DEVELOPMENT AND APPLICATION OF LEAN RAILROADING TO IMPROVE CLASSIFICATION TERMINAL PERFORMANCE

BY

JEREMIAH R. DIRNBERGER

B.S., Montana State University – Bozeman, 2003

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master's of Science in Civil Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2006

Urbana, Illinois

ABSTRACT

DEVELOPMENT AND APPLICATION OF LEAN RAILROADING TO IMPROVE CLASSIFICATION TERMINAL PERFORMANCE

Jeremiah R. Dirnberger, M.S.
Department of Civil and Environmental Engineering
University of Illinois at Urbana-Champaign
Christopher P. L. Barkan, Ph.D., Advisor

Although much attention has been focused on the growth of intermodal traffic over the past decade, manifest freight or carload traffic is a major revenue generator for railroads. In 2004, the top four commodity categories in terms of revenue per carload averaged $2,971 per carload compared to an average of $770 for intermodal carloads. The high potential profitability of carload traffic suggests that railroads should try to further grow this segment of traffic, especially in an era of limited railway capacity. To do this, they must meet the increasing logistical needs of their customers by providing more consistent, reliable service than has previously been the case for carload traffic. The classification terminal is a key determinant in the service reliability of manifest freight.

Terminal performance also impacts network efficiency. A regression analysis for six Class I railroads showed a statistically significant relationship between average system-wide terminal dwell and average manifest train speed. As average dwell time increased, average manifest train speed decreased. The R^2 values ranged from 20.7% to 34.7%. With demand for rail transportation service expected to continue to grow, increasing network efficiency will be vital to reliably handle the traffic.

Inadequate terminal capacity is viewed by many as a barrier to improved service reliability and network efficiency. Because railroad classification terminals can be considered

production systems, insight into the dynamics of a terminal system can be gained by adapting production management tools that have led to significant performance improvement in manufacturing. This work focused on utilizing the concepts of factory physics, lean, Theory of Constraints (TOC) and Statistical Process Control (SPC) as part of a new approach to improving terminal performance introduced here as "Lean Railroading." By combining scheduled railroading with a version of lean in their yards, Canadian Pacific Railway (CPR) reported average terminal dwell fell from 30.4 hours in March 2005 to 21.7 hours in March 2006 while average train speed rose 3.6 mph over the same period.

The most important manufacturing process analog to improving terminal capacity is the bottleneck. Improving the performance of the bottleneck is the best way to improve the performance of the entire terminal process. The train assembly (pull-down) process has been identified as the bottleneck in a majority of classification yards. The macroscopic evaluation method for determining pull-down capacity from Wong et al. (1981) is enhanced with additional equations and each component of the pull-down process is discussed in detail. The potential capacity improvement of several bottleneck management alternatives is discussed using Bensenville Yard (CPR) as the example. The alternatives increased estimated capacity between 2.6% to 8.3% over the baseline case of 541 cars per day.

For Bensenville, pull-down capacity can be improved without large labor or capital expenses through better management of the process and its interactions with the system. This also means that multiple improvement options can be accomplished together resulting in a capacity increase of 25% from the baseline case. One of the principal findings of this work is that the humping process should be subordinate to the pull-down process because the latter is the principal bottleneck in many yards. The hump should be managed and operated so that it

provides the bottleneck exactly what it needs when it needs it. If a track is "dirty," the pull-down process will have to conduct additional work to remove the out-of-place cars. This reduces capacity and can be eliminated through better sorting of the cars at the hump. A quality of sort metric called the Bowl Condition Monitor (BCM) is developed to monitor this interaction.

The BCM is designed to fill the current gap in most terminal control systems regarding the condition of the bowl and the performance of the pull-down process. It consists of two components, the Dirty Track Counter (DTC) and the Incorrect Sort Rating (ISR). The DTC is used to determine the expected capacity reduction for a given number of dirty tracks. The ISR is used to measure adherence to a static track allocation plan if one is in place. It is shown that as bowl volume increases, the number of incorrectly sorted cars as measured by the ISR also increases. It also appears this effect is greater at higher volumes. A method for implementing the ISR with front-line managers and operators using modified SPC control charts is presented.

To Becky,

"This world was never meant for one as beautiful as you."

ACKNOWLEDGEMENTS

First, I would like to thank all of my friends and colleagues for their sympathy and support for my family and I this past year. Above all, I would not have been able to complete this thesis without the love and unwavering support from my soul mate, Lisa. Our incredible journey has only just begun. Of course, I wouldn't be here without the support of my entire family. To my mom, thanks for always encouraging me to dream big and being there whenever my dreams needed some reinforcement. To my dad, thank you for teaching me the value of hard work and dedication. To Jessica, Nathan and Gwen, thanks for believing in your big brother. To Ben, thank you for being my mom's soul mate. And, to my late sister Becky, thank you for making the world a better place. You will always be with us.

During my graduate studies I am very fortunate to have been supported by a CN Fellowship in Railroad Engineering. This research would not have been possible without this support. This research also benefited greatly from the professional support of many people in the railroad industry. From CN, I would like to thank Mack Barker, Mike Meleskie, the MacMillan Yard team and Merle Metz. From CPR, I want to thank Jeff Adams, Rob Weir, Dan James, Michelle Jacobs and the rest of the YOP group, as well as all of the managers and crews at Alyth, Agincourt and Bensenville for making me feel so welcome. You're insights and experience greatly enhanced this work. Thank you to NS, BNSF and UP for letting me visit their yard facilities. I also want to thank a fellow "Lean Railroader," Prescott Logan of GE Yard Solutions, for taking the time to share his knowledge on this new approach.

All of my professors at Montana State University deserve recognition for giving me the Industrial Engineering skills that allowed me to complete this work. In particular, Dr. Sobek, who, in my opinion, is one of the best teachers in the field. At the University of Illinois, thank you to my undergrad assistants, Darwin Schafer, Ahsan Alvi and Justin Wood for all of your hard work. To my fellow railroad engineering graduate students, I am fortunate to know you both as great scholars and friends. And finally, to my advisor, Dr. Chris Barkan, thank you for giving me the opportunity and helping me to grow as a researcher and a railroader. I am also very fortunate to call you my friend.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: THE CLASSIFICATION TERMINAL AS PART OF THE RAILROAD NETWORK

Manifest (or carload) traffic has traditionally been one of the major sources of revenue for Class I railroads. Rail traffic consists of many commodity categories that I have grouped into three revenue tiers (Table 1.1). Traffic from categories in the top tier moves almost entirely in manifest trains. The same is true of traffic in the middle tier with the exception of grain, motor vehicles and other farm products, which tend to be moved in unit trains although some is also handled in manifest trains. The bottom tier includes coal, gravel and other bulk commodities as well as “All other” which is the principal category for intermodal traffic. In spite of the low average for intermodal traffic ($770 per carload), the industry has placed substantial focus on that sector due to its substantial growth (ca. 40%) in the past ten years (AAR 2006a).

Table 1.1 U.S. Class I revenue per carload 2004 (from AAR 2005)

AAR Line	Traffic Category	Average Revenue per Carload
586.	Lumber and Wood Products	$4,192
587.	Pulp, Paper and Allied Products	$2,587
588.	Chemicals and Allied Products	$2,567
589.	Petroleum Products	$2,537
	Top Tier	$2,971
590.	Stone, Clay and Glass Products	$2,225
592.	Metals and Products	$2,119
577.	Grain (Including Soybeans)	$2,098
593.	Motor Vehicles and Equipment	$2,094
584.	Food and Kindred Products	$2,075
578.	Other Farm Products	$2,027
583.	Grain Mill Products	$1,850
	Middle Tier	$2,070
594.	Waste and Scrap Materials	$1,319
591.	Coke	$1,264
580.	Coal	$1,185
595.	Forwarder and Shipper Association	$1,083
585.	Primary Forest Products	$992
579.	Metallic Ores	$935
582.	Non-Metallic Minerals	$869
596.	All Other	$770
581.	Crushed Stone, Gravel and Sand	$758
	Bottom Tier	$1,020

However, growth in carload traffic is also important, particularly in light of its profitability. Norfolk Southern (NS) recently reported that 43% of its rail revenue is derived from carload shipments and 18 of its top 50 customers are carload shippers (Bragman 2006). The Union Pacific (UP) reports that manifest business already accounts for 41 to 45% of revenue and it is their fastest growing business line (Gray 2006). A key operational distinction between traffic that runs in unit trains versus traffic in manifest trains is that the later will frequently visit several terminals between origin and destination. Service quality is thus determined as much by terminal performance as over-the-road operations. To meet the logistical requirements of shippers, railroads need to ensure reliable train connections and adequate terminal capacity (Martland et al. 1992).

1.1 Manifest Freight, Terminals and Reliability

Martland et al. (1992) names the classification terminal as a key determinant in the service reliability of general manifest (or carload) freight. "Cars spend most of their time in terminals, and that's where the service battle is won or lost for carload business" (Murray 2002). For example, Logan (2006a) studied 35,000 car records from a major North American railroad for October 2004 and reported a median value of 59% of transit time was spent in terminals (Figure 1.1). For the first nine months of 2004 the Canadian Pacific Railway (CPR) reported that the average freight car on its system spent 64% of the time in terminals (Figure 1.2) (CPR 2004). "This suggests that the reliability of car movements can be improved by reducing the time spent in those activities or by making them more reliable" (Kwon et al. 1995). The transition to scheduled operations by all of North America's Class I railroads has increased the interaction between terminal performance and service reliability (Kraft 2002c, 2002d) because "efficient high-throughput classification yards are vital to scheduled railroading" (Ytuarte 2001).

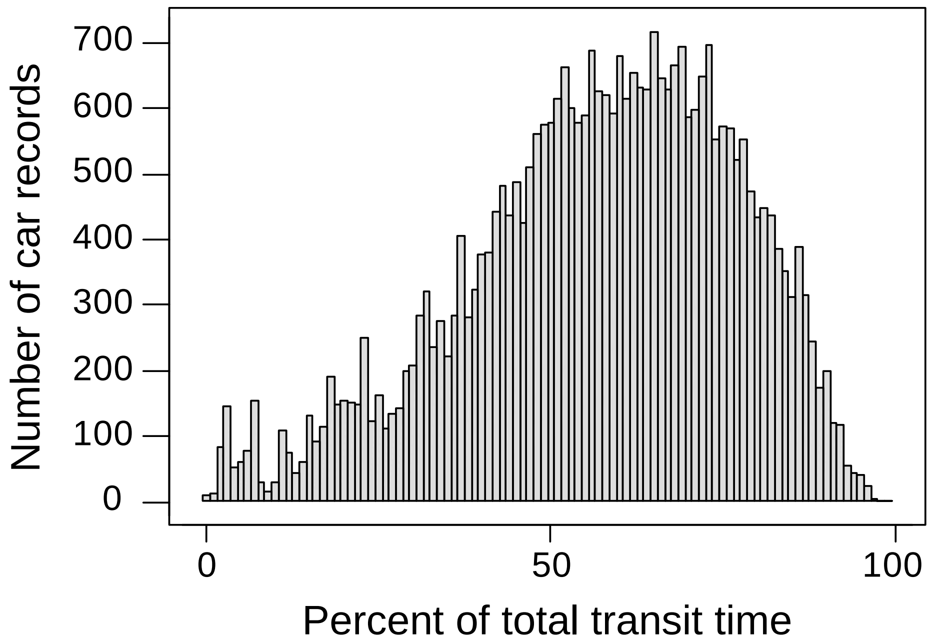


Figure 1.1 Terminal dwell as a percentage of total transit time (Logan 2006a)

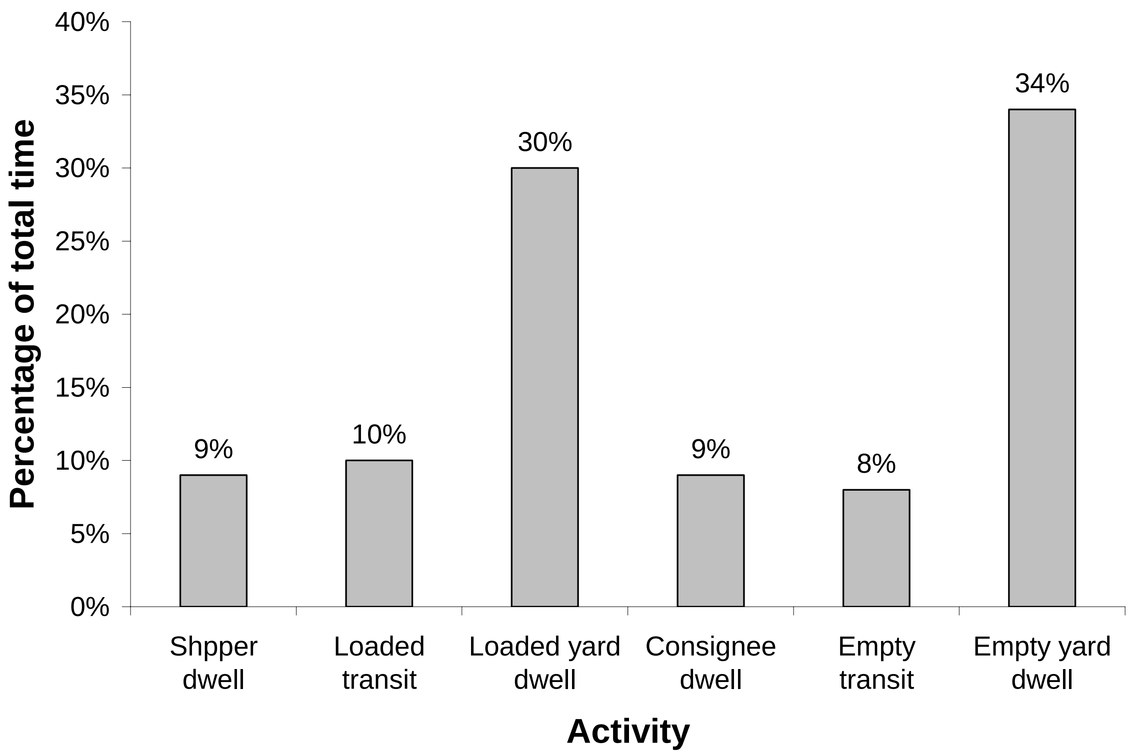


Figure 1.2 Distribution of freight car time on CPR, January through September 2004

1.1.1 Defining Yards and Terminals

The terms yard and terminal are used throughout this thesis, so understanding their definitions is important. Droege (1925) defines a yard as:

> "A system of tracks within defined limits provided for making up trains, storing cars and other purposes, over which movements not authorized by time table or by train order are made, subject to prescribed signals, rules or special instructions."

Closely related to yards are railroad terminals. A terminal is defined by Droege (1925) as:

> "An assemblage of facilities provided by a railway at a terminus or at intermediate points on its line for the purpose of assembling, assorting, classifying and relaying trains."

While these two terms are often used interchangeably, a yard is considered the physical track layout and a terminal usually consists of a major classification yard along with various support yards and facilities.

1.2 Railroad Efficiency and Reliability

There are three principal factors that contribute to railroad freight transportation efficiency:

1. The low coefficient of friction between the steel wheel and rail means low rolling resistance
2. A fixed guide-way for the movement of motive power and rolling stock that permits the operation of trains by a single operating crew
3. The strength of the infrastructure to support heavy loads permits economies of scale

The combination of these three factors allows railroads to spread their operating costs over a large number of heavily loaded cars, pulled by a locomotive operated by a single crew, in the form of a train (Armstrong 1998).

Railroads are a network industry serving and connecting customers spread over a wide geographic area. In order to recoup the high costs of maintaining that network, railroads seek to exploit the efficiencies afforded them by the three factors listed above. Running long trains as a single unit from origin to destination is generally the most efficient way to operate a railroad. However, many rail customers require carload rather than trainload quantities. In order to gain the economies of running long trains when serving these customers, cars must be gathered and consolidated into trains. This practice improves train operating efficiency, but it also tends to increase car cycle time (Table 1.2) and transit time as well as reduce reliability compared to unit train and intermodal service (Table 1.3).

Kwon et al. analyzed data from the Association of American Railroads' (AAR) Car Cycle Analysis System (CCAS) for a one-year time period: December 1989 to November 1990 for boxcars, and 1991 for covered hoppers and intermodal cars (Tables 1.2 and 1.3). The authors measured reliability in terms of the largest percentage of cars that arrived within a prescribed time window (Table 1.3). They define this variable as "maximum n-day-%" where n equals number of days. A higher n-day-% indicates more reliable service.

Table 1.2 Car cycle time for different train services in 1990 (from Kwon et al. 1995, p. 8)

	Boxcar (carload)	Covered hopper (unit train)	Double-stack (intermodal)
Loading time (days)	2.15	1.92	0.73
Loaded transit and yard time (days)	8.77	5.33	3.21
Unloading time (days)	1.48	1.27	0.22
Empty transit and yard time (days)	14.48	6.76	1.99
Total cycle time (days)	26.88	15.27	6.15

Table 1.3 Transit time and reliability performance of different train services in 1990 (from Kwon et al. 1995, p. 8)

	Boxcar (carload)	Covered hopper (unit train)	Double-stack (intermodal)
OD Pairs	477	102	20
Number of railroads traversed	2.11	1.42	n/a
Distance (miles)	788.1	831.0	n/a
Mean transit time (days)	7.16	5.25	2.53
Std dev of trip time (days)	2.62	2.04	0.50
Maximum 1-day-%	32.42%	41.90	89.20
Maximum 2-day-%	48.56%	60.95	n/a
Maximum 3-day-%	61.07%	73.21	n/a

1.3 Consolidation, Connections and Service Reliability

The lower level of service reliability for carload customers is due to the gathering and consolidation process that is typical of how this type of traffic is handled by railroads. Unlike unit trains or intermodal cars, a carload shipment does not move on one train directly between origin and destination. Instead, it moves on a series of trains determined by a trip plan that is generated when the shipment is ordered. The trip plan specifies the connections between trains that the freight car must make in order to arrive at the receiver at the scheduled time. These connections take place at classification terminals. To illustrate this process, a hypothetical carload is tracked from origin to destination on the BNSF Railway (Figure 1.3). Coors Brewing Company in Golden, Colorado, is the originator and a beverage distributor in Superior, Wisconsin, is the receiver, or consignee.

The car is picked-up by the Golden local, along with other freight cars originating in Golden, and taken to Denver, the nearest network hub (Figure 1.4). At Denver, the car is consolidated with other freight cars whose next intermediate destination is Lincoln. At Lincoln, the car is consolidated with other freight cars heading to Northtown (Minneapolis) where it is consolidated with other cars bound for Superior. Once in Superior, the car is placed in a local train and delivered to the consignee (Figure 1.5).

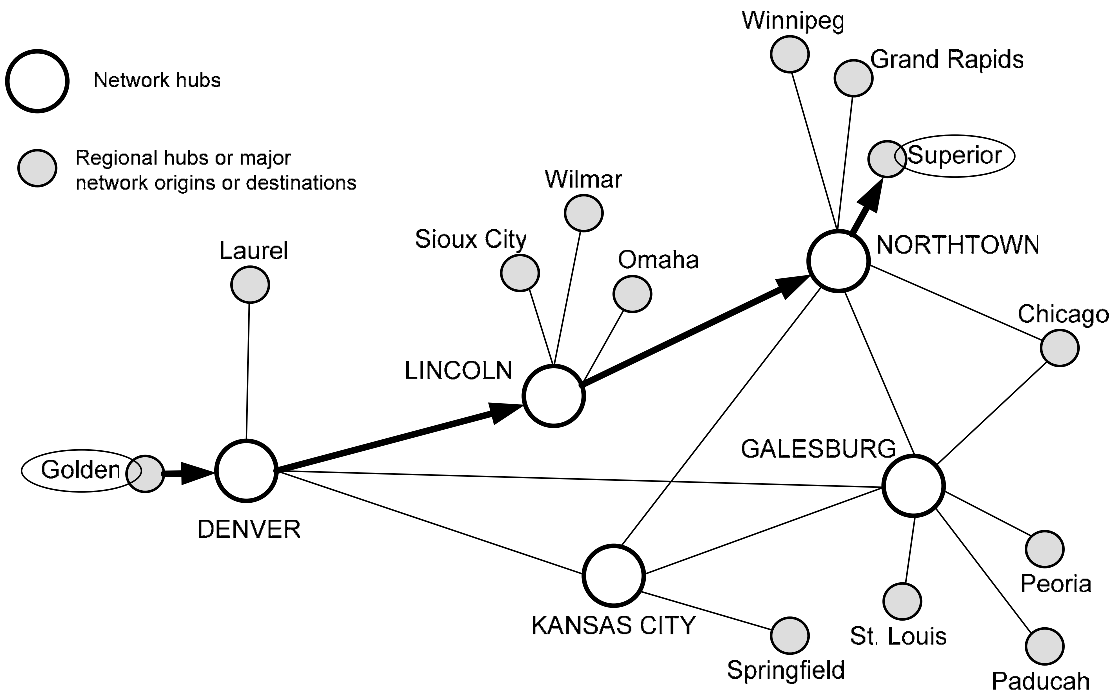


Figure 1.3 Portion of BNSF manifest freight traffic network with hypothetical trip plan from Golden, Colorado, to Superior, Wisconsin (network from Van Hattem 2004, pp 56-57)

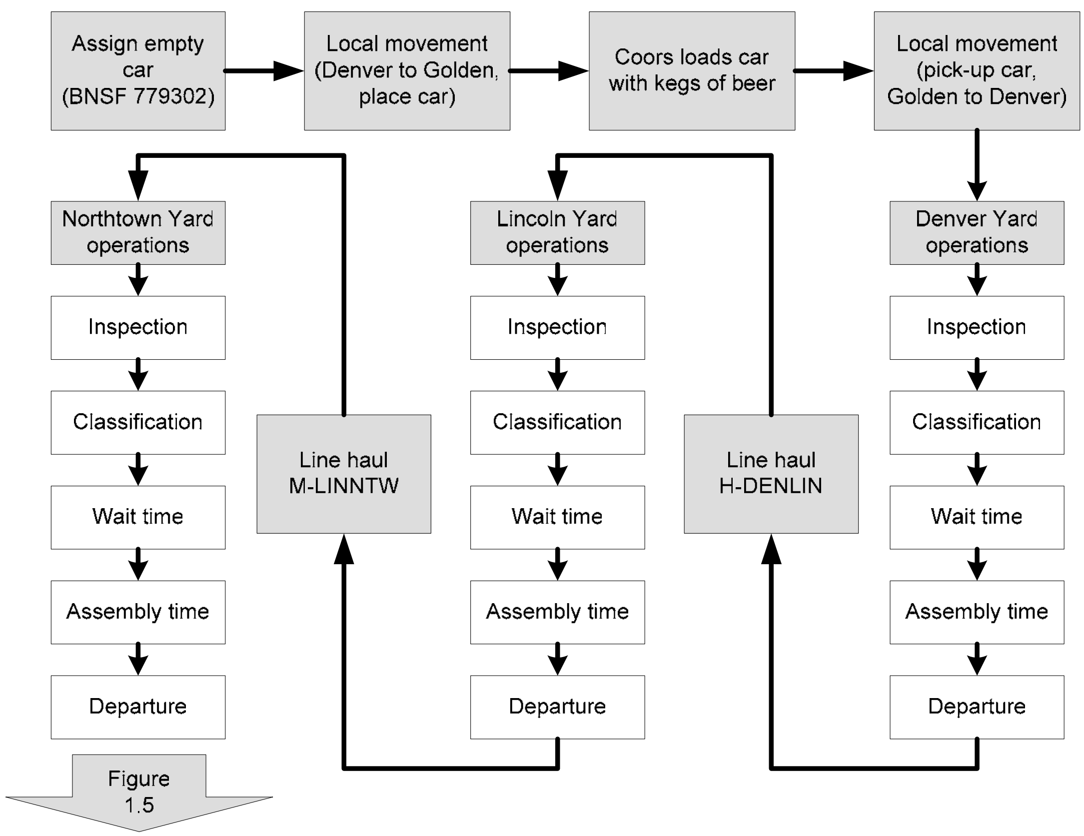


Figure 1.4 Sequence of activities for carload from Golden to Northtown on BNSF

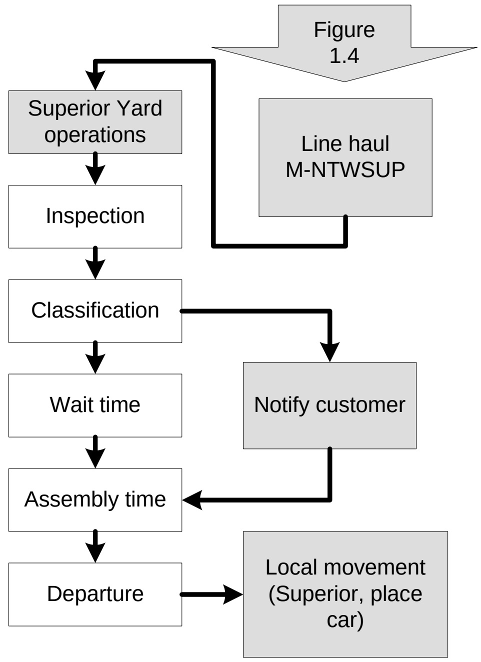


Figure 1.5 Final sequence of activities for carload from Northtown to Superior

The shipment moves through a series of four links from Golden to Superior (Figure 1.6). Assuming that the connection reliability is 90% for the first link and 85% for the remaining links, the combined reliability is the product of the individual connection reliabilities (Kraft 2000a). This results in an overall reliability of 55.27% (0.9 x 0.85 x 0.85 x 0.85). In other words, only about 55% of the shipments moving through this path will arrive on time. Of the remainder, 35% can be expected to miss one connection, 8% two connections and 1%, three or more connections (Figure 1.7).

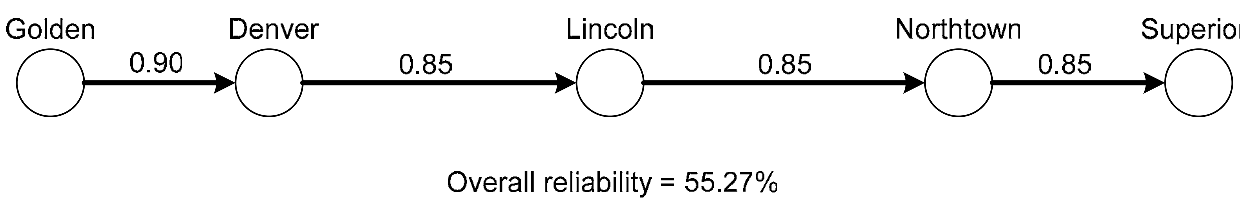


Figure 1.6 Four links in series with connection reliabilities for shipment example

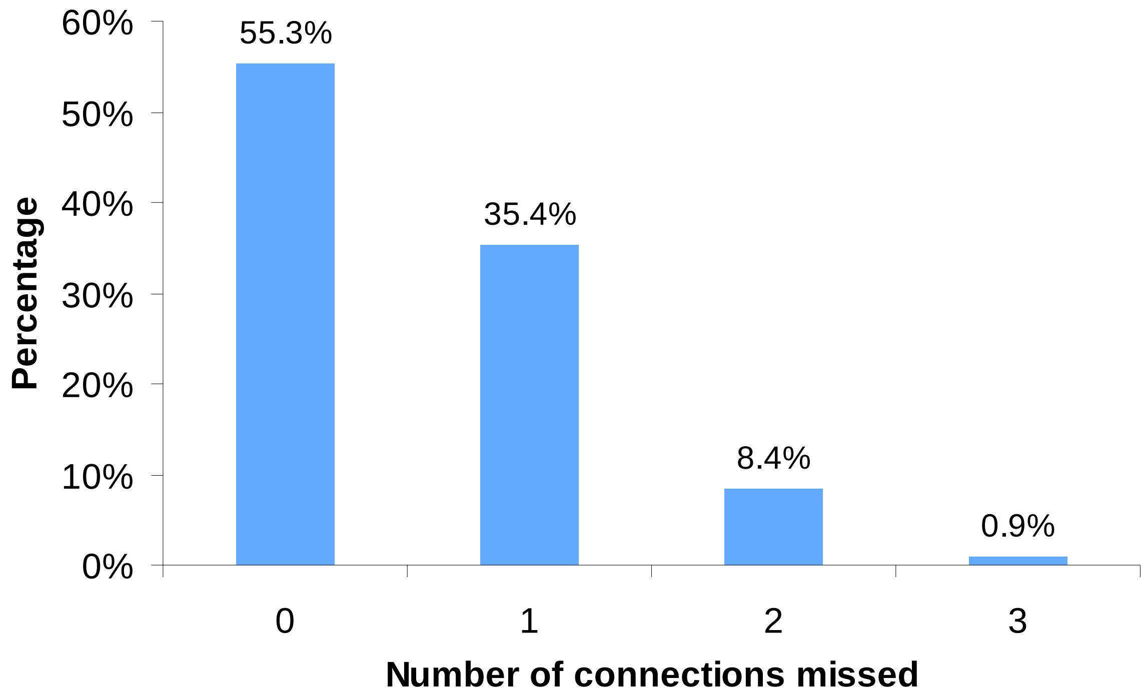


Figure 1.7 Arrival time distribution for shipment example

It is evident that the overall reliability of the system is considerably lower than that of any individual link because if the car misses one of its planned connections, it will have to wait in the classification yard until the next appropriate outbound train departs (usually 12-24 hours later). In order to achieve a 95% likelihood of no missed connections over the entire journey, each terminal would have to maintain a 98.7% level of reliability.

Contrast this to a typical truckload shipment. The truck carries one load per trip without any connections. This results in higher labor, fuel and equipment expenses per load, but it also results in high service reliability because the shipments do not interact and one delay does not affect others. The railroad freight consolidation process achieves much lower labor, fuel and equipment expenses per load than truck, but is inherently less reliable because of the need to make connections and the much greater interaction between shipments for a variety of reasons. Consequently, shipping by railroad is usually less expensive but has a lower level of service reliability. The result is that rail shippers and receivers will have lower transportation charges but higher total logistics costs because the higher variability in shipment arrival times means additional inventory must be carried by railroad customers to maintain a fixed level of customer service (Hopp & Spearman 2001, p. 88).

Extremely unreliable service will usually be more expensive to railroad customers than conversion to another mode. Because manifest traffic generates higher margins than intermodal traffic, this creates a necessity to deliver consistent, reliable service in order to keep the traffic on the rails and support the level of pricing that permits reinvestment in the railroad (Murray 2002). The trade off between high cost efficiency and reduced service quality is inherent to carload railroad operations. Overcoming it is at the core of the potential freight market share increase that will enable railroads to continue to grow this segment of their business.

1.4 The Importance of Terminals to Service Reliability and Network Efficiency

The railroad terminal and its interactions with the network is a complex system. Network fluidity affects overall system efficiency and economics because of its effect on the utilization of labor, equipment and infrastructure. It also affects service reliability which, in turn, affects traffic mix and volume (Figure 1.8). This impacts rail freight rates and revenues. While many activities affect either costs or income, service reliability affects both.

Sussman (1975) names the terminal as the key control point in the rail network and his work was a part of a series of studies by the Center for Transportation Studies, Massachusetts Institute of Technology (MIT), sponsored by the Federal Railroad Administration (FRA) and the AAR in the 1970s and early 1980s. Martland et al. (1992) summarized the major findings of these studies at the terminal level including:

"Rail service was unreliable because of missed train connections. A significant percentage of cars (10-30%) missed connections at each yard as a result of inbound train delays, yard congestion, or inadequate outbound train capacity. Bad orders, misroutes, and no-bills were relatively unimportant in their effect on connection reliability. . . While there have been few more recent studies, they have tended to support these conclusions."

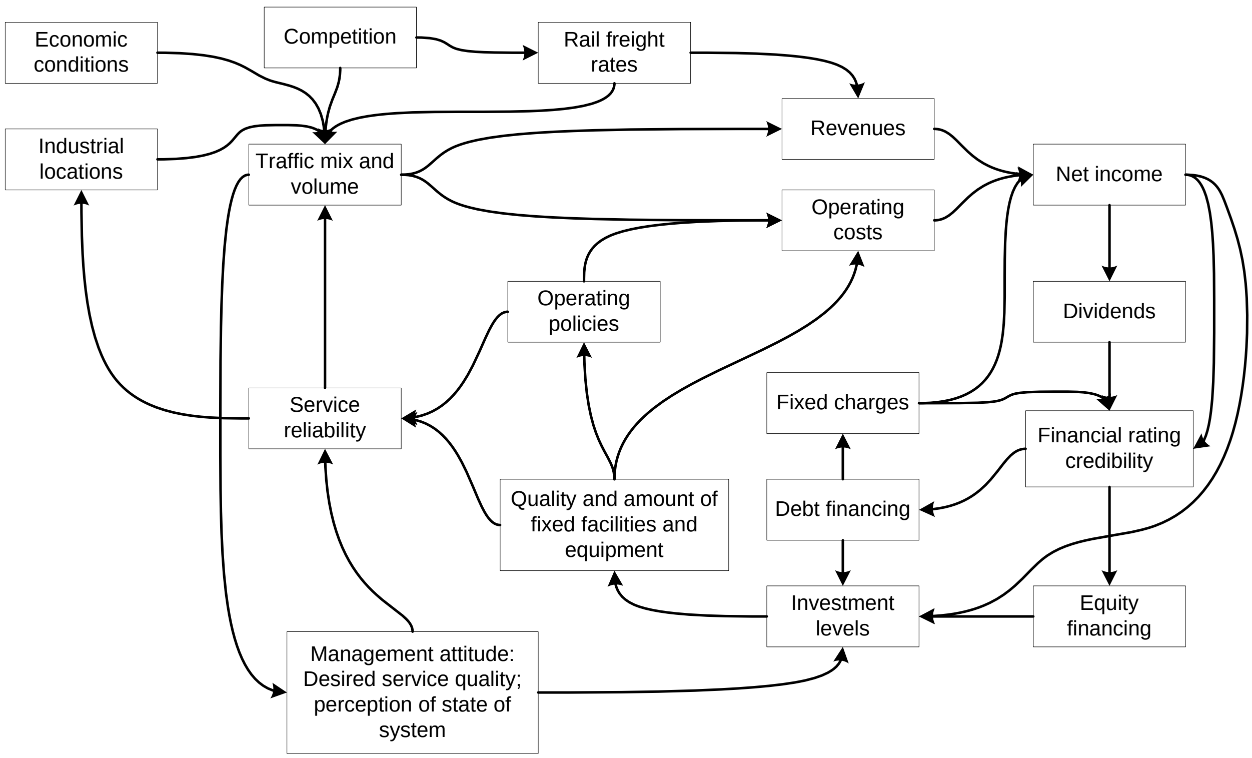


Figure 1.8 System diagram (Sussman 1975, pg D-2) modified to reflect deregulation

At the network level, Kraft (1998) cites the estimates from Martland et al. (1994) of which root-causes contributed the most delay in a sample of manifest traffic:

1. *Power availability delays* – include delays to trains due to locomotives not being in position to move the required tonnage (24.4% of all train delays)
2. *Terminal delays* – including yard congestion, cars not switched in time, cars moved on other than scheduled trains, etc. (20.2%)
3. *Train delays* – management decisions regarding which trains to run, and with what resources, including maximum tonnage limits, trains annulled due to lack of traffic, train consolidations, etc. (20%)
4. *Mechanical delays* – defects requiring repair of cars or locomotives (16%)
5. *Line delays* – track work, curfew, train meets, etc. (13.3%)
6. *Other* – derailments, unknown causes, no bills, etc. (6.1%)

Martland et al (1994) concluded that "even if the railroad had 'perfect' technology, only 30 percent of the delays would disappear; 65 percent of the delays required better management of resources (terminal management, train management, and power distribution)." As the key control points in the railroad network, terminal performance directly impacts all three of the resource management areas. Therefore, the MIT studies established the importance of the classification terminal to service reliability and demonstrated that reliability could be improved through modifications to the terminal operating plan or improvements in the terminal control system.

1.4.1 The Impact of Classification Terminal Operations on Network Efficiency

In addition to improving service reliability, better performing terminals result in more efficient railroad networks. A common measurement of network efficiency is average train speed. Train speed measures line haul movement between terminals and the average is calculated by dividing train-miles by total hours operated, excluding yard and local trains, passenger trains, maintenance of way trains, and terminal time (AAR 2006b). A higher system-wide average train speed indicates a more fluid network because trains spend more time moving. Therefore, fewer cars and locomotives are required to move the traffic because the equipment is cycled faster. Network efficiency can be thought of as a cycle (Figure 1.9) with terminal dwell linked directly to car velocity and indirectly linked to average train speed through a series of events.

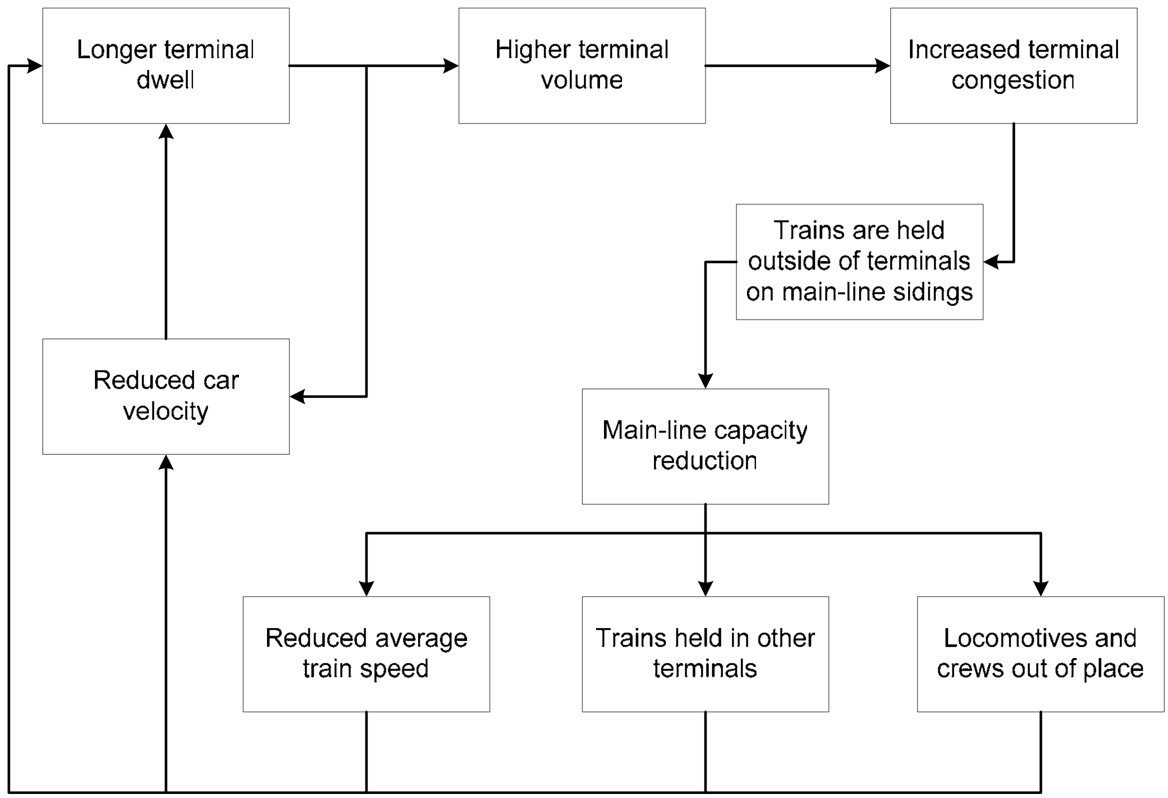


Figure 1.9 The network efficiency cycle

One estimate of the impact of improved terminal performance on carload velocity is provided by Logan (2006a). Every 15% reduction in system-wide average terminal dwell time results in an approximate increase in carload velocity of 2 mph. The exact relationship between terminal dwell and average train speed is more difficult to determine because of the complex interactions between the factors in Figure 1.9. However, an idea of the relationship can be obtained by analyzing performance measures for each railroad.

1.4.2 The Relationship between Terminal Dwell and Average Train Speed

In order to better understand the relationship between average terminal dwell and average train speed, performance measurement data were analyzed for all of the Class I railroads except Kansas City Southern (KCS). The data for BNSF, CSX, NS and UP were obtained from the AAR's (2006b) railroad performance measures website. The data for CPR and CN were obtained from their corporate websites (CPR 2006a and CN 2006). KCS was excluded because their incorporation of the Texas Mexican Railway into their system during the analysis timeframe may have skewed the results. For all of the railroads except CPR, average weekly manifest (carload) train speed and corresponding average weekly terminal dwell for the entire railroad was obtained for the weeks ending May 27, 2005 through May, 26, 2006 (53 weeks). CPR's website contained weekly averages for the weeks ending October 1, 2004, through May 26, 2006 (87 weeks). The Christmas week was excluded in each case because it created an extreme outlier. A simple linear regression model (Irwin 2005) was applied to each railroad's data and hypothesis tests at a 95% level of significance were performed. A scatter plot for the railroad with the most data, CPR, is shown in Figure 1.10. The scatter plots for the remaining railroads are found in Appendix A. Table 1.4 summarizes the regression results for each railroad.

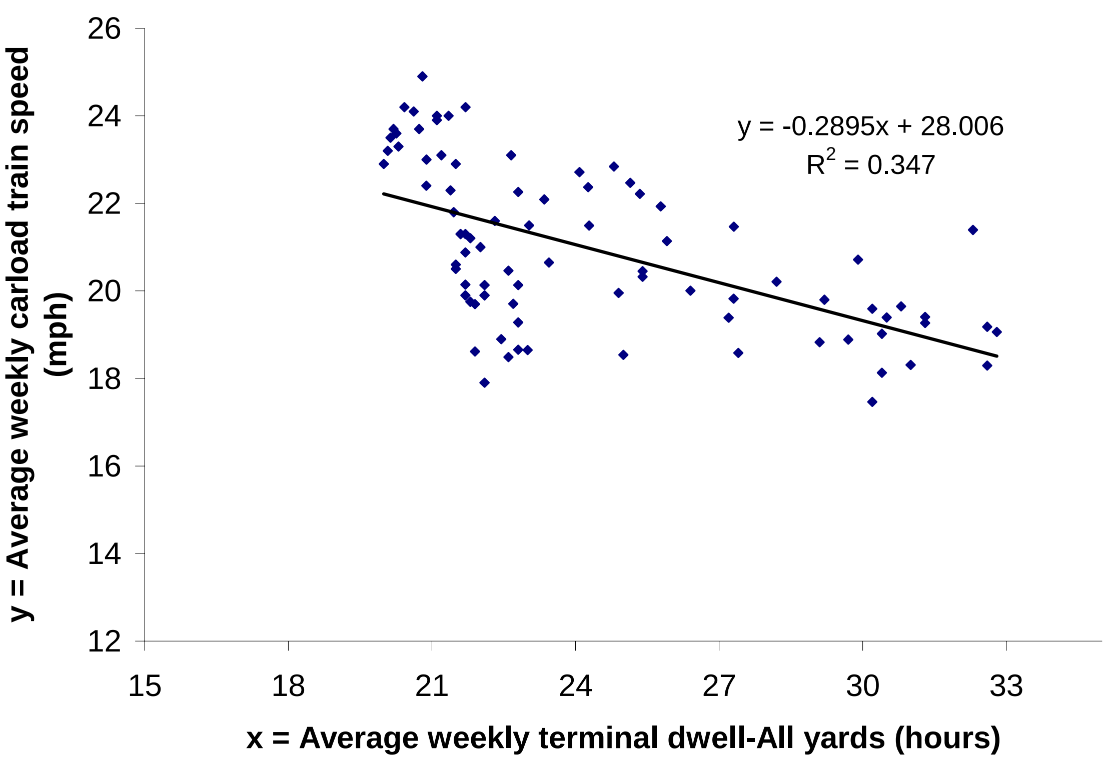


Figure 1.10 Scatter plot and regression results: CPR, weeks ending October 1, 2004 to May 26, 2006

Table 1.4 Regression analysis results for the relationship between average weekly terminal dwell (TD) and average weekly carload train speed (TS)

Railroad	Regression Analysis Results		Railroad	Regression Analysis Results	
CPR	TS = 28.006	- 0.2895TD	CSX	TS = 24.613	- 0.2064TD
t-stat	26.04	-6.64	*t-stat*	16.96	-3.91
p	< 0.001	< 0.001	*p*	< 0.001	< 0.001
R^2	0.347		R^2	0.234	
CN	TS = 28.290	- 0.285TD	NS	TS = 31.111	- 0.4579TD
t-stat	28.57	-3.61	*t-stat*	15.40	-5.18
p	< 0.001	< 0.001	*p*	< 0.001	< 0.001
R^2	0.207		R^2	0.349	
BNSF	TS = 32.395	- 0.4943TD	UP	TS = 24.829	- 0.1857TD
t-stat	11.56	-4.35	*t-stat*	15.32	-4.61
p	< 0.001	< 0.001	*p*	< 0.001	< 0.001
R^2	0.275		R^2	0.2468	

The potential network efficiency gain is seen in the regression results. For each railroad, sample information indicates a statistically significant relationship between carload train speed and terminal dwell. The R^2 values range from 20.7% to 34.7% indicating that this amount of the variation is explained by the models. The remaining variation in average train speed is explained by other factors such as locomotive availability, mainline train speeds, meets and passes, signal systems, weather, crew availability, line congestion, etc. Considering all of these other factors affecting system performance, the percentage of variability accounted for by the single variable, dwell time, may seem surprising. However, the high percentage of time that freight cars spend in terminals (Figures 1.1 and 1.2) and the 65% of root-causes of delays related to terminal management (Martland et al. 1994) point to the importance of terminal dwell time as a critical factor affecting railroad network efficiency.

1.4.3 Current Trends in Network Performance and Demand

For the weeks ending May 27, 2005 through May 26, 2006, (53 weeks) the weekly average manifest train speed for the same six Class I railroads was plotted over time to determine the current network performance trends (Figures 1.11 and 1.12). Four of the six railroads (CN, BNSF, NS and UP) have been experiencing a slight downward trend in average manifest train speed over the past year. CSX has been experiencing a slight upward trend while CPR saw a sharp increase early in 2006 which has since leveled off. Demand for rail transportation, measured in revenue ton-miles, has continued the growth that began in the early 1990's (Figure 1.13). If demand continues to increase, as the indicators suggest (Leilich 2006), average train speeds will continue to trend downward (for CN, BNSF, NS and UP) or level off (in the case of CPR and CSX) unless the network resources are managed better. As shown in the previous sections, the network resource that has one of the greatest impacts on network efficiency and service reliability is the classification terminal.

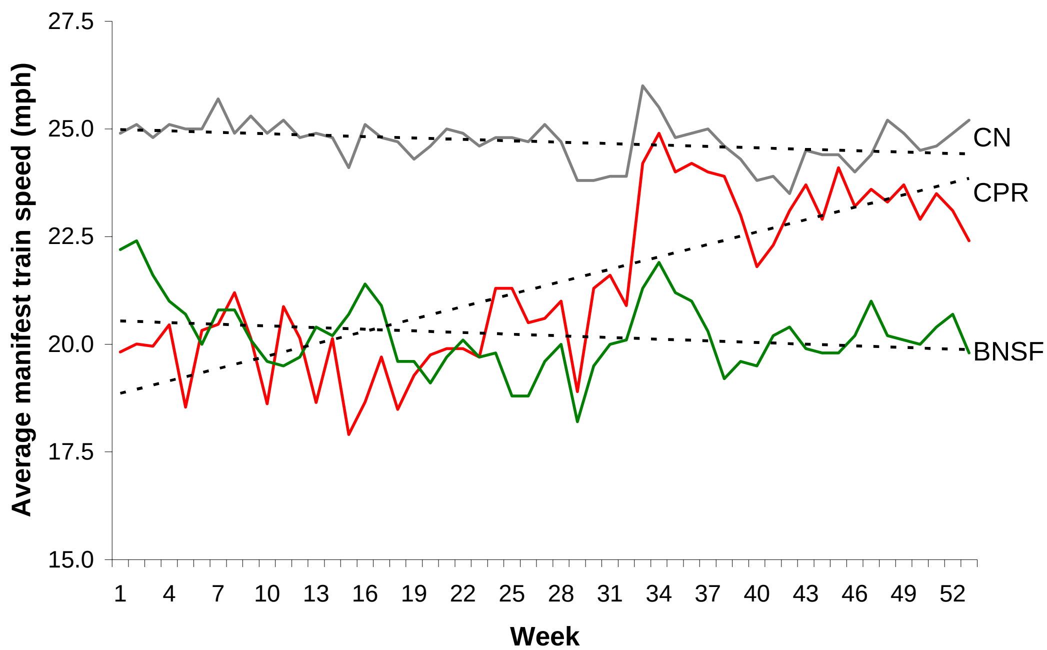


Figure 1.11 Weekly average manifest train speed trends for CN, CPR and BNSF May 2005 to May 2006

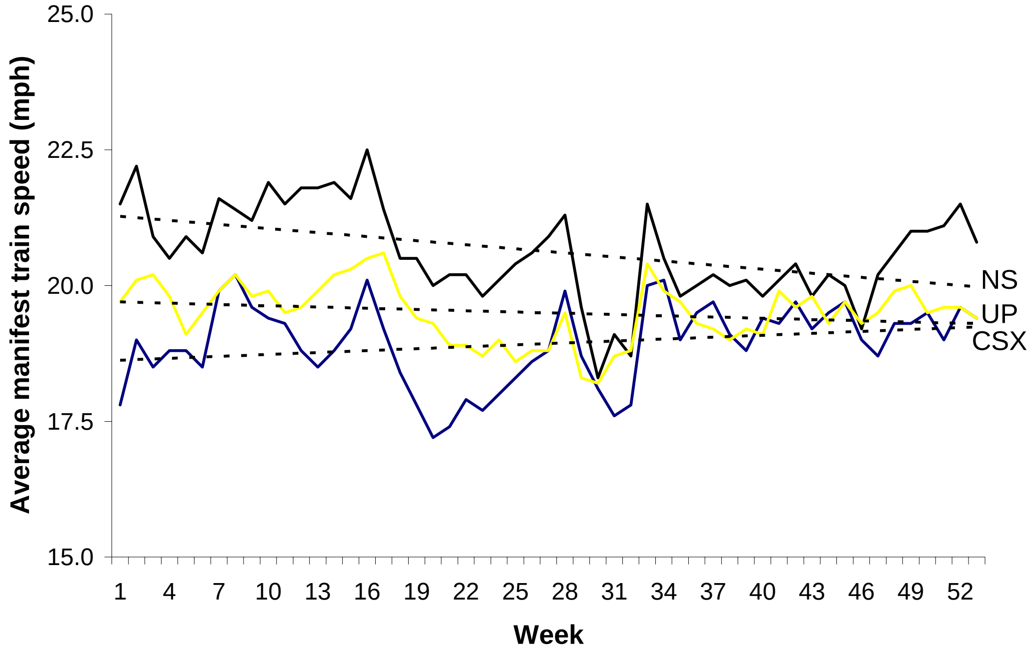


Figure 1.12 Weekly average manifest train speeds for NS, UP and CSX May 2005 to May 2006

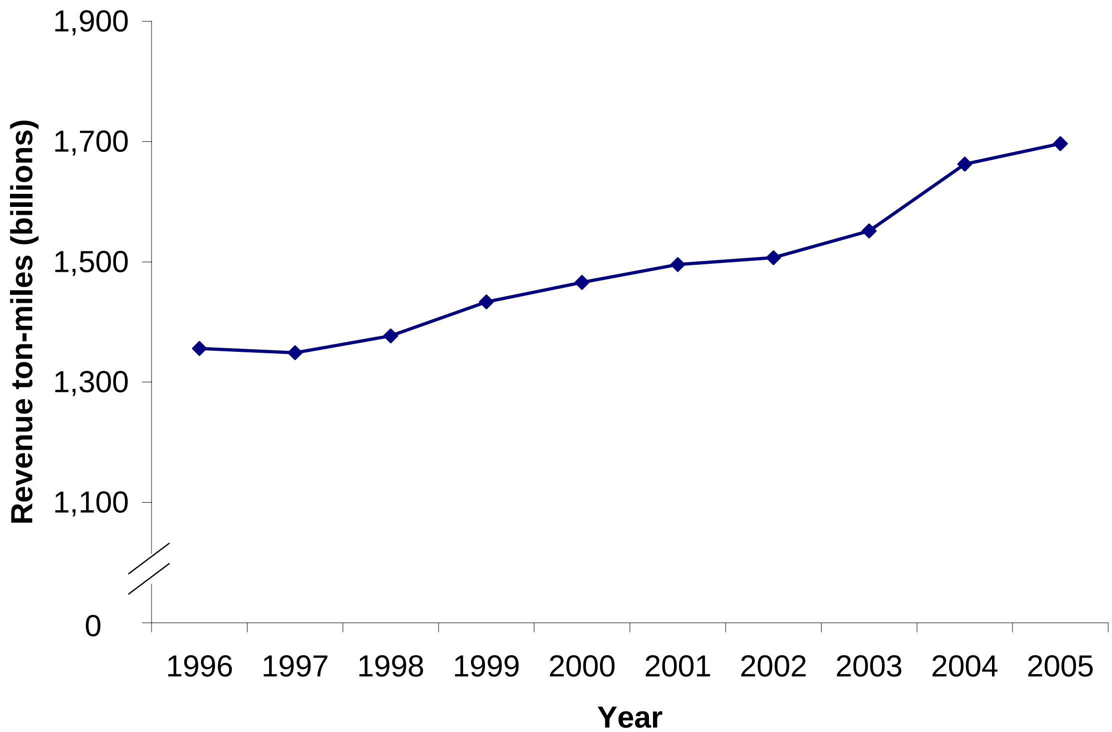


Figure 1.13 Revenue ton-miles for 1996 to 2005 (AAR 2002, 2003-2006)

1.5 Increasing Terminal Capacity

Kraft (2000b) raises the issue of inadequate terminal capacity as a barrier to improved service reliability.

> "A very serious concern is the level of terminal congestion. Under the guise of operating 'efficiency,' some carriers may have reduced terminal capacity too much. Apart from the need to run trains on a reliable scheduled basis, the most important thing railroads can do to improve reliability is to invest in adequate terminal capacity."

In addition, the importance of terminals to network efficiency leads to the conclusion that inadequate terminal space is also a barrier to improved network performance. However, the availability of capital and physical space to expand some yards may be constrained. Therefore, in addition to considering infrastructure expansion, railroads must also determine how to harness as much capacity from extant infrastructure as possible. This creates the need for new management and operational methods that will increase the capacity of existing facilities.

1.5.1 Problem Statement

Terminal capacity can be improved by reducing dwell time through the adoption of methodologies that have been successfully applied at the production control level in manufacturing such as elements of "Factory Physics," developed by Hopp & Spearman (2001), the Theory of Constraints (TOC), originally developed by Goldratt (1990), and Lean Manufacturing (Womack & Jones 2003).

In my thesis research I consider the classification terminal as a production system in order to utilize the production management methods referred to above and introduce the "Lean Railroading" approach to improving classification terminal performance. To illustrate the approach, I will demonstrate the theoretical importance of the bottleneck, identify the pull-down process as the bottleneck in a majority of hump yards, and provide a detailed analysis of the pull-down process. After enhancing the macroscopic evaluation method for determining pull-down process capacity first presented by Wong et al. (1981), I will compare several alternative bottleneck management methods. Then, I will introduce a new metric that has been developed to measure the performance of the sorting process because of its direct interaction with the pull-down process.

Currently, the performance of hump controllers is measured primarily by the number of cars humped per unit time (i.e. hour or shift). While management knows effective sorting is important, they do not have a quantitative way to measure the quality of sorting. To address that need and thereby better manage and understand the interaction between the pull-down process and its immediate upstream process (the hump), a quality of sort metric is developed and its relationship to bowl volume and pull-down throughput is established. These results will enable terminal management to add a much needed quality measurement component to the hump controller performance criteria that will facilitate an increase in pull-down process throughput and, ultimately, increase capacity through a reduction in terminal dwell time.

CHAPTER 2: PRODUCTION MANAGEMENT FOR RAILROAD CLASSIFICATION TERMINALS

2.1 Introduction to Production Management

The science of production management developed as a result of the need to manage and understand large manufacturing systems. These management tools and methods include time and motion study, engineering economic analysis, quality control, operations research and production/inventory management (Hicks 1994). These methods have been successfully applied outside of manufacturing in industries ranging from health care to transportation (Sobek & Jimmerson 2003, Paixao & Marlow 2003). A railroad transportation example is the study by Mundy et al. (1992) in which Statistical Process Control (SPC) techniques were used to analyze the performance of the Burlington Northern's Tennessee Yard facility in Memphis. SPC is a collection of quality tools that have been widely adopted in manufacturing and the work stems from a more general interest in the application of SPC in service industries. Renewed interest in improving terminal operations presents an opportunity for further use of production management techniques.

2.1.1 The Hierarchy of Problems in Production Management

Brandimarte & Villa (1995, p. 10) describe a three-level hierarchy of problems: production planning, production scheduling and production control. Each level in their hierarchy has a characteristic time horizon and type of activities associated with it that can be adapted for evaluation of railroad terminals (Table 2.1). The use of methods found in the production control level to increase railroad terminal capacity is a principal objective of my research. In this chapter I will use elements of the "science of manufacturing," established by Hopp & Spearman (2001) in *Factory Physics*, to consider the terminal as a production system and better understand its dynamics. The Theory of Constraints (TOC) approach, originally developed by Goldratt (1990), and Lean Manufacturing (Womack & Jones 2003) provide practical guidance and tools for improving production system performance. In Chapter 3, I will continue to build the understanding of the "Science of Lean Manufacturing" provided by Factory Physics and describe how to apply these techniques to railyard operations as part of the establishment of the Lean Railroading approach.

Table 2.1 Production management hierarchy (Hopp & Spearman 2001, Kraft 1998, Brandimarte & Villa 1995)

Level	Time Horizon	Manufacturing Decisions	Rail Decisions
Production planning	Long term (strategy)	Financial decisions	Financial decisions
	Greater than 6 months	Marketing strategies	Marketing strategies
		Product designs	Service design
		Process technology decisions	Locomotive acquistion
		Capacity decisions	Line capacity expansion
		Facility locations	Terminal expansion
		Supplier contracts	Supplier contracts
		Personnel development programs	Personnel development programs
		Plant control policies	
		Quality assurance policies	
Production scheduling	Intermediate term (tactics)	Work scheduling	Train routing
	Less than 3-4 months	Staffing assignments	Classification policy
		Preventative maintenance	Train make-up policy
		Sales promotions	Traffic routing
		Purchasing decisions	
Production control	Short term (operational)	Material flow control	Train scheduling
	Less than 1-2 weeks	Worker assignments	Locomotive distribution
		Machine setup decisions	Car scheduling
		Process control	Empty car distribution
		Quality compliance decisions	Crew scheduling
		Emergency equipment repairs	Terminal work plans

2.2 The Classification Terminal as a Production System

In order to successfully apply the necessary elements of the selected production management methods, classification terminals can be considered as production systems. The idea of comparing a terminal to a production system is not new. Ferguson (1980) used the analogy to contrast local decisions versus system decisions in a terminal control system.

> "This situation is analogous to a manager of an automobile assembly plant . . . In the railroad industry, the terminal superintendent is the plant manager and his function is to assemble inbound trains or parts of trains into completed outbound trains."

Stated another way, classification terminals are factories that produce outbound trains by sorting and properly assigning cars from various inbound trains. As such, they are subject to many of the same relationships observed in manufacturing systems.

2.2.1 Defining a Terminal System

The precise definition of a manufacturing system is provided by Hopp & Spearman (2001):

> "A manufacturing system is an *objective*-oriented *network* of *processes* through which *entities flow*."

Because classification terminals are production systems, a precise definition for a terminal system is possible by replacing the words "manufacturing system" with "railroad terminal system." The key words in the definition are italicized and are defined, in the terminal environment, as:

1. *The objective* – "The ultimate objective of a yard is to sort cars, and connect them reliably to the earliest possible candidate outbound train, while minimizing cost" (Barker 2005)
2. *The network* – The railroad terminal system is a network of interacting parts, often subject to complex interactions
3. *The processes* – Comprised of the physical processes (receiving, classification, pull-down, departure, etc.) and the support processes (crew assignment, switch list generation, maintenance, car data entry, etc.)
4. *The entities* – Freight cars, locomotives and the information associated with both
5. *The flow* – The way in which the entities are processed through the system

Figure 2.1 provides a graphical representation of the railroad terminal as a production system that transforms a set of inputs into the desired outputs. From this figure, the interaction between the terminal and the rest of the rail network can be seen and it serves as a starting point for a production level control analysis. While this research focuses on applying select production management techniques to hump yards because they more clearly resemble a factory-type assembly line with distinct breaks between the production processes, the techniques can also be applied to flat yards.

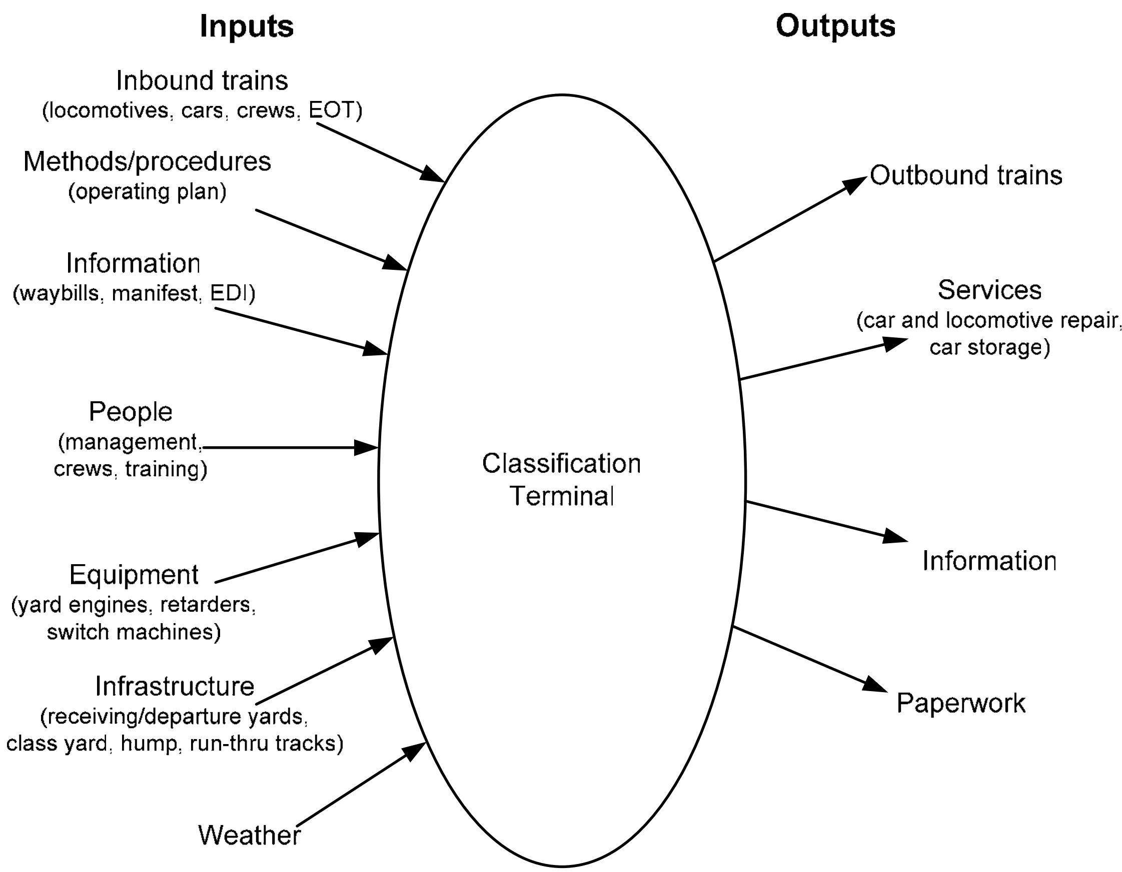


Figure 2.1 The classification terminal as a production process

2.3 Terminal Operations Review

There are two types of classification yards in North America, flat yards and hump yards. In 2003, of the 96 major classification yards in operation in North America, 39 were flat yards and 57 were hump yards (Wegner 2003). Before the differences between the two can be explained, some terminology must be defined using the definitions from Daganzo et al. (1983), modified where applicable.

1. *Cut* – Any set of cars that share a common destination track and, by chance or design, are sequenced together in an arriving train.
2. *Block* – A set of cars that has been purposefully sequenced together in a departing train because they share a common attribute (such as track destination). Within in any departing train, cars will have been sorted into one or more blocks.
3. *Switch* – The operation that separates two adjacent sets of cars, and sends the sets to their assigned classification tracks. Although every car must be sorted, not all require switches.

2.3.1 Flat Yards

Flat yards have a relatively flat profile and locomotives sort cars either by flat switching (Figure 2.2) or "kicking" (Figure 2.3).

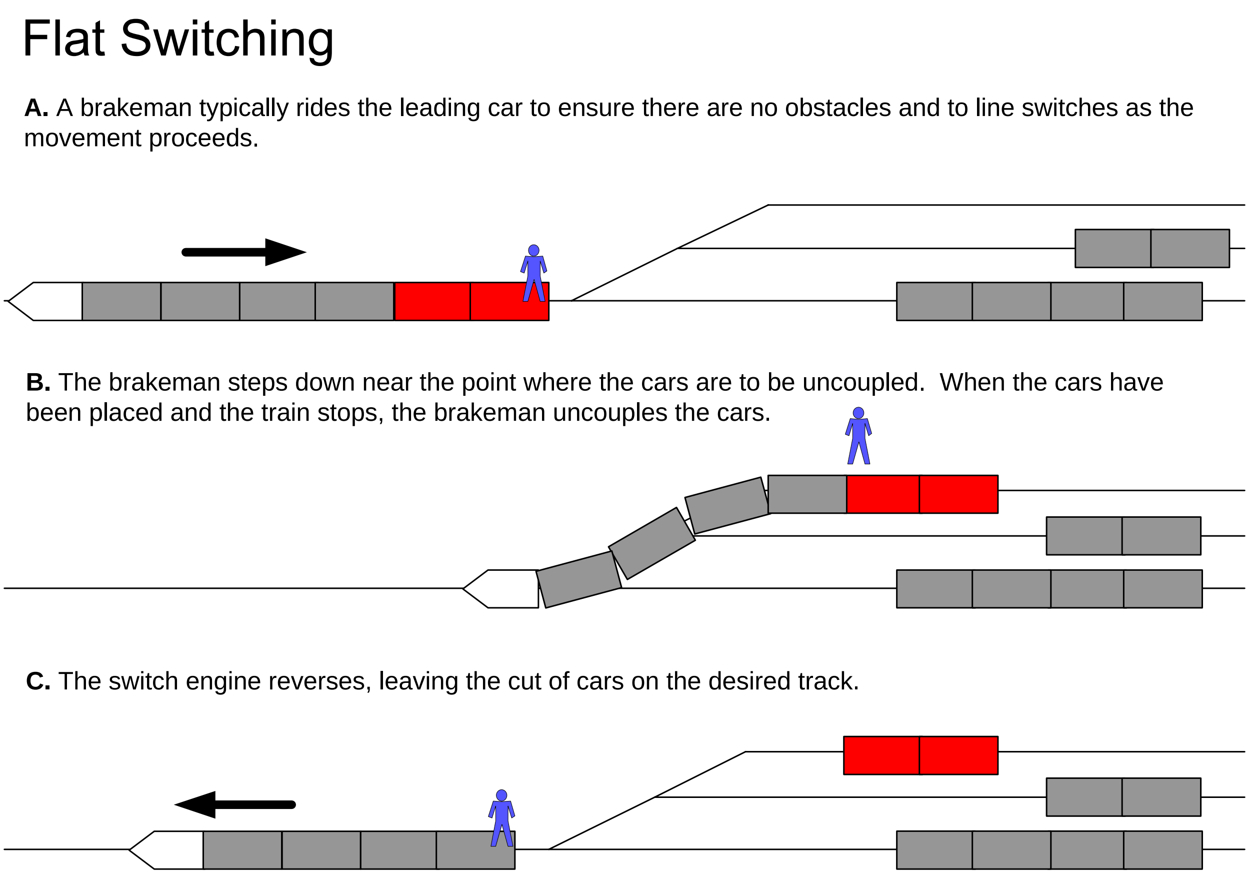


Figure 2.2 Flat switching cars in a flat yard (modified from Kraft 2002a, p. 52)

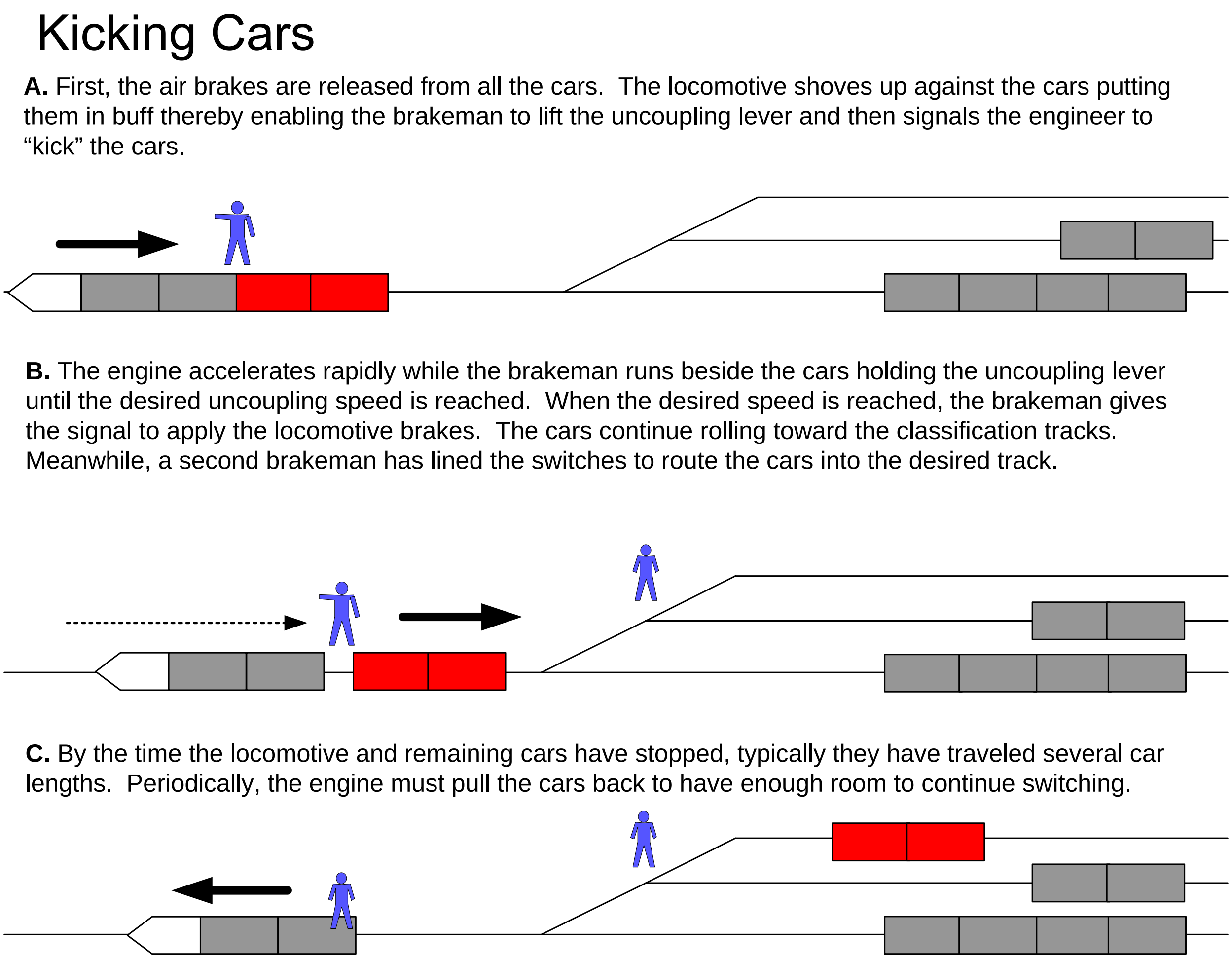


Figure 2.3 “Kicking” cars in a flat yard (modified from Kraft 2002a, p. 52)

If flat switching is used, the cars near the engine are subject to several extra handlings as the engine moves back-and-forth switching the cuts. “Kicking” attempts to reduce the amount of these movements by uncoupling the cuts when the forward momentum is sufficient to keep the cars rolling down the desired track. As shown in Figure 2.3, the engine must periodically back up to have enough room to continue switching. “In flat yards a switch is required between every pair of adjacent *cuts*, all cars in a cut being destined for the same classification track” (Daganzo et al. 1983). Because of this, flat yards are more efficient for large cuts of cars because switching single-car cuts creates the potential for many reverse movements before a car is placed on its proper track, thereby wasting inertial energy.

2.3.2 Hump Yards

Contrast this to a hump yard where inbound trains are shoved up a specially-built hill known as a hump in order to store potential energy to aid in the sorting process. At the crest of the hump, the cars are uncoupled and each car rolls into the proper track in the classification yard (or bowl because of its saucer-shaped profile) by gravity (Kraft 2002c). The tracks in the classification yard are often called class tracks or bowl tracks. Special braking devices, known as retarders, are mounted to the tracks and control the speed of the cars as they roll down the hump and into the bowl. There are two types of retarders: automatic and distributed piston. Automatic retarders use electro-pneumatic valves controlled by a computer system to slow the cars so they will couple at 4 mph or less ("walking speed"). Multiple retarders are needed to precisely control the speed: a master retarder is located directly past the hump crest, group retarders are located on the lead to each group of tracks and tangent-point retarders are located at the entrance to each classification track (Figure 2.4).

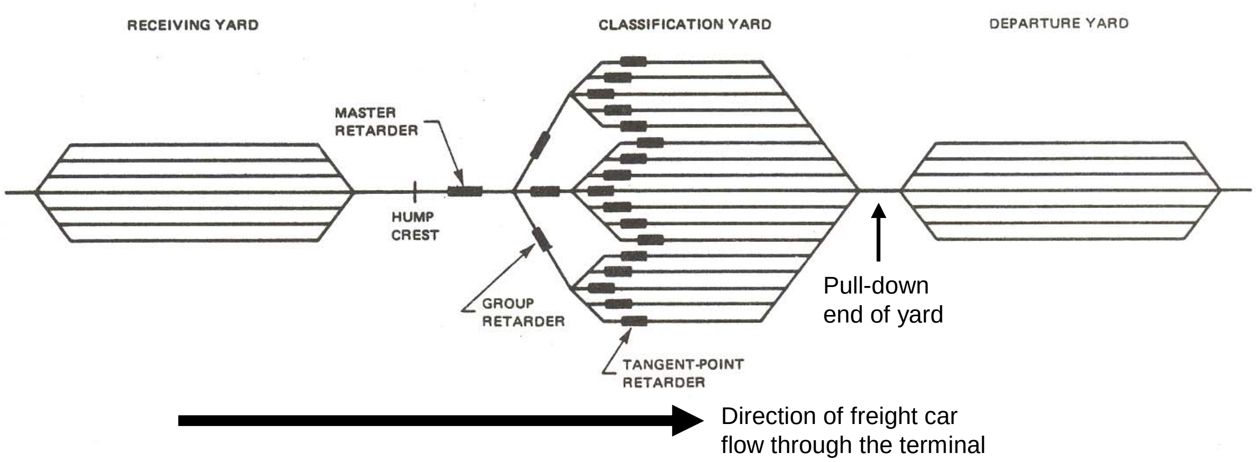


Figure 2.4 Hump yard schematic and processing sequence (Wong et al. 1981, Kraft 2002c)

A common type of distributed piston retarder is the Dowty retarder. Kraft (2002b, pp. 43-44) describes how Dowty retarders work.

> "These retarders, named for the British company that first developed them, look like mushrooms springing up next to the rails. A Dowty retarder consists of a hydraulic piston-and-cylinder so positioned that a wheel flange depresses the piston as the car passes over. A valve in the retarder slams shut if the cars speed exceeds a preset value, forcing hydraulic fluid through a narrow orifice. That resists the downward pressure of the wheel and slows the car. If the car is moving slower than the preset speed, the valve remains open so the cylinder compresses without resistance."

Adjustments to rolling speed are made by installing or removing retarder housing pairs during maintenance windows. The time of year affects the number of retarder housings needed. For example, yards with colder winter conditions will require fewer Dowty retarders because cars tend to roll more slowly when it is cold.

Hump yards use automatic-power switches to route the cars to the appropriate track (Wong et al. 1981). A hump controller supervises the classification operation and the switches are controlled by a computer system. "In hump yards, a switch is generally required between every pair of adjacent *cars* though it is possible to switch as many as eight cars at a time" (Daganzo et al. 1983). Hump yards are better at sorting cars one-at-a-time (Kraft 2002c) and this is the normal practice in North America.

2.3.3 Operating Costs versus Capital Costs

Wong et al. (1981) states, "In the case of a flat yard versus a hump yard, the economic analysis involves a tradeoff of the higher operating expenses of the flat yard versus the higher capital expenses of a hump yard." Hump yards tend to be high-volume operations because of their intensive capital requirements. Flat yards are more labor intensive; making them most applicable to small and medium-sized classification operations, in which fewer than 1,000 cars per day are processed (Wong et al. 1981). A hypothetical illustration of this tradeoff is shown in Figure 2.5.

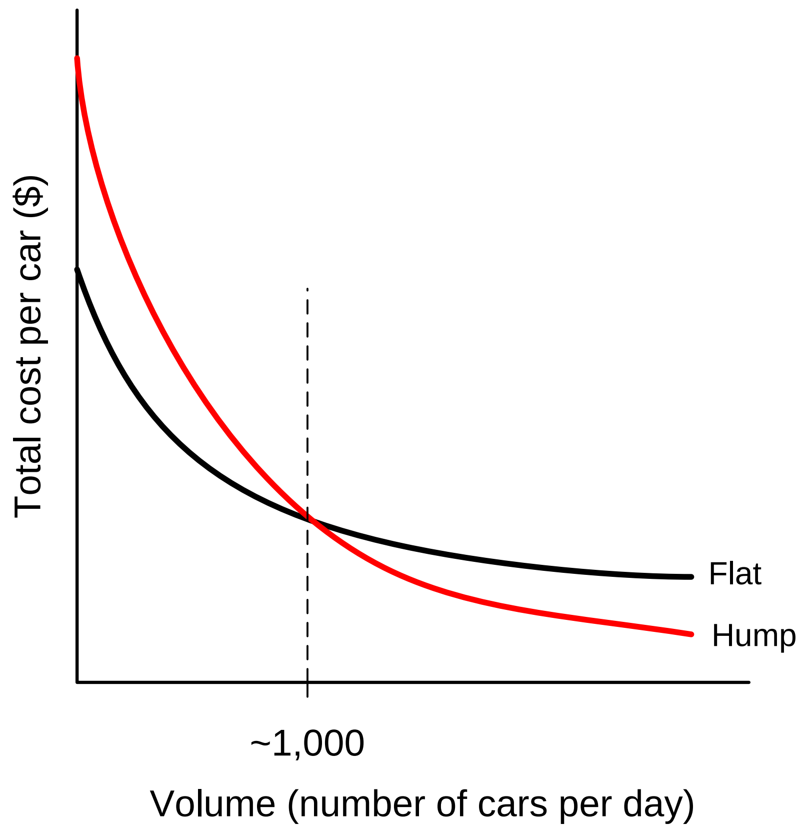


Figure 2.5 Hypothetical per unit cost curves as a function of volume for flat and hump yards

The introduction of Remote Control Locomotive (RCL) technology in the late 1990's and early 2000's has reduced the operating expenses for flat yards while the development of "mini-hump" designs, pioneered by the Southern Pacific (Wong et al. 1981), has reduced the capital expenses for small- and medium-sized yards. The availability of distributed piston retarders further reduces the capital and maintenance expenses of the mini-hump designs.

2.3.4 Freight Car Flows through the Terminal

Figure 2.6 shows an example of freight car flow through a typical hump yard. From a production standpoint, it supplies the flow of the entities (cars) between the primary processes and the basic process interactions. The cars brought by the inbound trains are subject to three different process flows. All cars go through the receiving process where they are inspected and defects are flagged for repair.

1. *Run-through cars* – Are usually found in unit trains that do not need to be classified en-route. These trains use the terminal to swap power, re-crew and, when necessary, receive the FRA mandated 1,000-mile inspection. Although they by-pass the classification and pull-down processes, they still use terminal resources and impact other car flows.
2. *Cars-to-be-swapped* – Arrive in a solid block of sufficient volume and/or priority to warrant bypassing the principal classification process altogether (known as "block swapping"). By definition, these blocks are moved as a group directly from an inbound train to the appropriate outbound train.
3. *Cars-to-be-processed* – Are the most prevalent in hump yards and are involved in the most processes. They are individual or small groups of cars making connections from an inbound train to an outbound train. Low priority cars, cars from small blocks, cars lacking proper blocking instructions and cars needing repair are often subject to being classified more than once (re-hump process).

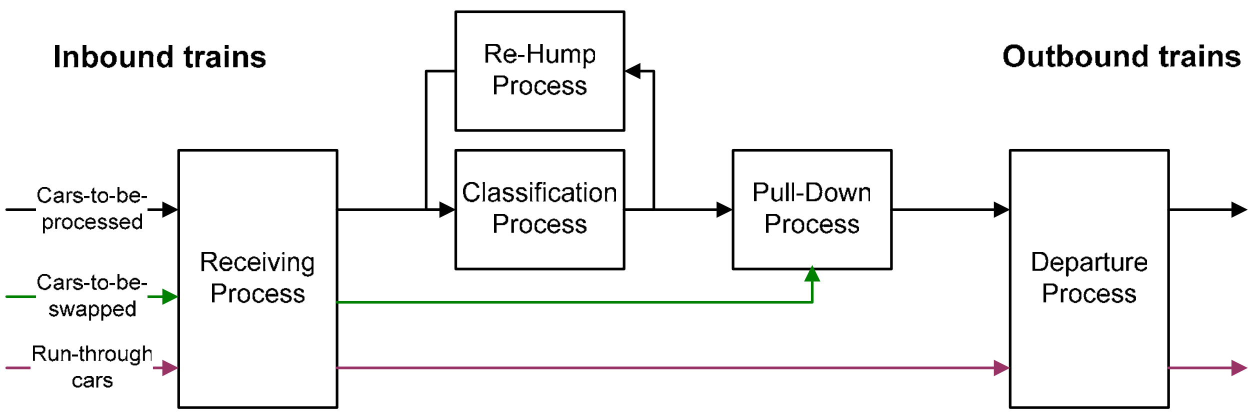
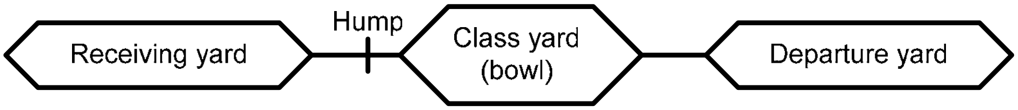


Figure 2.6 Freight car flows in hump yards

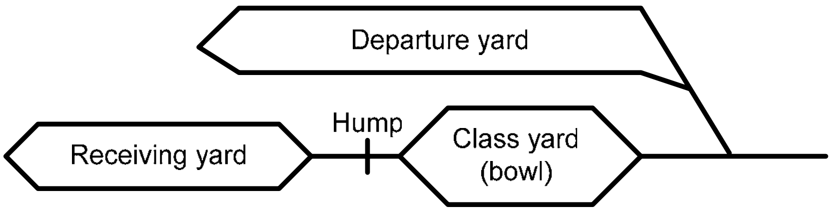
More detail for each of the processes in Figure 2.6 is found in the process flow charts in Mundy et al. (1992) and Wong et al. (1981 pp. 9-12) that provide an excellent tutorial on classification yards and their operations. The study of railroad transportation (Part II in *Studies in the Economics of Transportation*) by Beckmann et al. (1955) contains a detailed description of the differences between flat and hump yard operations and pioneering applications of Operations Research models to railroad systems. The articles by Kraft (2002a, 2002b) provide a more recent overview of yard operations, technologies, challenges and designs along with a good history of their development. The inline receiving/inline departure design shown in Figure 2.4 is one of several hump yard configurations (Figure 2.7).

Inline Receiving/Inline Departure



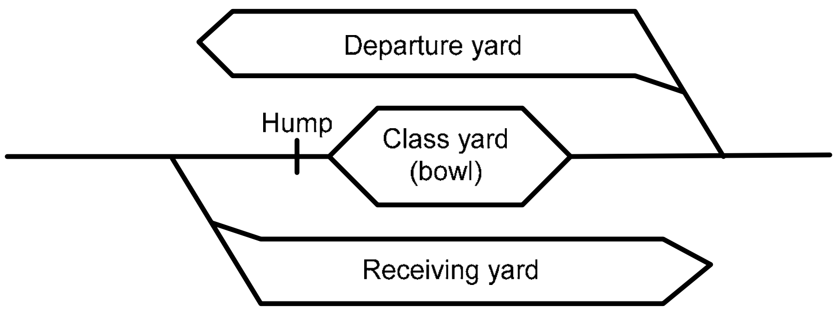
Examples: Bailey Eastbound Yard, North Platte, NE (UP),
DeWitt Yard, Syracuse, NY (CSX)

Inline Receiving/Parallel Departure



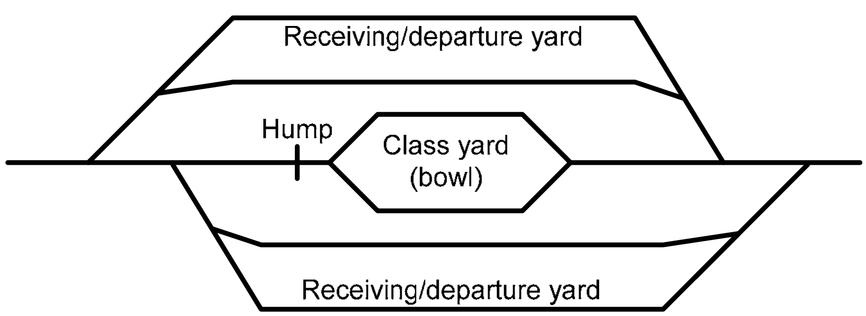
Examples: Northtown Yard, Minneapolis, MN (BNSF), Barstow Yard, CA (BNSF),
Selkirk Yard, NY (CSX), Spencer Yard, Linwood, NC (NS)

Parallel Receiving/Parallel Departure



Examples: Centennial Yard, Fort Worth, TX (UP), Argentine Yard, Kansas City (BNSF),
Avon Yard, Indianapolis, IN (CSX), Bensenville Yard, Chicago (CPR)

Combined Parallel Receiving/Departure



Examples: Alyth Yard, Calgary, AB (CPR), Agincourt Yard, Toronto, ON (CPR)

Figure 2.7 Most common hump yard design options (Wong et al. 1981) and examples (Rhodes 2003)

The inline receiving/inline departure and inline receiving/parallel departure designs are used where there is primarily one direction of traffic; from the left in this case. Consequently, very large terminals using these designs will most likely have one yard for each direction of traffic with each yard operated independently. The parallel receiving/parallel departure and combined parallel receiving/departure designs are used where traffic flows in both directions.

2.4 Production System Performance Measures

In order to effectively manage a production system, a set of measurements is needed to gauge the performance of the system over time and respond to changes. Railroads track this base set of production metrics at their classification terminals:

1. *Volume* – The standing car count in the terminal (or a sub-yard) at a particular point in time, the synonymous manufacturing term is Work-In-Process (WIP) inventory.
2. *Dwell time* – The average time taken for freight cars to move through the terminal (the time a car spends as part of the volume count), known as cycle time or throughput time in manufacturing.
3. *Throughput* – The average number of cars processed by the terminal (or a sub-process) per unit time (e.g., cars per day).

Because every car has a trip plan, there is also a quality aspect that must be considered when defining terminal throughput. This is done in one of two ways: measuring the number of cars whose dwell time was less than or equal to a standard established by the connection protocol (e.g., percent under x hours) or the percentage of scheduled connections met. As railroads have moved toward a scheduled operations approach these quality standards have become more important.

2.4.1 Performance Measurement Relationships

One of the first steps that Hopp & Spearman (2001) accomplish in Factory Physics is the establishment of fundamental relationships between the production system parameters. These relationships form the foundation for developing better production management methods. Queuing theory has provided a means for establishing the fundamental relationships both in manufacturing and in classification terminal operations. Petersen (1977a, 1977b) and Turnquist & Daskin (1982) have used queuing theory to model throughput rates and classification and connection delays. These models provide important insights about basic terminal dynamics, but because of their complexity, they are not effective as front-line management tools. Furthermore, some of the basic assumptions of queuing models do not apply in railroad terminal operations (Martland 1982).

However, there is one queuing theory relationship that is sufficiently intuitive to be used in a production control environment and it relates the three performance measures defined above: volume, dwell time and throughput. Hopp & Spearman (2001) use it as their first "Factory Physics" relationship. It is known as Little's Law (named for the man who provided the mathematical proof) and is written, using terminal performance measures, as:

$$\text{Volume (car count)} = \text{Throughput (cars per day)} \times \text{Dwell Time (days)} \quad (2.1)$$

Kraft (2002b) has graphed this relationship for terminals (Figure 2.8). Little's Law holds for all production systems over the long term, not just those with zero variability and it can be applied to a single station, a line or an entire plant (Hopp & Spearman 2001). This means that a terminal manager can use it at any level within the terminal, provided that all three quantities are measured in consistent units. For example, if a terminal manager knows the processing capacity of the terminal and the number of cars in the terminal, the expected average dwell time for those cars can be calculated.

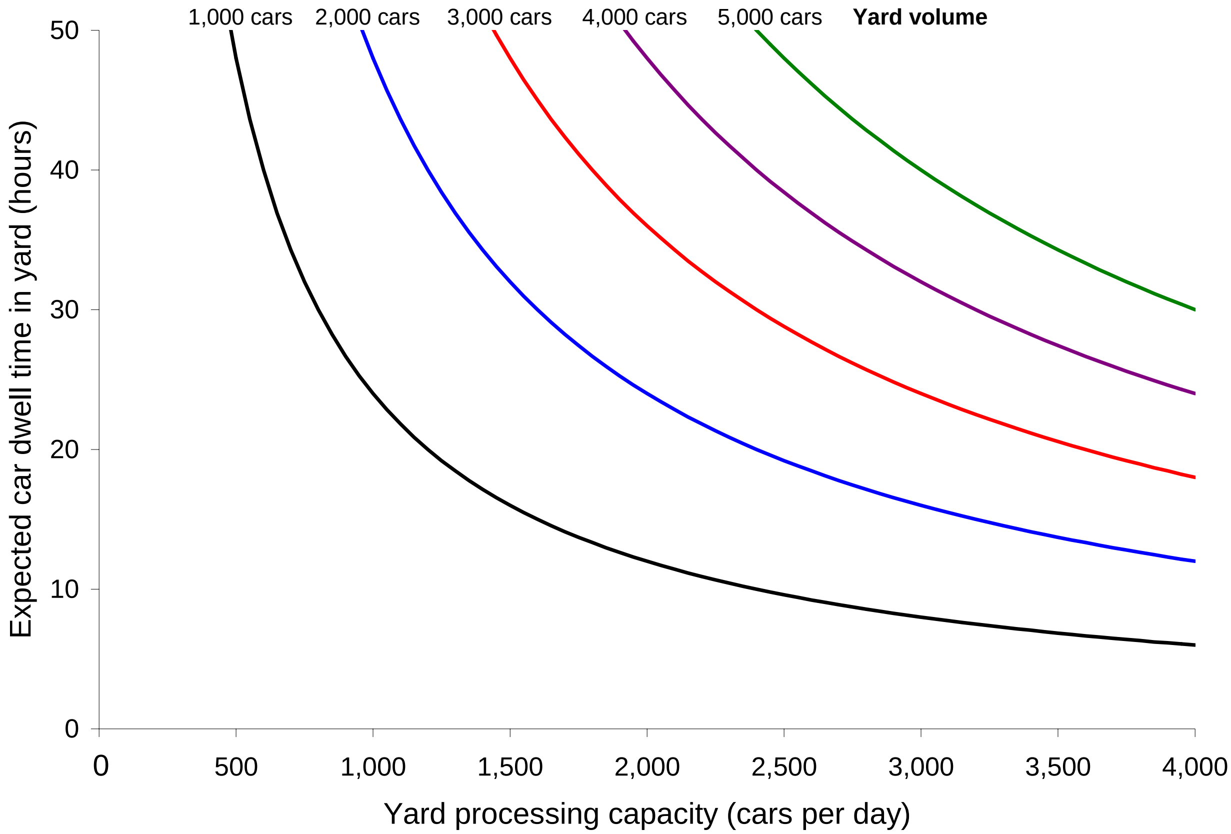


Figure 2.8 Little's law for terminals

2.5 Determining Capacity in a Production System

Capacity is defined as the upper limit on the throughput of a production process (Hopp & Spearman 2001). Releasing work into the terminal system at or above the processing capacity causes the terminal to experience unstable operations (i.e., build up volume above the standing capacity). Because every classification terminal has different design characteristics, operating practices and traffic mixes, it may be difficult to accurately determine a terminal's processing capacity. Experienced managers often have a relatively good understanding of the maximum number of cars that their terminal can process per day using current methods. In order to improve the processing capacity, managers need to better understand how capacity is determined in a production system. Factory Physics provides a means to understand processing capacity and guidance in improving terminal performance at the production control level. Knowing terminal processing capacity is a necessary part of effective terminal management.

2.5.1 The Theoretical Importance of the Bottleneck

Each process in a production system interacts with every other process in a variety of ways. The interactions can be simple (process 1 feeds parts to process 2) or complex (variability in train arrival times leads to a lack of motive power which leads to congestion in the departure yard). Having multiple product lines that use each process differently increases the complexity of the interactions. The science of production management provides insights that can increase capacity by focusing on the processes that will have the greatest impact on the system's performance.

In a production system, the bottleneck occupies an important position. The bottleneck is the process that limits the throughput of a production system. As such, the processing rate (throughput) of the bottleneck process establishes the long-term processing capacity of the entire system. The important relationship between the bottleneck rate and production system performance provides key insight into where improvement initiatives should be targeted. Because bottleneck rate and throughput have the same units (cars per day for terminals), Little's Law can be used to quantify the benefits of improving the bottleneck rate:

$$\text{Bottleneck rate} = \text{Terminal throughput (cars per day)} = \frac{\text{Volume (car count)}}{\text{Dwell time (days)}} \tag{2.2}$$

Increasing the bottleneck rate will reduce the dwell time for any given volume level in the terminal (Figure 2.8). Knowing this, management can reduce the complexity of managing a complex production system by focusing on improving the bottleneck process. This will lead to a reduction in terminal dwell time and a corresponding increase in terminal capacity.

2.5.2 The Estimated Benefits of Increasing Terminal Capacity

Little's Law (Equation 2.2) was used to illustrate the potential benefits of increasing capacity for each of the four yards most closely studied during this work. The average daily yard volume and average terminal dwell time for 2004 were calculated. Daily volume numbers were provided by railroad management (CPR 2005a, CN 2005a). The average terminal dwell for the CPR yards was calculated using the weekly dwell time results from the railroad performance measures (CPR 2006b). Average terminal dwell for MacMillan Yard was calculated using data provided by CN (2005a). The capacity of each yard at the current level of volume and dwell time was estimated using Equation 2.2 (Table 2.2). Then, volume was held constant while capacity was increased and the resultant terminal dwell was plotted (Figure 2.9).

Table 2.2 Yard performance measurements, 2004

Yard	Railroad	Average daily yard volume (cars)	Average terminal dwell (hours)	Estimated average capacity (cars per day)
Alyth (Calgary, AB)	CPR	1,988	33.9	1,407
Agincourt (Toronto, ON)	CPR	2,432	51.2	1,140
Bensenville (Chicago, IL)	CPR	1,476	25.3	1,400
MacMillan (Toronto, ON)	CN	4,509	28.1	3,851

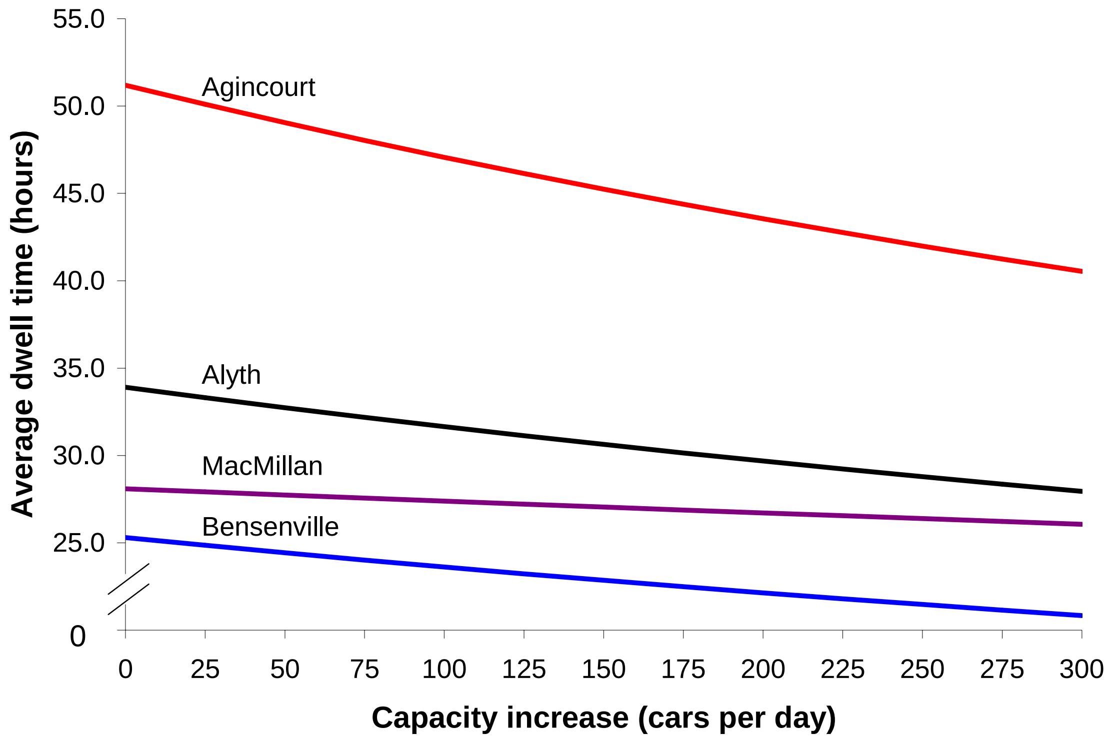


Figure 2.9 Estimated dwell time reduction as capacity is increased

In all cases, dwell time decreased as capacity was increased and the rate of change in dwell time was nonlinear. An explanation for the different rates of decrease for the dwell time curves can be understood by considering Figure 2.8. The starting dwell time and volume level for each yard is at a different point on the curves in Figure 2.8; consequently, the rate of decrease for each yard is also different. A 300 car per day capacity increase is estimated to reduce the average dwell time at Alyth by 6.0 hours, Agincourt by 10.7 hours, Bensenville by 4.5 hours and MacMillan by 2.0 hours. As discussed in Section 1.4.1, reductions in average system terminal dwell time are correlated with average manifest train speed. Dwell time is reduced by increasing capacity and increasing terminal capacity will also improve service reliability (Section 1.4). Therefore, management should work to improve terminal capacity by better management of resources by adapting production management methods. Chapter 3 presents a new approach for this improvement initiative.

CHAPTER 3: LEAN RAILROADING

In this chapter I introduce a new approach to improving terminal performance that combines the concept that a terminal is a production system with the foundation laid by Factory Physics and three production management methods: Lean Manufacturing, the Theory of Constraints (TOC) and Statistical Process Control (SPC). None of these production management methods are new, nor is their application to the transportation sector, as illustrated by the previously cited application of SPC to classification yards by Mundy et al. (1992) and by the concept of agile ports (Paixao & Marlow 2003).

> "The implementation of agility in ports supports itself on the concepts of lean, flexibility, just-in-time and business process redesign (or re-engineering) techniques (since their implementation has contributed to the achievement of good results in the manufacturing environment there is no reason to believe that the same results cannot be achieved in a service sector, such as the port industry)." (Marlow & Casaca 2003, p. 190)

What is new is the documented, integrated application of these methods to classification terminal systems in the railroad industry. At least two railroads (CPR and UP) have begun to apply the principles of lean to improving classification terminal performance. GE Yard Solutions has conducted lean terminal studies on the Belt Railway of Chicago, BNSF, NS and UP (Logan 2006b). BNSF has an active value engineering and "six sigma" (another term for SPC) group. And many of the "precision railroading" principles that CN has used to improve their operating performance can also be considered lean. In this chapter, I review this work by railroads and develop a summary description of "Lean Railroading."

The Lean Railroading approach to improving terminal performance will first be developed by providing a brief history of lean in relation to other production management systems and defined using Factory Physics. The tools used to identify and implement improvement initiatives will be described and applied to the terminal system. Then, to illustrate how to apply these concepts and tools, the TOC will be incorporated to provide guidance as to where the first improvement initiatives should be targeted. SPC methods and tools will be incorporated into the detailed description of the pull-down process in Chapter 4 and the development of the Quality of Sort metric in Chapters 5 and 6.

3.1 Lean Manufacturing

Lean Manufacturing has its roots in the Toyota Production System (TPS), an approach to manufacturing pioneered by the Toyota Company after World War II. TPS enabled Toyota to experience steady productivity and quality improvements throughout the 1970's while U.S. manufacturers were in the middle of the "MRP crusade" (Spearman 2002a). MRP (Material Requirements Planning) and its successor, MRP II (Manufacturing Resources Planning), were efforts to use computer software to control all aspects of material flow in a production system.

While sales of MRP continued to climb in the 1980's, many began to think MRP was a mistake. Less than 10% of 1,100 firms who were interviewed in 1980 were able to recoup their investment within two years (Hopp & Spearman 2001). U.S. automobile manufacturers, in an effort to catch up to Toyota, adopted one aspect of TPS in the 1980's, Just-In-Time (JIT). In spite of the benefits experienced by JIT firms, a new version of business application software emerged, Enterprise Resource Planning (ERP) that targeted all operations of a company. ERP was more complex and much more expensive than MRP II. In spite of the high cost and a growing number of "implementation horror stories" (see Hopp & Spearman 2001, p. 176), companies installed ERP in an effort to more effectively control their entire enterprise.

At the same time, TPS continued to be studied and receive attention. The term "Lean Production" was created by Womack et al. (1990) as part of the five-year study of automobile production methods conducted by the International Motor Vehicle Program (IMVP) at MIT. The study compared American, European, and Japanese automobile manufacturing techniques and concluded that the Japanese techniques, particularly those of Toyota, were superior (Spearman 2002a). JIT, along with other components of TPS, became Lean Manufacturing and companies who studied, implemented and embraced it, soon outpaced their ERP counterparts. MRP, MRP II and ERP did provide important contributions to the body of manufacturing knowledge, but the assumptions of infinite capacity and fixed lead times in the basic model underlying these systems were flawed (Hopp & Spearman 2001, p. 145, 174). Womack & Jones (2003) describe the application and results of the implementation of "Lean Thinking" at several manufacturers around the world (including Toyota, Lantech, Wiremold, Porsche and Pratt & Whitney) and explore possible applications outside of traditional manufacturing.

3.2 Defining Value for a Railroad Customer

The focus of lean is the identification and elimination of waste (known as *muda* in lean terminology). Waste is defined as any step or process in a production system that, from the standpoint of the customer, does not add value to the product (Rother & Shook 1999).

> "The critical starting point for lean thinking is *value*. Value can only be defined by the ultimate customer. And it's only meaningful when expressed in terms of a specific product (a good or a service, and often both at once) which meets the customer's needs at a specific price at a specific time." (Womack & Jones 2003, p. 16)

A railroad customer wants to purchase the product of freight mobility. The first step in defining value is to define the customer's ideal product. In other words, specify the product that provides the maximum value to the customer. For freight mobility, the ideal is:

1. Instantaneous delivery of the exact number and type of empty, clean and defect-free railcars specified when ordered
2. Instantaneous pick-up of those railcars when loaded and released
3. Non-stop movement of those railcars from origin to destination without delay
4. No damage to the freight being transported
5. Exact, on-time delivery of every railcar shipped (100% service reliability)
6. Instantaneous removal of railcars once unloaded
7. All of this done with the maximum level of safety and for the minimum cost

The purpose of defining the ideal product is to provide railroads with a concrete goal that will enable a culture of continuous improvement to pervade the company. With that in mind, waste is now defined as any activity or event that results in the ideal product not being provided. On a railroad, waste occurs when:

A. The railroad cannot deliver the product (i.e. it must turn away business that would otherwise help maximize profits)
B. There is a delay between order and delivery of empty railcars
C. The wrong number or type of empty railcars are delivered
D. The empty railcars are dirty or unusable because of defects
E. There is a delay between the release and pick-up of the loaded railcars
F. Anytime the railcars are not moving either in a train or in a terminal
G. A connection is missed
H. The railcar or the freight is damaged
I. There is a delay between the release of the unloaded cars and their removal
J. Any unsafe activity occurs
K. Any activity or asset that increases costs beyond the minimum required to deliver the product (i.e. excessive fuel consumption, unnecessary labor cost, overpowered trains, additional surge fleets of locomotives and freight cars, excessive or overbuilt track/signal systems, etc.)

Therefore, reducing the occurrence and/or magnitude of these activities must be a good thing. Unfortunately, it is never that simple and Spearman (2002b) explains why.

> "We seldom reduce any single type of waste . . . without increasing another. By reducing lot sizes, we reduce inventory (a form of waste) but also increase both the amount of material handling and the number of setups (also forms of waste). Likewise, the sign of the waste equation is unclear when we add safety stock inventory (waste) to reduce lost sales (also waste)."

A railroad could run shorter, more frequent trains to reduce yard inventory levels (and reduce dwell, Figure 2.8) but this would create a need for additional locomotives, crews and effectively reduce mainline capacity because of the fixed block length traffic control system typical of most railroad signal systems (Pachl 2002). Additional railcars could be purchased or leased as a safety stock (waste) to reduce lost sales (also waste).

Recalling the importance of classification terminals to service reliability and network efficiency discussed in Chapter 1, and the operations that occur in a terminal presented in Chapter 2, it is apparent that terminal operations directly or indirectly have the potential to affect nearly every type of waste in the railroad system. And because terminals can be considered production systems, they present the best opportunity to apply and illustrate lean techniques to railroaders. However, before the successful implementation of the concept of Lean Railroading can occur, an understanding of the root cause of waste must be established. Again, use of Factory Physics can provide this understanding.

3.2.1 The Root Cause of Waste

Naturally, some waste is the result of poor railroad practices. Unnecessary moves, mistakes that require an operation to be repeated, badly designed track layout that leads to excessive handling, inefficient dispatching, inadequate track maintenance and unsafe operations are just a few examples of railroad practices that represent direct waste. Focusing on these is important, but the goal of eliminating direct waste is as old as the railroad system itself.

Variability is a fundamentally different source of waste. Its importance is such that Hopp & Spearman (2001, p. 295) state, as a Law of Manufacturing, that, "Increasing variability always degrades the performance of a production system." Railroad terminals are no different. Consider what the *perfect* performance of a classification terminal system would look like. Trains would always arrive on schedule, each process would operate at 100% utilization (no wasted capacity) and would only work on processing railcars (i.e. no setup or down time). There would be no trains waiting to leave the departure yard and there would be no waiting for trains to be built (i.e. the mainline would ask for the train as soon as it was built). Similarly, there would be no inbound trains waiting in the receiving yard and processes would finish "just-in-time" for the next process to begin. The only railcars in the yard would be the ones being processed and there would be none waiting in a queue. No rework or car damage would ever occur. "Obviously, because there is always some variability (some under our control, some not) perfect performance is impossible" (Spearman 2002b).

While variability takes many forms, it can be divided into two classes: internal and external (Spearman 2002b). Some examples of internal and external variability experienced by terminals are presented in Tables 3.1 and 3.2 respectively.

Table 3.1 Examples of internal variability

Type	Examples
Planned outages	Weekly hump maintenance
Unplanned outages	Locomotive failure, retarder failure, injuries, accidents
Variable process times (including setup times)	Inspection, classification, pull-down
Rework	Planned and unplanned rehumps
Car damage	Broken knuckles, overspeed impacts
Sorting	Block-to-track assignment, misroutes, block splits
Workplace variation	Manager experience, methods, mentality
Crew variation	Crew experience, cooperation, motivation

Table 3.2 Examples of external variability

Type	Examples
Arrival variability	Late inbound trains, early inbound trains
Weather	Wind, rain, snow, cold
Mainline outage	Derailment, grade crossing accident, washout, rockslide
Traffic volume	Inbound trains of differing lengths, outbound trains exceeding capacity
Car damage	Cars on inbound trains damaged
Traffic flow	Shifts in directional flow of traffic
Yield management	High-priority cars, service differentiation
Departure	Bunching of depatures in plan, mainline blocked when train ready

Most of the examples of internal variability are true sources of waste and efforts should be taken to eliminate or reduce their occurrence. Dealing with external variability is more challenging because they can cause as much waste as internal sources but lie outside the control of the terminal. Eliminating these sources may be impossible (i.e. the weather) or dependent upon other entities within the railroad system (i.e. arrival variability). Coordination with the outside entities is imperative to reduce the waste caused by external sources, but many times the sources should not be eliminated (i.e. yield management).

Another Law of Manufacturing from Factory Physics is "Variability in a production system will be buffered by some combination of inventory, capacity and time" (Hopp & Spearman 2001, p. 295). In a classification terminal, an inventory buffer is seen in the form of railcars sitting in the arrival, classification or departure yards. A capacity buffer takes the form of a process throughput greater than the process demand. A time buffer is the extra time built into each car's trip plan in order to ensure that the connection will be made and is reflected in the terminal dwell time.

Spearman (2002b) states, "In many ways, the 'waste' discussed in Lean is the 'buffer' of Factory Physics. However, this is not always the case. If external variability creates the need for a buffer, is it waste?" Providing different service levels increases variability, but would the railroad be better off if it were to only offer one service level? "The point is that while not all variability is waste, all variability *will* lead to a buffer which indicates that *logistical* (but not necessarily financial) performance has suffered" (Spearman 2002b). As long as the increase in railroad revenue is greater than the increase in operating costs, profits will increase. Therefore, it becomes the task of terminal management to reduce internal variability and the task of network management to manage the external variability so that the bad sources (like arrival variability) are reduced and the good sources (like service level differentiation) increase profit.

3.3 Implementing Lean Railroading

Logan (2006a) suggests an integrated approach combining process (lean terminal studies), technology (remote control locomotives, remote control switches, yard control systems) and people (manager & yardmaster training, job manuals, on-the-job training) to improve terminal performance and assumes that the benefits are additive. CPR advocates a similar integrated approach combining people, technology, schedule adherence and proper metrics (CPR 2006b). These integrated approaches seek to ensure the effectiveness and sustainability of the change initiatives required by Lean Railroading.

With the integrated approach in mind, the steps for implementing Lean Railroading in terminals can be defined. The steps are adapted from Spearman (2002b) using the experiences of CPR's Yard Operations Performance Group (CPR 2006c) and Logan (2006b).

0. **Eliminate direct waste**: Take a fresh look at the terminal system by drawing a Value Stream Map (VSM) and try to eliminate obvious sources of waste. The list would include (but is not limited to):
   a. Rework
   b. Accidents, injuries and other safety problems
   c. Car damage
   d. Unnecessary motion
   e. Yard engine failure
   f. Long setups
   g. Unnecessary information collection
1. **Swap buffers**: Decrease the time buffer (dwell time) by reducing the idle time between processes. This is synonymous to enabling continuous flow. In order to accomplish this, develop a production control schedule based on the departure schedule and set standards for each process. Then, stack the processes in a fashion that compresses idle time. Increase the capacity buffer by focusing on improving the performance of the bottleneck.
2. **Reduce variability**: This is done in several ways:
   a. By addressing problems in sorting, rework, car damage, down time and setups. The reduced dwell time results in lower yard volume and this facilitates tracing problems to their source. The time between "defect creation" and "defect

detection" is reduced. Use of SPC or "six sigma" techniques is appropriate here (see Mundy et al. 1992).

b. Implement standardized work plans to reduce variability in process cycle times.

c. Work with network management to increase the on-time reliability of inbound trains. The yard must also do its part to ensure on-time departure of outbound trains.

d. Fix the network operating plan to "schedule no more 'core' trains than the carrier is committed to operate reliably every day" (Kraft 2002c, p. 94). This is the approach that CN has adopted and it reduces variability in demand as seen by the yard. Trains are never cancelled but extra trains can be added if the demand warrants. Then, the planned workload of the yard can be spread out over the entire day to minimize traffic peaks that exceed terminal processing capacity.

3. **Continuous improvement**: "Once variability is significantly reduced, we can reduce the capacity buffer while continuing to identify and eliminate variability. Only at this point do we begin to make real gains in productivity. If we do not reduce variability, we will not be able to reduce the capacity buffer without hurting customer responsiveness. The result is a system that continues to improve over time" (Spearman 2002b, p. 5).

While Lean Railroading can be applied to any terminal, the implementation will be more successful when scheduled operations are in place. The first Class I to fully adopt scheduled railroading in the 1990's was CN. Since then, the transition towards scheduled operations by all of North America's Class I railroads has resulted in additional capacity, increasing overall velocity and tonnage (Gormick 2005) along with on-time deliveries and cost savings. Kraft (2002c) explains why a scheduled railroad environment lays the foundation for improved terminal performance.

> "Operating trains closer to schedule is an important first step, since it helps terminal managers plan and control their operations and, by promoting a smoother flow of traffic, it may well reduce variability in traffic levels as well. But while a scheduled railroad produces conditions under which terminal operations may be improved, it does not by itself fundamentally change the nature of switching or car sorting within yards."

Scheduled operations reduce the external variability, which enables managers to better focus on eliminating direct waste, swapping buffers and reducing the internal variability.

3.3.1 The Value Stream Map

Identifying and eliminating direct waste involves the use of Value Stream Mapping techniques. A value stream is all of the actions required to bring a product through the three critical management tasks of any business: the problem-solving task, the information management task and the physical transformation task (Womack & Jones 2003). For railroads, the problem-solving tasks are decisions on where to run trains, service frequency, routing, blocking plan, what equipment to use, infrastructure requirements, etc. The information management task involves the reservation/car ordering system and the operating schedule. The physical transformation task is the actual operation of trains over specific routes, the classification terminal operations and the required maintenance to support the daily operation. As discussed above, terminals can be considered production systems; therefore, this level in the total value stream (consisting of the loaded and empty value streams, Figure 3.1) is a good place to begin the mapping and lean implementation effort.

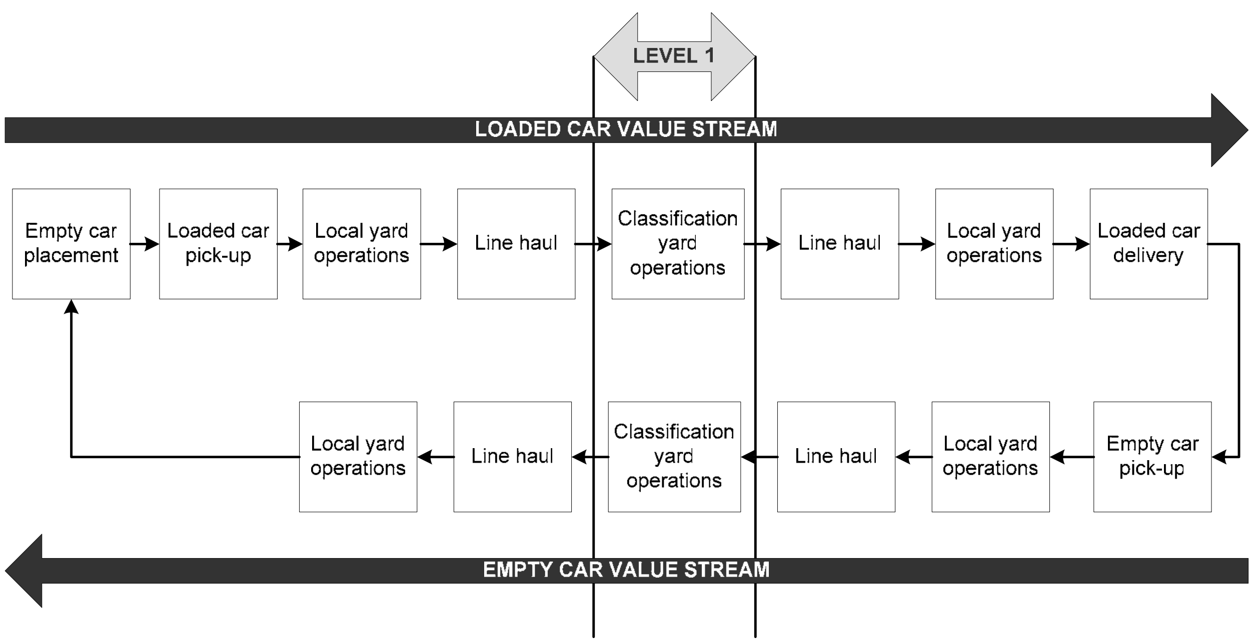


Figure 3.1 Total value stream for freight mobility as provided by a railroad

A VSM shows all of the processes and the related information flow that must occur in order to get the finished product to the customer. All of the processes are classified as either those that add value to the finished product (value-adding) or those that do not (non-value-adding). The total time that a product spends in the system is compared to the time that value is being added. Improvement efforts focus on reducing total processing time by reducing non-value-added time. At least two major North American railroads are using versions of VSM as part of their efforts to implement lean methods in their organizations (CPR and UP). Preparation of a VSM for a classification terminal enables one to focus on understanding the flow (Figure 3.2).

The current-state VSM allows for everyone on the team to see the entire value stream being studied and agree on its current level of performance. It "helps managers envision the initial flow *kaizen* needed to drastically compress the throughput time for the product, eliminate wasted steps, and rectify quality, flexibility, availability, and adequacy problems" (Womack & Jones 2003). *Kaizen* is a Japanese term meaning "continuous incremental improvement." Management then works in concert with the operators to develop the future-state VSM. The future-state VSM is used by the value-stream manager to develop the implementation plan that will enable the system to achieve the future-state. Follow-through initiatives are required to sustain the cycle of continuous improvement.

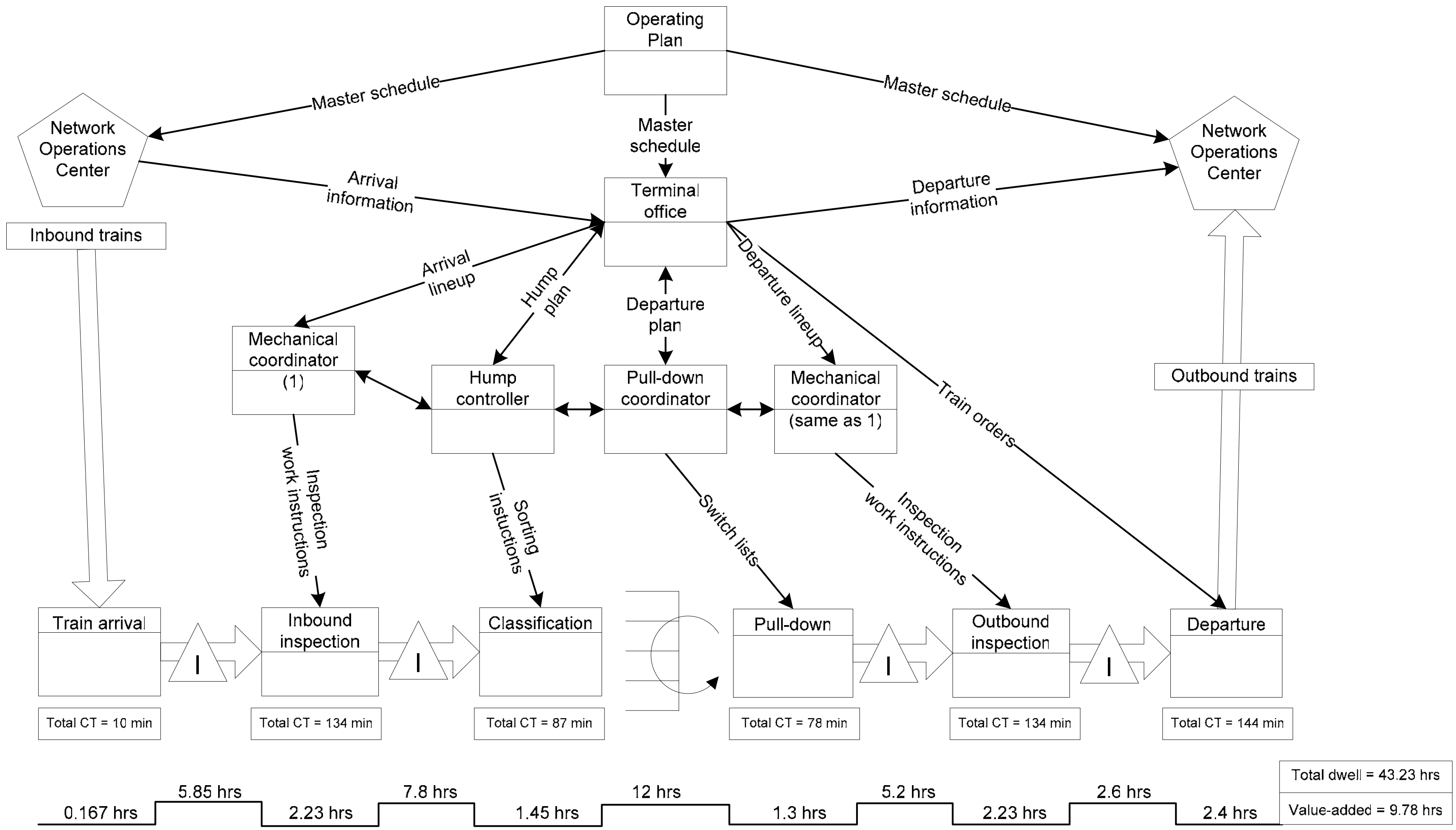


Figure 3.2 Example classification terminal current-state value stream map-simplified (Jacobs 2005, Rother & Shook 1999, McClish 2005)

3.3.2 The Theory of Constraints (TOC)

TOC provides a structured approach to improving production system performance by focusing on the system's bottleneck. Goldratt (1990) has established the general process in the TOC approach. For any production system, the TOC approach is:

1. Identify the system's constraint
2. Decide how to exploit that constraint
3. Subordinate the remaining resources to the decision in the previous step
4. Elevate the system's constraint
5. If in the previous steps the constraint has been overcome, go back to step one

Terminals have few actual constraints (although many more are often perceived) but must have at least one. Step 1 means identifying the actual constraints and focusing improvement efforts on the one that impacts the objective (or The Goal in TOC parlance) the most. From the Factory Physics standpoint, the most important constraint is the bottleneck. Exploiting the bottleneck (Step 2) means managing it in a way that maximizes its throughput. This goes hand-in-hand with Step 3 since the remaining resources (the non-constraints) should be managed so that they provide the bottleneck exactly what it needs and nothing more. Efforts should continually be made to elevate the bottleneck (Step 4) until it is broken and a new constraint becomes the most limiting to the system (Step 5). At this point, the process begins again at Step 1 as the new system constraint is identified.

3.4 Identifying a Terminal's Bottleneck from the VSM

For this study, the TOC approach was used with the current-state VSM to identify the bottleneck in the terminal system. The total terminal dwell time for this particular terminal (43.23 hours) and the value-added time (9.78 hours) are located in the lower right corner of the VSM (Figure 3.2). From the standpoint of the customer (the Network Operations Center), the products are the outbound trains and the value-added time is the time that the cars and blocks spend in processes that are required to build the outbound trains. These times are located in the lower sections of the timeline. Any other time spent in the terminal is non-value-adding (waste) and these times are located in the upper portions of the timeline. For this example, it is clear that a majority of time spent in this terminal (33.45 hours or 77.3%) is non-value-added idle time that the cars and blocks spend waiting for the next process to begin. A more detailed time-in-motion study by the GE Yard Solutions group of one classification yard provides a similar idle time figure of 71% of the 28.2 hour dwell time (Figure 3.3) and Logan (2006a) estimates that focusing on idle time reduction presents an opportunity for a 15-30% capacity improvement in terminals.

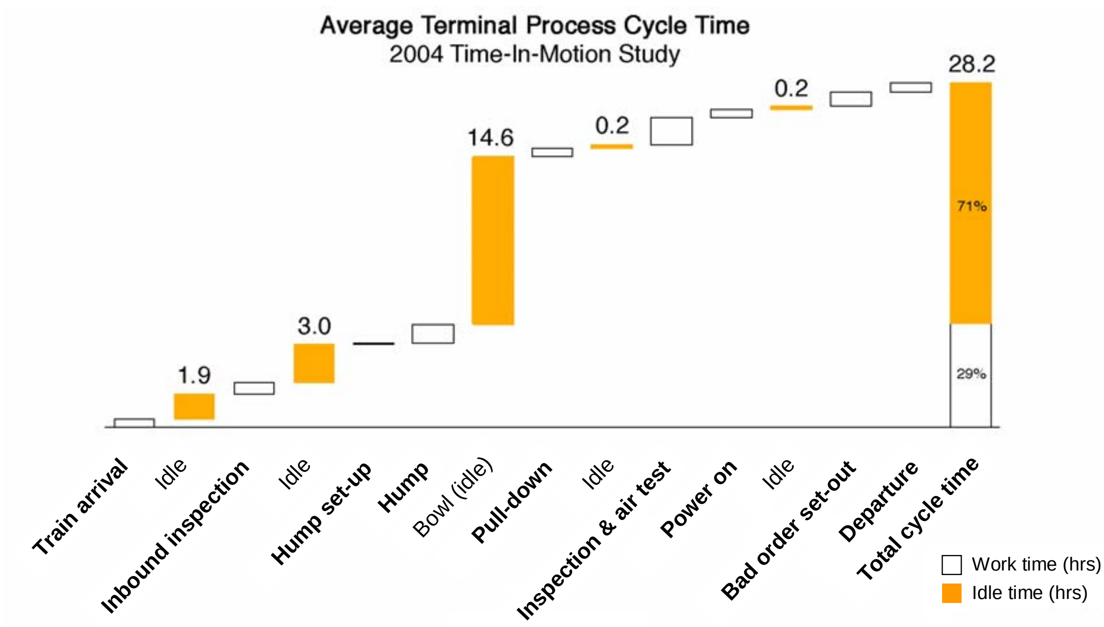


Figure 3.3 Average terminal process cycle time from GE (Logan 2006a)

The largest portion of the idle time (12 hours in Figure 3.2 and 14.6 hours in Figure 3.3) is spent in the classification yard (or bowl). A disproportionately long wait time immediately upstream from a production process is a good indicator that process is the bottleneck of the system (Sobek 2003). This suggests that the pull-down process is the bottleneck. This is corroborated by previous studies (Petersen 1977a, Petersen 1977b, Kraft 2002c) and by railroad management at several levels at CPR (2005b), CN (2005b) and UP (McClish 2005).

CHAPTER 4: THE PULL-DOWN PROCESS

In spite of the theoretical importance of the bottleneck in production systems, relatively little work has been done documenting and understanding the rail yard pull-down process in detail. Identification of the best approaches to improve bottleneck performance requires a thorough understanding of the process and the factors that determine its capacity. In this chapter I consider the pertinent parts of previous yard and terminal studies and combine them with information gathered during site visits to develop a detailed analysis of the steps required to assemble blocks into trains. My own site visits to gain a general understanding of yard operations were conducted at Champaign Yard (CN) in fall 2004, Decatur Yard (NS) in September 2004 and Denver Yard (BNSF) in October 2004. More detailed visits focusing on hump yard operations were conducted at Alyth Yard in Calgary (CPR) in March 2005, Agincourt Yard in Toronto (CPR) in July 2005 and Roseville Yard (UP) in November 2005.

The first site visit in which I focused primarily on the pull-down process was at MacMillan Yard in Toronto (CN) in July 2005. Most of the management and crew feed-back and observational time study data were gathered at Bensenville Yard in Chicago (CPR) in the spring of 2006. The insights of the Yard Operations Performance (YOP) group at CPR and Prescott Logan of GE Yard Solutions also enhanced this portion of the research.

4.1 Process Overview

The pull-down process (also called “trimming” or train assembly) consists of blocks of cars being pulled from the classification tracks (bowl) and placed together to form outbound trains in the departure tracks. Every terminal has an operating plan that details the blocks that comprise each outbound train and the order in which the blocks are to be placed in the train. A sample taken from the Alyth Yard operating plan is presented in Figure 4.1. Train 274 is comprised of seven different blocks and operates seven days a week. The blocks are listed in reverse order in the train (i.e. from rear to front) (Figure 4.2) and all of the blocks are built at Alyth (no block swaps). The train is ordered at 15:00 and is scheduled to depart at 17:00 for a total station time of 2:00 hours. The plan is frequently updated as the service planning department changes block and train routings to meet changing network requirements. The pull-down process follows the operating plan when assembling the outbound trains.

TRAIN 274	**Operates:**	**Order FROM REAR:**	**Order (train):**	**Departs:**	**Station time:**
Block name	SMTWTFS	ClassCode	1500	1700	2 hr 00 min
Alliston empties		**3410EA1**	Lift		
Toronto autos		**3173FA1**	Lift		
Toronto manifest		**3173MA1**	Lift		
Thunder Bay manifest		**5002MA1**	Lift		
Winnipeg manifest		**5200MA1**	Lift		
Regina manifest		**6016MA1**	Lift		

Figure 4.1 Train 274 block to train assignment from Alyth Yard operating plan (CPR 2005c)

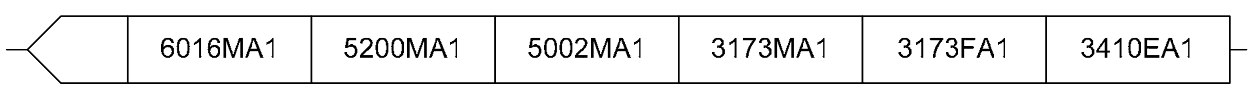


Figure 4.2 Block order for Train 274

The pull-down process takes place at the pull-down (also called pull-out or trim) end of the bowl (Figure 2.4) where the many tracks of the classification yard are funneled into a small number of switch leads. For example, Alyth Yard has 48 tracks in the bowl and three leads for use by pull-down jobs (Figure 4.3). As is the case with many yards, these pull-down leads also double as arrival and departure tracks for trains entering and exiting the yard. Because of this, interference among the pull-down jobs, and between those jobs and other trains, is a common occurrence.

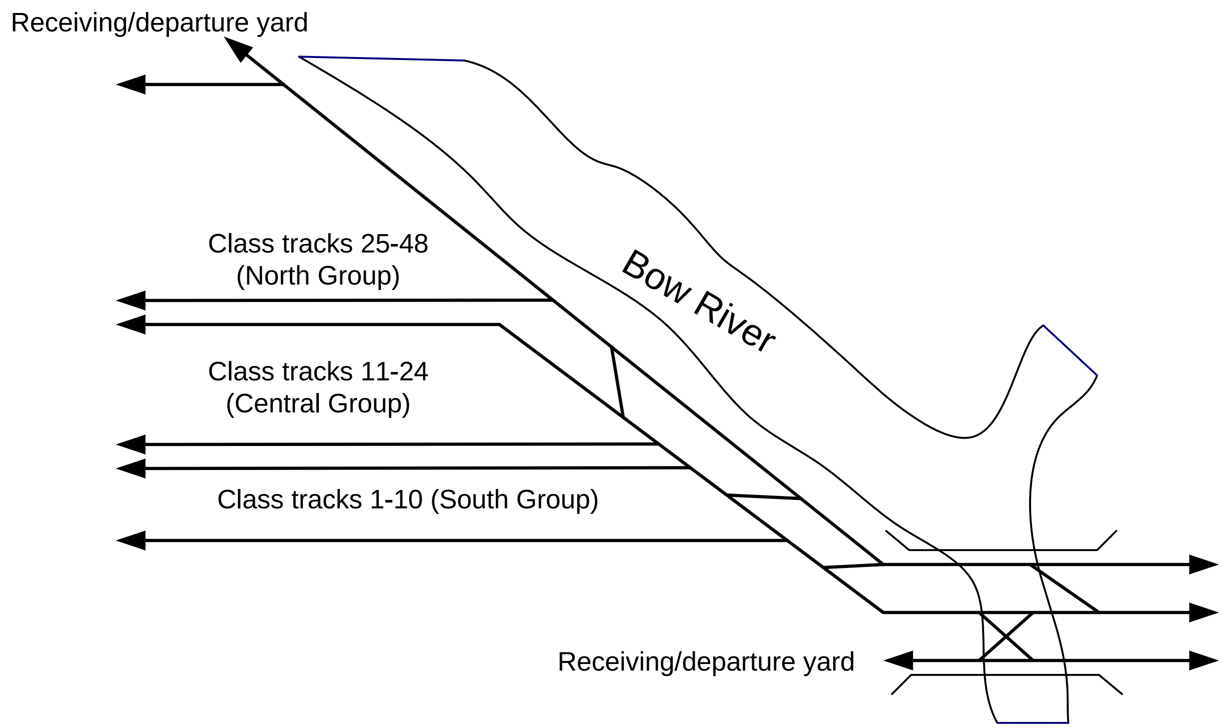


Figure 4.3 Pull-down area diagram for CPR's Alyth Yard in Calgary (Middleton 2003)

4.1.1 Crew Requirements and Equipment

There are two types of equipment used for pull-down operations: conventional and Remote Control Locomotive (RCL). In a conventional operation, each pull-down crew usually consists of two people, a locomotive engineer and a switchman. Each crew is assigned a switching locomotive that is operated by the engineer while the switchman handles all the work on the ground. They communicate via radio and hand signals. Depending on the size of the terminal and management preference, crews will either be directly supervised by a pull-down coordinator from a tower at the pull-down end, or by a yardmaster in the yard office. The yard office is usually located next to the hump. A General Yardmaster (GYM) coordinates all of the traffic within the yard limits. In some yards, they are assisted by utility switchmen who are assigned to an area of the yard or rove in vehicles to throw switches, spot cars and set hand brakes. Shifts are usually eight hours long and most major classification terminals will have multiple crews working during each shift.

Operations involving RCL technology blur the craft lines of locomotive engineers and switchman but have generally resulted in increased productivity, flexibility and safety (CN 2000). RCL is more accepted in Canada because the technology originated there and more flexible labor union work rules enabled CN and CPR to use it in yards and terminals. However, it is becoming the standard for a majority of yard and local switching work in the United States. As of February 2005, there were an estimated 1,500 RCL systems in use by North American carriers including 23 U.S. railroads (Luczak 2005). All the yards visited in Canada for this work used RCL equipment for humping and pull-down switching operations. However, Bensenville Yard, which was used for the work process flow diagrams and time studies, did not use RCL when this study was conducted.

4.2 Determining Pull-Down Capacity

A macroscopic evaluation method is presented in Wong et al. (1981) for use designing new yards or redesigning old yards. The method also serves as an excellent starting point for evaluating the potential impact of different improvement strategies for existing yards. Equation 4.1, from Wong et al. (1981), estimates the capacity of the pull-down end for a parallel departure yard design:

$$C_P = \frac{T_M \cdot N_E \cdot N_C}{(T_H + T_L + N_D \cdot T_D)\,(1.0 + C_F) + T_C} \tag{4.1}$$

where: C_P = Capacity of the pull-down end (cars/day)
T_M = Productive crew time (min)
N_E = Number of pull-down engines
N_C = Average number of cars per block pulled
T_H = Average travel time from the classification yard to the departure yard (min)

T_L = Average travel time from the departure yard to the classification yard (min)
N_D = Average number of doubling maneuvers to be made per pull
T_D = Average time required to complete a doubling maneuver (min)
C_F = Conflict coefficient
T_C = Average coupling time to couple an average size block (min)

This equation will be refined with additional detail to increase its robustness and used to evaluate the effectiveness of several bottleneck management improvement options in Chapter 5.

4.2.1 Operational Methods

The two major activities performed by pull-down crews are coupling cars on the classification tracks and pulling cars from the classification yard to the departure yard (Wong et al. 1981, p. 146). Pull-down operational methods are closely related to the design of the pull-down end of the yard and the orientation of the departure yard to the classification yard (Figure 2.7). In parallel departure yard designs, the method of making up trains can vary. The first method involves an engine pulling the cars on one track directly to the departure yard and will be referred to as "single pull". In the second method, engines pull cars from several tracks and then move them as a group to the departure yard, which will be referred to as "multiple pull". In inline design departure yards, trains are usually built using the multiple pull method.

> "The engine usually pulls one block from a classification track, then without pulling that block all the way to the departure track, shoves back to the next classification track with the first block still intact, and then doubles the blocks. In this manner the engine doubles all the blocks and makes up a train by doubling blocks in the trim-end area. After finishing a complete train, the engine then pulls the entire train to the departure yard. The trim engine travels back to the classification yard via an empty departure track." (Wong et al. 1981, p. 146)

In practice, railroads tend to use the method most appropriate to the current operational situation at the yard. The yardmaster will decide what method or combination of the two (i.e. the crew on engine 2026 builds Train 291 with multiple pull and the crews on engines 1543 and 4608 build

Train 287 with single pulls) will be employed based on his or her preference, experience of the pull-down crews, work load, track maintenance and potential interference on the switch leads.

The Conflict coefficient (C_F) in Equation 4.1 indicates the effect of pull-down-end geometry on conflict between a pair of engine trips and was derived on the assumption that the number of non-conflicting routes between the bowl and the departure yard is either one or two (Wong et al. 1980, p. 149). Every additional engine beyond two increases C_F by 1.0. It is assumed here that increases less than 1.0 with no engines added can be used to represent the impact of increased crew interference that may result from poor coordination.

A detailed analysis of the pull-down process was conducted at Bensenville Yard near Chicago. Bensenville has a parallel departure yard design and both operational methods are used. However, because single pull was the predominant method used, N_D (the average number of doubling maneuvers to be made per pull) and T_D (the average time required to complete a doubling maneuver), can be used to reflect a similar activity, namely rework.

4.2.2 "Clean" and "Dirty" Tracks

The terms "rework," "clean" track and "dirty" track need to be defined before the cycle time components are described. The term "rework" is used because the pull-down process must correct (rework) the sorting of the hump process. Rework occurs on the pull-down end when tracks are "dirty" because at least one car is out of place or does not belong on that track. Kraft (2000b) defines a dirty track as one that has more separations than blocks, with a separation defined as a group of cars in standing order having the same block. The number of separations is determined by looking at standing order of cars on the track and counting the number of times the block changes.

A slight modification of Kraft's definition is used here because his term "separation" might be confused with a gap between cars on the classification track so Daganzo et al's (1983) term "cut" (Section 2.3) is used instead. Cut is standard railroad terminology that can be used to describe any situation where a set of cars that share a common destination track are sequenced together. A cut is a set of cars in standing order all having the same block. When there are more cuts than blocks, at least one car must be out of place on that track (Figure 4.4). Here, the term switch is also used again. A switch is required between every cut. Because the pull-down process is assembling blocks into trains, anytime there are more cuts than blocks, extra switches will be required to accomplish the task of building a particular train. This is the same as doubling a track, which is why N_D and T_D can be used to reflect the impact of rework.

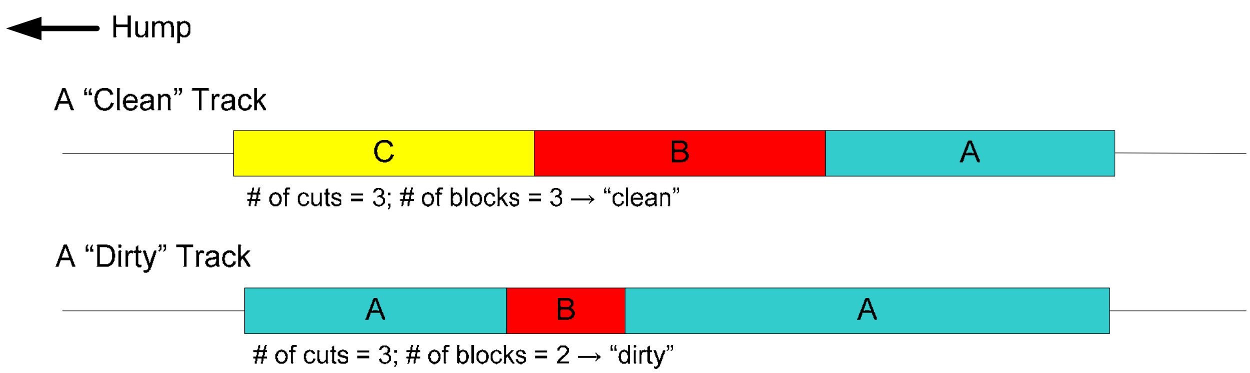


Figure 4.4 Modified definitions of "clean" and "dirty" tracks

4.3 Cycle Time Components

All of the time parameters in Equation 4.1 (T_M, T_H, T_L, T_D, T_C) need to be determined by conducting time studies. The first, the productive crew time (T_M), is the time that the crew is doing productive work. The maximum possible productive crew time is 1,440 minutes per day, minus the total minutes for meals and breaks (Wong et al. 1981). However, this value should be further refined to reflect real work conditions. Crews do not maintain a maximum pace every minute of the work day because of interruptions, fatigue and unavoidable delay (Niebel &

Freivalds 1999). Also, crews will exert different effort levels depending on a variety of factors such as skill level, motivation and age. The result is a reduction in the productive crew time and can be accounted for with Equation 4.2:

$$T_M = (1440 - M_B) \cdot P_F \tag{4.2}$$

where: T_M = Productive crew time (min)
M_B = Total meal and break time (min)
P_F = Performance factor

The remaining time parameters can be added together to calculate the cycle time of the pull-down process (Equation 4.3). The cycle time is the time it takes to complete one cycle of the process:

$$C_T = T_L + T_C + B_R \cdot T_D + T_H + D_R \cdot T_D \tag{4.3}$$

where: C_T = Average pull-down process cycle time (min)
T_L = Average travel time from the departure yard to the classification yard (min)
T_C = Average coupling time to couple an average size block (min)
T_D = Time required to complete a doubling maneuver (min)
T_H = Average travel time from the classification yard to the departure yard (min)
B_R = Bowl rework occurrence integer (0 or 1)
D_R = Departure yard rework occurrence integer (0 or 1)
$B_R + D_R \leq 1$

For the pull-down process, the cycle time begins when the crew receives the switch list from the yardmaster. It ends when the crew uncouples from the grouping of cars after placing them on the required track in the departure yard. The high-level process flow diagram in Figure 4.5 illustrates this procedure and breaks it into five cycle time components: setup (T_L), coupling (T_C), bowl rework ($B_R \cdot T_D$), transport (T_H) and departure yard rework ($D_R \cdot T_D$). It is assumed for this model that rework will occur at most one time per pull; either in the bowl or in the departure yard.

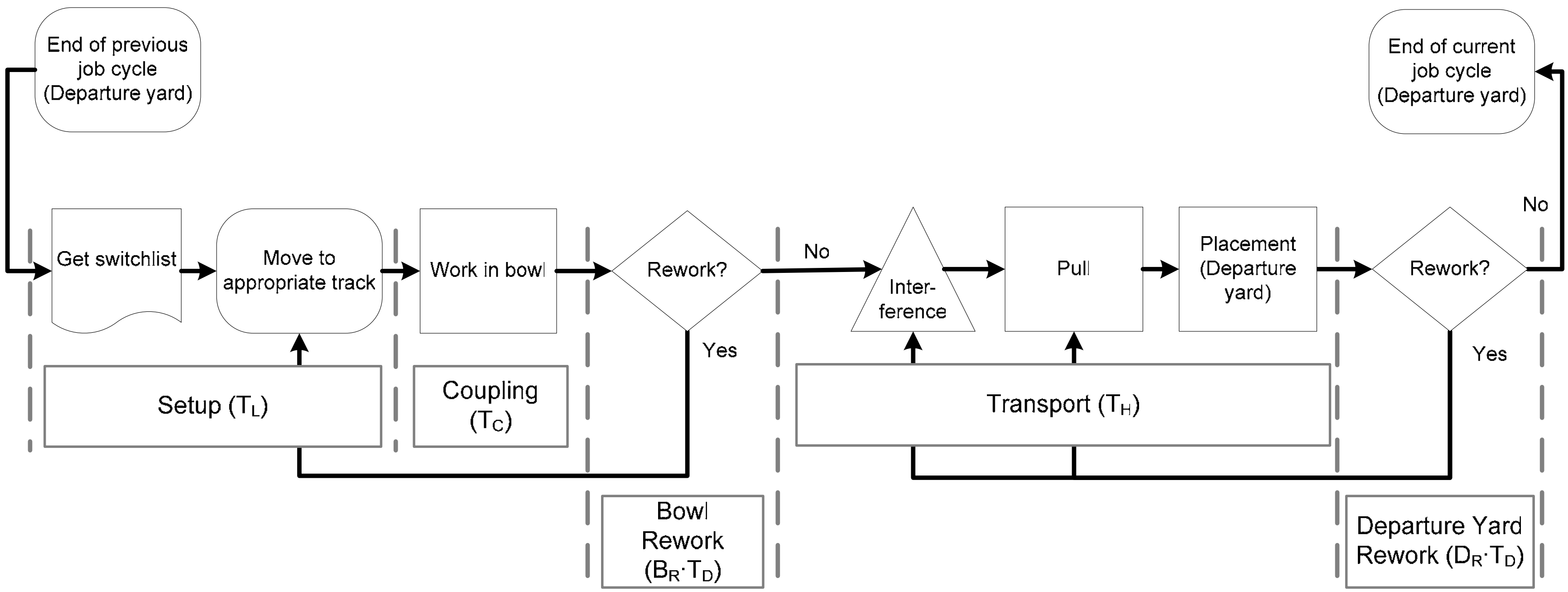


Figure 4.5 Pull-down process cycle time components

4.3.1 Setup

The crew begins work when they obtain the switch list (Figure 4.6). The switch list is the work order for the switchman. The upper left-hand corner tells him or her which track they will pull and how many cars will be pulled. The car closest to the pull-down end is Line 1. Moving from left to right, the switch list provides the identity of each car (reporting mark and number), whether the car is loaded or empty, the classification code (block) and which track or tracks in the departure yard the cars are to be placed. The final column gives the weight of each car in thousands of pounds. Switch lists are printed by the yardmaster and a paper copy is given directly to the switchman, transmitted to a printer or faxed for pick-up. The yardmaster in charge of the pull-down process coordinates the work of the crews using the switch lists and the radio. Figure 4.6 is an example switch list from Bensenville Yard and will be used to illustrate a typical pull-down job using conventional switching equipment. RCL operations follow a similar work flow but there are a few differences that will not be discussed in detail here.

PULLB ENGINE WORKLIST					PAGE 1 OF 1			
01120 LEAD: P2 TRACK: CT30					MAR 19 2006 0219EST			
LOW or EAST/SOUTH END								
TOTAL CARS: 18	LOADS:	EMPTIES: 18		TONS: 550	FEET: 1041			
LNE	EQUIPMENT		LE	CCODE	TO TRACK	LOCAL HAND	WEIGHT	PLATE
18	SOO	116437	E	4930EG1	BD08		031	
17	ITLX	020023	E	4930EG1	BD08		031	
16	SOO	118909	E	4930EG1	BD08		031	
15	SOO	119733	E	4930EG1	BD08		031	
14	SOO	119517	E	4930EG1	BD08		031	
13	SOO	122400	E	4930EG1	BD08		031	
12	SOO	116477	E	4930EG1	BD08		031	
11	NRLX	047302	E	4930EG1	BD08		032	
10	SOO	122233	E	4930EG1	BD08		031	
9	NRLX	047647	E	4930EG1	BD08		032	
8	NAHX	053600	E	4930EG1	BD08		027	
7	SOO	074543	E	4930EG1	BD08		031	
6	SOO	115687	E	4930EG1	BD08		031	
5	EEC	060752	E	4930EG1	BD08		031	
4	NRLX	100003	E	4930EG1	BD08		032	
3	GNWR	253288	E	4930EG1	BD08		031	
2	NOKL	821592	E	4930EG1	BD08		031	
1	NOKL	003120	E	4930EG1	BD08		032	
**	497	gp32	Spot to west/air		2026			

Figure 4.6 Example Bensenville Yard switch list, from March 19, 2006 (CPR 2006d)

In this case, the crew is assigned to pull 18 empty cars totaling 1,041 feet from track CT30 and place them on track BD08 in the departure yard. All of the cars are from the same block (4930EG1 = North Dakota empty grain) and destined for the same departure track (BD08):

Track CT30: # of cuts = 1; # of blocks = 1 → clean

Assuming there is enough space on track BD08, the crew will be able to move all 18 cars from CT30 to BD08 without any rework or doubling movements. In Equation 4.3, this means B_R and D_R will both equal zero. Therefore, the total cycle time (C_T) to pull the clean track is:

$$C_T = T_L + T_C + 0 \cdot T_D + T_H + 0 \cdot T_D \rightarrow C_T = T_L + T_C + T_H$$

The time to complete the first component of the cycle time, setup (T_L), is a function of the location of the engine and crew after the previous job, the location of the track in the bowl that the crew will be working, the speed, the route and interference from other engines or trains. Equation 4.4 can be used to calculate T_L:

$$T_L = TT_L + ST_L + DT_L \tag{4.4}$$

where: TT_L = Time spent moving from departure yard to required bowl track (min)
ST_L = Time spent throwing switches (min)
DT_L = Time spent not moving due to interference (min)

4.3.2 Coupling

Once the crew has its orders to pull all 18 cars from class track CT30 to track BD08 in the departure yard, they coordinate with the General Yardmaster (GYM) to move the engine to CT30 (Figure 4.7). Before they can enter CT30, they must contact the humpmaster and request that a block authority be placed on that track. A block authority is a safety measure built into the hump control system that does not allow any car to be routed to a track that a pull-down crew is working. Once the block authority has been granted, the crew couples the locomotive to the first car on CT30 (NOKL 003120) and the switchman begins walking along the track comparing the switch list to the standing order of the cars on the track.

Walking the track is necessary to ensure the cars that the crew has been instructed to pull are actually on that track and that no other cars are mixed into the cut. If an out-of-place car (sometimes called a “sleeper”) is discovered or a car on the switch list is missing, the GYM is contacted to resolve the discrepancy (Figure 4.8). The other task to accomplish while walking the track is to ensure all of the cars on the track are coupled together and that the “joints” (couplings) are secure. Every few minutes while the switchman is walking the track, he or she will radio the engineer (in a conventional operation) and request that the cars be “stretched” to pull out the slack and test the joints. If a joint has not been properly made, the switchman will inspect the coupler hardware to see if everything is in working order. If a drawbar appears to be "kinked" (out of alignment) the switchman will have the engineer pull back until there is at least 50 feet of separation between the cars and then request "3-point protection." This will ensure that the locomotive and the attached cars do not move while the switchman is centering the drawbar(s). Once the switchman is ready and clear, he or she will have the engineer push forward to "make the joint" (couple) and then continue walking the track.

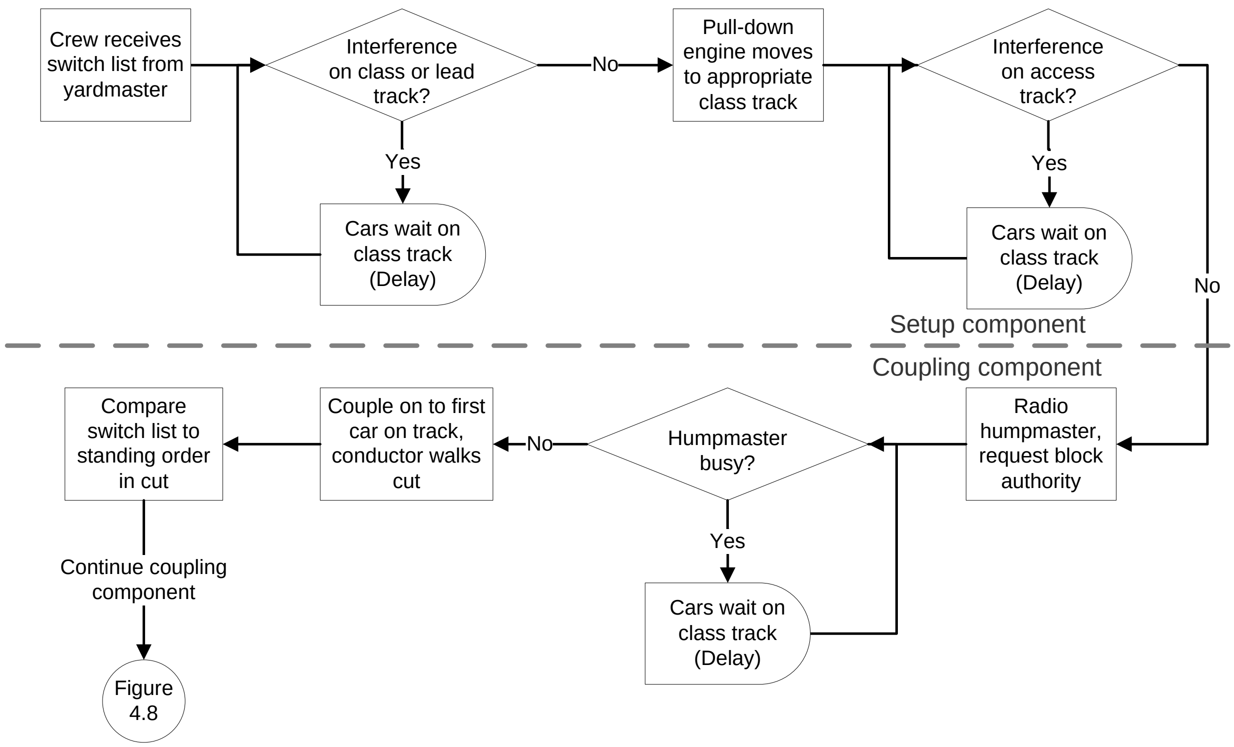


Figure 4.7 Detailed work process flow, setup component and first portion of the coupling component

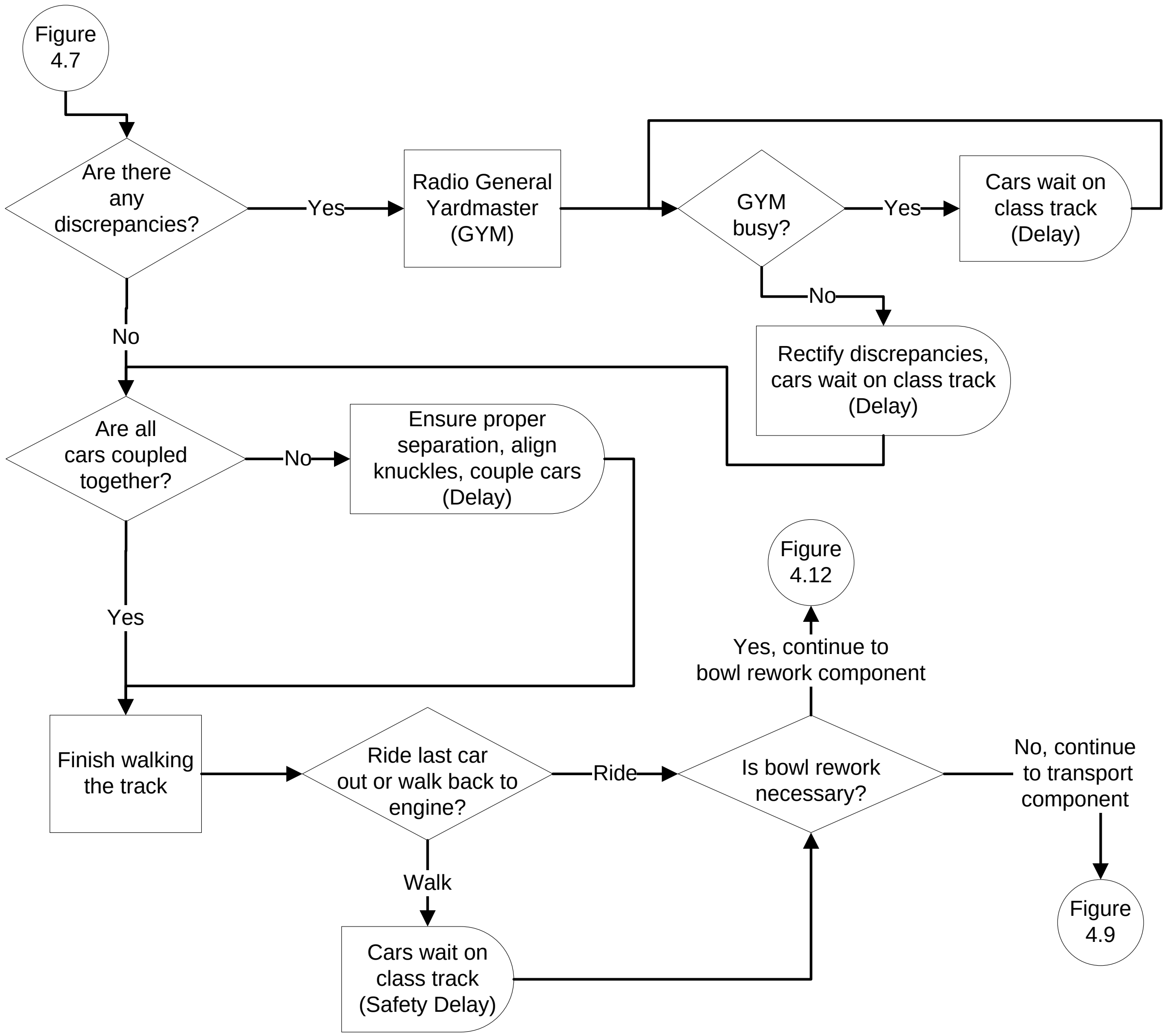


Figure 4.8 Detailed work process flow, second portion of the coupling component

In this example, all 18 cars were coupled properly and there were no discrepancies between the switch list and the standing order. When the switchman reaches the last car on the track (SOO 116437), he or she decides to either walk back to the engine, or ride the last car while the engine pulls the cars out of CT30. From a productivity standpoint, riding the last car out is quicker and yard management prefers this option. However, if the switchman thinks it is safer to walk back to the engine, he may do so. Equation 4.5 can be used to calculate the coupling component (T_C):

$$T_C = WT_C + CT_C + GT_C + ET_C \tag{4.5}$$

where: WT_C = Time spent walking the track (min)
CT_C = Time spent correcting switch list discrepancies (min)
GT_C = Time spent closing gaps between cars (min)
ET_C = Time spent exiting the track (min)

4.3.3 Transport

Because CT30 is a clean track, the crew will be able to complete the transport component and end the current cycle since no rework is necessary. The transport component begins when the grouping of cars has cleared the bowl track. At this point, the crew will radio the humpmaster and release the block authority that was given to them in order to work the track safely (Figure 4.9). The hump can now resume sending cars to CT30. The engine will continue to pull the cars to a point where they will be able to line the switches into the departure yard. Then the engine will push the cars into the departure yard, stopping when necessary to throw switches and wait for any interference to clear. When the grouping of cars is properly located on the departure track, the switchman sets the appropriate number of hand brakes and the engine uncouples. No departure yard rework is necessary and the crew can now move on to the next job. Equation 4.6 can be used to calculate the transport component of the cycle time (T_H).

$$T_H = TT_H + ST_H + DT_H + UT_H \tag{4.6}$$

where: TT_H = Time spent moving from track to track (min)
ST_H = Time spent throwing switches (min)
DT_H = Time spent not moving due to interference (min)
UT_H = Time spent setting hand brakes and uncoupling (min)

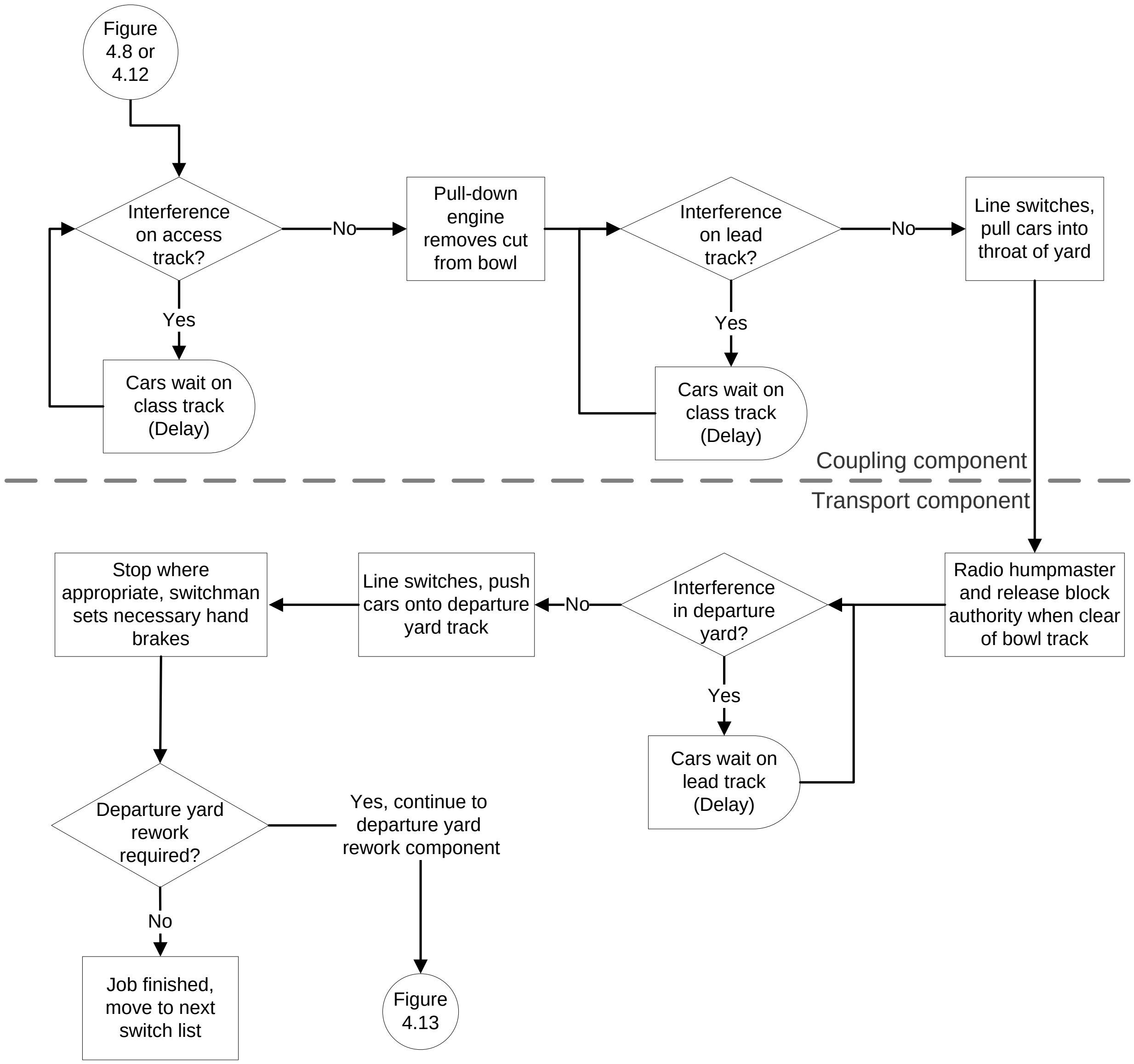


Figure 4.9 Final portion of work flow when track is clean

Using "six sigma" statistical techniques, Logan (2006b) determined that the time spent throwing switches was the most significant factor in the transport component. For the yard studied, every one minute increase in switch time resulted in a 5.36 minute increase in total transport time. The number of cars transported was moderately significant with total transport time increasing by 6.6 seconds for every additional car pulled from a track. Logan's unpublished study also analyzed the potential impact of installing remote control switches by conducting an experiment that simulated the remote control condition using a utility switchman who aligned switches for the pull-down crews. The experiment showed that in the yard studied, remote control switches would reduce transport time enough to cost-justify their installation. This is consistent with one of the design specifications for an efficient hump yard found in Wetzel (1985); power-operated switches between the classification and departure yards will often be cost-effective. Cost-effectiveness is dependent on the volume of traffic handled at the yard. Bensenville did not have remote control switches installed at the time of this study; however, to improve throughput and efficiency they did use a utility switchman dedicated to throwing switches between the classification and departure yards.

4.3.4 Rework

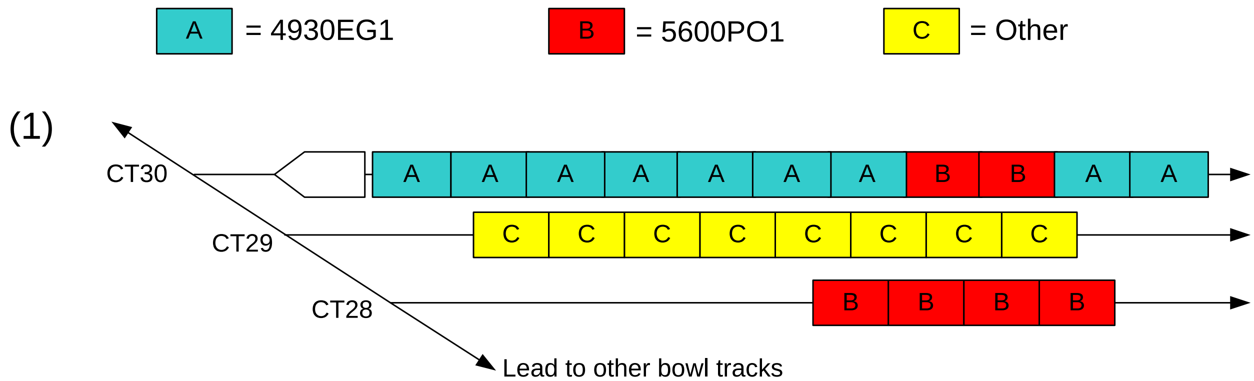
Based on Logan's study, it is reasonable to assume that time spent throwing switches will significantly impact total time spent in other portions of the pull-down process. There are two primary causes of extra switches being thrown: "cherry-picking" high-priority cars and rework. Cherry-picking is the pulling of high-priority cars located behind other cars on a bowl track (Kraft 2002c). Kraft (2000b, 2002c, 2002d) has extensively analyzed priority car classification and how to avoid cherry-picking. All of his solutions (prioritizing inbound trains, re-humping and rescheduling low-priority cars) follow the TOC approach by subordinating the upstream processes to the bottleneck and eliminating the need to cherry-pick cars, effectively increasing the capacity of the pull-down process.

The other cause of throwing additional switches is rework. If the track is dirty, the crew will have to double tracks thereby reducing efficiency. To illustrate the additional work that must be performed when a track is dirty, the switch list from the previous example was modified (Figure 4.10). The CCODE (class-code or block) of lines 8 and 9 were changed to indicate three cuts on the track, instead of one. CT30 is now dirty because the number of cuts (3) is greater than the number of blocks (2). The additional switching that is required when a track is dirty is similar to the work required when cherry-picking (Figure 4.11). Rework can occur in two locations: the bowl or the departure yard. In this case, the switch list instructs the crew to place the two cars from block 5600PO1 (Brandon Grain) on track CT28 in the bowl. This block is not needed for any outbound trains currently being built. If the cars were from a block that is currently being assembled into an outbound train, the rework of the track would occur in the departure yard. Figure 4.12 details the general bowl rework procedure.

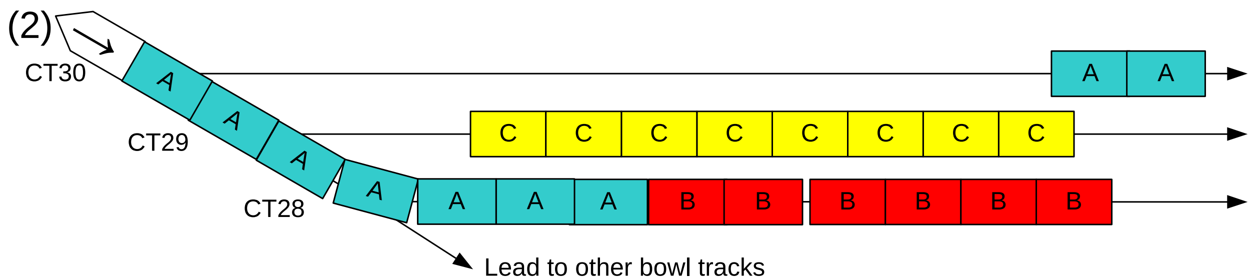
PULLB ENGINE WORKLIST PAGE 1 OF 1
01120 LEAD: P2 TRACK: CT30 MAR 19 2006 0219EST
LOW or EAST/SOUTH END
TOTAL CARS: 18 LOADS: EMPTIES: 18 TONS: 550 FEET: 1041

LNE	EQUIPMENT		LE	CCODE	TO TRACK	LOCAL	HAND	WEIGHT	PLATE
18	SOO	116437	E	4930EG1	BD08			031	
17	ITLX	020023	E	4930EG1	BD08			031	
16	SOO	118909	E	4930EG1	BD08			031	
15	SOO	119733	E	4930EG1	BD08			031	
14	SOO	119517	E	4930EG1	BD08			031	
13	SOO	122400	E	4930EG1	BD08			031	
12	SOO	116477	E	4930EG1	BD08			031	
11	NRLX	047302	E	4930EG1	BD08			032	
10	SOO	122233	E	4930EG1	BD08			031	
9	NRLX	047647	E	5600PO1	CT28			032	
8	NAHX	053600	E	5600PO1	CT28			027	
7	SOO	074543	E	4930EG1	BD08			031	
6	SOO	115687	E	4930EG1	BD08			031	
5	EEC	060752	E	4930EG1	BD08			031	
4	NRLX	100003	E	4930EG1	BD08			032	
3	GNWR	253288	E	4930EG1	BD08			031	
2	NOKL	821592	E	4930EG1	BD08			031	
1	NOKL	003120	E	4930EG1	BD08			032	
**	497	gp32	Spot to west/air		2026				

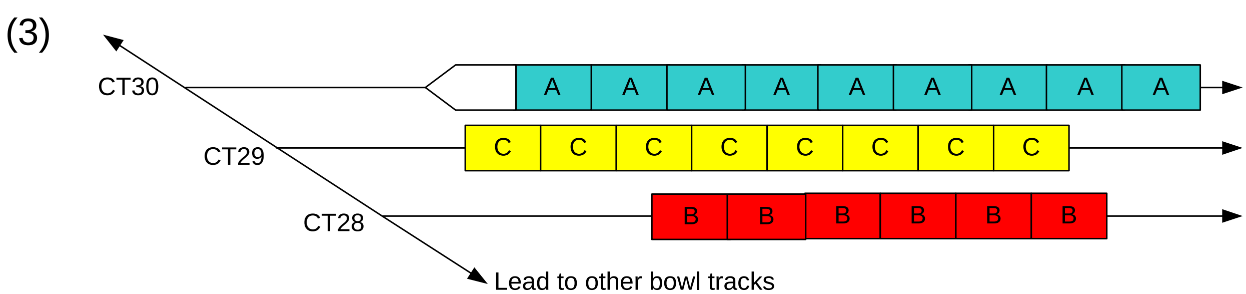
Figure 4.10 Modified switch list from Figure 4.6, track CT30 is now dirty



(1) The current outbound train needs all of the cars from block A, the cars from block B on the upper track must be removed or "dug" out and placed on CT28



(2) The out-of-place cars are moved to CT28 but the lead is blocked during this move, which can cause interference for other pull-down crews and slow them down



(3) The crew moves the cars from block A back to CT30 and couples to the last cars; only now can all cars from block A be pulled from the bowl

Figure 4.11 "Digging" out a car from a dirty track; the cars from block B on CT30 are out-of-place

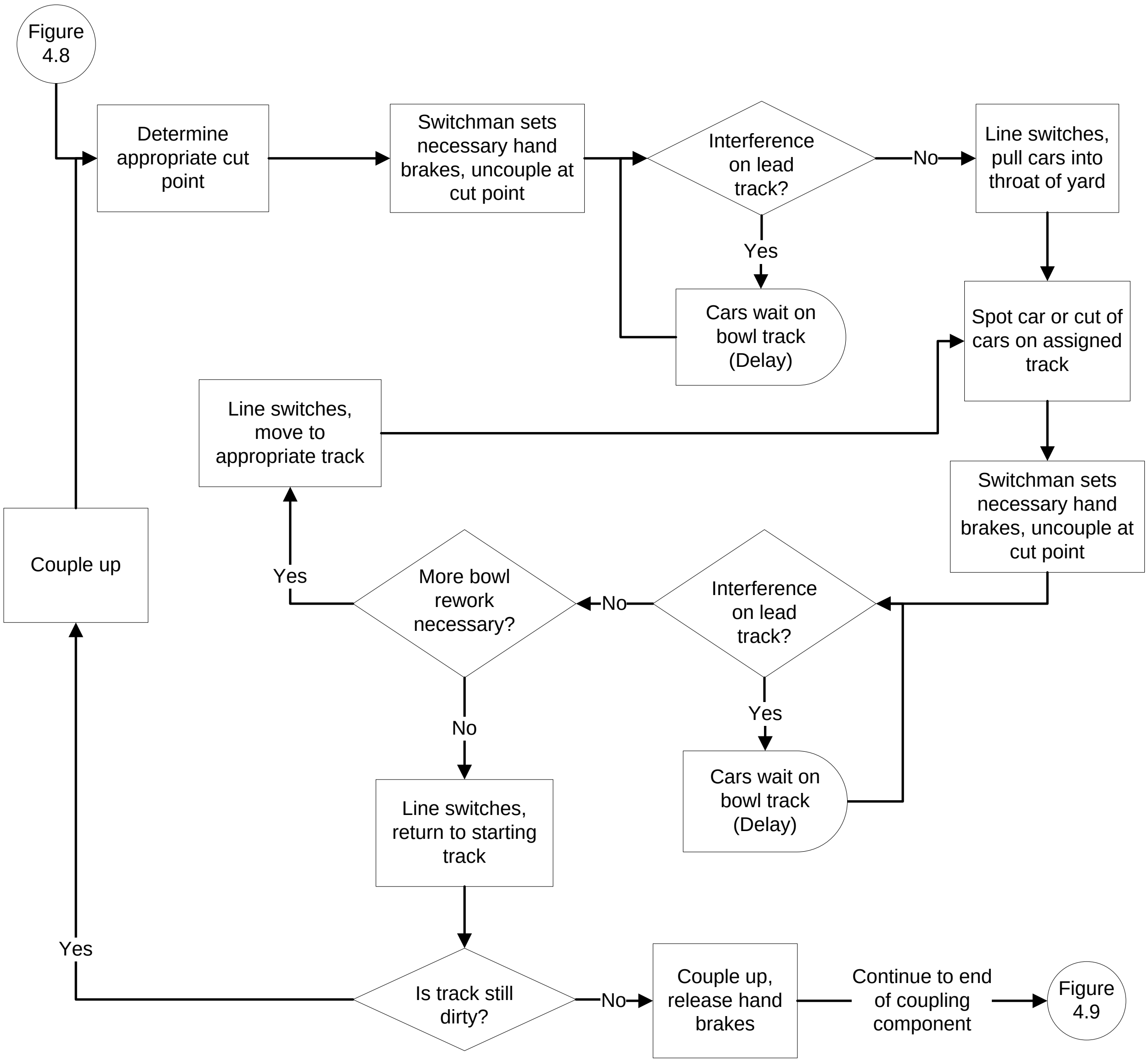


Figure 4.12 Detailed work process flow, bowl rework component

As shown in Section 4.2.2, the time required to complete the rework (T_D) will depend on the number of switching (doubling) movements during one pull (ND_D). ND_D is equal to the difference between the number of blocks and the number of cuts (Equation 4.7). The times to complete each movement are summed together to calculate the total time for bowl rework (Equation 4.8):

$$ND_D = C - B \tag{4.7}$$

where: C = Number of cuts on the track
B = Number of blocks on the track

$$T_D = \sum_{n=1}^{ND_D} (TT_D + ST_D + UT_D + DT_D)_n \tag{4.8}$$

where: TT_D = Time spent moving from track to track (min)
ST_D = Time spent throwing switches (min)
UT_D = Time spent releasing hand brakes, setting hand brakes, coupling and uncoupling (min)
DT_D = Time spent not moving due to interference (min)

For the switch list in Figure 4.10, ND_D = 1. To calculate the total cycle time for the bowl rework case, the bowl rework occurrence integer (B_R) is set to 1 in Equation 4.3:

$$C_T = T_L + T_C + 1 \cdot T_D + T_H + 0 \cdot T_D \rightarrow C_T = T_L + T_C + T_D + T_H$$

The additional time component, T_D, means that the cycle time for a pull down job with rework will be greater than a job without rework. As will be shown in Chapter 5, this reduces the capacity of the pull-down process. The procedure for departure yard rework is the same except that it occurs in the departure yard area (Figure 4.13). The departure yard rework occurrence integer (D_R) is set to 1 in Equation 4.3 and the cycle time for the departure yard rework case is also greater than the case without rework:

$$C_T = T_L + T_C + 0 \cdot T_D + T_H + 1 \cdot T_D \rightarrow C_T = T_L + T_C + T_H + T_D$$

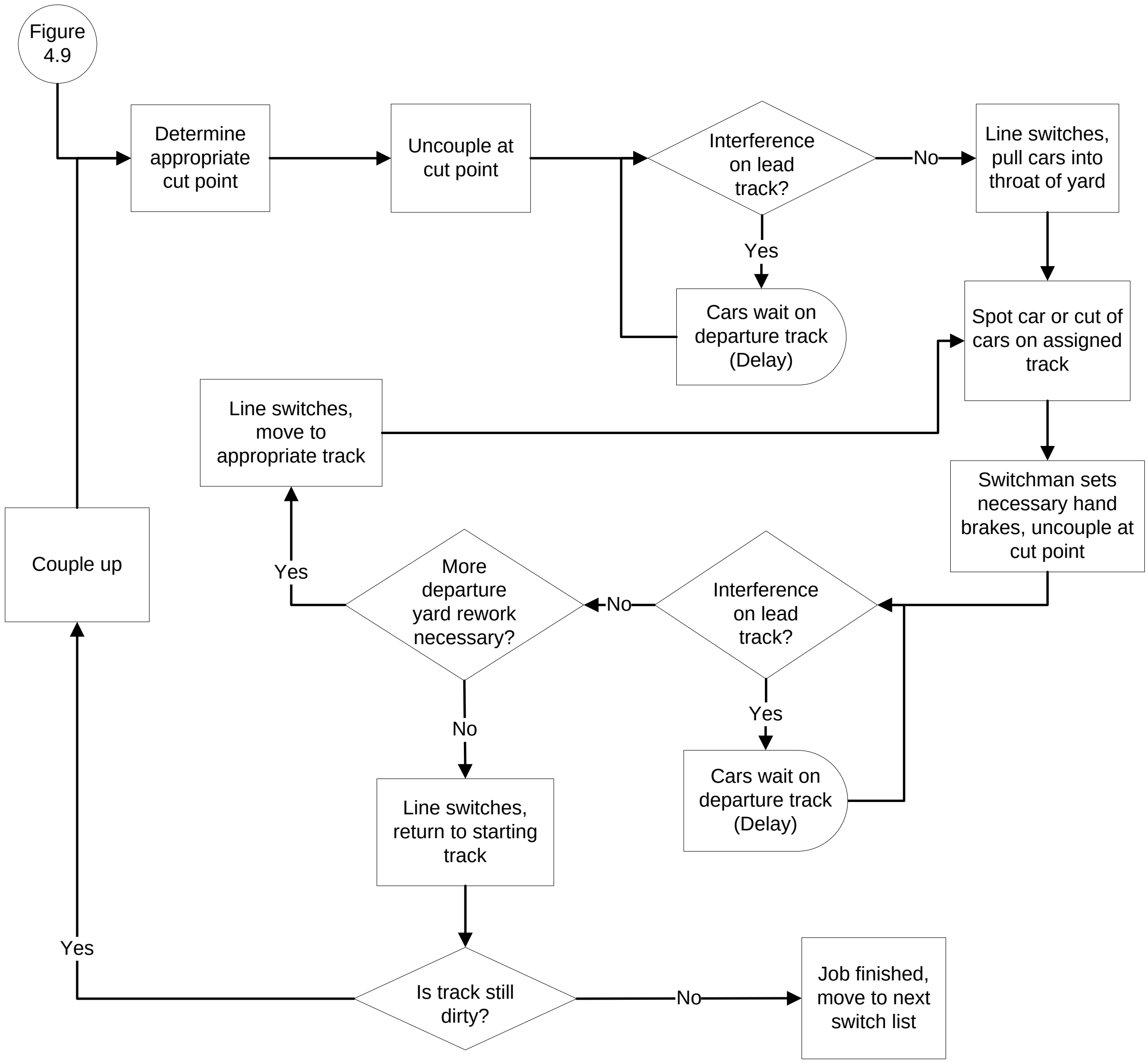


Figure 4.13 Detailed work process flow, departure yard rework component

4.4 Estimated Capacity using Current Operating Practices

Pull-down time studies were conducted at Bensenville over a four-day-period in March 2006. The time of day that the observations were gathered was different each day. A total of fifteen complete cycles were observed during the available time period. These data were used, along with other yard measurement data normally tracked by CPR, to calculate the parameters for Equations 4.2 and 4.3 (Table 4.1). This allowed Equation 4.1 to be used to estimate the capacity of the pull-down process for a baseline case.

Table 4.1 Calculated parameters for the baseline case, Bensenville Yard

Paramter	Value	Notes
M_B = Total meal and break time (min)	135	30 min lunch, 15 min breaks, 3 shifts
P_F = Performance factor	0.85	85% is the standard (Logan 2006b)
T_M = Productive crew time (min)	1,109	$T_M = (1440 - M_B) \cdot P_F$
C_T = Average pull-down process cycle time (min) (no rework)	79	Net travel time, no conflicts
C_T = Average pull-down process cycle time (min) (with rework)	98	Net travel time, no conflicts
T_L = Average travel time from the departure yard to the classification yard (min)	10	Net travel time, no conflicts
T_C = Average coupling time to couple an average size block (min)	48	Calculated using bowl authority logs, Dec. 26 to Jan. 2, for 22 cars
T_D = Time required to complete a doubling maneuver (min) = rework	19	Net travel time, no conflicts
T_H = Average travel time from the classification yard to the departure yard (min)	21	Net travel time, no conflicts
N_E = Number of pull-down engines	3	Three crews per shift on average
N_C = Average number of cars in a cut of a block (cars)	22	Dec. 24, 2005 to Jan. 10, 2006 average
N_D = Average number of doubling maneuvers to be made per pull	0.17	Over 4 day period, average of 17% dirty tracks pulled per day
C_F = Conflict coefficient	1.55	from Wong et al. (*18*) pg. 149, Configuration 1 with 3 engines

The baseline case has an estimated capacity of 541 cars per day. To check the accuracy of the estimate, the average daily process car count for 2004 was calculated. CPR defines process cars as those that go through all yard processes: arrival, classification, pull-down and departure. The average throughput was 521 cars per day. Average throughput should be less than theoretical capacity (Hopp & Spearman 2001); therefore, the estimate is acceptable. From the baseline case, the estimated capacity increase of several bottleneck management options will be discussed in Chapter 5.

CHAPTER 5: IMPROVING BOTTLENECK PERFORMANCE

The physical design of the pull-down end of a yard imposes a limit on the number of pull-down engines that can operate simultaneously, and the length of these moves. Adding new pull-down leads, lengthening existing leads and rearranging the track layout all may be useful, but are also expensive capital projects. Furthermore, they may be limited by geographic and environmental constraints adjacent to the yard (e.g. the Bow River in Figure 4.3). As TOC implies, developing management methods that exploit the bottleneck and subordinating other resources to cost-effectively improve operating efficiency should be undertaken before investing capital to expand infrastructure. In this chapter, several methods to improve bottleneck performance are considered and Equation 4.1 is used to estimate the capacity increase for each one.

5.1 Bottleneck Management Improvement Alternatives

Starting with the baseline case presented at the end of Chapter 4, individual parameters, and combinations of parameters, were modified to determine the potential capacity increase for each alternative (Table 5.1). The following parameters were modified:

1. Add one engine with crew at the pull-down end
2. Use hump engine to build trains when it would otherwise be idle
3. Increase performance factor (P_F) by 5%
4. Eliminate rework and "cherry-picking"
5. Reduce the interference between the pull-down engines
6. Reduce the component cycle times for T_H, T_L and T_D by one minute each
7. Reduce the coupling cycle time (T_C) by five minutes

The boxes in each column highlight the parameters that were changed from the baseline. The last row, C_P, gives the estimated capacity for each alternative. The following sections describe each parameter change in detail.

Table 5.1 Effect of various bottleneck improvement alternatives on estimated capacity; boxes indicate the parameters changed from the baseline for each alternative

Parameter	Baseline	(1) Add Engine at pull-down	(2) Use hump engine when idle	(3) Increase P_F by 5%	(4) No rework	(5) Reduce interference	(6) Cycle times 1 min less	(7) Coupling time 5 min less
T_M =	1,109	1,109	1,109	1,174	1,109	1,109	1,109	1,109
N_E =	3	4	3.25	3	3	3	3	3
N_C =	22	22	22	22	22	22	22	22
T_H =	21	21	21	21	21	21	20	21
T_L =	10	10	10	10	10	10	9	10
N_D =	0.17	0.17	0.17	0.17	0	0.17	0.17	0.17
T_D =	19	19	19	19	19	19	18	19
C_F =	1.55	2.55	1.55	1.55	1.55	1.45	1.55	1.55
T_C =	48	48	48	48	48	48	48	43
C_P =	541	576	586	573	576	555	564	562
% increase in C_P =	N/A	6.5%	8.3%	5.9%	6.5%	2.6%	4.3%	3.8%

where:

T_M = Productive crew time (min) = 1305·P_F

P_F = Performance factor

N_E = Number of pull-down engines

N_C = Average number of cars per block pulled

T_H = Average travel time from the classification yard to the departure yard (min)

T_L = Average travel time from the departure yard to the classification yard (min)

N_D = Average number of doubling maneuvers to be made per pull

T_D = Average time required to complete a doubling maneuver (min)

C_F = Conflict coefficient

T_C = Average coupling time to couple an average size block (min)

C_P = Capacity of the pull-down end (cars/day)

5.1.1 Add a Pull-Down Engine

One option to increase capacity at the pull-down end is to add another engine. This was the first alternative tested and it resulted in a capacity of 576 cars per day, a 6.5% increase compared to the baseline. The limiting factor when adding another engine is the increased conflict coefficient (2.55 vs. 1.55). Other yard designs may have higher conflict coefficients that will further limit effectiveness. While this option results in the one of the highest capacity levels, it is also the most expensive because of the additional engine and labor cost.

5.1.2 Pull from the Hump End when Idle

At Agincourt Yard (CPR) in Toronto, an option has been implemented that increases capacity without increasing interference or engine and labor costs. The hump engine is used to build trains when the hump is idle. This is done by placing the hump in trim mode (disabling the retarders), allowing the engine to enter the bowl and pull blocks from the hump end (Figure 5.1). Agincourt, like Bensenville, has parallel receiving and departure yards. Although quite effective, this solution would not be practical for yards with in-line designs.

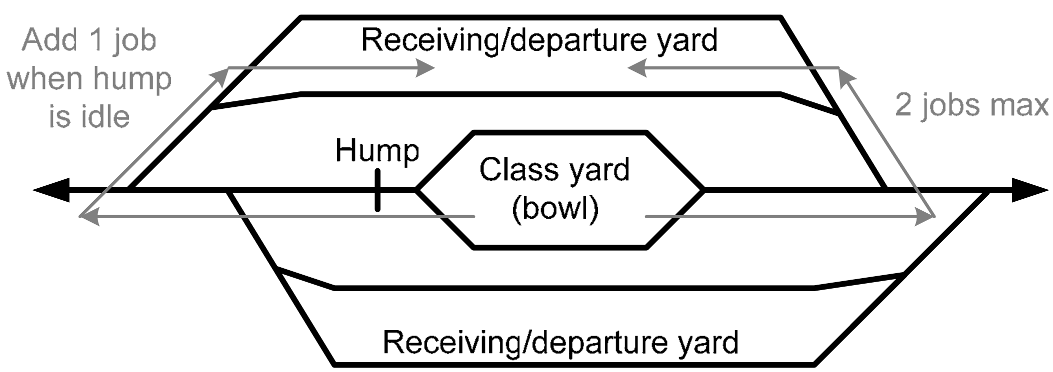


Figure 5.1 Adding pull-down capacity at the hump end of the yard schematic

This solution is consistent with the TOC approach; having identified the system's constraint, yard management was able to exploit the pull-down process by subordinating one of the other resources in the yard (the hump) to it. This is possible because hump productivity is generally governed by the number of cars available for sorting. "Productivity is usually less than the capacity of the hump" (Wetzel 1985, p. 38). At Agincourt, for the first six months of 2005, the highest monthly average hump utilization was approximately 56% (Table 5.2). Due to this low utilization rate, the hump could be used in trim mode part of the time and still be able to sort all of the required cars. If Bensenville, with a daily average hump utilization of 49% (CPR 2006e), implemented a similar solution, capacity would increase to 586 cars per day, an 8% increase. Because of the commitment to humping and other trim operations, it is assumed that using the hump engine would increase the number of engines to 3.25.

Table 5.2 Daily average hump statistics for Agincourt Yard, January to June 2005 (CPR 2005d)

Month	Hump time (percentage)	Idle time (percentage)	Trim time (percentage)	Hump count (cars)	Hump count (feet)
January	47.06	5.23	47.66	756	49,537
February	50.05	2.44	47.51	807	53,057
March	52.60	3.47	43.31	866	56,356
April	55.49	3.56	40.66	884	57,703
May	53.55	9.06	37.35	864	56,330
June	55.64	10.00	33.95	889	58,629

5.1.3 Increase the Crew Performance Factor

Workers with higher motivation tend to work harder. The third alternative reflects this by increasing the performance factor by 5%. This results in an increase of productive work time by 65 minutes and a capacity increase of almost 6% to 573 cars per day. Achieving this is non-trivial but management should explore means of increasing crew productivity because of the potential to increase capacity without capital or operational expense.

5.1.4 Eliminate Rework and "Cherry-Picking"

At Bensenville, over the four-day period the time studies where conducted, an average of 17% of the tracks pulled per day were dirty and required rework (Table 5.3). Therefore, for the baseline case, it is assumed that the average number of doubling moves made per pull (N_D) is 0.17. If tracks are kept clean, no rework will have to be performed ($N_D = 0$). This means crews will not have to dig cars out of tracks when they are assembling trains and capacity would increase to 576 cars per day, which is the equivalent of adding an engine on the pull-down end. Keeping the tracks clean requires analyzing the interaction between the hump and the pull-down processes.

Table 5.3 Dirty vs. clean tracks pulled at Bensenville Yard during observation period

Date	Dirty tracks pulled	Clean tracks pulled	Total pulled	% dirty
March 7, 2006	5	14	19	26%
March 21, 2006	4	18	22	18%
March 22, 2006	2	20	22	9%
March 23, 2006	4	24	28	14%
Average	3.75	19	22.75	16.97%

Cherry-picking is the most commonly accepted method of protecting connections that are in danger of being missed when outbound train capacity is exceeded. It "exacerbates the capacity bottleneck which already exists there, and reduces the throughput of the whole facility" (Kraft 2002c). Kraft has extensively studied the issue of connection-priority. He reported that time and budgetary constraints prevented the development of quantitative metrics on the performance of the algorithms resulting in conclusions based primarily on the subjective assessments of the author and the implementation team (Kraft 2002d, p. 117).

As explained in Chapter 4, the additional switching required for rework and cherry-picking is similar. Therefore, I was able to quantify some of the parameters needed to assess the impact of Kraft's proposed solutions on pull-down capacity by modifying N_D. If it is assumed that cherry-picking occurs in 10% of the pulls, then $N_D = 0.27$ for the baseline case (0.17 for rework, 0.10 for cherry-picking) and capacity is 522 cars per day. Eliminating cherry-picking would increase capacity 3.6% to 541 cars per day. In addition to the capacity increase, this will also improve connection performance.

5.1.5 Reduce Interference among Pull-Down Engines

Better coordination of the pull-down engines would reduce interference and reduce the conflict coefficient in Equation 4.1. This might be accomplished by having a dedicated pull-down yardmaster. Reducing the coefficient by 0.10 is estimated to increase capacity to 555 cars per day, a 2.6% increase.

5.1.6 Decrease Component Cycle Times

Faster cycle times result in increased process throughput. Cycle times can be reduced by eliminating unnecessary moves, throwing fewer switches, increasing engine speed, preventing engine breakdown, using experienced crews, etc. For this alternative, it was assumed that the average travel times from the bowl to departure yard (T_H) and departure yard to bowl (T_L) as well as the average time for rework (T_D) were all reduced by 1 minute. This would increase capacity approximately 4% to 564 cars per day.

5.1.7 Decrease Coupling Time

Several factors affect the time it takes to couple the cars on the bowl track, including walking speed, number of cars on the track, switch-list discrepancies and the number, spacing and location of gaps between cars. At Bensenville, crews also have the option of walking back to the engine or riding the last car out. In this analysis, I assumed that crews walked out of the bowl. There are several options available to decrease coupling time, including:

1. *Eliminate gaps between cars* – Anytime there is a gap between cars, the switchman will have to walk further to accomplish the same amount of work. It also means that the crew must couple the cars. Gaps can be reduced by better retarder control, humping multiple-car cuts when possible and taking time to push cars down the tracks from the hump end.

2. *Better track inventory control* – Discrepancies from the switch list slow the crew down while they contact the GYM and determine a course of action. This causes the actual cycle time to be longer than the planned time and introduces variability into the process. More accurate track inventory control can be achieved through better use of Automatic Equipment Identification (AEI) scanners and the inventory control system itself.
3. *Quicker correction of out-of-alignment drawbars* – Sometimes cars do not couple when they impact each other. This is often the result of the drawbars being out of alignment. Equipment exists to help the crew align the drawbars more quickly. Ensuring that all engines carry this equipment will make it easier for the crew to correct out-of-alignment drawbars and speed up the coupling component.

To illustrate the impact of incorporating these options, average coupling time was decreased 5 minutes. The resultant capacity increase was 21 cars per day, an improvement of just under 4%.

5.1.8 Comparing Improvement Alternatives

The estimated capacity improvements range from 2.6% to 8.3% over the baseline case of 541 cars per day. Some alternatives will be more cost-effective than others. Reducing interference by adding a yardmaster may only result in a 2.6% increase. Adding another engine at the pull-down end increases capacity by 6.5% but labor as well as equipment and maintenance costs will probably keep this solution from being cost-justified. However, eliminating rework results in the same 6.5% increase and using the hump engine when it is idle provides the largest individual increase of 8.3%. Better management of the process and its interactions with other processes will result in capacity gains without large labor or capital expenses. This also means that multiple improvement options can be accomplished together. If the hump engine is used, rework eliminated and cycle times are reduced in accordance with Table 5.1, capacity would increase by 25% to 678 cars per day.

Reducing cycle times can be accomplished without any additional crews by studying the process to improve the work flow. Chapter 4 can serve as a starting point for a cycle time reduction project. Section 5.1.2 explains how and why the hump engine can be used to increase pull-down capacity using the practice at Agincourt Yard as the example. The next section discusses the alternative that provides the same capacity increase as adding an engine and crew, eliminating rework.

5.2 Reducing the Occurrence of Rework

Because a classification yard is a system, managing the interactions between the processes is just as important as managing the individual processes. All but two of the options above improve only the pull-down process. Those involving the interaction between the pull-down and the hump (Alternatives 2 and 4) result in greater capacity increases. This is consistent with the Theory of Constraints.

One of the principal findings of this work is that the humping process should be subordinate to the pull-down process. Because the pull-down is the bottleneck, the hump should be managed and operated so that it provides the bottleneck exactly what it needs when it needs it. The practice of measuring hump performance based only on number of cars processed can, and often does, contribute to poor pull-down performance because it can lead to incorrectly sorted cars. In order to better manage the interaction between the hump and the pull-down processes, a measurement of how well the cars are being sorted is needed. The first step to develop a measurement is to better understand the interaction between the hump and the pull-down processes.

5.2.1 Understanding the Humping Process

The hump is the immediate process upstream from the pull-down (Figure 2.6). As discussed in Section 2.3, the purpose of the hump is to sort (classify) cars from inbound trains into blocks.

> "Trains arriving at a railroad classification yard consist of cars intended for many destinations, but each departing train carries cars going only to a few destinations. Thus, there is a need for sorting. In railroad jargon, sorting classifications are called blocks; blocks correspond to a final car destination, or to another yard serving several final destinations. A departing train is defined by the blocks it carries. To minimize the need for further sorting at other yards, blocks within a departing train are usually arranged in a specific order." (Daganzo 1986, p. 189)

Cars can also be sorted by contents, empty cars awaiting distribution instructions or mechanical defects so they are separated from the blocks departing on scheduled trains. The sorting problem has two levels: the block to track assignment and the number of sorting stages. There are two types of block to track assignment: static and dynamic. A static plan permanently assigns each block to a particular classification track (Daganzo 1987).

> "In most railroad yards, however, static blocking strategies are not used because most of the time tracks would then contain blocks that do not leave in the immediate future, and classification space would not be effectively utilized. In order to free some space on the classification tracks, cars are often classified only for a short time before their block is scheduled to depart. If a car does not arrive during this time interval, it is set aside for later sorting. Blocking strategies like these are called dynamic, because the blocks assigned to the classification tracks change with time. Dynamic blocking strategies require fewer classification tracks, but more switches than their static counterparts." (Daganzo 1987, p. 3)

Of the five hump yards visited during the course of this study, four (Alyth, Agincourt, MacMillan and Roseville) had static assignment plans in place. Bensenville did not have a written plan in place, so it could be considered dynamic; however, certain tracks were static. Table 5.4 summarizes the use of the classification tracks at each yard and the number of assigned blocks.

Table 5.4 Bowl track assignment characteristics (CPR 2005e, CPR 2005f, CPR 2006f, CN 2005c, UP 2005)

Yard	Railroad	Bowl tracks	Class tracks	Assigned blocks	Mechanical tracks	Re-hump tracks	Assignment Plan
Agincourt (Toronto)	CPR	72	60	51	11	1	Static
Alyth (Calgary)	CPR	48	42	44	4	2	Static
Bensenville (Chicago)	CPR	34	31	40	2	1	Dynamic
MacMillan (Toronto)	CN	75	71	60	3	1	Static
Roseville	UP	55	52	60	2	1	Static

The observed practices at these yards appear to contradict Daganzo's claim that static strategies are not used in most yards. However, having designated re-hump tracks is a simple dynamic technique used to set aside cars for later sorting. If the number of different blocks classified at the yard is greater than the number of tracks in the bowl, multistage sorting strategies must be used. These strategies include sorting-by-train, sorting-by-block, triangular and geometric (see Daganzo et al. 1983 and Daganzo 1986). However, these strategies were not observed in any of the yards listed in Table 5.4, most likely because the number of available tracks was greater than or very close to the number of assigned blocks.

Static plans have the advantage of being easier for the operators to manage. They also have the potential to reduce process time variability in pull-down cycle times for individual trains because the crews would go to the same tracks each time the train is built. The variability reduction effect may be greater with the multiple pull operational method. From a Lean Railroading standpoint, the potential to reduce variability should be explored. The tradeoff is that they may not use track space as efficiently. This is only a major problem when volume at the yard nears capacity. Then, the actual block assignment will start to be more dynamic as cars are sent to other tracks because their designated tracks are full. The process of temporarily moving block assignments to other tracks is known as "swinging." For every car humped, there are three tracks that the system records:

1. *Assigned track* – The track that the car is assigned to based on the block to track plan.
2. *Destination track* – The track that the car will be sent to when it reaches the crest of the hump. If the destination track is different than the assigned track, a swing has been entered in the computer for that car's block.
3. *Actual track* – The track that the car actually ended up on when it entered the bowl. If the actual track is different from the destination track, a misroute has occurred.

A misroute is the result of built-in safety features that are in place to prevent accidents during the sorting process. Some of the more common causes of misroutes are:

1. *Fouling protection* – A car stalls on a track circuit or a presence detector senses a car crossing the fouling point on a classification track. The next car routed into that circuit will be misrouted.
2. *Cornering protection* – A car catches up to the previous car in a switching area. "Catch-ups are caused by cars with wide variance in rolling resistance (usually an easy-rolling car requiring heavy retardation followed by a hard rolling car requiring no retardation)." (Wetzel 1985, p. 38)
3. *Block authority protection* – The destination track for a car has a block authority in place.
4. *Escape route protection* – An escape route for an engine working in the bowl has been entered and sending the car to its destination track would foul the route.

Misroutes cause the actual track to become dirty. If a misroute occurs, the operator must now decide whether or not to put the hump into trim mode and have the hump engine move the misrouted car to its destination track.

5.2.2 The Interaction between the Hump and the Pull-Down

At Bensenville, a total of 3,539 misroutes occurred during the period of January 1, 2005, to April 10, 2006 for an average of 7.6 times per day. Figure 5.2 shows the probability density of the number of misroutes per day for that time period with a distribution fitted to the data. The blue solid bars are the sample data frequency density and the red lines are the fitted distribution. Palisade's BestFit™ software (Palisade 2006) was used to determine the distribution that best fit the data. BestFit™ uses Goodness-of-Fit (GOF) algorithms to test up to 28 distributions on the selected data. The chi-squared test is used to rank the distributions' fit to the data.

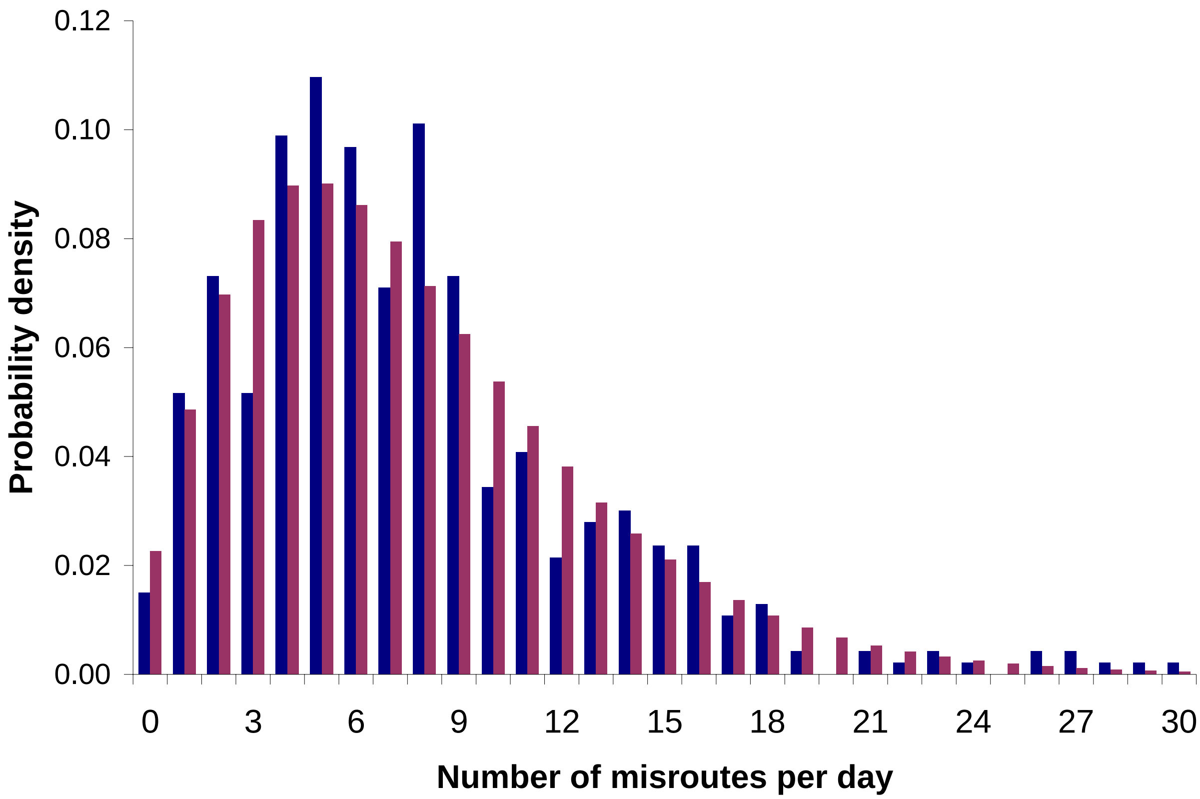
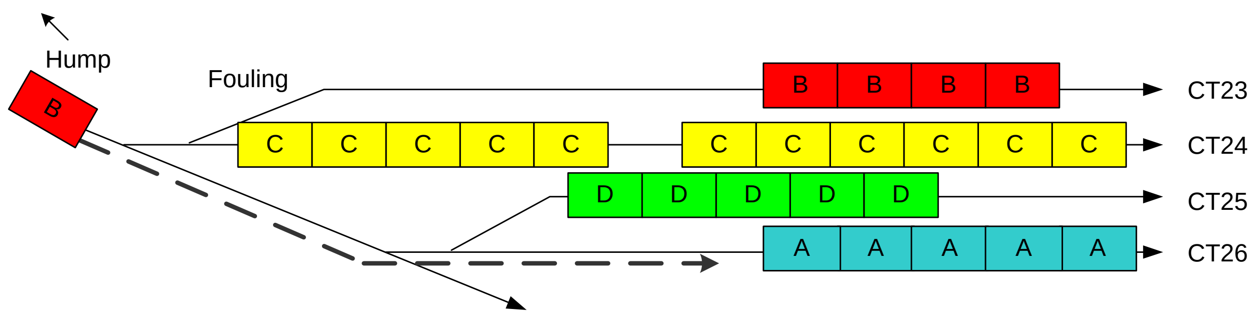


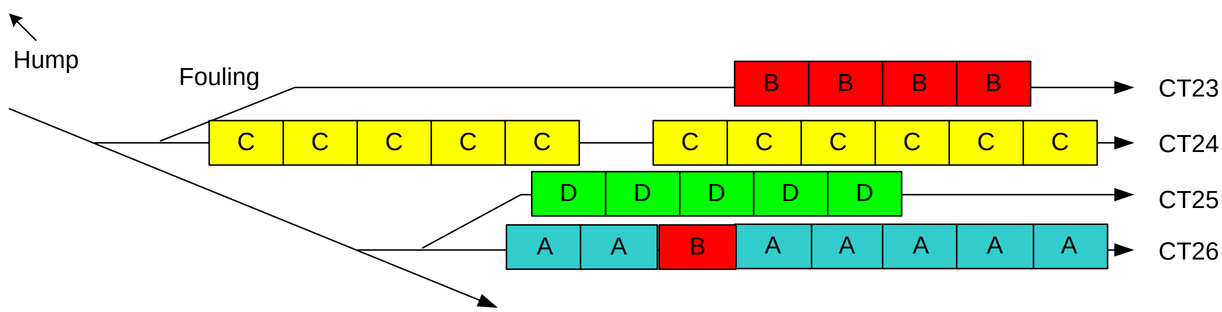
Figure 5.2 Bar chart with best fit distribution – Number of misroutes per day, Bensenville Yard, period of January 1, 2005 through April 10, 2006

I found that the Negative Binomial (x, p) distribution with parameters $x = 3$ and $p = 0.28273$, where x is the quantile or number of failures (misroutes) and p is the probability of a failure (Evans et al. 2000) the best fit to the data. The default method used by BestFit™ for defining the bins used in the chi-squared test is to adjust the bin sizes based on the fitted distribution so that each bin has an equal amount of probability. For discrete data, the bins are only approximately equal and for the misroute per day data, this resulted in 15 bins and 14 degrees of freedom. At a 0.05 level of significance, the chi-squared critical value is 23.685. The observed chi-squared value is 22.72, which is slightly less than 23.685. Although, we cannot reject the hypothesis that the Negative Binomial (3, 0.28273) distribution fits the observed data, there are some discrepancies. However, examination of the residuals did not indicate any consistent pattern of deviations.

If a misroute is left on the track, the track will become dirty when additional cars from the first block are sent there (Figure 5.3). This will cause the pull-down process to have to conduct rework when that track is pulled and will lead to a reduction in capacity.



(1) A car on CT24 is fouling the circuit and blocking access to CT23. The next car is destined for CT23, but the system misroutes the car to CT26



(2) The next two cars are destined for CT26. If the misroute is left on CT26, the track becomes dirty when the next two cars from Block A arrive.

Figure 5.3 How misroutes can cause dirty tracks

5.2.3 Eliminate Dirty Tracks when they occur

The lower average utilization rate of the hump enables the hump engine to be used to correct misroutes when they occur. Each day, the hump control system at Bensenville reports the time spent in each of the following modes: hump, trim and idle. From December 1, 2005, to April 10, 2006, the system spent an average of 700 minutes per day in hump mode, 418 minutes per day in trim mode and 310 minutes per day in idle mode. Because Bensenville has already adopted a policy of trying to correct misroutes when they occur, it can be assumed that a portion of the trim mode time is spent correcting misroutes. The rest of the time in trim mode is spent in three ways: removing stalled cars, eliminating gaps and creating room at the hump end by pushing cars down the classification tracks, or idle. The idle time is the result of controllers leaving the hump in trim mode after the trimming has been completed if there are no more cars to hump.

Assuming that the time to trim is the same as the average time for rework (19 min), the 7.6 misroutes per day would take just over 144 minutes to correct. The maximum number of misroutes observed in a day, 30, would take 570 minutes to correct. This is still below the total available time if the average idle time of 310 minutes is included. On average, the hump has the time available to correct all misroutes when they occur and still be able to meet its sorting requirements.

5.2.4 Other Causes of Dirty Tracks

Dirty tracks can also be the result of poor sorting by the hump controller. However, there is no measurement of how well the cars are being sorted. In order to better manage the interaction between the hump and the pull-down process (particularly during periods of high yard congestion), a sorting measurement was needed. The term "quality of sort metric" is introduced here to serve as a general title for measurements of this type.

5.3 A Quality of Sort Metric

The metric is called the Bowl Composition Monitor (BCM) and has two components: the Incorrect Sort Rating (ISR) and the Dirty Track Counter (DTC). The purpose of the ISR is to measure the adherence to a static assignment plan if one is in place. The DTC measures the number of dirty tracks and their heterogeneity. The DTC can be applied in yards that use either static or dynamic track assignment plans. The ISR applies only to yards using a static assignment plan. Separating the metric into two components increases its utility and results in a clearer understanding of the dynamics of bowl composition.

5.3.1 The Dirty Track Counter

The DTC measures the impact of dirty tracks on the capacity of the pull-down process. The higher the DTC, the more doubling moves are required. I used equation 4.1 to conduct a sensitivity analysis of the effect of the number of doubling moves on yard capacity using the Bensenville baseline case (Table 4.4). All of the parameters were kept constant except for the average number of doubling moves per pull (N_D) and I assumed that the maximum number tracks pulled per day is 24. The solid line in Figure 5.4 represents the reduction in capacity as N_D increases and the dashed line shows the impact of increasing interference as the result of additional rework; with C_F increasing 0.02 for each additional doubling move made. The light gray lines represent the decrease in the number of tracks pulled using the average of 22 cars per pull from Table 4.4. Thus, the figure can be used to determine the expected reduction in pull-down capacity as a result of increased rework.

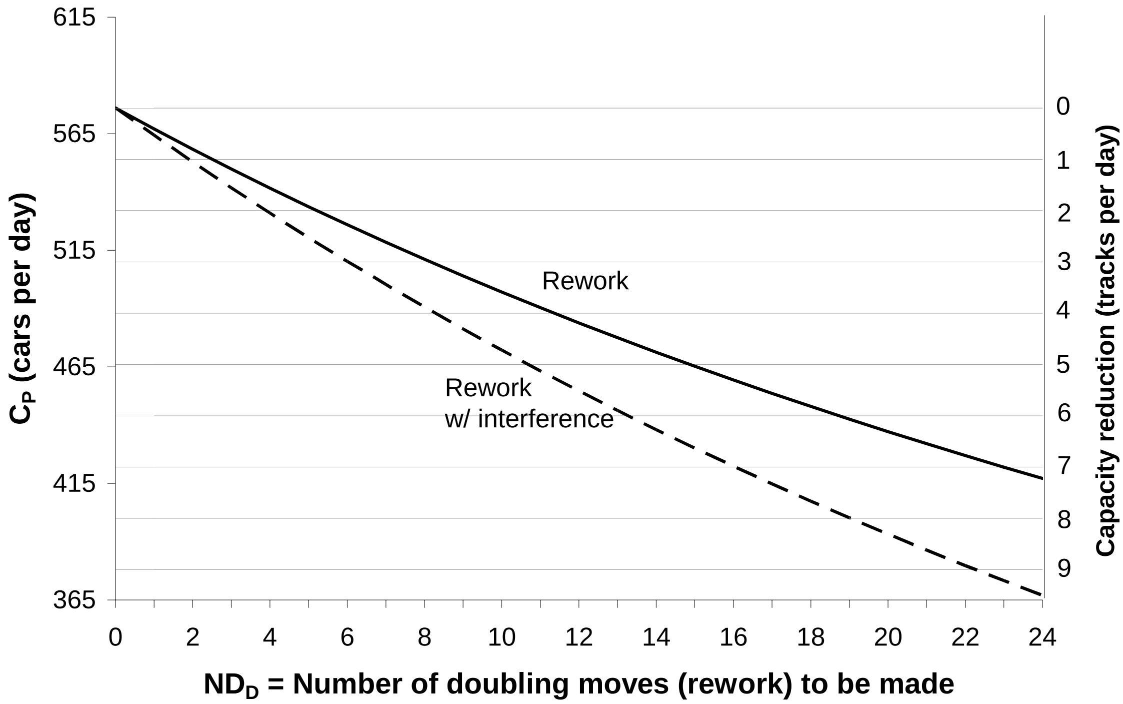


Figure 5.4 Estimated pull-down capacity vs. number of doubling moves for Bensenville Yard

The DTC is equal to the minimum number of doubling moves that will occur as a result of the current state of the sorting in the bowl. This value is the sum of the ND_D values for each track (Equation 5.1).

$$DTC = \sum ND_D \qquad \textit{for all tracks except special tracks} \qquad (5.1)$$

where: DTC= Dirty Track Counter
$ND_D = C – B$ (where: C = number of cuts, B = number of blocks) (Equation 4.7)

For example, if there are four dirty tracks in the bowl, and $ND_D = 1$ for each dirty track, then the DTC = 4. This means at least four doubling moves will be required when those tracks are pulled resulting in a reduction in capacity of 34 cars per day, due to rework alone, and 45 cars per day, when additional interference is accounted for as well (Figure 5.4). Yard management should expect to be able to pull two fewer tracks during the shift when those tracks are to be pulled. If any additional doubling moves are required because of the rework, additional capacity will be lost.

5.3.2 The Incorrect Sort Rating

The second component of the BMC measures the adherence to a static block to track assignment plan. The ISR is measured in number of cars and a low ISR indicates fewer incorrectly sorted cars. Every car that is humped into the bowl receives two ratings (Equation 5.1):

$$ISR = RT + RG \qquad s.t.\ RT = 0 \text{ or } \alpha;\ RG = 0 \text{ or } \beta;\quad \alpha + \beta = 1 \tag{5.1}$$

The first rating (RT) in Equation 5.1 is called Right Car-Right Track and it records if the car was sent to its assigned track. The second rating (RG) recognizes the need for flexibility of the static track allocation scheme and is called Right Car-Right Group. Yard management has the option of weighting each rating so that the desired level of flexibility is achieved. If stricter adherence to the static plan is the goal, α will be larger than β to emphasize the importance of placing cars on the correct tracks.

An example from Alyth Yard will be used to illustrate Equation 5.1. Car ICE 70512 has classification code 4850MA1 (St. Paul Manifest Block) and that block is assigned to track CT12 (Central Group) in the bowl. CT12 is the destination track. If the actual track that ICE 70512 is humped to equals the destination track (CT12), then RT=0 and RG=0. If actual track does not equal destination track, then RT=α and RG=0 if the actual track is in the Central Group; otherwise, RG=β.

The ISR is measured on three levels: car, track and bowl. Equation 5.1 provides the ISR value for each car. At the track level, the individual ISR values are summed together (Equation 5.2). A higher Track ISR indicates that more cars on that track are not on the correct track or in the correct group:

$$Track\ ISR = \sum_{i=1}^{C} ISR_i \tag{5.2}$$

where: C = number of cars on track n

The Track ISR value for designated mechanical and re-hump tracks is not calculated because the cars on those tracks will be out of place. Including them would artificially inflate the Bowl ISR.

The Bowl ISR is the highest level and reflects the overall performance of the hump controller in maintaining a "clean" bowl. The Bowl ISR is the sum of the Track ISR values for every track in the bowl except the designated mechanical and re-hump tracks (Equation 5.3):

$$Bowl\ ISR = \sum_{i=1}^{T} Track\ ISR_i \tag{5.3}$$

where: T = number of classification tracks (excluding mechanical and re-hump)

Before the Bowl ISR can be used to gauge a hump controller's performance and better manage the adherence to a static allocation plan, the expected Bowl ISR over a range of bowl volume levels needs to be known. In order to understand this relationship, a Bowl Replay program was developed to analyze yard event data. The development and details of the program are found in Appendix A. Alyth Yard was the primary yard used during the development of the program. The versions for Bensenville were built during the detailed time study period that provided the data used in the capacity calculations in Section 5.1. The final version of the program used event data from the yard control system to calculate and track the components of BCM and volume levels over several time periods for both yards. The output of the program is used to understand the relationships between bowl volume, DTC, ISR and capacity and provide insight into the implementation options for the metric. The relationships and implementation options are presented in Chapter 6.

CHAPTER 6: IMPLEMENTING THE QUALITY OF SORT METRIC

The BCM serves a dual role in the Lean Railroading implementation steps presented in Section 3.3. First, the DTC is part of the effort to exchange the time buffer for a capacity buffer by increasing the capacity of the bottleneck using the Theory of Constraints. As shown in Chapter 5, reducing the occurrence of rework through better management of the process interactions will increase the capacity of the pull-down process. Second, the purpose of the ISR is to reduce variability in the sorting process by tracking the adherence to a static track allocation plan. The BCM can also serve as part of the process of continuous improvement by providing performance feedback to hump controllers.

6.1 The Role of the BCM in a Terminal Control System

The BCM is designed to be incorporated into a terminal control system (Figure 6.1). Specifically, it can be used as part of the Terminal Performance Measurement System (TPMS) to enhance the planning and evaluation loop between the TPMS and the terminal superintendent and staff.

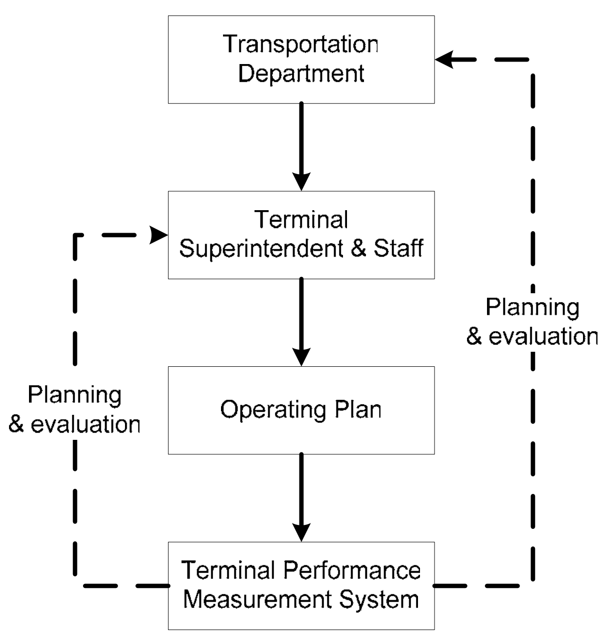


Figure 6.1 Components of a terminal control system (from Ferguson 1980, p. 13)

The role of the TPMS within a detailed existing control system is shown in Figure 6.2. The TPMS uses data from the yard management system to determine dwell times, connection performance, volume levels and throughput. The hump control system provides the TPMS with hump statistics such as number of cars humped, time in each mode (hump, trim, idle), number and type of misroutes, rolling speeds, etc.

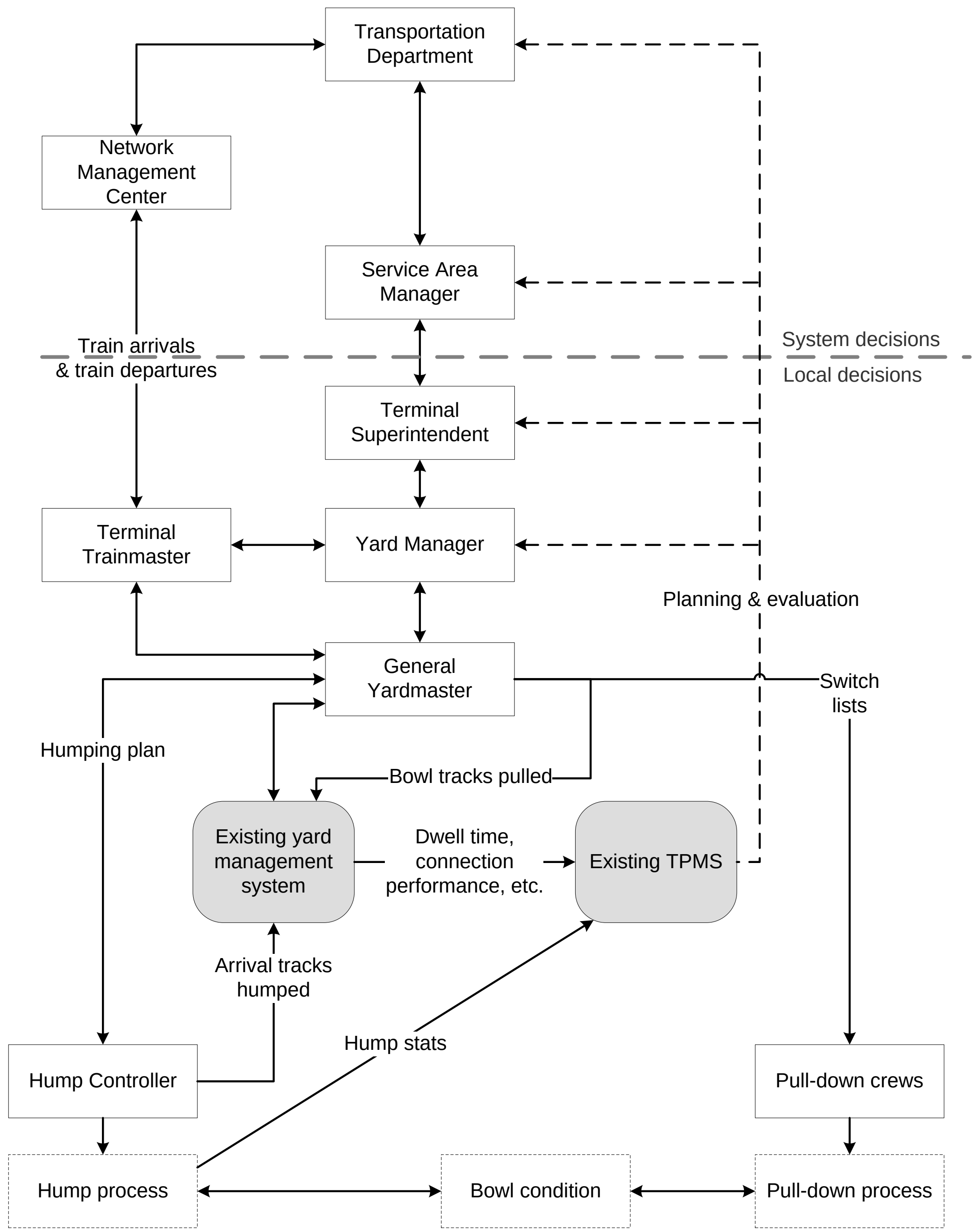


Figure 6.2 Detailed existing terminal control system

Notice the lack of information flow to the TPMS regarding the bowl condition and the pull-down process in Figure 6.2. The yard management system provides some of this information, but it is limited and usually requires an additional level of interpretation before it can be used. As the bottleneck, the performance of the pull-down process needs to be monitored on a regular basis. In addition, the importance of the interaction between the pull-down process and the hump process requires that the condition of the bowl also be known. Figure 6.3 shows a portion of a proposed terminal control system that incorporates the two new information flows into the TPMS: the BCM and pull-down cycle times. It also shows the new planning and evaluation flows (dashed lines): sorting performance feedback to the hump controller, a real-time bowl "picture" for use by the hump controller and performance feedback for the pull-down crews.

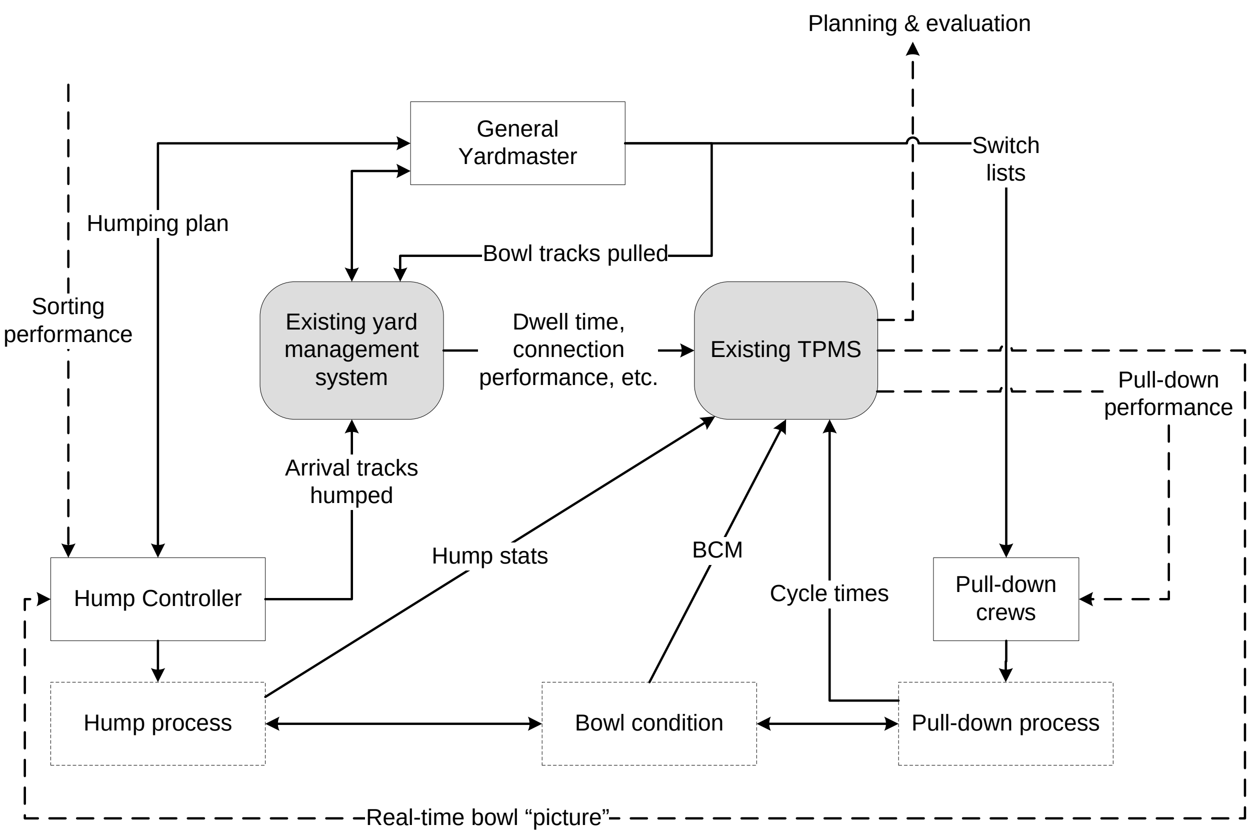


Figure 6.3 Terminal control system with proposed information, planning and evaluation flows (below the General Yardmaster level)

6.2 Implementing the Proposed Terminal Control System

This chapter will address the implementation issues associated with the proposed terminal control system. As the stand-alone component of the BCM, an implementation strategy for the DTC will be discussed first. For any quality measurement to be effectively used in a production system, it must be presented to both the operators and management in a manner that allows for quick assimilation of the information. The bowl “picture” is presented as a means to accomplish this using a visual interface. Once the DTC is implemented, the ISR will be calibrated to ensure that it accurately reflects the expected behavior of the sorting procedure. Statistical Process Control (SPC) concepts will be described and control charts that can be used with the ISR will be presented. The planning, evaluation and sorting performance flows will use SPC concepts and control charts. Finally, some options for gathering pull-down cycle times will be described.

6.2.1 Using the Dirty Track Counter and Bowl Picture

Yard inventory and event information from the existing yard management system is fed into the TPMS. This information is used to determine the current composition of the bowl, calculate the DTC and ISR (if necessary) and create the bowl picture. The Bowl Replay program described in Appendix A was used to simulate this procedure and serves as a proof-of-concept for it. The outputs include the DTC tracking screen (Figure 6.4) and the bowl picture (Figure 6.5).

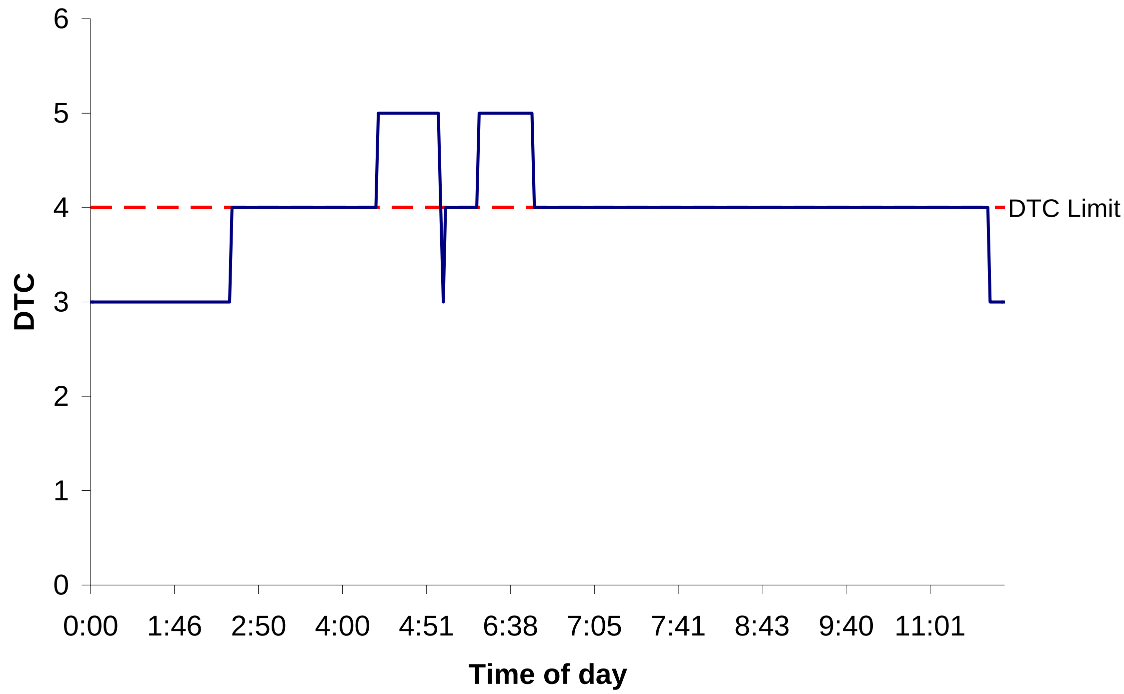


Figure 6.4 DTC tracking screen example for Bensenville Yard – March 7, 2006 0:00 to 12:00

Based on the capacity relationship information contained in Figure 5.4, yard management should set an allowable threshold level for the DTC. This level should be determined based on the yard's operating plan so that all of the required trains are still able to be built on time. In Figure 6.4, the DTC Limit is set at four which results in approximately two fewer tracks pulled per day. Any time that the DTC exceeds four, the dirty tracks must be identified and their composition analyzed. The bowl picture in Figure 6.5 can be used to quickly determine the composition of each track and the entire bowl. The coloring scheme is based upon the previously described dirty vs. clean track definition (Figure 4.4) and allows for quick identification of the cuts on each track. The first car on every track is colored green and every additional car from the same block on that track is also colored green. When a car from a different block is placed on that track, it is colored red. Any other cars from that second block are colored red. Cars from a third block are colored blue and so on.

	B	C	D	E	F	G	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI
1	TRACK	CAR COUNT	BLOCKS	CUTS	LENGTH	REMAIN																											
2	CT01	5	2	2	3116	2773	820M	UPPB	UPPB	UPPB	UPPB																						
3	CT02	2	2	2	3238	3085	5600	8579																									
4	CT03	2	1	1	3112	3004	5600	5600																									
5	CT04	11	3	4	3237	2604	5600	5600	5600	5600	3173	BRCX	5600	5600	5600	5600	5600																
6	CT05	12	1	1	3171	2557	CNHA	CNHA	CNHA	CNHA	CNHA	CNHA	CNHA	CNHA	CNHA	CNHA	CNHA	CNHA															
7	CT06	12	1	1	3170	2045	820F	820F	820F	820F	820F	820F	820F	820F	820F	820F	820F	820F															
8	CT07	15	1	1	3170	2271	0238	0238	0238	0238	0238	0238	0238	0238	0238	0238	0238	0238	0238	0238	0238												
9	CT08	2	2	2	3275	3127	820M	3174																									
10	CT09	19	1	1	3275	2370	3173	3173	3173	3173	3173	3173	3173	3173	3173	3173	3173	3173	3173	3173	3173	3173	3173	3173	3173								
11	CT10	36	2	2	3274	1061	820M	820M	ICBE	ICBE	ICBE	ICBE	ICBE	ICBE	ICBE	ICBE	ICBE	ICBE	ICBE	ICBE	ICBE	ICBE	ICBE	ICBE	ICBE	ICBE	ICBE	ICBE	ICBE	ICBE	ICBE	ICBE	ICBE
12	CT11	32	7	8	3274	574	BRCX	IMSX	IMSX	IMSX	IMSX	IMSX	IMSX	IMSX	IMSX	IMSX	IMSX	IMSX	IMSX	2304	STOP	STOP	STOP	STOP	STOP	STOP	STOP	493M	IMSX	IMSX	IMSX	IMSX	IMSX
13	CT12	41	3	4	3274	756	4817	4817	4817	4850	4850	4850	4850	4850	4850	4850	4850	4850	4850	4850	4850	4850	4850	4850	4850	4850	4850	4850	4850	4850	4850	4850	4850
14	CT13	2	1	1	3274	3155	STOP	STOP																									
15	CT14	17	2	2	3274	1694	2320	W02X	W02X	W02X	W02X	W02X	W02X	W02X	W02X	W02X	W02X	W02X	W02X	W02X	W02X	W02X	W02X										
16	CT15	12	1	1	3274	2473	3250	3250	3250	3250	3250	3250	3250	3250	3250	3250	3250	3250															
17	CT16	8	4	4	3275	2807	IMSX	STOP	STOP	NOBI	BRCX	BRCX	BRCX	BRCX																			
18	CT17	17	2	2	3274	1440	BRCX	BRCX	BRCX	BRCX	BRCX	BRCX	BRCX	BRCX	BRCX	BRCX	BRCX	BRCX	BRCX	BRCX	BRCX	BRCX	OCS3										
19	CT18	5	1	1	3206	2877	4817	4817	4817	4817	4817																						
20	CT19	15	2	2	3107	1191	3250	0581	0581	0581	0581	0581	0581	0581	0581	0581	0581	0581	0581	0581	0581												
21	CT20	30	3	3	3216	628	5600	4760	4760	3250	3250	3250	3250	3250	3250	3250	3250	3250	3250	3250	3250	3250	3250	3250	3250	3250	3250	3250	3250	3250	3250	3250	3250
22	CT21	10	2	2	3043	2335	520M	2320	2320	2320	2320	2320	2320	2320	2320	2320																	
23	CT22	2	2	2	3045	2931	3800	820M																									
24	CT23	34	2	2	2795	745	IMSX	520M	520M	520M	520M	520M	520M	520M	520M	520M	520M	520M	520M	520M	520M	520M	520M	520M	520M	520M	520M	520M	520M	520M	520M	520M	520M
25	CT24	20	1	1	2797	1536	0215	0215	0215	0215	0215	0215	0215	0215	0215	0215	0215	0215	0215	0215	0215	0215	0215	0215	0215	0215							
26	CT25	14	2	2	2547	1234	3250	W02X	W02X	W02X	W02X	W02X	W02X	W02X	W02X	W02X	W02X	W02X	W02X	W02X													
27	CT26	30	2	2	2548	585	0581	4760	4760	4760	4760	4760	4760	4760	4760	4760	4760	4760	4760	4760	4760	4760	4760	4760	4760	4760	4760	4760	4760	4760	4760	4760	4760
28	CT27	18	3	4	2299	929.36	493M	9684	9684	9684	9684	9684	6016	6016	493M	493M	493M	493M	493M	493M	493M	493M	493M	493M									
29	CT28	1	1	1	2174	2120	493M																										
30	CT29	25	2	2	1924	407	IHBX	IHBX	IHBX	IHBX	IHBX	IHBX	IHBX	IHBX	IHBX	IHBX	IHBX	IHBX	IHBX	IHBX	IHBX	IHBX	IHBX	IHBX	IHBX	IHBX	IHBX	IHBX	IHBX	IHBX	493E		
31	CT30	3	1	1	1925	1756	493E	493E	493E																								
32	CT31	16	2	3	1803	844	3800	3800	3800	3800	3800	3800	3800	3800	3800	493M	3800	3800	3800	3800	3800	3800											
33	CT32	21	14	19	1804	225	4865	4817	UPPB	820M	4760	4817	BRCX	4817	493E	4760	5600	4540	5600	FW19	FW19	FW19	3250	9684	4850	BRCX	STOP						
34	CT33	9	4	4	1804	1111	493M	0652	520F	520F	520F	520F	OCS4	OCS4	OCS4																		
35	CT34	14	9	12	4500	3598	3173	520M	820M	E41X	EJSP	EJSP	3250	HMTY	5600	E18X	EJSP	E18X	EJSP	EJSP													
36	Volume	512	62	67	99494	62848																											
37																																	
38	Time	Mar/07/2006 06:41																															
39																																	
40					401																												

Figure 6.5 Screenshot of the bowl picture from the Bowl Replay program 06:41 March 7, 2006 (Version 3.2.6) Bensenville Yard

There are five dirty tracks shown in Figure 6.5: CT04, CT11, CT12, CT27 and CT31. Each track has one more cut than block. The events that caused the dirty tracks occurred up to 24 hours earlier. Therefore, the impact of the increased DTC will not be felt until those tracks are pulled. Depending on their compositions and when the tracks are to be pulled, the most appropriate corrective action should be taken. At least one track should be cleaned; otherwise the yard's ability to meet its production schedule will be jeopardized during a later shift. Reviewing the event data provides details about the causes of the increases in the DTC. Figure 6.5 represents the situation at 06:41. The DTC exceeded the limit at 04:38 and at 06:02. In both cases, the subsequent reductions in DTC were the result of rework at the pull-down end.

6.3 Adding the ISR when a Static Track Allocation Plan is in Place

As part of their Lean Railroading improvement initiatives, the YOP group at CPR implemented a static track allocation plan at Alyth. The purpose of the plan is to reduce variability in the sorting process by assigning each block to a bowl track. Therefore, Alyth can serve as the implementation example for yards with similar static plans in place such as Agincourt Yard (CPR), MacMillan Yard (CN) and Roseville Yard (UP). The first step in this process is to ensure that the metric reflects the expected behavior of the sorting process.

6.3.1 The Expected Behavior of the Sorting Process

It is the opinion of several terminal and network managers that higher bowl volume results in a "dirtier bowl" (CPR 2005g, Barker 2005, UP 2005). A bowl with more cars has less room to put the cars that need to be classified. It becomes harder to maintain the block to track assignment plan, because it becomes more difficult to find track assignments with enough capacity to start new train blocks. Therefore, blocks tend to be split onto more than one classification track when a yard is congested (Kraft 2002d). If a static assignment plan is in place, it will be harder to keep all cars on their assigned tracks. Therefore, in order for the metric to provide useful feedback on the performance of the hump controller, it is necessary to know the relationship between bowl ISR and bowl volume. This will help establish reasonable performance goals based on the condition of the yard and verify that the metric can be used in the proposed terminal control system.

6.3.2 The Relationship between Bowl Volume and Bowl ISR

To test if the metric was formatted correctly and to aid with program development, weighted values were assigned to the ISR ratings. The wrong track and wrong group ratings were rated equally with α and β assigned values of 0.50. When cars are pulled from the bowl, their ISR values are removed from the totals. The impact of rework and trim events are also reflected in the ISR subject to the two ratings. Bowl rework cars have an added penalty for the extra work. Cars trimmed at the hump end also have a penalty built in, but it is smaller than the bowl rework penalty because the hump has a lower utilization and can take the time to correct sorting errors. The program records the bowl volume and bowl ISR every time a car is humped and when the last car of a cut being pulled leaves the bowl. Each time a car is humped, there are three possible outcomes:

1. *Right track and right group* – The car is sent to the right track according to the static plan; therefore, it is automatically in the right group and ISR = 0.
2. *Wrong track and right group* – The car is sent to the wrong track but it is still in the right group; ISR = 0.5.
3. *Wrong track and wrong group* – The car is sent to the wrong track and the wrong group; ISR = 1.

To develop the first relationship presented here, a bowl replay was developed for Alyth Yard, using event data for a five-day period. A total of 5,060 observations of bowl volume and corresponding ISR were recorded. The observations were grouped by volume level and any volume level with less than ten observations was discarded. Averages for the remaining observations were calculated and plotted (Figure 6.7). A second group of event data for a two-day period was also analyzed. In the second case, 2,563 observations were recorded and volume levels with less than ten observations were discarded. Averages for the remaining observations were calculated and plotted (Figure 6.8). A linear trend line was fitted to the data in both figures to develop equations for the relationships. In both cases, the expected trend of a higher volume resulting in a “dirtier” bowl is observed. In Figure 6.7, the volume level ranges from 388 to 605 cars and the slope of the line is 0.1969. In Figure 6.8, the volume level ranges from 557 to 691 and the slope is 0.3741. Therefore, it appears that a higher volume levels, the ISR tends to increase faster. This is the behavior predicted by railroad management (Barker 2005, CPR 2005g, 2006c) but further analysis is required before this behavior can be confirmed.

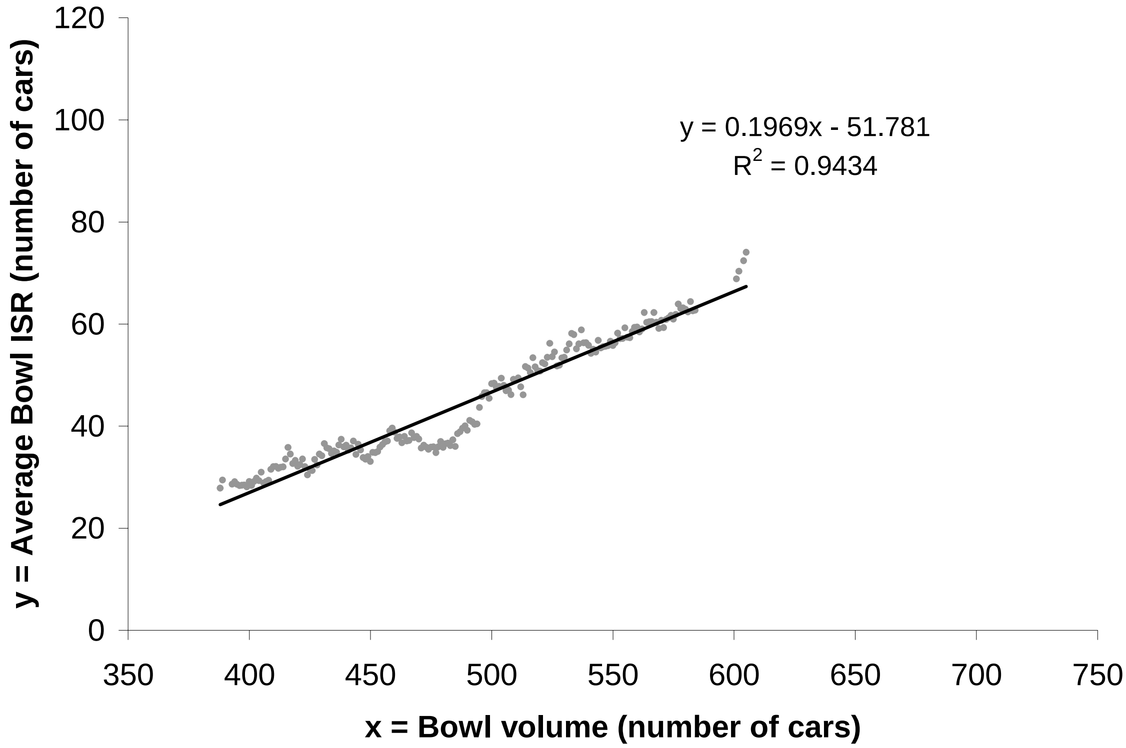


Figure 6.7 Average ISR vs. bowl volume for Alyth Yard, September 13 to 17, 2005

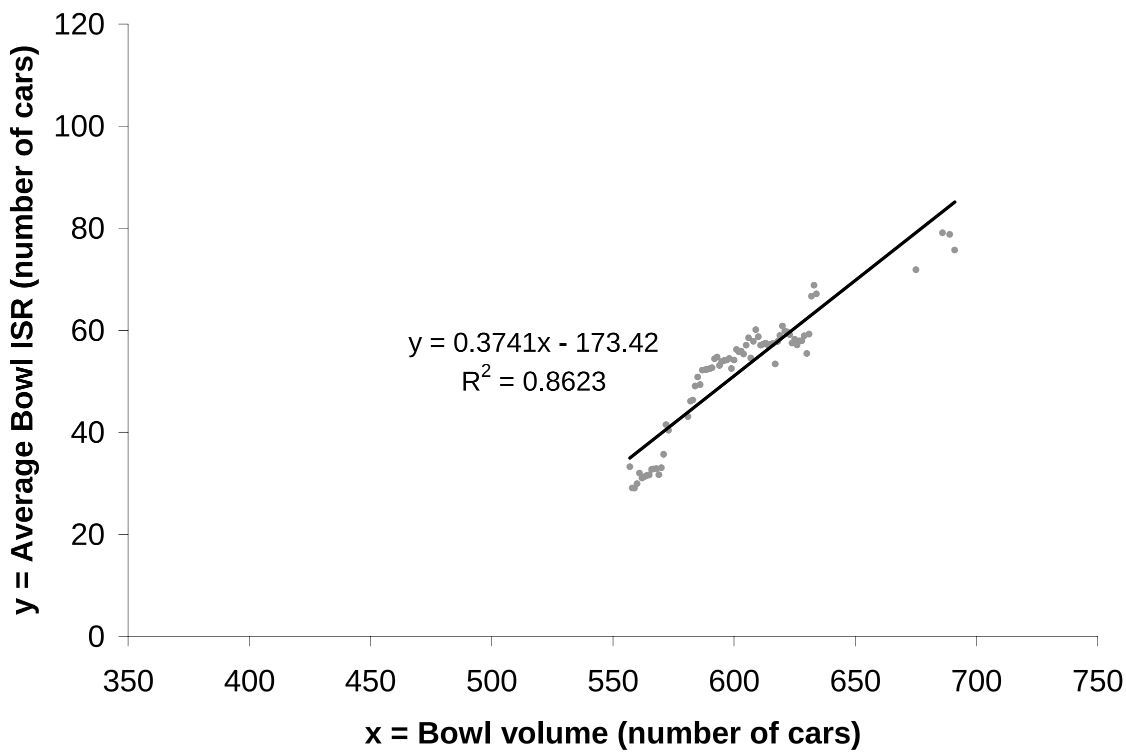


Figure 6.8 Average ISR vs. bowl volume for Alyth Yard, November 19 to 20, 2005

6.3.3 Interpretation and Potential Use of the Results

The results of the relationships can be used to measure controller performance. For example, if bowl volume is at 600 cars, the expected Bowl ISR is 66.36 cars in Figure 6.7. If the actual Bowl ISR is 55 cars, then the operator has kept the ISR below the expected ISR at that volume level. Therefore, the operator has done an acceptable job of sorting. However, this would only be applicable during the week that the data were collected. If Figure 6.8 is used, the expected Bowl ISR is 51.04 cars and an actual Bowl ISR of 55 cars is above the expected value. Another difficulty in this method is the high amount of variation in the ISR values at each volume level. One of the problems with using the average ISR for each volume level is that it assumes that the error is the same at every point. While the results do provide insights into the expected behavior of the sorting process as bowl volume increases, they may not be useful to front-line managers and operators on a per shift basis. Because there is variation in the output, and a need to track the metric over time, concepts adapted from Statistical Process Control (SPC) may be a more effective way to implement this metric in the terminal environment.

6.4 Adapting SPC Charts to Control and Improve the Sorting Process

As cited in Section 2.1, Mundy et al. (1992) studied the applicability of SPC methods in classification yards. Using the Burlington Northern's Tennessee Yard in Memphis as their case study, they first developed flowcharts for the processes being studied. Control charts were then created for the weekly percentages of cars over the connection standard of 24 hours. The third technique applied was Pareto analysis, which is the assignment of causes to failures. This allows the root causes to be identified so that improvement efforts can be focused on their elimination. Notice that SPC was used to help identify the causes of problems, not to solve the problems. This is why SPC requires a management team committed to the philosophy of continuous, gradual improvement (Mundy et al. 1992) such as the Lean Railroading approach presented in Chapter 3.

"To use statistics, one must have both something to measure and a means of measurement. The something should be thought of as the output of a process" (Mundy et al. 1992, p. 54). Here, we are measuring the output of the sorting process using the ISR. The impact of bowl volume is accounted for by dividing each Bowl ISR value by the corresponding bowl volume to obtain the percentage of incorrectly sorted cars (%-ISR). SPC assumes that the sample means of the process tend to follow a normal distribution. However, the %-ISR values do not follow a normal distribution and this prevents the complete SPC approach from being used. But, the concepts and process of SPC still provide valuable insight into the planning and evaluation methods needed in the proposed terminal control system.

6.4.1 Adapting the Process of SPC

The implementation process of SPC can be explained using the classical feedback control system perspective (Figure 6.9). The feedback control system perspective is not unique to SPC; therefore, it can be used with the ISR. Each of the stages in the control loop will be explained and details for each stage are found in the next section, with the sorting process and ISR as the example.

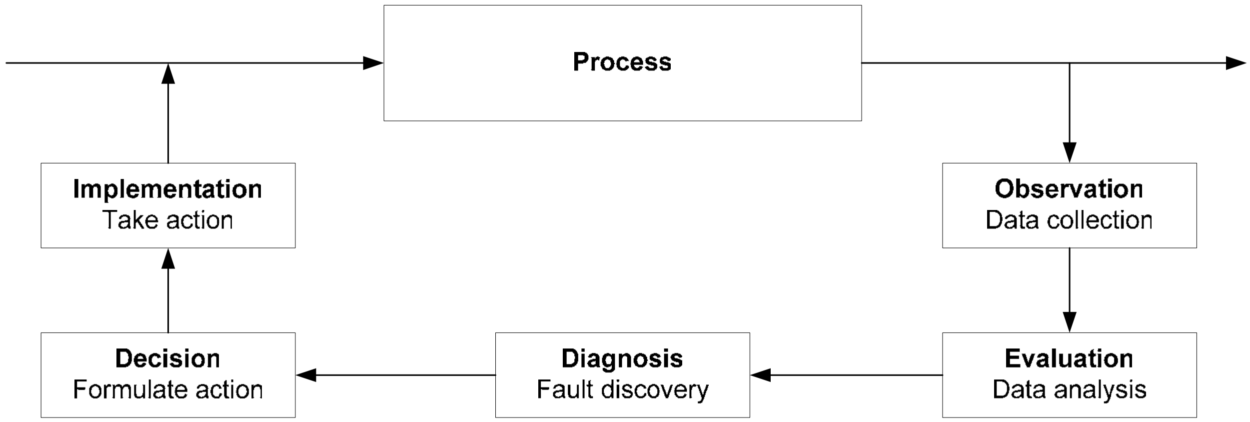


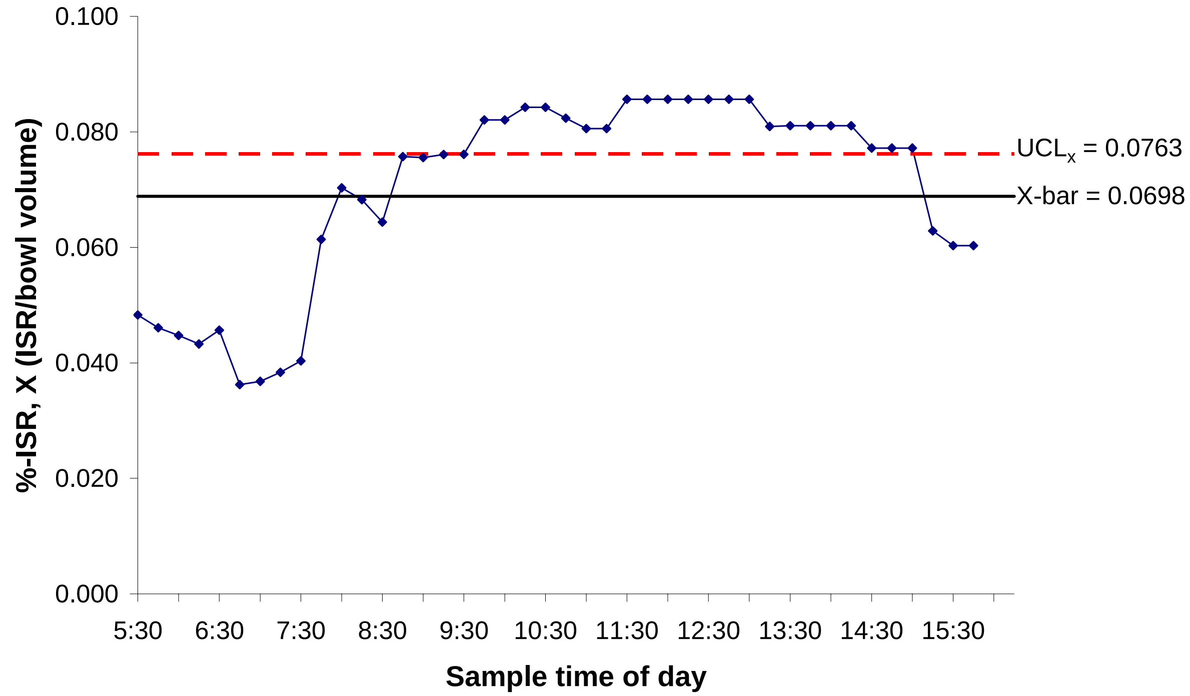
Figure 6.9 Classical control system view of SPC implementation (from DeVor et al. 1992, p. 134)

1. Observation – The first aspect of observation is determining what to observe and how to observe it. The second aspect is the statistical sampling issue of how many measurements to take, how often, etc.
2. Evaluation – A chart is needed that shows us how the process should look if it is in control. The data are compared to the model to identify the "signals," which may tell us that a specific opportunity for improvement exists.
3. Diagnosis – This stage is identified by DeVor et al. (1992) as the most crucial because "being able to move from comparison to diagnosis marks the difference between making charts and solving problems." The key is to be able to associate an out-of-control signal with the physical state(s) of the system. Some of the fault diagnosis aids include Pareto charts, cause-and-effect diagrams and decision trees for failure modes effects analysis.
4. Decision – Once the root cause of the fault is found, determine the appropriate action to eliminate the fault.
5. Implementation – Define the specific means to make the correction happen. Study the full implications of the action on the entire system to ensure that the action taken will be lasting. "*Holding the gains* over the long term requires that the system truly accept the total value of the action." (DeVor et al. 1992, p. 135)

6.4.2 Using SPC Concepts with the ISR

The first aspect of the observation stage has already been accomplished by identifying the bottleneck in the terminal, applying the Theory of Constraints to better understand the interaction between the pull-down process and the sorting process, and developing a quality of sort metric and Bowl Replay program to observe the relationship. The second aspect, the statistical sampling issue, involves the design and collection of the samples or subgroups. In this case, the type of control charts used will be modifications of the X and moving range (R_m) charts. The X and R_m charts are used for individual measurements where X is the value of the individual measurement and R_m is the range of a group of *n* consecutive individual measurements combined to form a subgroup of size *n* (DeVor et al. 1992, p. 347). In traditional SPC control charts, an Upper Control Limit (UCL) and a Lower Control Limit (LCL) are calculated at ±3 standard deviations from the mean. These are used to identify when special causes of variability have entered the system. Since the goal is to keep the %-ISR as low as possible, and because the data are not normally distributed, there is no need for a LCL. The UCL is kept to illustrate the use of an upper limit on the acceptable ISR. We are most concerned with large increases in ISR between samples and these are shown in the modified control charts (Figure 6.10). The same data from the Bowl Replay program used in the previous section is used to construct the control charts. The charts are constructed for one shift and each shift lasts 12 hours. To construct the control charts for the 0530 to 1730 shift on September 13, 2005, %-ISR measurements were taken every 15 minutes for a total of 49 observations. A size of $n = 2$ is used to calculate the moving ranges. The steps for constructing the X and R_m control charts are found in DeVor et al. (1992, pp. 348-349).

In Figure 6.10, large increases between sample points indicate that a higher percentage of cars in the yard are incorrectly sorted according to the ISR. The moving range chart shows the relative size of the increases and decreases. In Figure 6.10, ISR exceeded the upper limit from 09:00 to 16:45 with extreme moving range increase at 07:45. From this evaluation, the event data from the yard management system can be reviewed with the hump controller to diagnose the cause of the %-ISR increase and a decision can be made to prevent the situation from occurring again. The solid black line indicates the average %-ISR value for the shift, which can also be used to gauge the controller's performance.



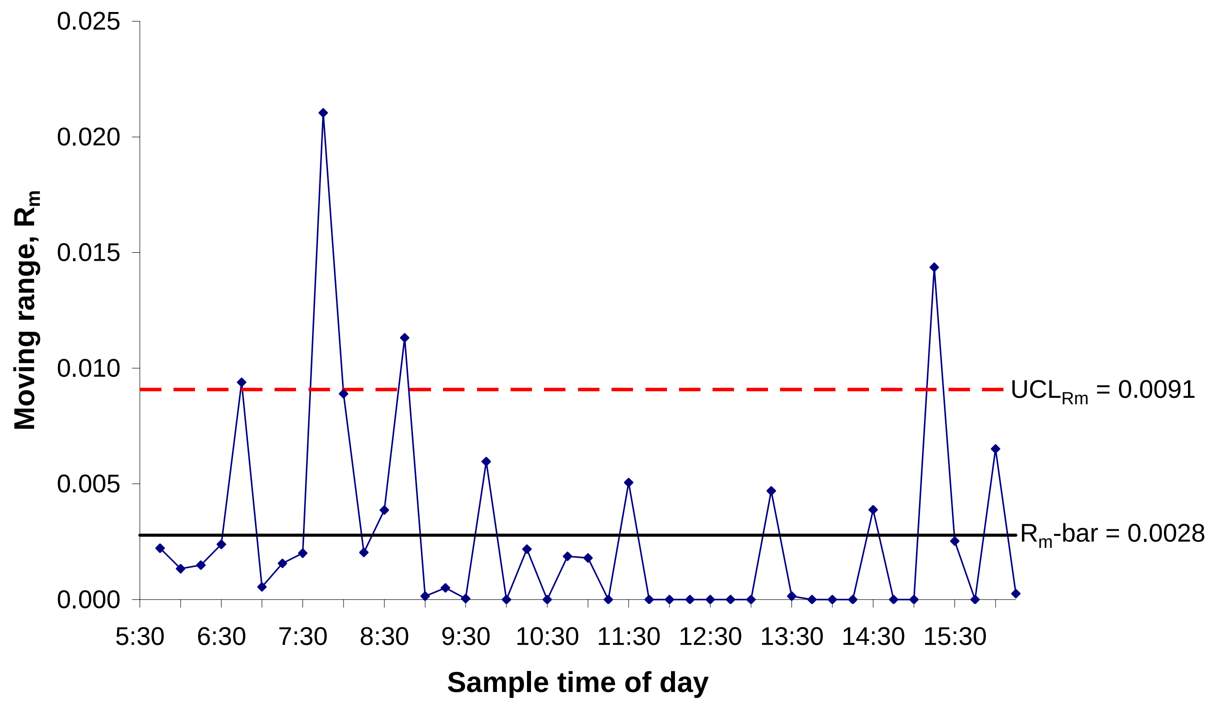


Figure 6.10 Modified X and R_m control charts for %-ISR, 0430 to 1630 shift, Alyth Yard

6.5 Measuring Pull-Down Performance

Several options have already been presented for improving the performance of the pull-down process. However, the ability of most yards to gather and record the cycle times for the components of Equation 4.3 is limited. For this research, direct observation was used to gather the component cycle times but this solution is not practical for gathering large amounts of data. There are also safety concerns with having an extra person in an active switching area as well as labor and management workload issues. Most yard managers do not have the time to spend gathering time-study data. Bringing in outside consultants is a possibility, but they come with a cost and are not permanent. On the labor side, most workers are reluctant to have a manager watching their every move with a stopwatch. Initial studies can be conducted in this manner, but long-term performance tracking for use in planning and evaluation requires a more automated approach.

The existing hump control system offers a means to gather one component of the pull-down cycle time. In Section 4.3.2, the method for pull-down crews to obtain a block authority is described. Whenever a crew is working on a bowl track, a block authority must be in place to prevent any cars from being humped to that track. The hump control system records the time the block was placed and the time it was released. This can be used as a reasonable approximation for the time spent in the coupling component. To calculate the remaining components (setup, transport and rework), additional systems must be installed. One possibility is using Global Positioning System (GPS) units to track the movement of engines, cars or crews as they go through the pull-down process. The cycle times for each component can be extrapolated from the time the GPS unit spends in each area: the bowl, lead tracks and departure yard. These data can then be used to evaluate the performance of the pull-down process. It can also be used with the event data, DTC and bowl picture to more accurately calculate the capacity reduction caused by dirty tracks.

Once a system is in place to gather and analyze the pull-down cycle times, the proposed terminal control system will be complete. The system will eliminate an information gap in the current systems regarding the performance of the pull-down process, the interaction between the pull-down and the sorting process and the condition of the bowl. As the bottleneck in the system, proactive monitoring of the pull-down process and its interactions is necessary to help increase overall terminal capacity as part of the Lean Railroading approach.

CHAPTER 7: FURTHER RESEARCH AND CONCLUSIONS

In the course of this research, several topics were identified as potential areas for further research. These areas include:

1. *Updated car cycle time analysis* – The study conducted by Kwon et al. 1995 used data from 1990 and 1991 to analyze the reliability for three types of rail service. An update to this study using current data would be a necessary precursor to several other studies in this list.

2. *Better understanding of the impact of various factors on network efficiency* – A more detailed statistical analysis of the relationship between network efficiency, as measured by average train speed or another measurement, and the components in Figure 1.9 would help identify the best areas to target improvement efforts.

3. *More accurate calculation of terminal capacity* – The complex interactions between a terminal and the rest of the rail network, as well as the dynamics of the process interactions within the terminal, make determining the actual capacity difficult. The insight gained from factory physics provide a rough estimate of terminal capacity. Further refinement of the capacity calculation may help the network planning and recovery problems.

4. *The impact of new technology on terminal operations and design* – Remote Control Locomotive (RCL) technology has reduced labor costs and increased productivity for many types of switching operations including classification at flat yards, pull-down in both hump and flat yards, and has been used with hump engines in certain locations. Remote switch machines have also increased crew productivity. This has shifted the hypothetical cost curves in Figure 2.5 and made flat yards more economical for higher

volumes. At the same time, the advent of inert retarders has made "mini-hump" designs more feasible for lower volumes. Full development of the cost curves in Figure 2.5 as well as a curve for "mini-humps" would enable better decision-making regarding use of the appropriate designs and technology to improve the efficiency of particular yards.

5. *More extensive pull-down time studies* – Chapter 4 should only serve as the beginning of a larger effort to document and understand the pull-down process. Because the pull-down is the most common bottleneck, increases in pull-down capacity generally will have the greatest impact on overall terminal capacity. A larger-scale project should be undertaken to conduct time studies at several terminals to build averages and expected variance for each component of pull-down cycle time. This information can be used to develop the most appropriate improvement options by identifying the "bottleneck within the bottleneck" (Logan 2006b). Cost-benefit analyses for each alternative can also be conducted.

6. *Terminal simulation models* – Several companies including Simul8 and TranSystems have developed simulation models and templates for hump yards (Dronzek 2006). CN has also provided the University of Illinois with copies of several yard simulation models. These models could be used to conduct sensitivity analyses on bottleneck management methods or to develop better understanding of the process interactions.

7. *"Dynamic use" of block to track allocation plans* – Algorithms for dynamic assignment plans and multiple-stage sorting techniques have been around since the 1980's (see Daganzo et al. 1983, Daganzo 1986, Daganzo 1987). However, it appears that these plans are not regularly used. If a yard is dynamic (like Bensenville), it is because it does not have a static plan in place. The blocks are dynamically assigned by the operator on a

short term basis, not as part of a larger plan. With the technology available to implement the algorithms, the time has come to study how to best implement dynamic assignment plans and multiple-stage sorting techniques as a function of yard volume, inbound traffic flow and pull-down capacity.

8. *Optimization methods for use with the BCM* – The BCM was designed to help the hump controller make better decisions during the sorting process. A natural extension would be decision support tools based on the BCM that would tell the operator the best track to send a car based on the current condition of the bowl. A more advanced version could integrate the dynamic track assignment plan or multi-stage sorting techniques.

7.1 Conclusions

The importance of railroad classification terminals to service reliability is well established (Martland et al. 1992, Martland et al. 1994, Kwon et al. 1995). Although the impact on network efficiency is more challenging to quantify because of complex interactions in the railroad network, estimates have been calculated. For example, Logan (2006a) estimated that every 15% reduction in system-wide average terminal dwell time results in an increase in carload velocity of approximately 2 mph. In Section 1.4.2, I provide an estimate of the relationship between system-wide average terminal dwell time and average manifest train speed for six Class I railroads. For each railroad, there was a statistically significant relationship between average terminal dwell time and average manifest train speed. The R^2 values ranged from 20.7% to 34.7% indicating that this amount of the variation is explained by the models. Considering all of the factors affecting system performance, the percentage of variability accounted for by the single variable of dwell time further underscores the importance of terminal performance on network efficiency and the potential benefits of improvements in terminal operations.

Inadequate terminal capacity is viewed by many to be impeding the ability of the railroads to increase their service reliability and network efficiency (Martland et al. 1992, Kraft 2000b, Barker 2006, CPR 2005g, Logan 2006a). However, methods to increase terminal capacity should be thoroughly explored in parallel with consideration of infrastructure investment. In Chapter 2, it was shown that classification terminals can be considered production systems and additional insight into their operations can be gained through the use of Factory Physics. This allows application of a new approach to terminal performance improvement called Lean Railroading. Lean Railroading is the adaptation of proven manufacturing management methodologies including lean, Theory of Constraints and Statistical Process Control as discussed in Chapter 3.

The most important manufacturing process analog to improving yard capacity is the bottleneck. In a production system the bottleneck is the process that limits its throughput. As such, the processing rate of the bottleneck sets the rate for the entire system. Improving the performance of the bottleneck is the best way to improve the performance of the entire terminal process. The train assembly (pull-down) process has been identified as the bottleneck in a majority of classification yards. Chapter 4 provides a detailed analysis of the pull-down process using SPC flowcharts and enhancing the macroscopic evaluation method for the pull-down process (Wong et al. 1981). The pull-down process was broken into five components and cycle time equations were developed for each one. Time studies at Bensenville Yard were used to establish a baseline capacity level of 541 cars per day. The potential capacity improvement of several bottleneck management alternatives over the baseline case was discussed in Chapter 5.

Of the bottleneck management alternatives discussed, the two that exploited the interaction between the hump and the pull-down process resulted in the greatest capacity improvements. This is consistent with the Theory of Constraints. For Bensenville Yard, the first method, using the hump engine to build trains when it would otherwise be idle, would result in an estimated 8.3% capacity increase. The second, eliminating rework at the pull-down end, would result in an estimated 6.5% pull-down capacity increase; the same as adding another pull-down engine. Other yards could expect to obtain similar results.

Production systems often focus too much on quantity and not enough on quality and hump yards are no exception. Hump controllers are rated primarily on the number of cars humped during their shift with little emphasis placed on how well they have sorted those cars. This is because no measurement of sorting performance was available. A quality of sort metric called the Bowl Condition Monitor (BCM) was developed and presented in Chapter 5. The BCM consists of two components: the Dirty Track Counter (DTC), which seeks to measure the number of dirty tracks and their heterogeneity, and the Incorrect Sort Rating (ISR), used to measure adherence to a static track allocation plan if one is in place. The DTC can be used in any yard while the ISR can only be used in yards with static allocation plans. Chapter 6 described how the BCM would fit into an existing terminal control system and presented charts and methods for implementing the DTC and the ISR. Sustaining this quality emphasis will require management focus to shift from the hump to the pull-down process. Finally, the need for better pull-down performance monitoring was discussed with GPS proposed as a possible solution.

Pull-down capacity at Bensenville Yard can be increased an estimated 25% by improving the process and its interactions without adding any engine or labor expense. The lean emphasis on reducing idle time between all yard processes will further increase capacity. By combining scheduled railroading with a version of Lean Manufacturing in their yards, CPR reported average terminal dwell fell from 30.4 hours in March 2005 to 21.7 hours in March 2006 (CPR 2006a). Assuming a constant terminal volume of 1,500 cars, this results in an estimated average terminal capacity increase of 475 cars per day (Equation 2.1), a 40% increase. Over the same time period, average train speed increased 3.6 mph.

REFERENCES

Armstrong, J.H. (1998). *The Railroad: What It Is, What It Does*, 4th ed. Simmons-Boardman, Omaha.

Association of American Railroads (2002). Analysis of Class I Railroads.

Association of American Railroads (2003-2006). Green Book Analysis for 2002 – 2005 [Electronic].

Association of American Railroads (2006a). Rail Intermodal Transportation. Background Papers [Online], Available from: http://www.aar.org/GetFile.asp?File_ID=143. [July 17, 2006].

Association of American Railroads (2006b). Weekly Railroad Performance Measures [Online]. Available from: http://www.railroadpm.org. [May 31, 2006].

Barker, M. (2005). Assistant Vice President – Network Solutions, CN. Personal communications regarding CN's yard initiatives.

Bragman, T. (2006). Norfolk Southern Perspective on the Future of Single-Car Railroad Shipments, *Transportation Research Board 85th Annual Meeting* presentation, January 25, 2006.

Brandimarte, P. and A. Villa (1995). *Advanced Models for Manufacturing Systems Management*. CRC Press, Boca Raton, FL.

Canadian Pacific Railway (2004). Rendering Value from IOP Compliance, a "How To" Guide: Part 1 – The Hidden Truths, [Electronic] *CPR Internal* document, [October 21, 2004].

Canadian Pacific Railway (2005a). 10 Yards Daily Data 2004 to Date, [Electronic] *CPR Internal* document, [March 28, 2005].

Canadian Pacific Railway (2005b). Research trip to company headquarters, Calgary, AB, March 21-22, 2005.

Canadian Pacific Railway (2005c). IOP Marshalling Alyth November 7th, 2005, [Electronic] *CPR Internal* document, [November 29, 2005].

Canadian Pacific Railway (2005d). Toronto Report with Shifts, [Electronic] *CPR Internal* document, [July 12, 2005].

Canadian Pacific Railway (2005e). Alyth Yard site visits, March 23-24, 26, 2005.

Canadian Pacific Railway (2005f). Agincourt Yard site visit, July 18, 2005.

Canadian Pacific Railway (2005g). Yard Operations Performance Group. Personal communications regarding CPR's yard initiatives.

Canadian Pacific Railway (2006a). Key Metrics [Online], Available from: http://www8.cpr.ca/cms/English/Investors/Key+Metrics/default.htm. [May 31, 2006].

Canadian Pacific Railway (2006b). YOP Update to SAMs on March 2nd, [Electronic] *CPR Internal* document, [March 3, 2006].

Canadian Pacific Railway (2006c). Yard Operations Performance Group. Personal communications regarding CPR's yard initiatives.

Canadian Pacific Railway (2006d). Copies of pull-down switch lists, obtained during site visits, *CPR Internal* documents, [March 7, 20-22, 2006].

Canadian Pacific Railway (2006e). Bensenville hump statistics, [Electronic] gathered from *CPR Internal* documents, [May 12, 2006].

Canadian Pacific Railway (2006f). Bensenville yard site visits, December – March 2006.

CN Railway (2000). Canadian National Experience with Locomotive Remote Control Technology. Submitted to the *Federal Railroad Administration, Docket No. FRA 2000-7325*, November 16, 2000.

CN Railway (2005a). MacMillan Yard data, [Electronic] *CN Internal* document, [June 1, 2005].

CN Railway (2005b). Research trip to Network Operations Centre, Edmonton, AB, March 25, 2005.

CN Railway (2005c). MacMillan Yard site visit, July 19-20, 2005.

CN Railway (2006). Key Performance Measures [Online], Available from: http://www.cn.ca/about/investors/perfmeasures/en_InvPerfMeasures.shtml. [May 31, 2006].

Daganzo, C.F., R.G. Dowling and R.W. Hall (1983). Railroad Classification Yard Throughput: The Case of Multistage Triangular Sorting. *Transportation Research-Part A*, Vol. 17A, No. 2, pp. 95-106.

Daganzo, C.F. (1986). Static Blocking at Railyards: Sorting Implications and Track Requirements. *Transportation Science*, Vol. 20, No. 3, pp. 189-199.

Daganzo, C.F. (1987). Dynamic Blocking for Railyards: Part I. Homogeneous Traffic. *Transportation Research-Part B*, Vol. 21B, No. 1, pp. 1-27.

DeVor, R.E., T. Chang and J.W. Sutherland (1992). *Statistical Quality Design and Control: Contemporary Concepts and Methods*. Prentice Hall, Upper Saddle River, NJ.

Droege, J.A. (1925). *Freight Terminals and Trains*, 2nd Edition. McGraw-Hill, New York.

Dronzek, R. (2006). Transportation Modeling Tools. *2006 Institute of Industrial Engineers Annual Conference* presentation, May 23, 2006.

Evans, M., N. Hastings and B. Peacock (2000). *Statistical Distributions*, 3rd Edition. John Wiley & Sons, Inc., New York.

Ferguson, W.L. (1980). Improving Railroad Terminal Control Systems: A Case Study of Southern Railway's Brosnan Yard, *Studies in Railroad Operations and Economics*, Vol. 28.

GE (2006). PROYARD II Process Control System. *GE Transportation* product brochure [Online], Available from: http://wwwgetransportation.com/general/Global_signaling/systems/detail/proyard.asp. [April 11, 2006].

Goldratt, E.M. (1990). *Theory of Constraints*. North River Press, Great Barrington, MA.

Gormick, G. (2005). How fluid is your yard? *Railway Age*, August 2005, pp. 18-28.

Gray, J.T. (2006). Union Pacific Perspective on the Future of Single-Car Railroad Shipments, *Transportation Research Board 85th Annual Meeting* presentation, January 25, 2006.

Hicks, P.E. (1994). *Industrial Engineering and Management: A New Perspective*, 2nd Edition. McGraw-Hill, New York.

Hopp, W.J. and M.L. Spearman (2001). *Factory Physics: Foundations of Manufacturing Management,* 2nd Edition, McGraw-Hill, Boston.

Irwin, S.H. (2005). ACE 562: Applications of Regression Models I, *Class Notes*. University of Illinois Urbana-Champaign, Fall 2005.

Jacobs, M. (2005). Alyth Value Stream Map, *CPR Internal* document.

Kraft, E.R. (1998). *A Reservations-Based Railway Network Operations Management System*. PhD dissertation, University of Pennsylvania, Philadelphia.

Kraft, E.R. (2000a). Implementation Strategies for Railroad Dynamic Freight Car Scheduling. *Journal of the Transportation Research Forum*, Vol. 39, No. 3, pp. 119-137.

Kraft, E.R. (2000b). A Hump Sequencing Algorithm for Real Time Management of Train Connection Reliability. *Journal of the Transportation Research Forum*, Vol. 39, No. 4, pp. 95-115.

Kraft, E.R. (2002a). The Yard: railroading's hidden half (part I). *Trains*, June 2002, pp. 46-67.

Kraft, E.R. (2002b). The Yard: railroading's hidden half (part II). *Trains*, July 2002, pp. 36-47.

Kraft, E.R. (2002c). Priority-Based Classification for Improving Connection Reliability in Railroad Yards Part I of II: Integration with Car Scheduling. *Journal of the Transportation Research Forum* published in *Transportation Quarterly*, Vol. 56, No. 1, pp. 93-105.

Kraft, E.R. (2002d). Priority-Based Classification for Improving Connection Reliability in Railroad Yards Part II of II: Dynamic Block to Track Assignment. *Journal of the Transportation Research Forum* published in *Transportation Quarterly*, Vol. 56, No. 1, pp. 107-119.

Kwon, O.K., C.D. Martland, J.M. Sussman, and P. Little (1995). Origin-to-Destination Trip Times and Reliability of Rail Freight Services in North American Railroads. *Transportation Research Record 1489*, pp. 1-8.

Leilich, R.H. (2006). Railroad Capacity Issues. *Transportation Research Board 85th Annual Meeting* presentation, January 24, 2006.

Logan, P. (2006a). People, Process, and Technology – Unlocking Latent Terminal Capacity. *Transportation Research Board 85th Annual Meeting* presentation, January 24, 2006.

Logan, P. (2006b). Business Leader, GE Yard Solutions. Personal communication and internal documents.

Luczak, M. (2005). 'The power of remote control', *Railway Age*, February 2005, pp. 25-27, 32.

McClish, R. (2005). 2005 Analyst Meeting Presentation. *Union Pacific Railroad* [Online], http://www.up.com/investors/attachments/presentations/2005/analysts_conf/mcclish.pdf [July 12, 2005].

Marlow, P.B. and A.C. Paixao Casaca (2003). Measuring lean ports performance. *International Journal of Transport Management*, Vol. 1, pp. 189-202.

Martland, C.D. (1982). PMAKE Analysis: Predicting Rail Yard Time Distributions Using Probabilistic Train Connection Standards. *Transportation Science*, Vol. 16, No. 4, pp. 476-506.

Martland, C.D., P. Little, O.K. Kwon, and R. Dontula (1992). Background on Railroad Reliability. *AAR Report No. R-803*. Association of American Railroads, Washington, DC.

Martland, C.D., P. Little, and J.M. Sussman (1994). Service Management in the Railroad Industry, in Conference Proceedings 2 *Railroad Freight Transportation Research Needs- July 12-14, 1993*, National Academy Press, pp. 89-103.

Middleton, J.G. (2003). 'Calgary Area Track Maps', produced by Middleton Cartographics, Maple Ridge, BC, (*commissioned by Canadian Pacific Railway*).

Mundy, R.A., R. Heide, and C. Tubman (1992). Applying Statistical Process Control Methods in Railroad Freight Classification Yards. *Transportation Research Record 1341*, pp. 53-62.

Murray, T. (2002). Wrong train running. *Trains*, July 2002, pp. 30-35.

Niebel, B.W. and A. Freivalds (1999). *Methods, Standards, and Work Design*, 10th Ed., WCB/McGraw-Hill, Boston.

Pachl, J. (2002). *Railway Operation and Control*. VTD Rail Publishing, Mountlake Terrace, WA.

Paixao, A.C. and P.B. Marlow (2003). Fourth generation ports – a question of agility? *International Journal of Physical Distribution & Logistics Management*, Vol. 33, No. 4, pp. 355-376.

Palisade Corporation (2006). BestFit™. [Online] Available from: http://www.palisade.com/bestfit/default.asp [July 22, 2006].

Petersen, E. R. (1977a). Railyard Modeling: Part I. Prediction of Put-Through Time. *Transportation Science*, Vol. 11, No. 1, pp. 37-49.

Petersen, E. R. (1977b). Railyard Modeling: Part II. The Effect of Yard Facilities on Congestion. *Transportation Science*, Vol. 11, No. 1, pp. 50-59.

Rhodes, M. (2003). *North American Railyards*. MBI Publishing, St. Paul.

Rother, M. and J. Shook (1999). *Learning to See: Value Stream Mapping to Create Value and Eliminate Muda – Version 1.2*. The Lean Enterprise Institute, Brookline, MA.

Shughart, L. and H. Nulkar (2005). Improving Railroad Financials: The Facts on Pricing, Network Efficiency, and Cost Control. *INFORMS Annual Meeting 2005* presentation, November 13, 2005.

Sobek, D.K. (2003). IME 458: Production and Engineering Management, *Class Notes*. Montana State University, Spring 2003.

Sobek, D.K and C. Jimmerson (2003). Applying the Toyota Production System to a Hospital Pharmacy. *Proceedings of the Industrial Engineering Research Conference 2003*.

[Online] Available from: http://www.coe.montana.edu/ie/faculty/sobek/IOC_Grant/papers.htm [July 6, 2006].

Spearman, M.L. (2002a). To Pull or Not To Pull, What is the Question? Part I: What is Lean?', *White Paper Series*. Factory Physics, Inc., College Station, TX. [Online] Available from: www.factoryphysics.net [August 20, 2004].

Spearman, M.L. (2002b). To Pull or Not To Pull, What is the Question? Part II: Making Lean Work in Your Plant. *White Paper Series*, Factory Physics, Inc., College Station, TX. [Online] Available from: www.factoryphysics.net [August 20, 2004].

Sussman, J.M. (1975). The Relation of Terminal Performance to Level of Service in the Freight Industry. *A Seminar on Terminal Performance Measurement Systems- May 1975*, Labor/Management Task Force on Rail Transportation, October 10, 1975, pp. D-1-D-17.

Turnquist, M.A. and M.S. Daskin (1982). Queuing Models of Classification and Connection Delay in Railyards. *Transportation Science*, Vol. 16, No. 2, pp. 207-230.

Union Pacific Railroad (2005). Roseville Yard site visit and interview, November 11, 2005.

Van Hattem, M. (2004). Map of the month: BNSF's hub and spoke. *Trains*, January 2004, pp. 56-57.

Wegner, R. (2003). Map of the month: major yards. *Trains*, April 2003, pp. 42-43.

Wetzel, J.A. (1985). Optimal Use of Classification Yards. *Transportation Research Record 1029*, pp. 36-41.

Womack, J.P., Jones, D.T. and D. Roos (1990). *The Machine that Changed the World*. HarperPerennial, New York.

Womack, J.P. and D.T. Jones (2003). *Lean Thinking: Banish Waste and Create Wealth in Your Corporation*. Free Press, New York.

Wong, P.J., Sakasita, M., Stock, W.A., Elliott, C.V. and M.A. Hackworth (1981). Railroad Classification Yard Technology Manual – Volume I: Yard Design Methods. *FRA/ORD-81/20.I, Final Report*. National Technical Information Service, Springfield, VA.

Ytuarte, C. (2001). Getting over the hump. *Railway Age*, October 2001, pp. 29-30.

APPENDIX A: REGRESSION RESULTS FOR CN, BNSF, CSX, NS and UP

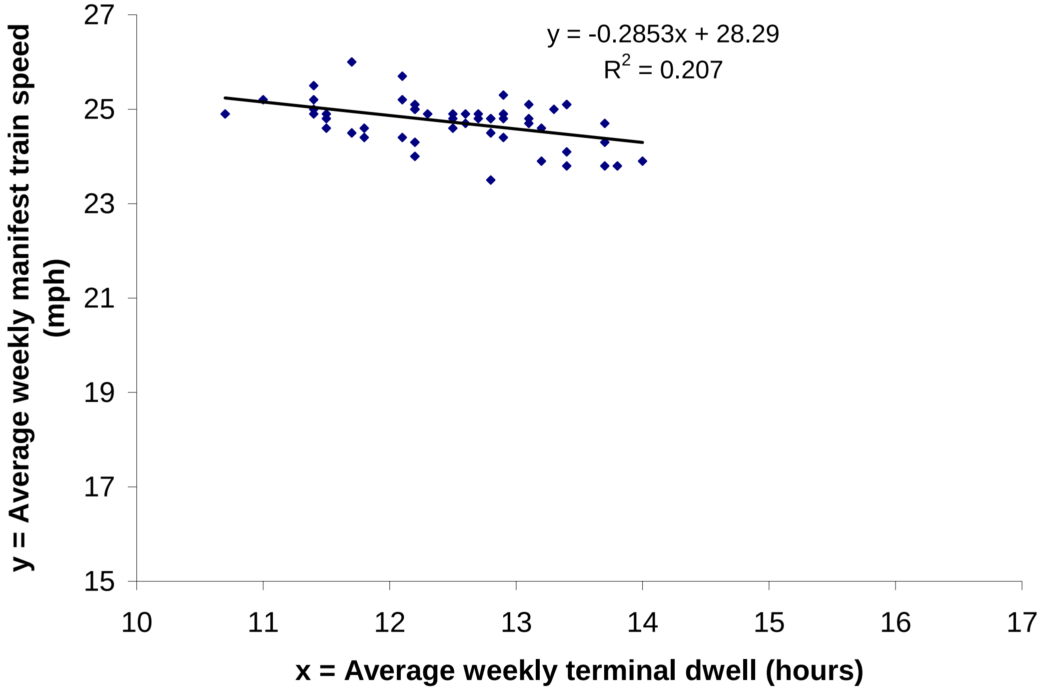


Figure A.1 Scatter plot and regression results: CN, weeks ending May 27, 2005, to May 26, 2006

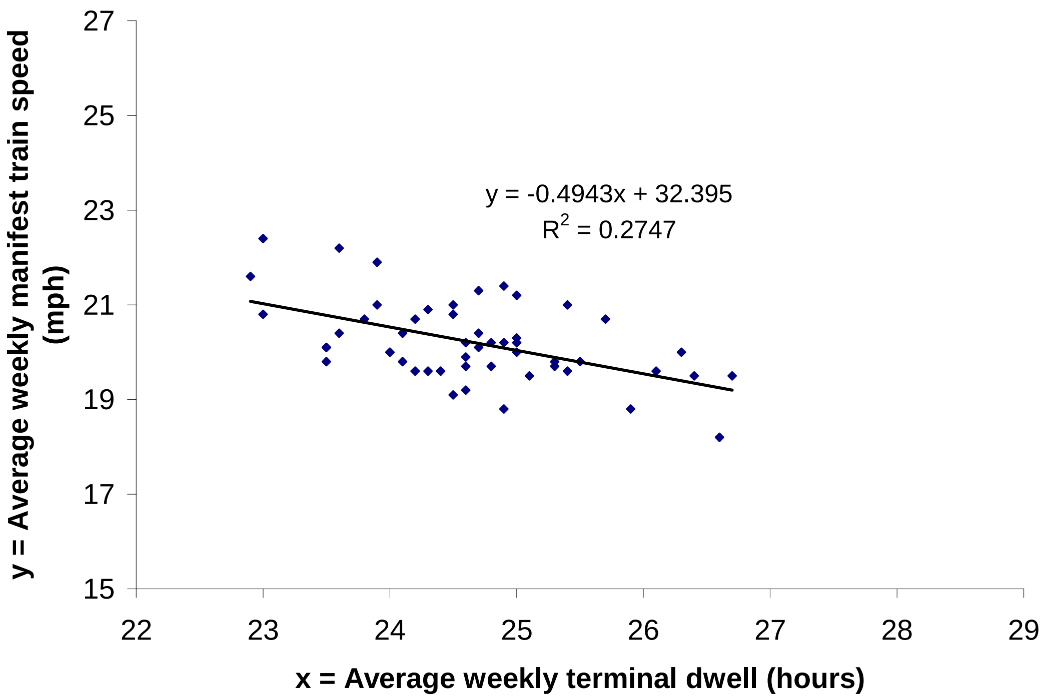


Figure A.2 Scatter plot and regression results: BNSF, weeks ending May 27, 2005, to May 26, 2006

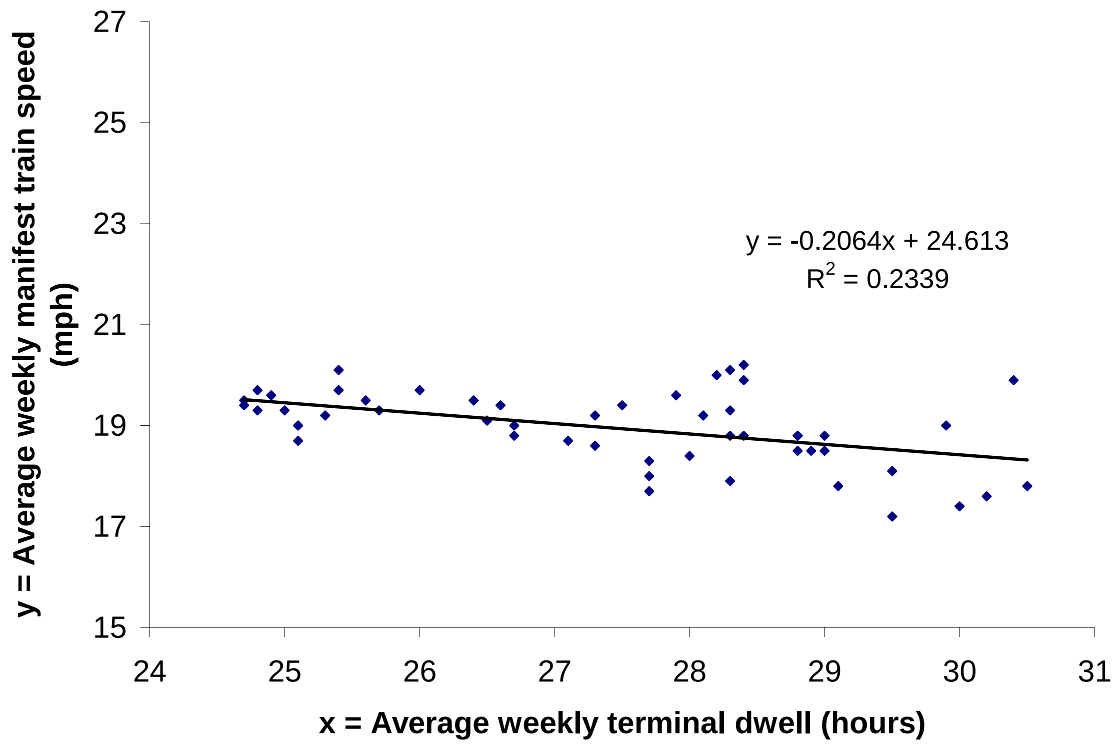


Figure A.3 Scatter plot and regression results: CSX, weeks ending May 27, 2005, to May 26, 2006

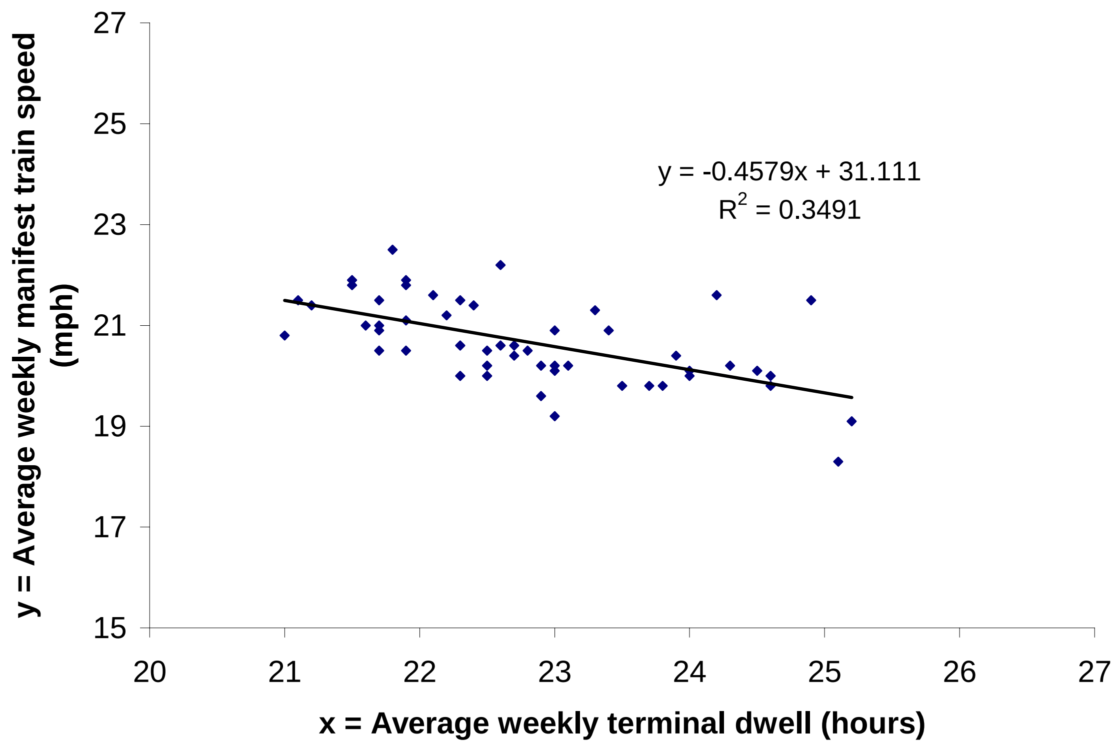


Figure A.4 Scatter plot and regression results: NS, weeks ending May 27, 2005, to May 26, 2006

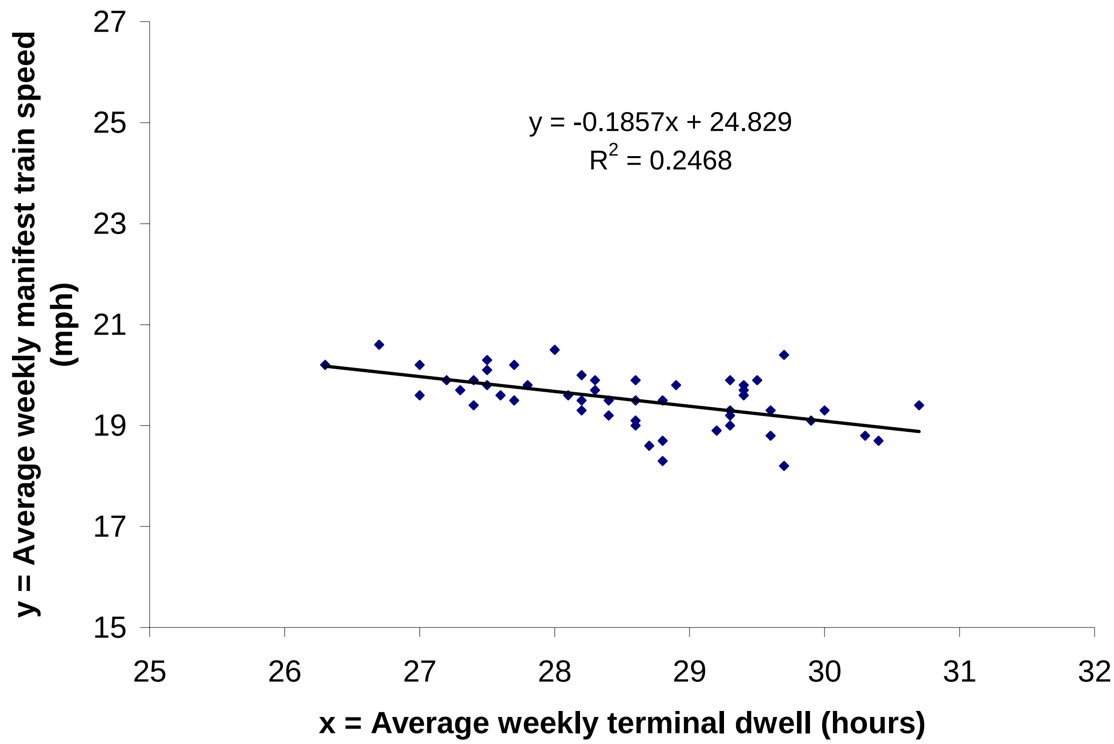


Figure A.5 Scatter plot and regression results: UP, weeks ending May 27, 2005, to May 26, 2006

APPENDIX B: THE BOWL REPLAY PROGRAM

The Bowl Replay program was developed to fill a specialized need not met by the yard simulation models currently available. While it might have been possible to adapt an existing simulation model, writing a specialized program resulted in more efficient data analysis based on the objectives of the program, which were:

1. Use existing yard event data, already captured by yard control systems, to accurately replay the events that occur in the bowl of a hump yard
2. Automatically calculate and track the ISR at the three levels (car, track and bowl) throughout the run-time of the program
3. Automatically record the bowl volume and the corresponding Bowl ISR for use in development of that relationship
4. Explore effective ways to present bowl status and ISR values to yard management

The program is not a simulation because no attempt is made to extrapolate the future state of the bowl. It only seeks to analyze past events in order to gather enough information about the dynamics of the bowl to model the impact on the pull-down workload.

B.1 Program Overview

The Bowl Replay Program is an Excel VBA program that uses event data normally captured by the terminal control systems of CPR combined with operational information gathered during site visits. The program works by recording the starting location of every car in the bowl at midnight on the day in question and the program runs for the number of events as entered by the user (Figure B.1). Excel VBA was chosen because the Excel spreadsheet format would allow for the relatively easy display of a graphical representation of the bowl. The main interface screen (Figure B.2) shows the current state of the bowl with each row representing a track and each colored cell representing a car in the bowl. The coloring scheme is based upon the previously described dirty vs. clean track definition (Figure 4.4) and allows for quick identification of the separations on each track. CN uses a similar coloring scheme for the bowl integrity screen in its SMARTYARD system (CN 2005c). For the Bowl Replay, the first car on every track is colored green and every additional car from the same block on that track is also colored green. When a car from a different block is placed on that track, it is colored red. Any other cars from that second block are colored red. Cars from a third block are colored blue and so on.

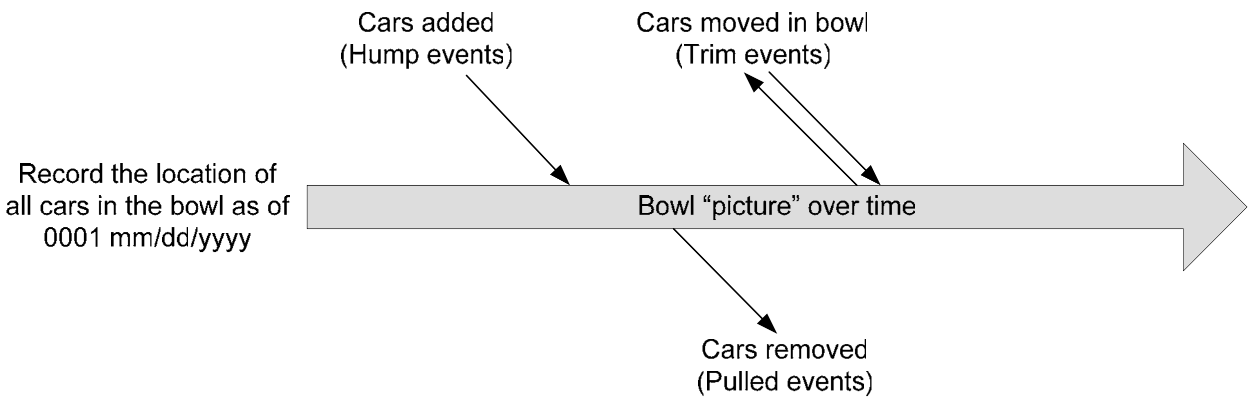


Figure B.1 Bowl Replay program event flow

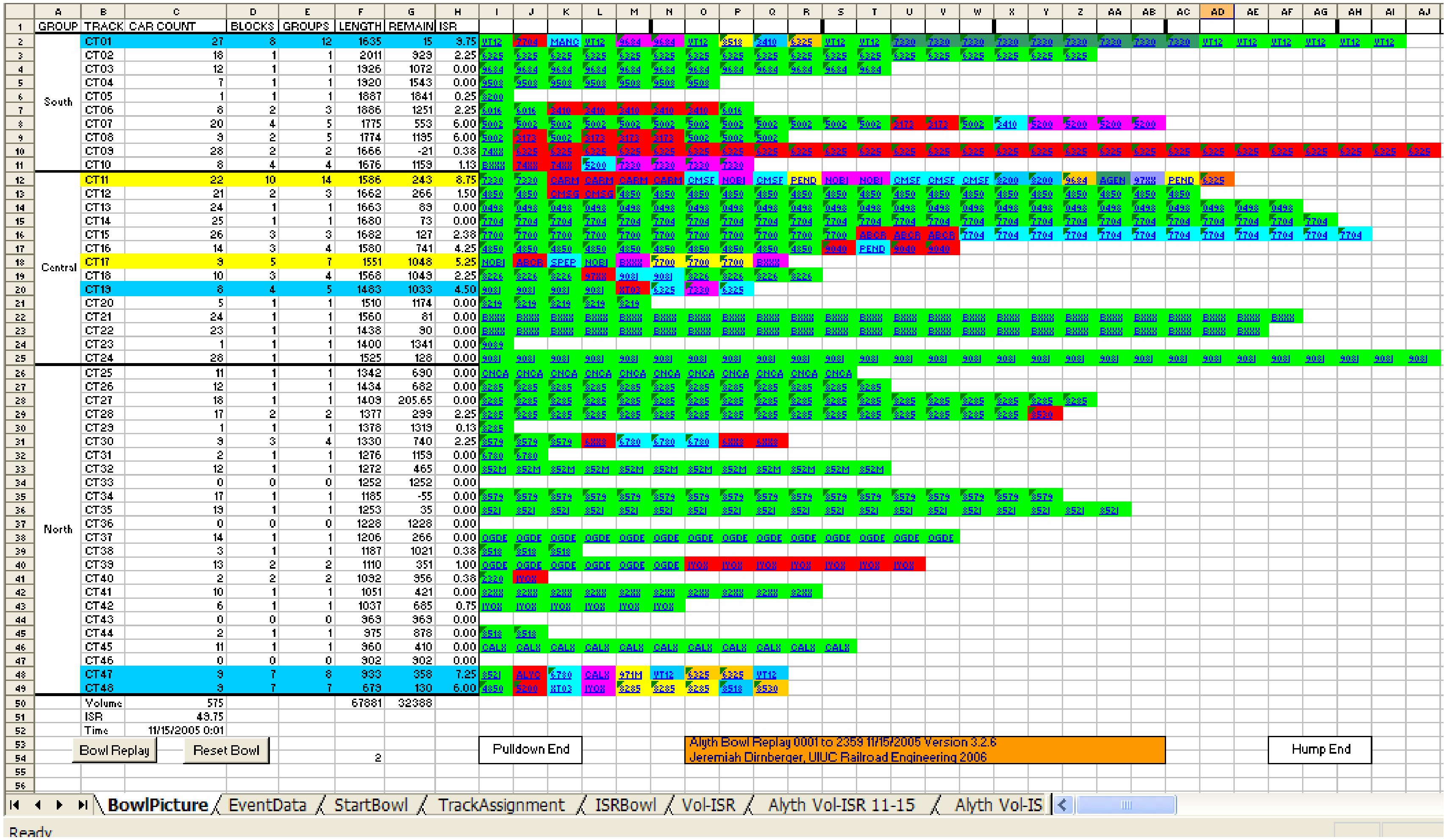


Figure B.2 Screenshot of main interface screen for the Bowl Replay program, Version 3.2.6, Alyth Yard

Macros were written to read the event data and execute a series of commands that would calculate the Car ISR, number of blocks on a track, number of groups on a track and update the "bowl picture" based on the type of event. The program is discrete-event with seven different event types:

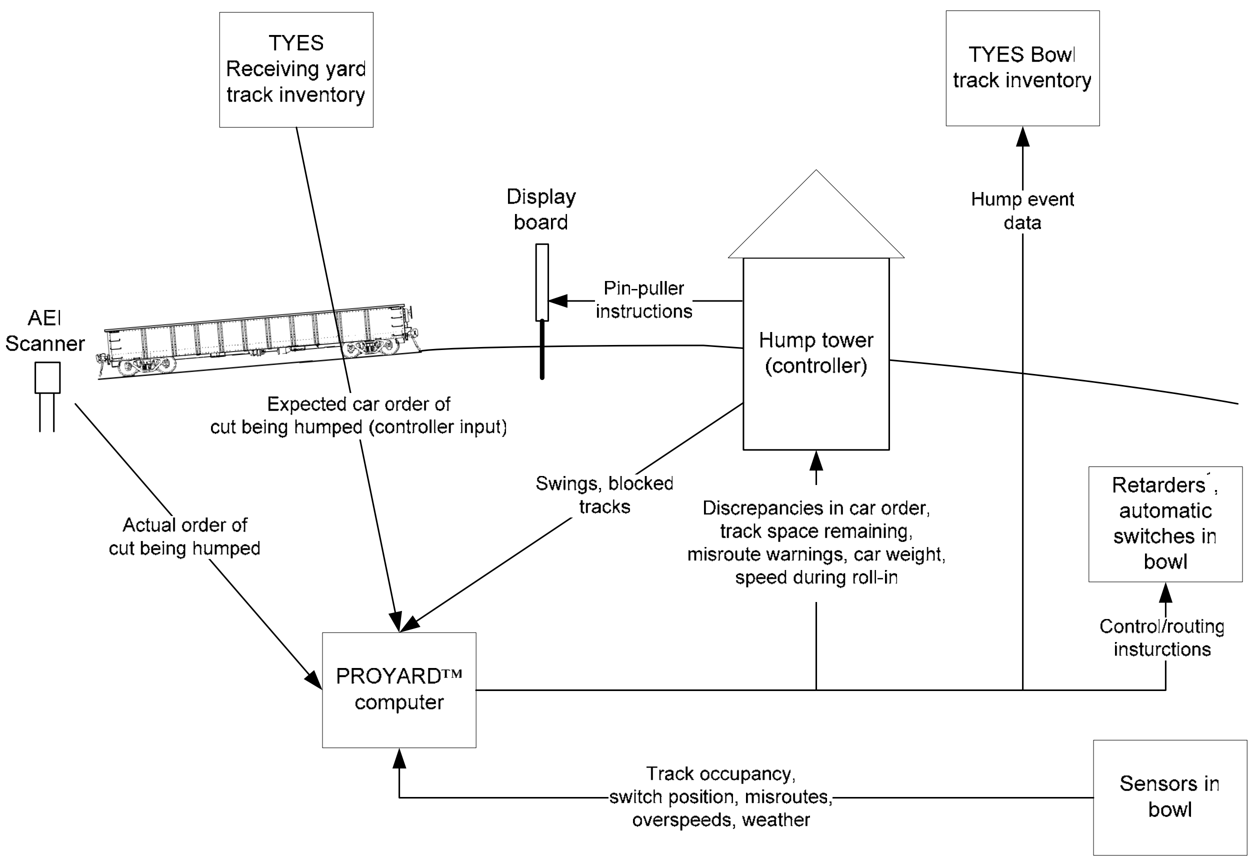
1. Hump (H) – The car is humped into the bowl
2. Pulled (P) – The car is removed from the bowl at the pull-down end
3. Hump-end trim move (Y) – The car is moved from one track to another at the hump end
4. Pull-down-end trim move (T) – The car is moved from one track to another at the pull-down end
5. Pull-down-end replacement (S) – The car is placed on the same track that it was previously removed from at the pull-down end
6. Hump-end pulled (D) – The car is removed from the bowl at the hump end
7. Hump-end replacement (R) – The car is placed on the same track it was previously removed from at the hump end

The hump, pulled, hump-end trim move and pull-down-end trim move events come directly from the TYES data. The other events are added manually to reflect movements not captured by TYES.

B.2 Program Event Data

By utilizing data that are normally captured by current yard control systems, the event data could be acquired without any modifications to the computer systems and would not create an additional burden for local yard management. The Yard Operations Performance (YOP) group at CPR provided starting bowl count and car location, hump, pull-down and class track-to-class track movement event data from the Train Yard Enterprise System (TYES). TYES is the yard management system that manages shipment connections from train to train within a terminal. It was originally developed by NS, who then sold a version of it to CPR. CPR completed implementation of TYES in 2005. The YOP group queried the central TYES mainframe in Calgary and e-mailed the necessary event data in Excel spreadsheet form.

TYES obtains the hump event data from the PROYARD™ process control system (Figure B.3). PROYARD™ is marketed by GE Transportation and is used to automate classification yards. It is used by nearly every Class I railroad in North America (GE 2006). The system identifies and measures railcars as they roll over the hump, automatically routes them to the appropriate classification track and controls railcar speed to a target of 4.5 mph ("walking speed") for coupling (GE 2006). There is an approximate delay of 2-4 minutes from the actual time a car went over the hump to the time the event is recorded in TYES. The Bowl Replay program used the time stamp from TYES.



1. Only applicable in yards with automatic retarders (i.e. Alyth, Agincourt). Dowty retarders (used in Bensenville) have no computer control.

Figure B.3 Example CPR TYES/PROYARD™ hump control system data flow

The pulled events are the result of a controller moving cars from a bowl track inventory to a departure track inventory in TYES after the pull-down crews have physically moved the cars. The class track-to-track events are the result of a similar manual procedure. Possibly as a result of this manual procedure, TYES orders a cut of cars pulled together alphabetically by reporting marks then by car number, not by standing order on the track. This presented some challenges while validating the event data during the program development.

B.3 Program Development

The first two versions of the program utilized hump event data obtained from the hump control system at MacMillan Yard (CN) in Toronto. CN only provided hump event data. Subsequent versions of the program utilized the CPR event data described in the previous section. Version 3.2.3 was the first fully-functional, bug-free version of the program. Version 3.2.6 was the final version. Table B.1 summarizes the development of the program.

Alyth Yard was the primary yard used during the development of the program. The versions for Bensenville were built during the detailed time study period that provided the data used in the capacity calculations in Section 5.1. Version 3.2.6 was used to provide the necessary data to understand the relationships between bowl volume, ISR and capacity. It also provided insight into the implementation options for the metric. The relationships and implementation options are presented in Chapter 6.

Table B.1 Bowl Replay program development summary

Version	Yard(s)	RR	New Features	Car label	Coloring scheme	Bugs
1.0	MacMillan	CN	Hump events only, Block and car counters for each track, Bowl volume counter, 1st version of track coloring macro	Destination track	Blue-Yellow	Track block coloring/count macro coloring function incorrect
2.0	MacMillan	CN	Coloring function corrected, Hump events only, No starting car location in bowl or pulled events provided	Destination track	Light Blue-Pink	Track block coloring/count macro unable to go past column Z
3.0	Alyth	CPR	Starting car location (Reset Bowl macro), hump and pulled events (H, P)	Destination track	Light Blue-Pink	Track block coloring/count macro unable to go past column Z
3.1	Alyth	CPR	Track car capacity estimator	Destination track	Light Blue-Pink	Track block coloring/count macro unable to go past column Z
3.2	Alyth	CPR	Replace track car capacity estimator with track length remaining, Car event hyperlinks, (2) Class track-to-class track events (Y, T, S) added	Destination track	Green-Red	Track block coloring/count macro unable to go past column Z
3.2.1	Alyth	CPR	Macro buttons on bowl replay sheet, Screen updating off	(1)ClassCode, (2)Destination track	Green-Red	Track block coloring/count macro unable to go past column Z
3.2.2	Alyth	CPR	Screen updating improved (event-by-event on bowl picture)	Destination track	Green-Red	Track block coloring/count macro unable to go past column Z
3.2.3	Alyth	CPR	Screen updating further improved, Track ISR added, Bowl volume and bowl ISR continually recorded	Destination track	Green-Red	
3.2.4	Alyth	CPR	User determined continuous screen updating	Train block	Green-Red	
3.2.5	(1)Alyth, (2)Bensenville	CPR	Track space remaining continually recorded, Additional hump end events (D, R) added	Train block	Green-Red	
3.2.6	(1)Alyth, (2)Bensenville	CPR	Final version of Track ISR, Groupings counter added	Hybrid (ClassCode with Local Train Blocks)	Green-Red	