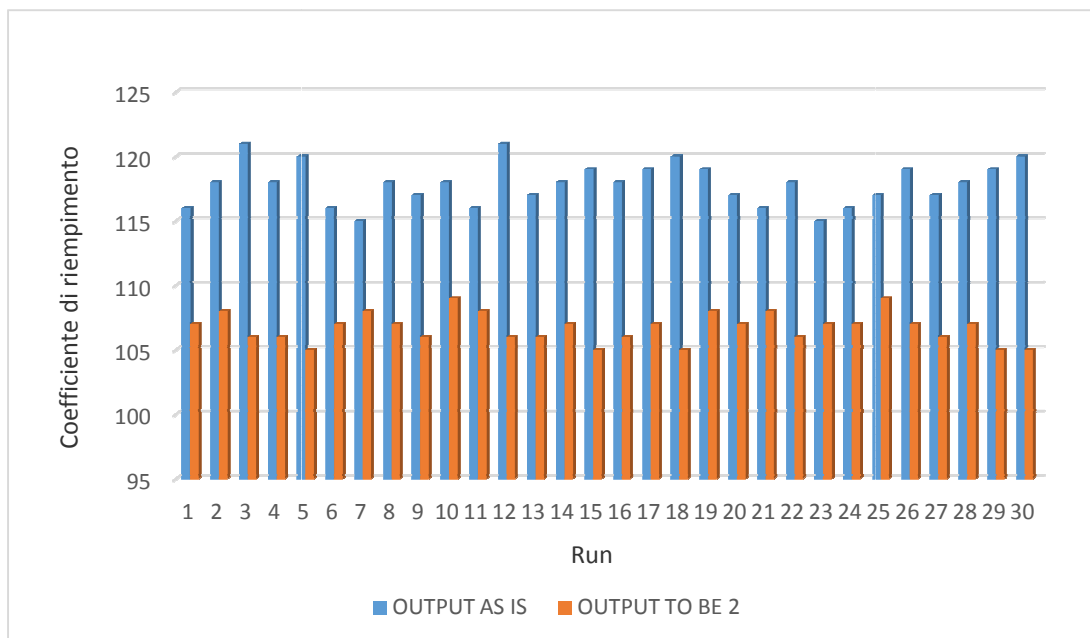


Figura 48: Confronto tra i valori del coefficiente di riempimento nei due scenari (TO BE 2)

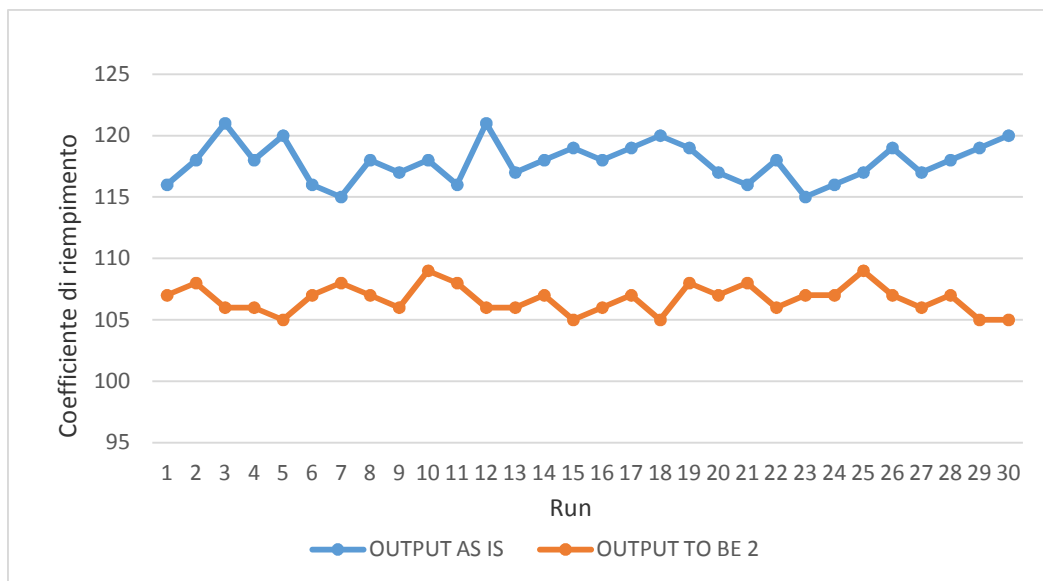


Figura 49: Andamento del coefficiente di riempimento per 30 run nei due scenari (TO BE 2)

Dalla figura 50 è possibile notare che la curva raffigurante l’andamento del coefficiente di riempimento dei posti letto nella configurazione TO BE 2 è nettamente al di sotto di quella registrata in configurazione AS IS.

Per valutare il tempo di permanenza media nella struttura ospedaliera in configurazione agile delle risorse è possibile osservare i valori registrati nella seguente tabella.

Tabella 12: Confronto tempi di permanenza media presso il DEA nei due scenari (TO BE 2)

RUN	OUTPUT AS IS	OUTPUT TO BE 2
1	7	6
2	6	5
3	7	6
4	6	4
5	5	5

6	6	5
7	7	6
8	5	5
9	6	5
10	6	6
11	5	6
12	6	4
13	7	6
14	6	6
15	6	5
16	5	5
17	7	5
18	6	4
19	8	6
20	6	6
21	6	5
22	5	5
23	5	4
24	7	6
25	6	5
26	6	6
27	7	5
28	5	5
29	7	5
30	6	6

Dalla tabella 12 si evince che, attraverso la configurazione agile delle risorse, i tempi di permanenza medi presso il DEA sono diminuiti. Tale risultato è in accordo con quello registrato dalla valutazione del coefficiente di

riempimento dei posti letto poiché appunto un tempo di permanenza minore comporta una diminuzione del coefficiente a seguito di dimissioni più frequenti.

La diminuzione dei valori del tempo di permanenza dei pazienti presso la struttura ospedaliera, ottenuta in configurazione agile delle risorse, è visibile mediante il grafico che segue.

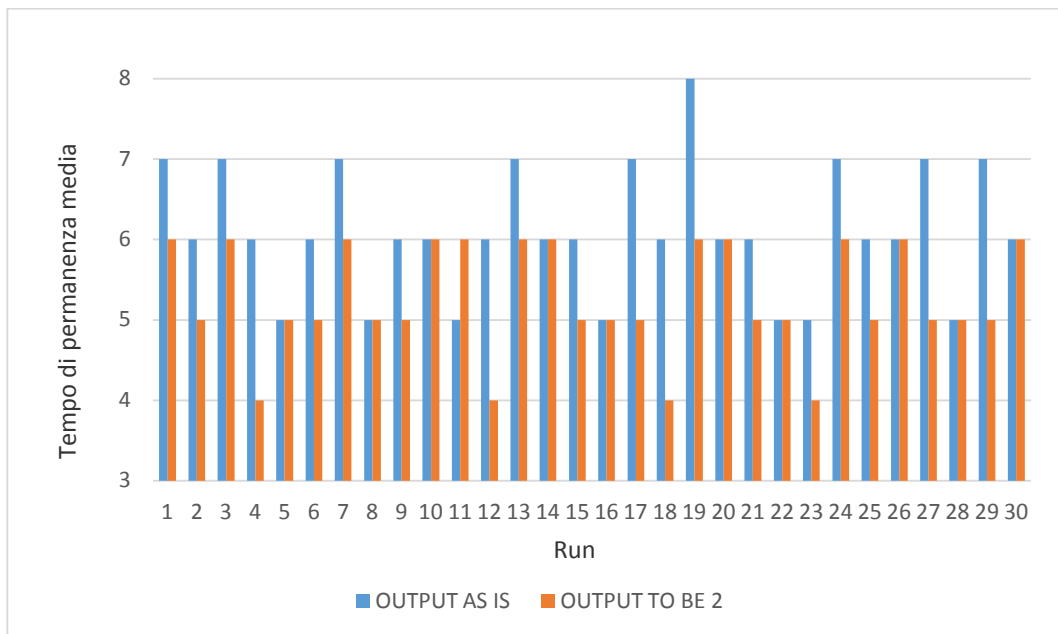


Figura 50: Grafico di confronto dei tempi di permanenza nei due scenari (TO BE 2)

La media dei valori registrati dal modello di simulazione in configurazione AS IS è circa 6 giorni mentre invece la media in configurazione TO BE 2 è pari a 5 giorni. Quindi si è ottenuta in media una diminuzione di 1 giorno di degenza

In tabella 13 sono riportati i valori del massimo tempo di attesa per essere sottoposti ad intervento chirurgico registrati su 30 run nei due scenari.

Tabella 13: Confronto massimo tempo di attesa intervento nei due scenari (TO BE 2)

RUN	OUTPUT AS IS	OUTPUT TO BE 2
1	450	356
2	458	359
3	436	357
4	442	350
5	459	358
6	460	349
7	448	354
8	449	360
9	444	357
10	457	352
11	453	356
12	448	354
13	445	353
14	451	348
15	453	357
16	449	346
17	448	352
18	452	354
19	458	357
20	450	351
21	446	356
22	447	350
23	443	357
24	451	354
25	450	352
26	454	357

27	449	359
28	453	354
29	448	350
30	450	352

Dai valori riportati in tabella 13 si evince che in configurazione TO BE 2 si registra una netta diminuzione del massimo tempo di attesa del paziente affinché possa essere sottoposto ad intervento chirurgico. Il valore medio che assume l'indicatore su 30 run è pari a 450 minuti per la configurazione parzialmente livellata delle risorse, mentre invece in configurazione agile tale valore diminuisce portandosi a circa 354 minuti.

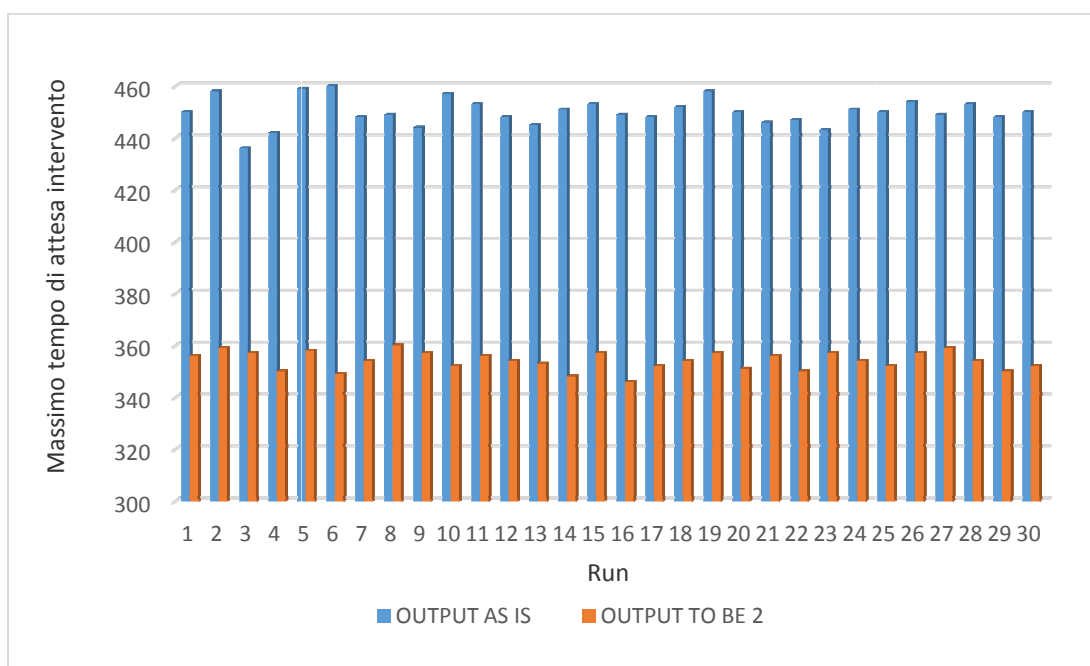


Figura 51: *Grafico dei max tempi di attesa per l'intervento registrati su 30 run nei 2 scenari (TO BE 2)*

Infine i valori registrati nel corso dei 30 run dell'indicatore tempo medio di refertazione nei due scenari sono riportati di seguito.

Tabella 14: Confronto tra le medie dei tempi di refertazione nei due scenari (TO BE 2)

RUN	OUTPUT AS IS	OUTPUT TO BE 2
1	36	35
2	38	39
3	35	34
4	37	37
5	38	38
6	34	35
7	39	38
8	41	40
9	37	38
10	38	38
11	36	37
12	37	36
13	38	38
14	37	37
15	40	39
16	39	40
17	38	38
18	39	40
19	36	35
20	37	37
21	38	38
22	37	36
23	36	37
24	37	37
25	38	37
26	35	36
27	37	38

28	36	35
29	35	35
30	37	37

Dalla tabella 14 è possibile vedere che in configurazione TO BE 2 non si registrata una variazione del tempo medio di refertazione.

Nel dettaglio, dai 30 run si evince che in configurazione AS IS il tempo medio di refertazione è pari a 37 minuti e tale valore si ottiene anche in configurazione TO BE 2.

I valori registrati nei 30 run sono stati rappresentati mediante il grafico che segue per rendere facilmente visibile che non è stata ottenuta alcuna variazione di tale indicatore.

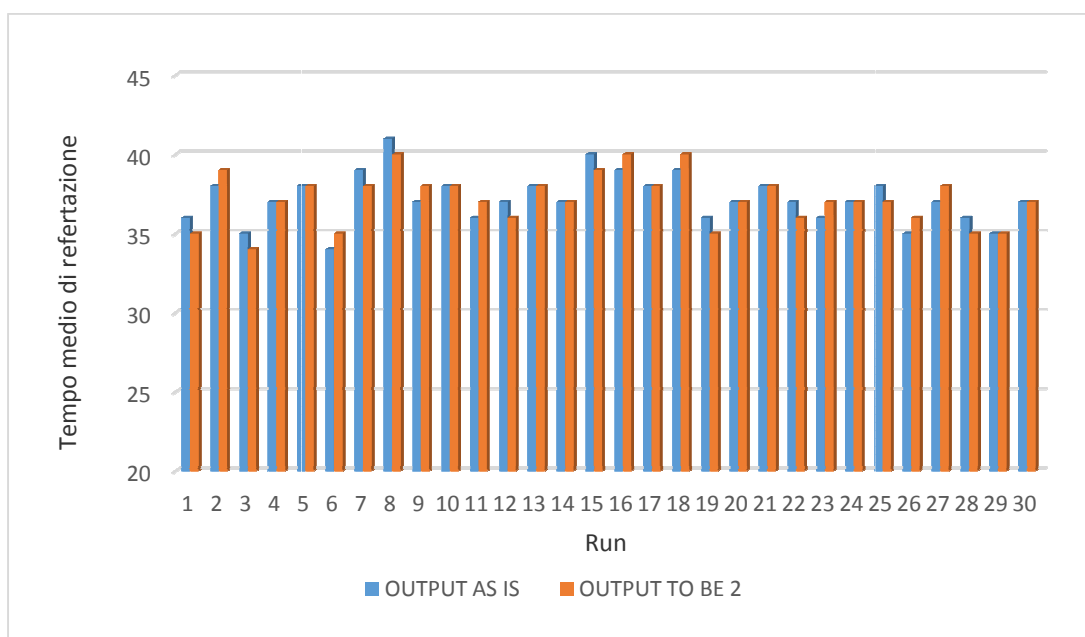


Figura 52: Grafico dei tempi medi di refertazione registrati in 30 run nei due scenari (TO BE 2)

Quindi applicando una configurazione agile delle risorse presso il blocco operatorio non si sono registrati miglioramenti per quando riguarda il tempo medio di refertazione ma è stato possibile riscontrare miglioramenti su tutti

gli altri indicatori, ottenendo quindi una parziale soluzione al problema del sovraffollamento delle strutture d'urgenza del DEA.

Un ulteriore miglioramento che si propone riguarda la gestione dei farmaci. Si pensa di introdurre il cosiddetto *armadietto di reparto* che consente, attraverso l'utilizzo della cartella clinica elettronica, di richiedere ogni giorno la quantità di farmaci strettamente necessaria alla stabilizzazione dei pazienti in degenza. In questo modo, tramite semplice consultazione, è possibile organizzare gli ordini, e quindi gli scarichi, con cadenza giornaliera, e non più settimanale, così da ridurre il numero di farmaci non utilizzati e ridurre la possibilità che i farmaci vadano in scadenza prima di poter essere utilizzati. Dunque è possibile minimizzare gli sprechi e quindi l'inefficienza derivante dalla gestione dei replenishments.

L'armadietto di reparto verifica automaticamente la presenza, nei singoli armadi distributori, di farmaci "fermi" (non prescritti o non utilizzati) da più di un determinato periodo o in eccesso rispetto al fabbisogno reale.

Il sistema garantisce la razionalizzazione delle risorse e permette di gestire informaticamente tutte le attività di prescrizione, somministrazione, approvvigionamento e conservazione dei farmaci in reparto, garantendo una maggiore sicurezza e la completa tracciabilità delle attività svolte. Inoltre gestisce le anagrafiche dei pazienti e le relative terapie con la possibilità di impostare dei profili terapeutici personalizzati. In qualsiasi momento il sistema conosce puntualmente la reale consistenza della giacenze nell'armadio automatico e con una certa precisione le quantità presenti sui carrelli delle terapie del reparto (dove rimangono presenti le confezioni parzialmente utilizzate) inoltre è a conoscenza dell'effettivo fabbisogno espresso in singole dosi per ciascun paziente.

Il sistema dispone della esatta conoscenza delle quantità conservate all'interno dell'armadio distributore automatico ed il fabbisogno di farmaci

nei giorni seguenti, pertanto è in grado di stabilire con precisione il quantitativo di farmaci da ordinare alla farmacia, giorno per giorno.

Il sistema inoltre traccia tutte le attività effettuate dagli operatori (inserimento e modifica di prescrizioni, sostituzioni di farmaci, somministrazioni) e questo garantisce la massima trasparenza delle attività e la tutela della professionalità degli operatori medesimi, infatti, in caso di errore si potrebbero accertare immediatamente le responsabilità. Inoltre la tracciatura puntuale delle attività consente di effettuare analisi dettagliate sulla qualità e sui costi della cura, studi comparativi sul tipo di terapia erogata per una medesima patologia, ecc.

10.7 - Sviluppi futuri

Nel seguente elaborato di tesi si è considerato un sistema di produzione monoprodotto, poiché non si è fatta distinzione tra i pazienti ma questi sono stati considerati indifferenziati, e multiprocesso.

In tal caso il ritmo produttivo assume configurazione vettoriale, in particolare sarà pari ad un vettore riga del tipo:

$$[\rho_1 \ . \ . \ . \ \rho_m]$$

È stato stabilito, attraverso l'ausilio della simulazione, quali fossero le attività collo di bottiglia, ovvero quelle attività in cui si registra una congestione delle richieste e in quali orari si presentano tali criticità.

Dopo di che si è cercato di rimuovere il collo di bottiglia attraverso una riconfigurazione delle risorse impiegate passando da una configurazione parzialmente stazionaria ad una configurazione agile delle risorse.

A questo punto è necessario ricavare una formulazione analitica, da confrontare con ciò che è stato fatto in via sperimentale, ovvero attraverso

simulazione, che consenta di capire la configurazione ottimale delle risorse da impiegare per bilanciare le accelerazioni registrate e quindi i picchi di domanda.

Pertanto è necessario calcolare l'aumento di produttiva del sistema, in termini di risorse, che consente di soddisfare le variazioni della domanda.

In precedenza è stata definita l'agilità, ovvero la variazione di ritmo produttivo, come:

$$\alpha(t) = \frac{dK}{dt}. \quad (10.7-1)$$

Tale variazione è dipendente da k.

Ma l'agilità è funzione dei fattori produttivi impiegati e quindi delle risorse per cui è possibile definirla anche indipendentemente da k e assume la seguente forma:

$$\alpha(t) = \alpha[\varphi(t)] \quad (10.7-2)$$

con $\varphi(t)$ definita come funzione dei fattori produttivi impiegati.

Il ritmo produttivo $\rho(t)$ è stato definito come:

$$\rho_A(t) = \rho_{\min}(t) + \Delta\rho(t) = \rho_L(t) + \Delta k(t) \quad (10.7-3)$$

ma può essere definito, alternativamente, come segue:

$$\rho_A(t) = \frac{P(t)}{\tau} \quad (10.7-4)$$

dove $P(t)$ è la produttività mentre τ è pari ad una costante temporale.

Derivando rispetto al tempo il ritmo produttivo si ottiene:

$$\rho(t)' = \frac{1}{\tau} \frac{dP(t)}{dt} \quad (10.7-5)$$

e tale quantità risulta essere pari alla derivata rispetto al tempo di k.

Quindi:

$$\rho(t)' = k(t)' = \frac{1}{\tau} \frac{dP(t)}{dt} \quad (10.7-6)$$

da cui si ricava che:

$$\tau \rho(t)' = \frac{dP(t)}{dt} \quad (10.7-7)$$

A questo punto si definisce la derivata rispetto al tempo della produttività come:

$$\frac{dP(t)}{dt} = \tau \psi \frac{d}{dt} \varphi(t) \quad (10.7-8)$$

con ψ caratteristica del sistema, variabile nel tempo, che assume però valore costante quando le componenti intrinseche del sistema non variano.

Sostituendo la (7) nella (8) si ricava che:

$$\tau \rho(t)' = \psi \frac{d\varphi(t)}{dt} \quad (10.7-9)$$

La derivata rispetto al tempo di $\rho(t)$ è stata precedentemente definita pari a α , pertanto si ha che:

$$\alpha = \frac{\psi}{\tau} \frac{d}{dt} \varphi(t) \quad (10.7-10)$$

ma essendo $\rho(t)' = k(t)'$ si ottiene che:

$$k(t)' = \varphi(t)' \frac{\psi}{\tau} \quad (10.7-11)$$

per cui

$$\varphi(t)' = k(t)' \frac{\tau}{\psi} \quad (10.7-12)$$

Passando all'integrale si ottiene:

$$\varphi(t) = \frac{\tau}{\psi} \int k(t)' dt + c \quad (10.7-13)$$

dove c rappresenta la costante di scorrimento.

Si conclude quindi che:

$$\varphi(t) = \frac{\tau}{\psi} k(t) + c \quad (10.7-14)$$

ovvero che i fattori produttivi impiegati dipendono da $k(t)$, dalla caratteristica del sistema ψ e dalla costante temporale τ nonché dalla costante di scorrimento c .

A questo punto, per via simulativa, essendo $k(t)$ noto, è stato definito $\varphi(t)$ ovvero il valore dei fattori produttivi impiegati, quindi il numero di risorse operanti. Essendo noti questi due valori, sperimentalmente è possibile ricavare il valore della caratteristica del sistema.

Per via analitica invece, supponendo noti $k(t)$ e ψ , è possibile calcolare $\varphi(t)$ ovvero il dimensionamento delle risorse che consente di compensare le variazioni della domanda.

Pertanto una volta calcolato analiticamente $\varphi(t)$, è necessario verificare se tale valore coincide con quello ricavato sperimentalmente attraverso il modello di simulazione.

È possibile fare tutto ciò nell'ipotesi in cui il sistema di produzione abbia configurazione Agile.

Nel caso in esame il confronto può essere fatto quando si riesce ad operare una diminuzione del coefficiente di riempimento dei posti letto tale da portarlo ad un valore almeno pari al 100%, quindi in condizione di perfetta efficacia del servizio.

CONCLUSIONI

Il corso di Dottorato nell'ambito del quale è stato sviluppato il lavoro di tesi qui presentato, è stato caratterizzato da un'attività di ricerca, i cui risultati assumono contorni di apprezzabile interesse, soprattutto in relazione alle conclusioni raggiunte in parziale controtendenza con gli orientamenti attualmente rilevabili sia nel mondo industriale, sia (sebbene più limitatamente) nella comunità scientifica internazionale (come mostrato nella corposa *literature review* riportata nella Parte Seconda).

Il lavoro trae origine dall'obiettivo di applicazione sistemica (e, quindi di integrazione) dei più avanzati modelli di gestione (previsti in materia di produzione di beni e servizi), alle risorse di natura tecnologica ovvero umana, che caratterizzano gli ambiti della sanità, al fine di generare soluzioni operative ad elevata performance.

Alla richiamata ricerca bibliografica, ha fatto seguito una successiva applicazione nella raccolta di dati ed informazioni sul campo, al fine di acquisire gli elementi necessari alla modellazione dei sistemi di suo interesse ai fini dello studio.

In tal senso ha assunto significativa rilevanza, la scelta della logica di modellazione, nonché l'ideazione dei costrutti e delle metodiche di realizzazione degli esperimenti su laboratorio simulato.

Non essendo possibile, per ovvie motivazioni, studiare "in vivo" la bontà e l'evoluzione di talune scelte, si è reso necessario procedere immaginando una logica di analisi efficace per valutare l'interazione sistemica dei fattori, che concorrono alla produzione dei servizi in sanità.

In tale ambito è stata messa a punto una logica simulativa ibrida, in grado di far coesistere gli elementi ad integrazione continua, con i fenomeni di natura discreta che caratterizzano l'evoluzione sistemica di un processo di erogazione complessa.

In modo specifico, si è ricorsi all'implementazione di modelli operanti in logica dinamica (System Dynamics), i cui loop iterativi fossero modellizzabili mediante un insieme di strutture "evolutive" (i processi sistemici, appunto), in grado di elaborare una materia "evolvente" opportunamente discretizzata in funzione dei valori assunti, tempo per tempo, dalle proprietà che le caratterizzano.

La logica originale, fondata su formalismo matriciale, che ne è derivata, ha consentito la realizzazione di un laboratorio di simulazione, appositamente costruito per i sistemi sanitari ad emergenza/urgenza, che costituisce il primo risultato della ricerca: le pubblicazioni scientifiche inerenti i reparti ospedalieri, costruite su strumento simulato, si fermano – infatti – all'analisi di un unico reparto, raramente interessandosi della sua interazione con altro sistema ospedaliero.

L'aver rivolto la propria attenzione e l'aver realizzato un *tool* di analisi per un intero DEA (che per la visione globale in materia nosocomiale, costituisce un ospedale nell'ospedale, quindi un'entità impiantistica autonoma), rappresenta un risultato già di per sé concreto e innovativo.

A margine delle considerazioni metodologiche, fin qui esposte, è opportuno evidenziare come la simulazione non rappresenti l'unico approccio di ricerca operativa, a cui si è puntualmente ricorsi; le problematiche di valutazione-rischi sono state, ad esempio, risolte mediante applicazione markoviana in grado di offrire uno strumento versatile per l'analisi di sensitività delle grandezze coinvolte nel fenomeno, più che di restituire una valorizzazione del rischio espressa mediante misurazione numerica classica (frequenza per magnitudo),

Risolte, nel senso esposto, le problematiche strumentali del percorso di studio, la ricerca ha maturato interessanti risultati anche in termini ontologici, relativamente alla reale manifestazione dei fenomeni gestionali, che emergono dalle differenti configurazioni ingegneristiche, cui è possibile ricorrere per la tipologia di impianto produttivo analizzato.

Anzitutto si è definito il concetto di sistema produttivo snello di tipo Agile, proponendone un modello analitico innovativo, ontologicamente duale al ben più noto e codificato sistema Lean.

Successivamente si è valutata la configurazione più opportuna del sistema produttivo in funzione della domanda, mediante la quale viene sollecitato l'impianto di produzione (nel caso trattato, i reparti di un DEA, sebbene - di principio - il paradigma valga per ogni sistema produttivo).

Quindi sono stati lanciati i run di simulazione, si è validato il modello virtuale e sono state interpretate e valutate le evidenze sperimentali, a cui si è giunti, sia in configurazione "as is", sia a seguito del reengineering dell'impianto.

La prima conclusione di rilievo consiste nella capacità dei sistemi di produzione Agile di trasformare onde di sollecitazione del sistema (andamento della domanda nel tempo) non livellate, in prestazioni di maggior regolarità e performance.

Pertanto, nell'ipotesi in cui ci si trovi ad operare con vincoli di irregolarità della sollecitazione produttiva dell'impianto e necessità di performance stabilizzata (tipico caso della emergenza/urgenza, ma anche di larga parte del restante mondo produttivo), si conclude (in controtendenza rispetto al credo industriale corrente) che un sistema Agile, risponde più efficacemente di un equivalente (per risorse) sistema Lean, pur permanendo la produzione (ovvero l'erogazione di un processo) in ambiente snello e governata da una logica produttiva di tipo pull.

Ulteriore elemento acquisito dallo studio, è consistito nella codifica di questo risultato (valido ben oltre l'architettura del modello di simulazione) mediante procedura operativa di reengineering delle risorse, nonché mediante proposizione di un approccio analitico in grado (in linea di principio) di bilanciare queste ultime in funzione dei valori caratteristici degli "arrivi poissoniani", variabili nel tempo, che contraddistinguono la sollecitazione.

In ultimo da un punto di vista applicativo, in ottica ospedaliera, si sancisce che ai fini della performance delle unità di urgenza, risultano critici i reparti “di produzione” medica e non quelli “di supporto” diagnostico. Le unità di emergenza, invece, (in controdeduzione rispetto a quanto ci aspetteremmo) vedono quali risorse critiche ai fini delle proprie performance, i reparti “di supporto” diagnostico e non quelli “di produzione” medica. Ciò induce, in configurazione Agile, ad un differente bilanciamento dei fattori della produzione, sia tecnologici che umani, all’interno di uno stesso DEA.

BIBLIOGRAFIA

- [1] (2005), “*LEAN MANUFACTURING IN TOYOTA: Un’evoluzione continua*“, Convegno Unindustria Treviso
- [2] (2011), “*Così «Lean» taglia gli sprechi*”, Il sole 24 ore
- [3] Agarwal A. - Shankar R. - Tiwari M.K, (2007), “*Modeling agility of supply chain*”
- [4] Aitken J.M. - Christopher M., (2002), “*Understanding, Implementing and Exploiting Agility and Leanness*”, International Journal of Logistics: Research and Applications
- [5] Antony J., (2011), “*Reflective Practice Six Sigma vs Lean: some perspectives from leading academics and practitioners*”, International Journal of Productivity and Performance Management
- [6] Bhasin S., (2008), “*Lean and performance measurement*”, Journal of Manufacturing Technology Management
- [7] Bhim S. - Garg S.K. - Sharma S.K., (2010), “*Lean implementation and its benefits to production industry*”, International Journal of Lean Six Sigma
- [8] Biroli M., (1992), “*Process analysis o process management*” in *Sistemi e Impresa*
- [9] Biselli D., (2013), “*L’organizzazione del Sistema Sanitario Italiano*”
- [10] Borgonovi E., (2000), “*Gli indicatori di efficacia e di efficienza nella sanità*”, Rivista trimestrale di scienza dell’amministrazione
- [11] Bottani E., (2010),” *Profile and enablers of agile companies: An empirical investigation*”, International Journal of Production Economics
- [12] Bovenzi F.M., (2011),” *Ospedale per intensità di cura: scegliere leggero, un compito pesante*”, pagine 21-22
- [13] Brailsford S.C, (2008), “*System Dynamics: What’s in it for healthcare simulation modelers*” Winter Simulation Conference
- [14] Braunscheidel M. - Suresh N., (2009), “*The organizational antecedents of a firm’s supply chain agility for risk mitigation and response*”, Journal of Operations Management
- [15] Bruzzi S., (1997), “*Finanziamento e gestione delle aziende ospedaliere*”, Giuffrè Editore
- [16] Busca C.-Giovannoni E.- Riccaboni A., (2009), “*Controllo di gestione*”, IPSOA
- [17] Cherubini, (2000), “*Il Marketing nei Servizi. Per lo Sviluppo Competitivo e la Customer Satisfaction*”, FrancoAngeli Editore
- [18] Christopher M., (2000),” *The Agile Supply Chain: Competing in Volatile Markets*”

- [19] Coyle R.G., (1996), *“System Dynamics modelling: a practical approach”*, CRC Press
- [20] D.M.Berwick, (2003), *“Improvement, trust, and and the healthcare workforce”*, Qual Saf Health Care
- [21] De Risi P., (2002), *“Introduzione alla gestione per processi nelle organizzazioni”*, Università degli studi di Pisa
- [22] Debra A. Elkins, Ningjian Huang, Jeffrey M. Alden, (2004), *“Agile manufacturing systems in the automotive industry”*, Int. J. Production Economics 91, pagine 201-214
- [23] *“Definizione degli standard qualitative, strutturali, tecnologici e quantitativi relativi all’assistenza ospedaliera”*, Ministero della Salute
- [24] Dombrowski U. - Mielke T., (2013), *“Lean Leadership - fundamental principles and their application”*, Conference on Manufacturing Systems
- [25] Donia Sofio A., (2000), *“Microeconomia sanitaria e politiche di intervento”*
- [26] Fillingham D., (2011), *“Lean healthcare. Trasformare la sanità a partire dall’esperienza del paziente”*, FrancoAngeli Editore
- [27] Fiorani G., (2009), *“System thinking, System dynamics e politiche pubbliche”*, Egea
- [28] Fliedner G. – Vokurka R., (1997), *“Agility: The Next Competitive Weapon”*
- [29] Frenk J., (2010), *“The Global Health System: strengthening national health systems as the next step for global progress “*
- [30] Galgano A., (2006), *“IL SISTEMA TOYOTA NELLA SANITA’: più Qualità meno Sprechi”*
- [31] Giusepi I., (2012), *“La gestione operativa in sanità - Operations management”*
- [32] Gligor D.M. - Holcomb M.C, (2013), *“A Multidisciplinary Approach to Supply Chain Agility: Conceptualization and Scale Development”*
- [33] Goldman S.L.-Nagel R.N., (1993), *“Management, Technology, and Agility: The Emergence of a New Era in Manufacturing”*, Interscience Enterprises
- [34] Goldman, S. L. - Nagel R.N.- Preiss K., (1995), *“ Agile Competitors and Virtual Organizations: Strategies for Enriching the Customer”*
- [35] Gunasekaran A., *“Design and implementation of agile manufacturing systems”*
- [36] Hallgren M.- Jan Olhager, (2009), *“Lean and agile manufacturing: external and internal drivers and performance outcomes”*, International Journal of Operations & Production Management
- [37] Homer J.B - Hirsch G.B, *“System dynamics modelling for Public Health: Background and Opportunities “*
- [38] Imai Masaaki, (1997), *“Kaizen -Lo Spirito Giapponese del Miglioramento,”* Il Sole 24 Ore Libri

- [39] Inman R.A- Sale R.S- Kenneth W. Green Jr- Dwayne Whitten, (2011), “*Agile manufacturing: Relation to JIT, operational performance and firm performance*”
- [40] Joosten T. – Bongers I. – Janssen R., (2009), “*Application of Lean Thinking to HealthCare: issues and observation*”, International Journal for Quality in HealthCare
- [41] Juan Martín García, (2006), “*Theory and Practical Exercises of System Dynamics*”
- [42] Karlsson A.- Persson T., (1998), “*Powersim*”
- [43] Kidd P.T., (1994),” *Agile Manufacturing: Forging New Frontiers* “
- [44] Liker J.- Hoseus M.,(2010), “ *Human resource development in Totota Culture*”, International Journal of Human Resources Development and Management
- [45] Liker J.K, (2004),”*The Toyota Way: 14 Management Principles from the World’s Greatest Manufacturer*”, McGraw-Hill
- [46] Martin Christopher “*The Agile Supply Chain: Competing in Volatile Markets*”
- [47] Mason Jones R., (2000), “*Lean, Agile or Leagile? Matching your supply chain to the marketplace*”, International Journal of Production Research
- [48] Mills J.-Platts K.-Bourne M.-Richards H., (2002), “*Strategy and Performance: competing through competences*”
- [49] Mitton C.- Donaldson C., (2006), “*Come stabilire le priorità in sanità*”, Il Pensiero Scientifico Editore
- [50] Monden, Y., (1983),” *Toyota Production System: A Practical Approach to Production Management. Industrial Engineers and Management*”
- [51] Moura D.A.-Botter R.C., (2012), “*Can a shipyard work towards lean shipbuilding or agile manufacturing?*”, Rizzuto & Guedes Soares (eds)
- [52] Moyano Fuentes J.,(2012), “*Learning on lean: a review a thinking and research*”, International Journal of Operation and Production Management
- [53] Munno A., “*APPROFONDIMENTI DI ECONOMIA SANITARIA: Efficienza ed equità in ambito sanitario*”
- [54] Nicosia F., “*Il nuovo ospedale è snello*”, FrancoAngeli Editore
- [55] Ohno T., (1988), “*Toyota Production System: Beyond Large-Scale Production*”
- [56] Pammolli F. - Salerno N.C, (2010),”*Alla ricerca di standard per la sanità federalista*”
- [57] Pammolli F. – Salerno N.C., (2009), ”*La sanità in Italia - dentro la 'scatola nera' delle differenze regionali*”
- [58] Quinn R.D.-Causey G.C.-Merat F.L.-Sargent D.M.-Barendt N.A-Newman W.S.-Velasco V.B-Podgurski A.-Sterling L.S., (1997), “*An Agile Manufacturing workcell design*”

- [59] Santillo L.C., Dispensa del Corso di Gestione della Produzione Industriale, A.A 2008/2009
- [60] Santillo L.C., Dispensa del Corso di Sistemi di Produzione di Beni e Servizi, A.A. 2008/2009
- [61] Schonberger, R.J., (1986), *“World Class Manufacturing: The lessons of simplicity applied”*
- [62] Serpelloni G.- Simeoni E.- Gomma M., *“LA DEFINIZIONE E LA RAPPRESENTAZIONE DEI PROCESSI: PRINCIPI DI BUSINESS PROCESS REENGINEERING (BPR)”*
- [63] Shah R. - Ward P.T., (2007), *“Defining and developing measures of lean production”*, Journal of Operations Management
- [64] Shingo S. (1987), *“Il sistema di produzione giapponese Toyota”*, Milano: Franco Angeli
- [65] Shook J.- Marchwinski C.,(2014), *“Lean Lexicon: A Graphical Glossary for Lean Thinkers”*
- [66] Swafford P.M. - Ghosh S.- Murthy N., (2008), *“Achieving Supply Chain Agility through it Integration and Flexibility”*
- [67] Tom Rust, Khalid Saeed, Isa Bar-On, Oleg Pavlov, *“Adapting Agile Strategies to Healthcare Service Delivery”*
- [68] Van Hock R.I. – Harrison A. - Christopher M., (2001), *“Measuring agile capabilities in the supply chain”*, International Journal of Operation and Production Management
- [69] Vignati E.- Bruno P., (2003), *“Organizzazione per processi in sanità”*, Franco Angeli Editore
- [70] Womack J. - Jones D, (1990), *“The Machine That Changed the World: The Story of Lean Production”*, Maxwell Macmillan International
- [71] Womack J. P.- J. D.,(2006), *“Lean Thinking. Come creare valore e bandire gli sprechi”*, Milano: Guerini e Associati
- [72] Wul P.J- Feng C.M- Chen Y.F, (2011), *“Adaptable distribution management for agile supply chains”*
- [73] Yauch C.A., (2007), *“Team-based work and system balance in the context of agile manufacturing”*
- [74] Yusuf Y.Y.-Sarhadi M. - Gunasekaran A., (1999), *“Agile manufacturing: The drivers, concepts and attributes”*
- [75] Yusuf, (1999), *“Agile manufacturing: The drivers, concepts and attributes”*
- [76] E. Vignati, P. Bruno, *Organizzazione per processi in sanità*, Franco Angeli, 2003.
- [77] Radnor, Zoe J., Matthias Holweg, and Justin Waring. "Lean in healthcare: the unfilled promise?." *Social Science & Medicine* 74.3 (2012): 364-371.

- [78] Smith, Mark S., and Craig F. Feied. "The next-generation emergency department." *Annals of emergency medicine* 32.1 (1998): 65-74.
- [79] G. Casati, E. M. (2006, Aprile). La gestione dei processi clinico assistenziali. *Caleidoscopio italiano*.
- [80] Pierantozzi, D. (s.d.). *La gestione dei processi nell'ottica del valore*. EGEA.
- [81] Gori Francesca (s.d.). *La gestione per processi in ambito sanitario*.
- [82] SIMEU-FIMEUC. (2011). Standard organizzativi delle Strutture di Emergenza-Urgenza.
- [83] Lazzi, Gabriele. "Reingegnerizzazione dei processi." (1999).
- [84] Raimondo, Carmen. "*Lean Innovation nelle Aziende sanitarie. Un approccio per ridurre gli sprechi e migliorare la qualità.*"
- [85] Nicosia, Francesco, R. Tramalloni, and A. Lagostena. "Ospedale "lean" per intensità di cure." *Management della sanità* 12 (2008): 36-40.
- [86] Bartezzaghi, Emilio. *L'organizzazione dell'impresa: Processi, progetti, conoscenza, persone*. Etas, 2014.
- [87] Pierantozzi, David. *La gestione dei processi nell'ottica del valore: miglioramento graduale e reengineering: criteri, metodi, esperienze*. Egea, 1998.
- [88] Muntlin, Åsa, Lena Gunningberg, and Marianne Carlsson. "Patients' perceptions of quality of care at an emergency department and identification of areas for quality improvement." *Journal of clinical nursing* 15.8 (2006): 1045-1056.
- [89] Fernandes, Christopher, and James M. Christenson. "Use of continuous quality improvement to facilitate patient flow through the triage and fast-track areas of an emergency department." *The Journal of emergency medicine* 13.6 (1995): 847-855.
- [90] Andersson H. , Wireklint Sundström B., Nilsson K., Jakobsson Ung E "Management of everyday work in Emergency Departments – [An exploratory study with Swedish Managers](#)" *International Emergency Nursing, Volume 22, Issue 4, October 2014, Pages 190-196*
- [91] Hoot, Nathan R., and Dominik Aronsky. "Systematic review of emergency department crowding: causes, effects, and solutions." *Annals of emergency medicine* 52.2 (2008): 126-136.
- [92] Fernandes, Christopher, Ann Price, and James M. Christenson. "Does reduced length of stay decrease the number of emergency department patients who leave without seeing a physician?." *The Journal of emergency medicine* 15.3 (1997): 397-399.
- [93] Chan, Theodore C., et al. "Impact of rapid entry and accelerated care at triage on reducing emergency department patient wait times, lengths of stay, and rate of left without being seen." *Annals of emergency medicine* 46.6 (2005): 491-497.

- [94] Weiss, Steven J., et al. "Relationship between the National ED Overcrowding Scale and the number of patients who leave without being seen in an academic ED." *The American journal of emergency medicine* 23.3 (2005): 288-294.
- [95] Johnson, Michele, et al. "Patients who leave the emergency department without being seen." *Journal of Emergency Nursing* 35.2 (2009): 105-108.
- [96] Bambi, Stefano, et al. "Characteristics of patients who leave the ED triage area without being seen by a doctor: a descriptive study in an urban level II Italian University Hospital." *Journal of Emergency Nursing* 37.4 (2011): 334-340.
- [97] Pérez, María Martínez, et al. "Application of RFID technology in patient tracking and medication traceability in emergency care." *Journal of medical systems* 36.6 (2012): 3983-3993.
- [98] Chowdhury, Belal, and Rajiv Khosla. "RFID-based hospital real-time patient management system." *Computer and Information Science, 2007. ICIS 2007. 6th IEEE/ACIS International Conference on. IEEE, 2007.*
- [99] Dobson, Ian, Quynh Doan, and Geoffrey Hung. "A Systematic Review of Patient Tracking Systems for Use in the Pediatric Emergency Department." *The Journal of emergency medicine* 44.1 (2013): 242-248.
- [100] Vest, Nola, Nancy K. Rudge, and Gwen Holder. "ED whiteboard: an electronic patient tracking and communication system." *Journal of Emergency Nursing* 32.1 (2006): 8.
- [101] Amber, Richard, and Valerie B. Everett. "Emergency department patient tracking: a cost-effective system using bar code technology." *Journal of Emergency Nursing* 22.3 (1996): 190-195.
- [102] Sonnenberg, Frank A., and J. Robert Beck. "Markov models in medical decision making a practical guide." *Medical decision making* 13.4 (1993): 322-338.
- [103] Reinhard, Jean-Marie. "On a class of semi-Markov risk models obtained as classical risk models in a Markovian environment." *Astin Bulletin* 14.01 (1984): 23-43.