Performance Evaluation of a Decision Support System in a Networked Enterprise:

Balancing Local Objectives and Network Relations

Jordan Srour
RSM Erasmus University

Tamas Mahr
Almende, Delft University of Technology

Mathijs de Weerdt
Delft University of Technology

Rob Zuidwijk
RSM Erasmus University

Hans Moonen
RSM Erasmus University

ABSTRACT

Measuring the performance of a decision support system implemented within a networked organization is neither a simple nor straightforward activity. Besides the traditional objectives that can be measured quite precisely, in a networked enterprise it is also of utmost importance that short and long-term customer and supplier relations are considered. These factors are, however, much harder to measure and to compare. We show how expert domain knowledge can be modeled with fuzzy logic, and used to find a balance between multiple quantitative system metrics and less tangible satisfaction measures. This approach is illustrated within freight logistics for the evaluation of different planning support systems.

1. INTRODUCTION

In their 1994 review of logistics performance literature, Chow *et al* indicate that while multiple methods exist for both qualitative and quantitative evaluation of logistics performance, few methods exist to assess performance in settings with a multiplicity of goals [1]. This theme is similarly documented in Krauth *et al*, in which the need to include multiple points of view in the performance measurement of logistics systems is motivated from the observation that in many cases there are conflicting needs and desires of all parties involved [2]. A basic example is that of a service provider who prefers to charge high prices and deliver a low-quality low-cost service in contrast to a customer who desires a lower price and a high-quality service. One instance in which a formalized mechanism is proposed in order to balance financial and non-financial results across both short- and long-term horizons is that of the balanced scorecard [3]. The development of the balanced scorecard must, however, be undertaken by managers and is thus capturing only a snapshot of management's perceived critical measures. As businesses move toward complex network structures it is even more important that the measures on which decision support systems are evaluated incorporate multiple performance dimensions.

As indicated by the balanced scorecard construct, decisions within a company should not be (and usually are not) based solely on performance metrics like costs and benefits. Although many things can be definitively measured, a large number of indicators may not be easily measured. Usually, the trade-off between direct or measurable costs and benefits against intangible benefits such as the satisfaction of other companies (and also employees, etc.) is understood by the human planners alone. Over years of experience, these planners learn the preferences of all the involved business entities. This paper establishes a framework to incorporate expert knowledge in order to measure the performance of a decision support system deployed in a setting with multiple business actors and relations.

The remainder of this paper is organized as follows: the next section presents the background and some pertinent literature on the history of multi-dimension evaluation constructs as well as mechanisms to model human reasoning. Section 3 describes the evaluation framework. In Section 4 we present the vehicle routing problem and the application of the evaluation framework to this vehicle routing case. Finally, in Section 5, the paper concludes with a discussion of the results and directions for future research.

2. BACKGROUND

Kaplan and Norton [4] argue that multiple performance measures are required to manage an organization. They propagate the balanced scorecard as a means to translate a strategy into a comprehensive and useful overview of business performance involving a mix of financial and non-financial measures. They consider a strategy to be a set of hypotheses about cause and effect, such as: "If we improve our logistics services, our customers will be more satisfied." Ultimate effects usually are in terms of financial performance, such as cash flow. A well constructed balanced scorecard should incorporate those performance measures that represent critical cause and effects relationships.

A system of performance measures is subject to a large number of explicit and hidden dependencies requiring simultaneous consideration. For example, one needs to balance the costs of keeping an inventory with the availability of products to meet service requirements. On the other hand, investing in IT infrastructure and advanced control mechanisms may result in reducing such operational costs while maintaining service requirements. The strategy of the organization or the network in terms of cause and effect relationships may help to pinpoint focus areas and managerial levers to improve performance, both in the short and long run. A performance measurement system in a networked environment must meet additional challenges due to the influence of multiple stakeholders. Indeed, meaningful metrics need to be defined for a network (e.g. supply chain [5]). Moreover, one needs to address local optimization behavior by considering intra- and inter-organizational coordination systems [6].

A partial view on enterprise performance may result in biased decision making with possibly undesired effects. As indicated above, optimizing individual performance while neglecting system-wide performance may result in poor decisions. The discussion of performance measurement in networked organizations is different and in principle more complicated than the local or myopic case. Even in a coordinated system, such as a supply chain, the challenge of deriving robust performance measures may be significant [3]. A bias towards specific "hard" performance metrics (i.e. readily quantified measures) may cause the neglect of critical, but difficult to quantify, satisfaction measures, such as customer satisfaction. If such a bias is propagated the customer relationship may deteriorate. It is relevant to note that in a network, many so-called "hard" metrics may coincidentally capture some satisfaction elements. This may be due to the fact that certain measures (such as costs) are often subject to intangible factors such as price negotiations (costs actually being tariffs) [8]. It is useful to distinguish direct metrics (cash flows) from derived measures (allocated costs) in order to properly assess ambiguity that may misguide decision making [9].

As mentioned previously, a proper understanding of the cause and effects between hard metrics and satisfaction measures is critical. For example, consider the cause and effect relationship: “If my trucks arrive on time at the customer's, my customer will be satisfied”. This cause and effect relationship proposes a positive relationship between timeliness of trucks at the customer site and customer satisfaction. Most people will agree with the proposal, but as such it does not provide a managerial lever to improve customer satisfaction through improving truck timeliness. There are several ambiguities present in the cause and effect relationship as indicated.

First, the customer could have measured the timeliness of trucks in terms of average amount of minutes too late, where minutes too early are neglected. The customer may just as well base her satisfaction on an extreme case or on the delivery last week. In other words, the customers' appreciation may be associated with a derived aggregate measure from the set of measurements of all deliveries, explaining her level of satisfaction. Secondly, the level of satisfaction itself is a “soft” measure that may be expressed in an ordinal scale. Each individual may respond in a different way, not only based on the actual state (e.g. level of satisfaction), but also on the understanding of the measure (e.g. what “good” stands for).

Within our framework we recommend that a module to capture such soft factors be built to represent each actor in a logistics business network. For example, in our case study of vehicle routing, a satisfaction evaluation component is built for the three classes of drivers (more generally this could be employees), customers, and society. Note that “managers” are treated as a special case given their specific business rules. Generally, it is difficult to model satisfaction of these parties properly, due to vague (verbal) boundaries of evaluation classes (e.g. “good”, “ok”, “bad” service). To tackle such issues, we propose the use of fuzzy logic for the design of these satisfaction measures.

Fuzzy logic is a well-established way of modeling human reasoning [10]. Admittedly, early uses of fuzzy logic were limited, however, with the advent of modern computing fuzzy logic has seen an increase in attention by the research community. One example of an early application of fuzzy logic may be seen in the 1977 paper by Bass and Kwakenaak in which they apply fuzzy logic to the problem of choosing between multiple alternatives [11]. The 1990s witnessed significant growth in the field including the publication of several text books on the matter [12].

The appeal of fuzzy logic is the ability to express verbal performance descriptions by so-called membership functions. These membership functions may then be combined via fuzzy rules to obtain an output set. These fuzzy rules can represent expert knowledge to express relations between both soft factors and hard measures. Finally, in a step usually called "defuzzification", the output set can be used with a pre-defined function to derive a single score denoting, in this paper, a level of satisfaction.

The objective of this paper is to illustrate a new evaluation framework by which networked organizations may score the impact of a new planning support system. This approach is premised on careful selection of performance metrics along with the conversion of such metrics into satisfaction measures. The selection of performance metrics is undertaken with a robust view of the organization – including metrics of importance across multiple stakeholder groups. The conversion these metrics to a measure of satisfaction is achieved through the application of fuzzy logic.

3. THE EVALUATION FRAMEWORK

Our evaluation framework is premised on the assumption that hard metrics may capture the performance of a decision support system in a networked setting, but only from the standpoint of one party in the network. Such measures fail to capture the satisfaction of all parties, and hence the sustainability of the system in a networked setting. As such, we have, at the core of our evaluation framework, a mechanism for the extraction of satisfaction measures from the hard metrics.

This section presents the evaluation framework as a four-step process. These four steps are stakeholder identification, key performance indicator (KPI) identification, construction of fuzzy rules to derive satisfaction scores, and application of the framework to the system under evaluation. This evaluation framework may be viewed in Figure 1.

When comparing two or more decision support systems as applied to problems in a larger network of players, it is important to evaluate the system in a multi-dimensional manner from the perspective of all stakeholders. The first step in the evaluation process is to identify all of the stakeholders. To facilitate this step we utilize the stakeholder categories described in Krauth *et al.* [2]. These categories are management, employees, customers, and society. Within these large classifications there may be one or more specific stakeholder groups that should be specifically designated. For example, in the category of employees there may be a plethora of stakeholder groups such as drivers, maintenance personnel, dispatchers, etc.

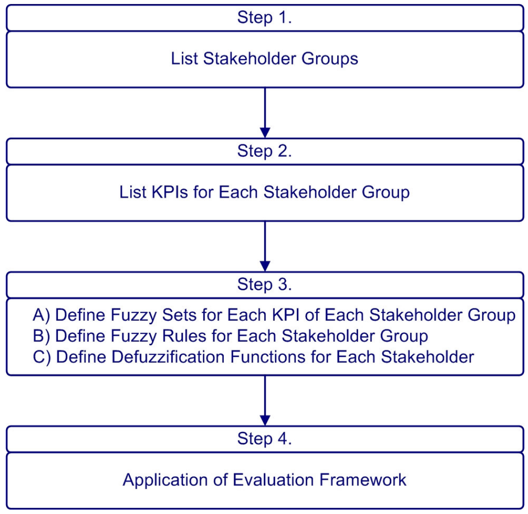

Figure 1. Overview of steps in constructing the evaluation framework.

Once all stakeholder groups have been identified, a list of key performance indicators should be constructed for each stakeholder group. These performance indicators represent key factors on which stakeholders are judging the implemented decision support system. The KPIs should also be such that they can be extracted or derived from the output of the decision support system or from the record of plan execution maintained following the implementation of the plan. Note that each stakeholder group will have one or more KPI – if no logical KPI can be defined for a stakeholder group then that stakeholder group should be removed from the set of stakeholders.

As indicated earlier, the indicators to be identified need to reflect system-wide (e.g. supply chain) performance besides individual performance of stakeholders and relate to the underlying system and stakeholder strategy. Surveying priorities of individual stakeholders may result in a bottom-up approach that does cover system-wide metrics. In practice, one needs to check how the constructed list of key performance indicators relate to the overall strategy.

Following the construction of the KPI lists it is necessary to define each indicator in terms of satisfaction. To achieve this we recommend the use of fuzzy logic. As presented in the literature review, fuzzy logic is a commonly accepted means by which to model human reasoning. The use of fuzzy logic is a three step process requiring the definition of fuzzy sets for each KPI for each stakeholder group; the construction of fuzzy rules to combine multiple fuzzy measures; and finally the definition of defuzzification functions to translate the fuzzy result into a quantifiable score. For a more detailed exposition of these three steps, the reader is advised to look in a text book on fuzzy logic; for example, *A First Course in Fuzzy Logic* by H.T. Nguyen and E. A. Walker [12], or see for example the extensive description at *Wikipedia.org*.

Fuzzy sets in effect allow the translation of a quantified metric into a verbal description of performance or satisfaction – i.e. "good", "ok", or "bad". Making this translation allows the meaning of a hard metric to be captured in a unique manner for each stakeholder. For example, considering the case of schedule deviation, in the context of order delivery, some customers may rate their satisfaction as "good" if an order arrives two minutes late. However, if the order is ten minutes late they may consider their satisfaction to be "bad"; but in the case of five minutes late – this may fall into the verbal "gray area" somewhere between "good" and "ok". Thus, fuzzy sets must be carefully constructed for each stakeholder and KPI combination.

After fuzzy sets have been constructed, fuzzy rules must be defined to merge all "fuzzified" metrics into a single fuzzy measure of satisfaction. In order to use this fuzzy measure as part of an aggregate score with non-fuzzy metrics, the final step of the fuzzy logic process is "defuzzification". In this way we can move from a perception of satisfaction to a quantifiable representation of satisfaction.

Following the definition of all the component parts of the evaluation framework, as listed in steps one through three, the framework should be implemented to determine the score of the decision support system as applied in a networked business setting. In the application of the defined evaluation framework, the final score is taken to be the linear combination of the score assigned to the various stakeholder groups (the equation for the final score may be seen in Figure 7, Section 4.2.4). Weights are assigned to each measure in the final score in order to reflect the relative importance of each stakeholder in the eyes of the entity implementing the decision support system.

To illustrate the use of this evaluation framework we apply it to a specific case in the freight logistics industry. The following section, Section 4, provides the details of this case and demonstrates how the evaluation framework may be tailored to such a setting.

4. APPLICATION OF THE EVALUATION FRAMEWORK

In this section we demonstrate how the evaluation framework set out in Section 3 may be tailored to the specific case of vehicle routing. We first provide some detail on the vehicle routing case under examination and then follow with a description of how the evaluation framework may apply.

4.1 Description of the Vehicle Routing Case

We illustrate our approach in the transportation domain. The primary task of a transportation company, often referred to as a logistics service provider (LSP), is to pickup goods from one location – generally a supplier – and to deliver them to another location – generally the customer. The LSP works directly for either the supplier or the customer. While delivering the goods efficiently is an important factor in the success of such a company, it certainly is not the only important factor. Most notably, the maintenance of good relations with customers, suppliers and sub-contractors is critical to the survival of such a company. In this section we first define the problem of delivering goods efficiently as it has been used in many optimization tools thus far, and after that, we describe the specific vehicle routing case used in this paper.

The problem of delivering goods efficiently is called the vehicle routing problem, and is defined as follows (see e.g. [7]). Assume that the following is given:

- a set of vehicles (usually initially located at a starting location called the depot, or the home base of the drivers) with a certain capacity, and
- a set of orders with a certain size, consisting of a pickup and a delivery location, and often also with a pickup and delivery time window.

What then is the best allocation of orders to vehicles, and in which order should the vehicles visit the pickup and delivery locations, such that all orders are executed while minimizing the traveled distance? As noted before, besides minimizing the direct costs by minimizing the traveled distance, there are other important factors:

- the satisfaction of each of the customers that gave the orders (and will give more orders in the future),the satisfaction of the employees, and the satisfaction of society (which may also be seen as potential customers).

In practice, such a problem is not static, but changes continually. Not only new orders may arrive, but orders may change, or vehicles may get delayed or even break down. To test our framework, we study a specific case of this problem.

Presently, researchers from the RSM Erasmus University, the TU Delft, the Free University Amsterdam, and the Center for Applied Mathematics and Computer Science (CWI) are working together with industrial partners Post-Kogeko, Vos Logistics, Almende, and CarrierWeb on the application of agent-based technologies to the vehicle routing problem. Specifically, decision support systems are developed to support the transport of containers over the road by the LSP, Post-Kogeko. Post-Kogeko is a mid-size LSP active in a several sectors, one being the transport of import and export containers, generally of the merchant haulage type. Post-Kogeko has a fleet of around 40 trucks active in this sector, handling around 100 customer orders each day.

The process of executing an order starts with the reception of an order, generally one day before required execution. The order is the request from a customer to Post-Kogeko to pickup a container at a container terminal (in case of an import container) and transport it to the customer, with delivery within a certain time window. Arriving at the customer requested location, the container is then unloaded, and the empty container is brought back to the same or another container terminal or empty depot – depending on the contract the customer has with the ocean carrier or shipping agent. This concludes the order, and the truck is ready for its next order. The process is reversed for export containers. What complicates matters is that not all containers are available at the start of operations early in the morning: either they have not physically left the ship yet, or they are delayed for administrative reasons – often due to an unsettled payment or customs. Post-Kogeko can only transport containers that have been released, and are allowed to leave the container terminal. For this reason it is hard to optimize the system in a traditional sense, since not all information is known beforehand, and will only become available sometime during the day. A large variety in the work per day, i.e., the number of orders per day and the distance to travel per order, complicates the planning process.

The planning and control of operations is currently performed manually by a team of three human planners, who take care of the order intake, the capacity planning for the next day – which means arranging the proper amount of trucks based upon required workload, and the assignment of currently executing orders to trucks. Given the primarily manual method of operations, the addition of a decision support system may greatly benefit the profitability and scalability of Post-Kogeko's operations. Researchers from the above mentioned institutes are working on three different agent based decision support systems for this highly dynamic case of the vehicle routing problem. To determine which system performs the best in a simulated setting, we propose evaluating the simulation output for these three systems using both traditional or hard metrics (e.g. empty distance traveled) and soft or satisfaction measures (e.g. customer satisfaction).

4.2 Evaluation Framework for the Vehicle Routing Case

Recalling the four step evaluation framework construction process detailed in Section 3, this section describes the construction and application of the evaluation framework used in the specific problem of vehicle routing. This section is divided into four subsections organized around the four framework steps.

4.2.1 Step 1: Identification of Stakeholders

Using the four general stakeholder classifications identified in Krauth *et al*. [2], a careful consideration of the stakeholder groups within each category was undertaken. Considering the vehicle routing problem, within the management category we consider only the managers for the company owning the fleet of vehicles to which the routing decision support system will be applied. The category of employees could encompass many stakeholder groups within the context of the vehicle routing problem. We, however, restrict our framework to only one stakeholder group – drivers. This decision was made based on the understanding that the drivers will be the most impacted by the plan emerging from the routing decision support system. Additionally, driver retention is often stated as a management goal; hence, it seems that accounting for their satisfaction is important. In practice each customer would be treated differently, receiving their own set of rules (or adaptation of a master set). In this case study, however, there is only one customer stakeholder group defined as all of the companies contracting with the fleet management company for the delivery of a container to/from the Port of Rotterdam. Finally, society in this case is considered to be all citizens impacted by the performance of the vehicle routing company. Thus, we have defined the following four stakeholder groups: Management, Drivers, Customers, and Society.

4.2.2 Step 2: Identification of Key Performance Indicators

In developing a list of KPIs for each stakeholder group, we first carefully considered the output of the vehicle routing decision support system simulation. The output is expected to include the routing plan and an archive of what occurred in execution. Considering this output, we first began by listing all measures that could be derived from such output; the KPIs identified in the paper by Krauth *et al*. [2] served as a basis for this listing. Once this preliminary list was constructed, we then focused on each stakeholder group assigning to them the KPIs from the list that were considered most important and brainstorming additional KPIs that may have been previously overlooked. This process yielded ten metrics of importance; a depiction of the KPIs and how they are split across the four stakeholder groups may be seen in Figure 2.

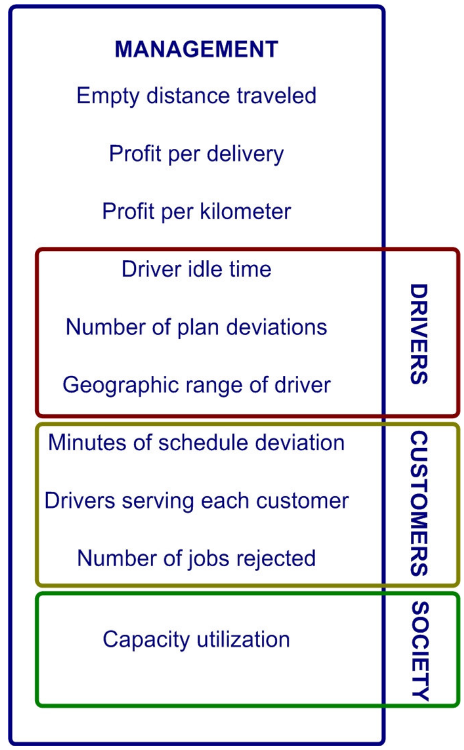

Figure 2. Division of KPIs across stakeholder groups.

4.2.3 Step 3: Definition of Satisfaction via Fuzzy Logic

As described in Section 4.1, the modeling of satisfaction is done through the use of fuzzy logic. Defining a fuzzy model requires selecting sets to model fuzzy concepts, defining connectives to combine measures via rules, and choosing a defuzzification procedure. This process can be tedious as it must be repeated over all the KPIs and for each of the stakeholder groups. As such, this section presents only the KPIs affiliated with the stakeholder group, “Drivers”, as an example of constructing a fuzzy model; the fuzzy sets and rules for all ten KPIs may be viewed in Appendix A.

<u>Step 3a) Selecting sets to model fuzzy concepts.</u>

In considering the satisfaction of the four stakeholder groups we first consider how each group might rate their experience with one KPI. We simplify this first step of set construction by considering only a limited number of linguistic terms – "good", "ok", and "bad". Note, throughout this paper we assume the following ordering: good is better than ok, and ok is better than bad. We now turn our attention to constructing the membership functions of these fuzzy terms for each KPI. Also, note that not all KPIs map to all three linguistic terms, some KPIs map only to good or bad with no descriptor for ok. This is justified as there are some measures that people would only classify as good or bad and never ok. For example, when considering order rejections in the eyes of the customer any level of rejections above zero will be bad; no rejections would be good.

Looking at the first Driver KPI, *driver idle time*, we construct three functions $D_{1,g}(x)$, $D_{1,o}(x)$, $D_{1,b}(x)$ that takes the total hours of driver idle time normalized by the total number of hours the driver is on duty, x, and returns the value of the function $D_{1,g}$, $D_{1,o}$, $D_{1,b}$, representing the degree that x falls into the verbal categories, "good", "ok", and "bad", respectively. In practice the structure of these functions may be derived via a combination of expert opinion and common sense, or even automatically learned and updated. In this example, however, for simplicity's sake we assume all functions to be of a triangular form.

In the case of vehicle routing, we consider *driver idle time* to be any time the driver is on duty, but not at a customer location and not driving. This measure is then normalized by the total number of hours the driver is on duty. Further, we assume that the drivers prefer more idle time over less idle time. Using these assumptions combined with the assumption of triangle membership functions, we obtain the following functional forms; depicted graphically in Figure 3.

$$D_{1,g}(x) = \begin{cases} 0 & \text{if } x \le .5 \\ 2x-1 & \text{if } .5 < x \le 1 \end{cases} \quad D_{1,o}(x) = \begin{cases} 2x & \text{if } 0 < x \le .5 \\ 2-2x & \text{if } .5 < x \le 1 \end{cases} \quad D_{1,b}(x) = \begin{cases} 1-2x & \text{if } 0 < x \le .5 \\ 0 & \text{if } .5 < x \end{cases}$$

Similarly, we construct membership functions, $D_2(x)$ and $D_3(x)$, for the remaining two Driver KPIs, number of plan deviations and geographic range of driver, respectively. The number of plan deviations is measured as the number of en-route diversions a given driver experiences in execution (i.e. the number of times that a driver receives instructions to carry an alternate load, after he is already in the process of driving to a previously specified order). This metric is normalized by the total number of orders that driver receives over the full horizon of execution. The geographic range of a driver is measured by examining the list of zip codes visited by a given driver as compared to the list of zip codes the driver prefers to visit – as such, this measure can range from 1 (indicating a 100% match between the two lists) to 0 (indicating a 0% match). Graphical depictions of the membership functions for these two measures may be seen in Figures 4 and 5.

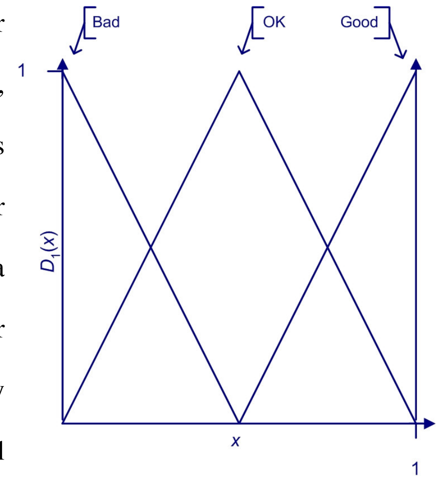

Figure 3. Graphical depiction of driver idle time membership functions.

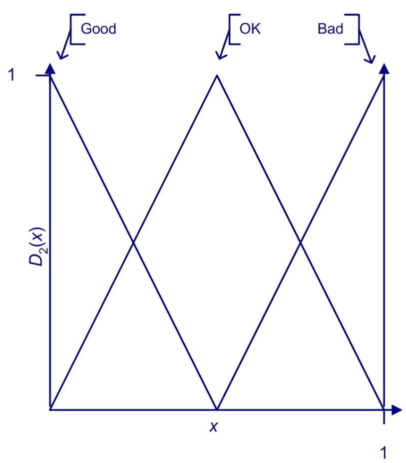

Figure 4. Graphical depiction of en route diversion membership functions.

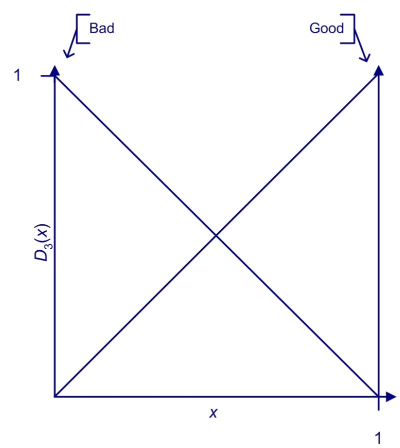

Figure 5. Graphical depiction of driver geographic range membership functions.

Moving from these three fuzzified driver satisfaction measures, encompassing a total of eight functions, to a representation of driver satisfaction requires the definition of connectives. Connectives are rules indicating the functional form of the output set or sets; defined as subset(s) of the full output space, for driver satisfaction. The definition of these connectives is the focus of Step 3b.

Step 3b) Defining connectives to combine measures via rules.

Within this example we combine the eight functions spread across the three KPIs in such a way that two subsets of the output space emerge – these represent the linguist terms "happy" and "unhappy". It is these measures of happiness that are then translated (in Step 3c) to a single score of satisfaction. Developing the rules to map the functions of Step 3a to the sets "happy" and "unhappy" is not a precise science. In practice, these rules should be derived by asking a pool of drivers which KPI influences their happiness the most or which interaction of KPIs produce a happy or unhappy result. In this case, however, we developed the rules using common sense alone.

The rules we constructed are as follows:

1. If (*idle time* is good AND d*eviations* is good AND *geographic range* is good) OR (*idle time* is ok AND *deviations* is bad AND *geographic range* is good) OR (*deviations* is ok AND *geographic range* is good) then the driver is happy.
2. If (*idle time* is bad OR *deviations* is bad OR *geographic range* is bad) OR (*deviations* is ok AND *geographic range* is bad) then the driver is unhappy.

Mathematically, these rules become:

1) Happy := $\max\{\min\{ D_{1,g}(x), D_{2,g}(x), D_{3,g}(x)\}, \min\{ D_{1,o}(x), D_{2,b}(x), D_{3,g}(x)\}, \min\{D_{2,o}(x), D_{3,g}(x)\}\}$

2) Unhappy := $\max\{\max\{ D_{1,\mathrm{b}}(x), D_{2,b}(x), D_{3,b}(x)\}, \min\{D_{2,o}(x), D_{3,b}(x)\}\}$

In this way we can map the driver experience with the designated KPIs into two sets describing a level of happiness with the system's performance. The next step addresses how we convert emerging levels of happiness/unhappiness into a single score of satisfaction.

Step 3c) Choosing a defuzzification procedure.

We are now at the point where the overall fuzzy output must be summarized in a single value. In this case, that single value represents driver satisfaction. Defuzzification is a process that maps multiple partial-membership values to one value. The defuzzification procedure that we apply for all stakeholder groups is called center-of-area.

In our case two fuzzy variables are to be defuzzified to derive the driver satisfaction measure - Happy and Unhappy. As the result of the rules defined in the previous step Happy and Unhappy will have values in [0, 1] expressing happiness and unhappiness of the driver simultaneously. Since happiness and unhappiness are the two extremes of the satisfaction spectrum we chose to define them as in Figure 6.

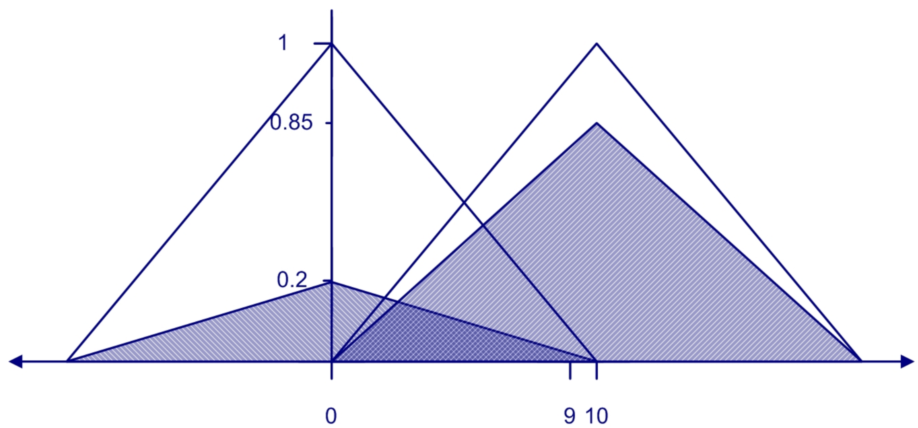

Figure 6. Illustration of "defuzzification" procedure.

Here, Happy is defined by the triangle on the points, (0,0), (10,1), and (20,0) at the upper end of the satisfaction spectrum (around the value 10). Unhappy, defined by the points (-10,0), (0,1), (10,0) is at the lower end (around the satisfaction value 0). The center-of-area defuzzification method works as follows. First the triangles representing the fuzzy values are discounted (in height) proportional to the actual values of Happy and Unhappy, in this example, 0.85 and 0.2, respectively. The point that divides the combined area of the two discounted triangles equally is returned as the result. Note that the satisfaction value will always be between 0 and 10. In the event that either Happiness or Unhappiness is zero, the satisfaction score will be either 0 or 10 respectively - regardless of the value of the other variable.

The fuzzy rules within this vehicle routing case are always defined (see APPENDIX A) to result in two fuzzy sets expressing the happiness and unhappiness of the stakeholder groups. In this regard, the same method of defuzzification is always applicable to derive the satisfaction scores of all three stakeholder groups.

4.2.4 Step 4: Application of Evaluation Framework

As noted in Section 4.1, the decision support systems our evaluation framework is aimed to judge are three different (and competing) agent-based planning systems designed and prototyped for the specific case of container transport at Post-Kogeko. The Free University Amsterdam works on a system that uses an advanced market place architecture to negotiate deals between the transportation company and its customers. Almende and the TU Delft work on a flexible planning system, where trucks and containers negotiate contracts that can be changed in case future events make this necessary/beneficial. RSM Erasmus University works on a system that focuses on real-time assignment – which means a focus on operational real-time control in a dynamic environment instead of planning beforehand, and re-arranging when events occur. The application of our proposed evaluation framework to these three systems allows for a fair comparison of the systems while also permitting the identification of their strengths and weaknesses.

The output of each of the three systems when run in simulation is a routing plan and a record of execution (that is a record detailing how the plan was actually carried out once implemented). An overview of how these output items are used in the evaluation framework is presented in Figure 7. These two items serve as the basis for the derivation of the ten KPIs identified in Section 4.2.2. Once, each of these ten metrics is derived they are fed into the fuzzy model (as described in Section 4.2.3). Note, the fuzzy model depends on a set of customer, driver, and society preferences stored external to the simulator. In this study, we define preferences at the group level, that is each individual member of a stakeholder group has the same set of preferences. Key to this application of the evaluation framework is that the management KPIs are not translated into a measure of satisfaction. This decision was made as management is usually concerned with viewing the hard metrics as an indicator of profit and performance there is thus little reason to fuzzify the metrics before including them in the final score.

Once the satisfaction scores have been obtained and defuzzified, they must be combined with the management score (an aggregation of hard metrics) to obtain a total score for the system. We recommend the use of a linear function to combine the management and stakeholder satisfaction scores such that each score is weighted by a term, α, denoting the relative importance of each stakeholder group's satisfaction in the eyes of management. We recommend that the weights be derived via organized focus groups with management.

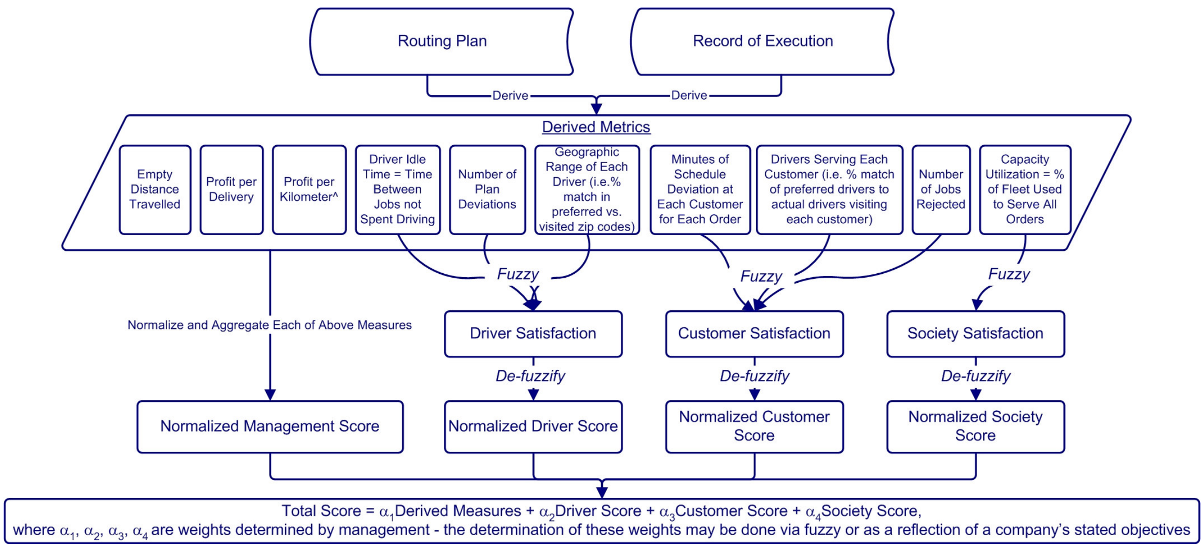

Figure 7. Application of the evaluation framework to the vehicle routing decision support system simulation output.

In this paper we propose a framework for the evaluation of networked business performance that includes both soft and hard measures. In our experiments we use this framework to evaluate the performance of competing software prototypes for logistical planning. Based on planning and execution data the evaluation framework derives scores for the soft measures by fuzzifying certain measured values, applying fuzzy rules that are defined based on expert knowledge, and finally defuzzifying the happiness/unhappiness fuzzy sets. The strength of the framework is that it formalizes the underlying business logic in a human readable way; in fact humans are primary sources in defining the right measures and rules. The next section discusses future directions for continuing this line of research.

5. DISCUSSION

This paper demonstrates the potential for a generalized evaluation framework to be tailored and applied to the problem of measuring the performance of disparate decision support systems in a freight logistics environment. The evaluation framework is unique in that it incorporates, via fuzzy logic, measures of employee, customer, and society satisfaction. The implications for this evaluation framework are significant.

Besides the use of this framework as a pure evaluation tool, we expect its usefulness also to lie in the decision support domain. A decision support tool integrated with a tailored evaluation framework will be able to suggest decisions and plans optimized on an acceptable balance of hard metrics and satisfaction scores calculated on the basis of domain knowledge gained from long-term relations with other organizations in the business network. Additionally such a system can explain these recommended decisions using human-readable rules.

Given the potential of this research, we would like to further investigate mechanisms by which to extract and model domain knowledge from experts in the logistics industry. In this paper we use fuzzy logic, but we remain open to other models. Additionally, we are interested in deriving a more realistic image of human reasoning and satisfaction from performance history data within a networked enterprise, using concepts known from the datamining and business intelligence fields.

Developments in economies around the globe impact enterprises and organizational structures in many different ways. The role of modern information and communication technologies is important in this context having a vast impact organizational processes. Competition becomes a 24/7 business, requiring real-time decision support systems. In parallel, companies increasingly operate in (supply) chains or business networks, requiring inter-organizational enterprise systems instead of traditional single-company focused systems. Performance evaluation and management, of individual companies and networks, thus becomes a crucial topic; which is surprisingly limited by existing research. The world around us is colored by perceptions and conceptions and may not be summed up by hard metrics alone. We therefore struggle with “measuring the unmeasurable”, which is likely to culminate into “controlling the uncontrollable” – a major challenge, and interesting domain for future research.

7. REFERENCES

[1] Chow, G., Heaver, T. D., Henriksson, L. E. (1994), “Logistics Performance: Definition and Measurement”, *International Journal of Physical Distribution and Logistics Management*, Vol. 24, No. 1, pp. 17-28.

[2] Krauth, E.; Moonen, H.; Popova, V.; Schut, M. (2005), Performance Measurement and Control in Logistics Service Providing; *The ICFAIAN Journal of Management Research*; Vol. 4, No. 7. August 2005.

[3] Brewer, P.C. & Th.W. Speh, (2000) Using the Balanced Scorecard to Measure Supply Chain Performance, *Journal of Business Logistics*, Vol. 21, No. 1, pp. 75-93.

[4] Kaplan, Robert S. and Norton, David P. (1996), Linking the Balanced Scorecard to Strategy, *California Management Review,* Vol. 39, No. 1, pp. 53-79.

[5] Lambert, Douglas M., and Pohlen,Terrance L. (2001), Supply Chain Metrics, *The International Journal of Logistics Management,* Vol. 12, No. 1, pp. 1-19.

[6] Ballou, Ronald H. , Gilbert, Stephen M. , and Mukherjee, Ashok (2000), New Managerial Challenges from Supply Chain Opportunities, Industrial Marketing Management, Vol. 29, pp. 7-18.

[7] J.-F. Cordeau, M. Gendreau, G. Laporte, J.-Y. Potvin, and F. Semet (2002), A guide to vehicle routing heuristics, Journal of the Operational Research Society 53, pp. 512–522

[8] Cokins, Gary (2001), Measuring Costs Across the Supply Chain, Cost Engineering, Vol. 43, No. 10.

[9] van Damme, D.A., and van der Zon, F.L.A. (1999), Activity Based Costing and Decision Support. *International Journal for Logistics Management*, Vol. 10, No. 1, pp. 71-82.

[10] Zadeh, L.A. (1965), Fuzzy sets, *Information Control*, Vol. 8, 338-353.

[11] Bass, S.M., H. Kwakenaak. 1977. Rating and ranking of multiple aspect alternatives using fuzzy sets. *Automatica* 13 47-58.

[12] Nguyen, H.T. & E.A. Walker, (1997), *A First Course in Fuzzy Logic*. CRC Press, NY.

APPENDIX A

KPI	Normalized	Fuzzy Set Functions	Fuzzy Rules
Empty distance traveled	Empty distance traveled/total miles traveled	N/A	N/A
Profit per delivery	N/A	N/A	N/A
Profit per kilometer	N/A	N/A	N/A
Driver idle time	Drive idle time/Total driver on-duty time	$D_{1,g}(x) = \begin{cases} 0 & \text{if } x \leq .5 \\ 2x - 1 & \text{if } .5 < x \leq 1 \end{cases}$ $D_{1,o}(x) = \begin{cases} 2x & \text{if } 0 < x \leq .5 \\ 2 - 2\text{x} & \text{if } .5 < x \leq 1 \end{cases}$ $D_{1,b}(x) = \begin{cases} 1 - 2\text{x} & \text{if } 0 < x \leq .5 \\ 0 & \text{if } .5 < x \end{cases}$	1. Happy := max{min{ D1,g(x), D2,g(x), D3,g(x)}, min{ D1,o(x), D2,b(x), D3,g(x)}, min{D2,o(x), D3,g(x)}} 2. Unhappy := max{max{ D1,b(x), D2,b(x), D3,b(x)}, min{D2,o(x), D3,b(x)}}
Number of plan deviations	En route diversions/total orders served	$D_{2,g}(x) = \begin{cases} 1 - 2\text{x} & \text{if } 0 < x \leq .5 \\ 0 & \text{if } .5 < x \end{cases}$ $D_{2,o}(x) = \begin{cases} 2x & \text{if } 0 < x \leq .5 \\ 2 - 2\text{x} & \text{if } .5 < x \leq 1 \end{cases}$ $D_{2,b}(x) = \begin{cases} 0 & \text{if } x \leq .5 \\ 2x - 1 & \text{if } .5 < x \leq 1 \end{cases}$	
Geographic range of driver	% of zipcodes visited in execution matching the driver's preferred list of zipcodes	$D_{3,g}(x) = x$ $D_{3,b}(x) = -x$	

KPI	Normalized		Fuzzy Set Functions	Fuzzy Rules
Minutes of schedule deviation[1,2]	Early	1. Number of early orders/Total number of orders	$C_{1,g}(x) = x$ $C_{1,b}(x) = -x$	1. Happy := min{min{ max{ $C_{3,g}(x)$, min{ $C_{1,g}(x)$, $C_{2,g}(x)$}, $C_{4,g}(x)$}, max{ $C_{7,g}(x)$, min{ $C_{5,g}(x)$, $C_{6,g}(x)$}, $C_{8,g}(x)$}, $C_{10,g}(x)$}, $C_{9,g}(x)$} 2. Unhappy := max{ $C_{10,b}(x)$, min{$C_{10,g}(x)$, max{ $C_{5,b}(x)$, $C_{6,b}(x)$, $C_{7,b}(x)$, $C_{8,b}(x)$}},min{$C_{10,g}(x)$, max{ $C_{7,g}(x)$, min{ $C_{5,g}(x)$, $C_{6,g}(x)$}, $C_{8,g}(x)$}, max{ max{ $C_{1,b}(x)$, $C_{2,b}(x)$, $C_{3,b}(x)$, $C_{4,b}(x)$}, $C_{9,b}(x)$}
		2. Number of orders in the maximum span of consecutively early orders/Total number of orders	$C_{2,g}(x) = x$ $C_{2,b}(x) = -x$	
		3. Number of orders in the maximum span of consecutively on-time orders/Total number of orders	$C_{3,g}(x) = -x$ $C_{3,b}(x) = x$	
		4. Total minutes of earliness attributable to the maximum span of consecutively early orders/Total minutes of earliness for full execution	$C_{4,g}(x) = x$ $C_{4,b}(x) = -x$	
	Late	5. Number of late orders/Total number of orders	$C_{5,g}(x) = -x$ $C_{5,b}(x) = x$	
		6. Number of orders in the maximum span of consecutively late orders/Total number of orders	$C_{6,g}(x) = -x$ $C_{6,b}(x) = x$	
		7. Number of orders in the maximum span of consecutively on-time orders/Total number of orders	$C_{7,g}(x) = x$ $C_{7,b}(x) = -x$	
		8. Total minutes of earliness attributable to the maximum span of consecutively early orders/Total minutes of earliness for full execution	$C_{8,g}(x) = -x$ $C_{8,b}(x) = x$	
Drivers serving each customer	% of drivers visiting a customer in execution matching the customer's preferred list of drivers		$C_{9,g}(x) = -x$ $C_{9,b}(x) = x$	
Number of jobs rejected	Number of jobs rejected/ Total number of orders		$C_{10,g}(x) = -x$ $C_{10,b}(x) = x$	

[1] Minutes of schedule deviation is split into two categories spanning 4 metrics each, in order to better capture how customers judge their satisfaction based on order delay (or earliness). Notice, in the expanded measures a customer's perception of satisfaction may be based on a period of time in which many orders were late or many orders were on-time.

[2] Note, in this example we consider it better to be early than late. This may not be the case for every customer; in practice we recommend the use of unique fuzzy sets for each individual customer.

KPI	Normalized	Fuzzy Set Functions	Fuzzy Rules
Capacity utilization	Total number of trucks used/Total number of available trucks	$S_{1,g}(x)=\begin{cases}1-2\mathrm{x} & \text{if } 0<x\le .5\\ 0 & \text{if } .5<x\end{cases}$ $S_{1,o}(x)=\begin{cases}2x & \text{if } 0<x\le .5\\ 2-2\mathrm{x} & \text{if } .5<x\le 1\end{cases}$ $S_{1,b}(x)=\begin{cases}0 & \text{if } x\le .5\\ 2x-1 & \text{if } .5<x\le 1\end{cases}$	1. Happy := max{$S_{1,g}(x)$, $S_{1,o}(x)$} 2. Unhappy := $S_{1,b}(x)$

.