Using Visualization to Support Investment & Divestment Decisions in the Canadian Armed Forces

Mark Rempel [1*]

[1] *Centre for Operational Research and Analysis-Defence Research and Development Canada*

Abstract: Many nations' defence departments use capability−based planning to guide their investment and divestment decisions. This planning process involves a variety of data that in its raw form is difficult for decision−makers to use. In this paper we describe how dimensionality reduction and partition clustering are used in the Canadian Armed Forces to create visualizations that convey how important military capabilities are in planning scenarios and how much capacity the planned force structure has to provide the capabilities. Together, these visualizations give decision−makers an overview of which capabilities may require investment or may be candidates for divestment.

Key words: clustering, dimensionality reduction, strategic planning, visualization.

1. Introduction

Prior to the end of the Cold War, many defence departments based their strategic planning process on the threat−based planning paradigm (NATO Re-search and Technology Organization, 2003). This paradigm is generally equipment− centric and tends to create capabilities designed for operations in a single environment. In recent years this approach has been deemed inadequate for several reasons, such as an increase in the diversity of environments in which militaries operate[1] (e.g., combat, humanitarian, counter−terrorism). As a result, many defence departments have migrated to the Capability-Based Planning (CBP) paradigm, whose focus is not only on equipment but also on personnel, research & development, infrastructure, training, etc. In contrast to threat−based planning, CBP tends to create capabilities that are: a system−of−systems; suitable to a wide range of environments rather than a single environment; and robust under uncertain conditions (Davis, 2002, Joint Systems and Analysis Group Technical Panel 3, 2005).

---

[*] Corresponding author.

[1] See examples in recent defence white papers by the Commonwealth of Australia (2013), the Government of Canada (2008), and the New Zealand Government (2010).

A component of CBP that has received much attention is how to communicate to decision–makers the capabilities' importance and the planned force structure's capacity to provide the capabilities. Christopher et al. (2009) used a weighted average approach to aggregate capability assessments from multiple scenarios and then ranked the aggregate assessments to determine the most important capabilities. Hristov et al. (2010) used the Analytical Hierarchy Process (Saaty, 2001) to determine each capability's importance and the gap between the planned force structure's capacity and the required capacity. The output was a list of capabilities that required investment, prioritized by their importance. Rempel (2012) proposed using dimensionality reduction (Salkind, 2007a, Skillicorn, 2007) and partition clustering (Jain et al., 1999, Salkind, 2007a) to visualize the importance of capabilities and show those that are required in all or only a few planning scenarios. Lastly, Pelletier (2013) investigated using the tau-x consensus ranking method (Emond and Mason, 2002) to create a ranked list of capabilities by importance.

In this collection of work two themes are evident: first, the majority of these approaches focus on the importance of capabilities rather than the planned force structure's capacity; and second, the approaches can be classified as either ranking or visualization. The visualization approach is attractive for two reasons: first, our brains understand graphics easier than words or numbers; and second, graphics are better at communicating multidimensional data (Economist, 2010).

In this paper we describe how the visualization approach is used in the CAF current CBP system. Similar to Rempel (2012), we use dimensionality reduction and partition clustering to visualize the importance of capabilities, however we extend this approach by using these methods to visualize the planned force structure's capacity and apply algorithms to select which dimensions and how many clusters to use in the visualizations. The result is two types of visualizations: first, a type that shows the importance of capabilities and groups those with similar importance; and second, a type that shows the capabilities' planned capacity and groups those with similar planned capacity. Together, these visualizations give decision–makers an overview of which capabilities may require investment or may be candidates for divestment.

The remainder of this paper is organized as follows: Section 2 describes how the importance of capabilities and the planned force structure's capacity to provide them are assessed in the CAF current CBP process; Section 3 presents a description of the dimensionality reduction and partition clustering techniques used; Section 4 discusses an example application; and Section 5 presents a conclusion.

2.   Capability and Force Structure Assessment

The assessment of the capabilities' importance and planned force structure's capacity requires four components: a set of planning scenarios, a set of capabilities, a planned force structure, and an assessment method. We now briefly present how each of these is constructed in the CAF CBP system.

The planning scenarios are instantiations of the six missions described in the Canada First Defence Strategy (Government of Canada, 2008). Each planning scenario uses

Canada's foreign policy, defence policy, and the future security environment (Chief of Force Development, 2009) to help specify its details (i.e., location, state and non–state actors, timeline, etc.). The planning scenarios are assumed to be valid for 30 years.

Each capability to be evaluated consists of several components – people, process, technology, etc. – and each is defined as 'a particular ability that contributes to a desired effect in a given environment within a specified time period and the sustainment of that effect for a designated period' (Rempel, 2010). It is important to note that a capability does not define how something will be done, rather what needs to be done. For example, 'Indirect Fire' is a capability, whereas a fighter aircraft is a component of a force structure that can perform indirect fire.

The planned force structure, which is the existing force structure plus all planned changes over the next 30 years, is separated into two sets: one that includes all changes within the 5–15 year horizon and a second that includes all additional changes within the 15–30 year horizon. For each scenario in each time horizon a subset of the horizon's planned force structure is identified as available. Only the available subsets are considered when the capacity assessments are performed.

The assessment of the capabilities' importance and the available force structure's capacity to provide the capabilities is based on qualitative judgements of subject matter experts. The importance of each capability in each scenario and time horizon is assessed to be in one of four categories: 'Critical', 'Essential', 'Routine', and 'Not Required'. The categories are described in Table 1. Likewise, the capacity of the available force structure in each scenario and time horizon to provide each capability is assessed to be in one of five categories: 'Excess', 'Matched', 'Ad Hoc', 'No Capacity', and 'Not Required'. The categories are described in Table 2. It should be noted that the capacity category 'Not Required' is only assigned to a capability when the importance category 'Not Required' is assigned to the capability.

These assessments are compiled into a set of matrices; one set of capability importance matrices and one set of capability capacity matrices. Each matrix in each set represents data for a single time horizon.

Table 1: Qualitative capability importance category definitions.

Assessment Level	Definition
Critical	Capabilities that deliver effects directly to the battlespace considered critical to mission success.
Essential	Capabilities that directly enable or support mission critical effects.
Routine	Capabilities that support the mission and will normally be deployed/employed as part of the overall package.
Not Required	Capabilities that are not required for a mission and course of action selected.

Table 2: Qualitative force structure capacity category definitions.

Assessment Level	Definition
Excess	Force structure capacity is over abundant for the scenario.
Matched	Force structure capacity is able to produce the required output.
Ad Hoc	Force structure capacity exists within the CAF but is ad hoc in nature and will constitute additional risk to force generation or sustainment.
No Capacity	Force structure capacity does not exist, or exists at such a level that it will not be able to produce required outputs for the scenario.
Not Required	Force structure capacity is not required in the scenario.

3. Methods

In this section we discuss the dimensionality reduction and partition clustering techniques used in this paper. Since the CAF CBP approach uses qualitative assessments to measure the capabilities' importance and planned force structure's capacity, we use Multiple Correspondence Analysis (MCA) to perform the dimensionality reduction. We use a variety of partition clustering techniques (k–means, k–medoids, fuzzy clustering) and cluster validation techniques (Silhouette Width, Dunn's Index, Connectivity) to suggest a number of clusters. The final number of clusters is based on qualitative inspection of the suggested solutions and selection of the most appropriate.

3.1 Multiple Correspondence Analysis

MCA is a dimensionality reduction technique that is used for categorical data. It accomplishes this reduction by identifying factors in the data, which are uncorrelated linear combinations of the attributes, that represent as much variation as possible in the data. Using this techniques makes it possible to visually assess similarities and differences between capabilities and determine whether they can be grouped. Details of MCA are found in Nenadic and Greenacre (2005) and Salkind (2007b). MCA was applied in this work as follows:

(1) **Create an indicator matrix:** Given a capability importance matrix $A(m \times n)$ or a capability capacity matrix $B(m \times n)$, where m is the number of capabilities and n is the number of planning scenarios, for a given time horizon with x possible categories for each assessment, an indicator matrix $Z(m \times (n \cdot x))$ is created. The indicator matrix is then scaled as $\hat{Z} = \tau^{-1}Z$, where τ is the sum of Z.

(2) Compute the Singular Value Decomposition: The singular value decomposition is computed as:

$$D_\phi^{-\frac{1}{2}} \left(\widehat{Z} - \phi\omega^T \right) D_\omega^{-\frac{1}{2}} = USV^T \tag{1}$$

where ϕ is a vector of the row sums of $\widehat{Z}$, ω is a vector of the column sums of $\widehat{Z}$, D_ϕ is diag(ϕ), D_ω is diag(ω), U is $m \times (n \cdot x)$, S is a $(n \cdot x) \times (n \cdot x)$ diagonal matrix with non-negative, non-increasing values $s_{1,1}, s_{2,2}, \ldots, s_{(n \cdot x),(n \cdot x)}$, and V^T is $(n \cdot x) \times (n \cdot x)$. The capabilities are given by the rows of U, the scenario/assessment category combinations are given by the rows of V, and the factors' coefficients are given by the olumns of V.

(3)Compute the variance that is represented by each factor: Due to the expansion of assessment categories in Z to scenario/assessment category combinations, the singular values provide a pessimistic estimate of the variation explained in each factor. The values are adjusted and expressed as percentages as follows:

$$s_{\alpha,\alpha}^{adj} = \frac{\left(\frac{n}{n-1}\right)^2 \left(s_{\alpha,\alpha} - \frac{1}{n}\right)^2}{\left(\frac{n}{n-1}\right) \left(\sum_{\beta=1}^{(n \cdot x)} s_{\beta,\beta}^2 - \frac{J-n}{n^2}\right)} \tag{2}$$

where $\alpha \in 1, 2, \ldots, (n \cdot x)$ and J is the total number of scenario/assessment category combinations. The adjusted values are computed for those singular values that satisfy the inequality $s_{\alpha,\alpha} \geq 1/n$.

(4) Determine which factors to retain: There are several approaches to determine the factors that should be used to visualize the data – scree plot, cumulative percentage of variance, broken stick model, Bartlett's test, cross – validation, etc. – although no single method is best. For a review of these methods see Jolliffe (2002).

   In this work we use the broken stick model, due to that it works well in practice, does not require distributional assumptions that are unrealistic, and is not computationally intensive. The model works as follows: if a stick of unit length is broken at random into Λ segments, then the expected length of the gth segment is:

$$l_g^* = \frac{1}{\Lambda} \sum_{\lambda=g}^{\Lambda} \frac{1}{\lambda} \tag{3}$$

   The proportion of variance explained by the gth factor is compared to l_g^*, and those which are greater than l_g^* are used to visualize the data.

(5) Interpret the factors:
   Interpreting the retained factors is performed by visually assessing a plot of the scenario/assessment category combinations and finding the groups of extreme values. When

two combinations from different planning scenarios are close to each other (e.g., 'Scenario 1/Excess' and 'Scenario 2/Excess'), capabilities tend to have these assessments together. When two combinations from the same planning scenario are close to each other (e.g., 'Scenario/1 Excess' and 'Scenario 1/Ad Hoc'), capabilities that have these assessments tend to have similar assessments in the remaining scenarios.

3.2 Partition Clustering

Partition clustering is an approach that is used to group objects with similar attributes. It accomplishes this by measuring the degree of similarity between objects, and based on their similarity assigns objects to clusters. Using this approach makes it possible to summarize the similarities and differences between the capabilities.

In this paper we use three commonly employed partition clustering techniques: $k-$means, $k-$medoids, and the fuzzy method $c-$means. We use multiple methods since this allows different solutions to be compared. Details of the methods are found in Bezdek (1981), Hastie et al. (2003), and Theodoridis and Koutroumbas(2006). These techniques have a variety of assumptions: the data must be on the same scale; the clusters' shapes are convex, in particular hyperspherical; the clusters are well−separated; and the data has no outliers (Jain et al., 1999, Halkidi et al., 2001). These assumptions arise from the similarity metrics used in the techniques. If these assumptions are not met, these techniques may fail to find the appropriate clustering.

A summary of how the methods compute their centroids and configuration cost is shown in Table 3. In these equations k is the number of clusters, h is the cluster index, $K(h)$ is the set of capabilities assigned to cluster h, p_i is the coordinates of capability i in the retained factors, c_h is the centroid of cluster h in the retained factors, $\pi(m \times k)$ is a capability − cluster membership matrix, and f is a fuzzifier where higher values lead to fuzzier sets. A typical value for f is 2.

The capability membershipsare computed as:

$$\pi_{h,i} = \frac{1}{\sum_{l=1}^{k} \left(\frac{||p_i - c_h||_2}{||p_i - c_l||_2} \right)^{\frac{2}{f-1}}} \tag{4}$$

Table 3: Partition clustering techniques.

Algorithm	Centriod Computation	Configuration Cost
k-means	$c_h = \frac{1}{|K(h)|} \sum_{i=1}^{m} p_i$	$w_k = \sum_{h=1}^{k} \sum_{i \in K(h)} ||p_i - c_h||_2^2$
k-medoids	$c_h = \text{Medoid}$	$w_k = \sum_{h=1}^{k} \sum_{i \in K(h)} ||p_i - c_h||_1$
c-means	$c_h = \frac{\sum_{i=1}^{m} (\pi_{h,i})^f p_i}{\sum_{i=1}^{m} (\pi_{h,i})^f}$	$w_k = \sum_{h=1}^{k} \sum_{i=1}^{m} (\pi_{h,i})^f ||p_i - c_h||_2^2$

Partition clustering methods separate capabilities into a pre−determined number of clusters; however, these methods do not identify the number of clusters that best describe the data. The preferred number of clusters may be identified through external, internal, or relative validation criteria (Jain et al., 1999, Halkidi et al., 2001). External criteria assess the validity of a clustering solution based on a priori information. Internal criteria assess the validity of a given clustering solution based upon the data itself. Relative criteria compares two solutions and determines their relative merit. In this work we use internal criteria as the aim is to select a cluster solution without a priori information.

We use three commonly used criteria: Dunn's index, Silhouette Width, and the Connectivity metric. The first and second criterion aggregate the compactness and separation of a given cluster solution into a single value. The third metric measures the connectedness of a solution. The Dunn's index (Dunn, 1974) of a clustering solution with k clusters is given as:

$$D_k = \min_{\substack{p,q \in \{1, ..., k\} \\ p \neq q}} \left(\frac{\delta_{p,q}}{\max_{h \in \{1, ..., k\}} \Delta_h} \right) \tag{5}$$

where Δ_h is the maximum distance between a pair of capabilities in cluster h (compactness), and $\delta_{p,q}$ is the minimum distance between a pair of capabilities in clusters p and q (separation). The range of values for D_k is between zero and infinity. Clustering solutions that consist of compact and well separated clusters have high D_k values.

$$S_k = \frac{1}{m} \sum_{i=1}^{m} \frac{b_i - d_i}{\max\{b_i, d_i\}} \tag{6}$$

where d_i is the mean distance of capability i to the remaining capabilities in its assigned cluster, b_i is the minimum mean distance to the remaining clusters in the solution (i.e., compute the mean distance to all capabilities in each cluster, and select the minimum). The denominator is a normalization term. The range of values for S_k is between -1 and 1, where larger values represents solutions that are compact (i.e., low d_i) and well separated (i.e., high b_i).

The connectivity metric (Handl et al., 2005) of a clustering solution with k clusters is given as:

$$Conn(K) = \sum_{i=1}^{m} \sum_{l=1}^{L} x_{i,nn_{i(l)}} \tag{7}$$

where K is a clustering solution, L is the number of nearest neighbours that contribute to the metric, $nn_{i(l)}$ is the lth nearest neighbours of capability i, and $x_i, nn_{i(l)}$ is zero if i and $nn_{i(l)}$ are in the same cluster and $1/l$ otherwise. The range of values for $Conn(K)$ is between zero and infinity, where smaller values represent preferred solutions.

4. Example

In this section we present an analysis and visualizations of a data set that describes the importance of a military's capabilities and its capacity to provide them. The data were sampled from a real data set — however, the capability and scenario names have been removed and replaced by randomly assigned integers. Although the data are a sample, their characteristics are representative of complete data set.

4.1 Data

The data used in this example are qualitative assessments as described in Section 2. The data consists of importance assessments of 100 capabilities in ten planning scenarios and capacity assessments of a planned force structure in a given time horizon. The capabilities are labelled 1,…,100 and the scenarios are labelled 1,…,10. The distribution of the importance and capacity assessments in the scenarios are shown in Table 4.

Table 4: Distribution of capability importance and capacity assessments.

Category	Scenario									
	1	2	3	4	5	6	7	8	9	10
Importance (%)										
Not Required	8	9	28	14	16	14	8	5	16	20
Routine	50	38	29	33	28	40	19	17	45	26
Essential	29	35	38	43	43	42	58	61	27	42
Critical	13	18	5	10	13	4	15	17	12	12
Capacity (%)										
Not Required	8	9	28	14	16	14	8	5	16	20
No Capacity	1	3	1	1	0	1	5	5	1	0
Ad Hoc	20	37	22	40	33	26	38	33	35	25
Matched	71	51	47	45	47	55	44	57	46	51
Excess	0	0	2	0	4	4	5	0	2	4

4.2 Multiple Correspondence Analysis

We ran a MCA on the capability importance and capacity data sets separately. Table 5 shows the adjusted variance and cumulative variance for the first eight factors in each analysis. Using the Broken Stick model in equation (3), we determined which factors to retain for further analysis. Table 5 shows that the model suggests factors one, two, and three be retained for capability importance analysis and factors one through four for the capacity analysis.

We interpreted the factors using the distribution of the scenario/assessment category combinations within them. The interpretations of the capability importance factors are shown in Table 6. The distribution of the combinations along Factor 1 is such that those with 'Routine' or 'Essential' assessments have low values and those with 'Not Required' or 'Critical' assessments have high values. This factor is interpreted as 'Moderate vs Extreme Importance'. With regards to Factor 2, combinations with 'Not Required' or 'Routine' assessments have low

values and combinations with 'Essential' or 'Critical' assessments have high values. This factor is interpreted as 'Low vs High Importance'. Lastly, while the Broken Stick model indicated that Factor 3 should be retained, we were unable to find a simple interpretation and thus did not consider it for further analysis.

Table 5: Multiple Correspondence Analysis results.

	Factor							
	1	2	3	4	5	6	7	8
Importance								
Adjusted Variance (%)	0.48	0.23	0.11	0.03	0.01	0.01	0.00	0.00
Cumulative Variance (%)	0.48	0.71	0.82	0.85	0.86	0.87	0.87	0.87
Broken Stick Model	0.11	0.08	0.07	0.06	0.05	0.05	0.05	0.04
Retain Factor?	T	T	T	F	F	F	F	F
Capacity								
Adjusted Variance (%)	0.36	0.31	0.16	0.07	0.03	0.02	0.01	0.01
Cumulative Variance (%)	0.36	0.67	0.83	0.90	0.93	0.96	0.97	0.97
Broken Stick Model	0.09	0.07	0.06	0.05	0.05	0.04	0.04	0.04
Retain Factor?	T	T	T	T	F	F	F	F

Table 6: Capability importance factor description.

Factor	Description	Notes
1	Moderate vs Extreme Importance	Low values – capability assessments are generally 'Routine' or 'Essential'; High values – capability assessments are generally 'Not Required' or 'Critical'
2	Low vs High Importance	Low Values – capabilities are generally 'Not Required' or 'Routine'; High values – capabilities are generally 'Essential' or 'Critical'

The interpretations of the capacity factors are shown in Table 7. The distribution of the combinations along Factor 1 is such that those with 'Not Required' or 'No Capacity' assessments have higher values and those with 'Ad Hoc' or 'Matched' assessments have lower values. This factor is interpreted as 'High vs Low Utilization'. Regarding Factor 2, those combinations with 'Excess' assessments have higher values; this factor is interpreted as 'Low vs High Excess Capacity'. The distribution of combinations in Factor 3 are such that those with 'Matched' or 'Excess' assessments have lower values and those with 'No Capacity' have higher values; this factor is interpreted as 'High vs Low Capacity'. Factor 4 does not have a simple interpretation, and is not included in our analysis.

4.3 Partition Clustering

Using the results of the MCA, we created four combinations of factors as shown in Table 8: one combination of capability importance factors and three combinations of capacity factors. We applied the clustering algorithms and validation metrics described in Section 3 to the capabilities

Table 7: Capacity factor descriptions

Factor	Description	Notes
1	High vs Low Utilization	Low values – capabilities are being used and are available for use; High values – capabilities are not being used or they are not available for use
2	Low vs High Excess Capacity	Low Values – capabilities are provided with low excess capacity; High values – capabilities are provided with high excess capacity
3	High vs Low Capacity	Low values – capabilities are provided with a high degree of capacity; High values – capabilities are provided with a low degree of capacity

in each combination of factors. We set the upper limit on the number of clusters to be 30 (roughly one third of the number of capabilities), For each combination of factors and validation metric, we selected the clustering algorithm and number of clusters that optimized the validation metric. The results are shown in Table 8.

Table 8: Partition clustering results.

	Validation Metric		
	Connectivity	Dunn	Silhouette
Importance (1 & 2)			
Score	13.42	0.29	0.56
Method	k–medoids	k–means	k–medoids
Clusters	2	24	27
Capacity (1 & 2)			
Score	5.50	0.40	0.74
Method	k–means	k–means	k–means
Clusters	2	2	2
Capacity (1 & 3)			
Score	2.93	0.76	0.72
Method	k–means	k–means	k–means
Clusters	2	2	2
Capacity (2 & 3)			
Score	5.48	0.27	0.76
Method	k–means	k–means	k–means
Clusters	2	2	2

The suggested algorithms and number of clusters differs between the validation metrics in the capability importance factors. Inspection of the clusters reveals that two clusters split the capabilities into those with an emphasis on moderate importance assessments (i.e., 'Routine', 'Essential') and those with an emphasis on extreme importance assessments (i.e., 'Not Required', 'Critical'). As the number of clusters increases, these two clusters are split into finer groupings. Given 24 or more clusters, as suggested by the Dunn's Index and mean Silhouette Width, are likely too many for a decision−maker to comprehend, we selected the 10 clusters solution from the $k-$medoids algorithm to group the capabilities. This selection is a compromise between the suggested solutions.

The suggested clustering algorithms and number of clusters agreed in each of the capacity factor combinations. As a result, we used these suggested groupings.

4.4 Visualization

Using the results of the MCA and partition clustering analysis, we created four visualizations: Figure 1 shows the importance of capabilities and groups those with similar importance; Figure 2 shows the force structure's capacity and highlights those capabilities with excess capacity; Figure 3 shows the force structure's capacity and highlights those capabilities with low capacity; and Figure 4 combines the low and excess capacity measures in a single figure. Each grey circle is a capability, the number to the right is the capability's number. Markers are shown to assist with interpreting the capabilities' location:

- in Figure 1, ■ is a capability with all 'Not Required' assessments; ● is a capability with all 'Routine' assessments; ▲ is a capability with all 'Essential' assessments; and * is a capability with all 'Critical' assessments; and

- in Figure 2 through Figure 4, ■ is a capability with all 'Not Required' assessments; ● is a capability with all 'No Capacity' assessments; ▲ is a capability with all 'Ad Hoc' assessments; * is a capability with all 'Matched' assessments; and ⊞ is a capability with all 'Excess' assessments.

Given the descriptions of the axes and the markers' locations, the visualizations are easily interpreted. For example, in Figure 1 capabilities that are generally not required are in the lower right, capabilities that are generally assessed as 'Routine' or 'Essential' are on the left, and capabilities that are generally critical are in the upper right. Capabilities between these regions are mixes of these categories; for example, capabilities midway between the ● and ▲ markers are a 50/50 split between 'Routine' and 'Essential'. The visualizations in Figure 2 through Figure 4 are interpreted in a similar manner.
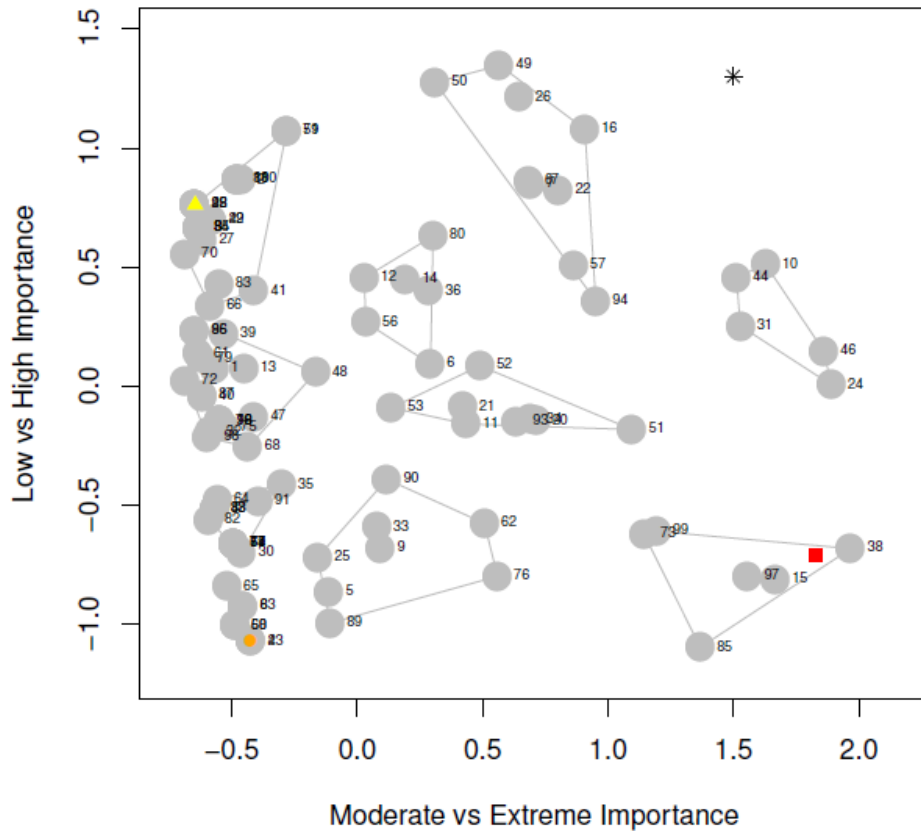
Figure 1: Capability importance and capacity visualizations – capability importance.

Recalling the importance and capacity level definitions in Table 1 and Table 2, four characteristics of the capabilities can be extracted from the visualizations:

- **Enabling Capabilities:** Capabilities in the left region of Figure 1 are more likely to be assessed as 'Routine' or 'Essential', which are defined as supporting/enabling capabilities. Without the capacity to provide these enabling capabilities, critically important capabilities may not be able to deliver their outputs;

- **Special Purpose Capabilities:** Capabilities in the right region of Figure 1 are more likely to be assessed as 'Critical' or 'Not Required', where for the former is defined as being critical to mission success. These capabilities may only be required in a few scenarios, and thus can be interpreted as special purpose. Regardless of their lower utilization, these capabilities are vital to achieving success in the planning scenarios;
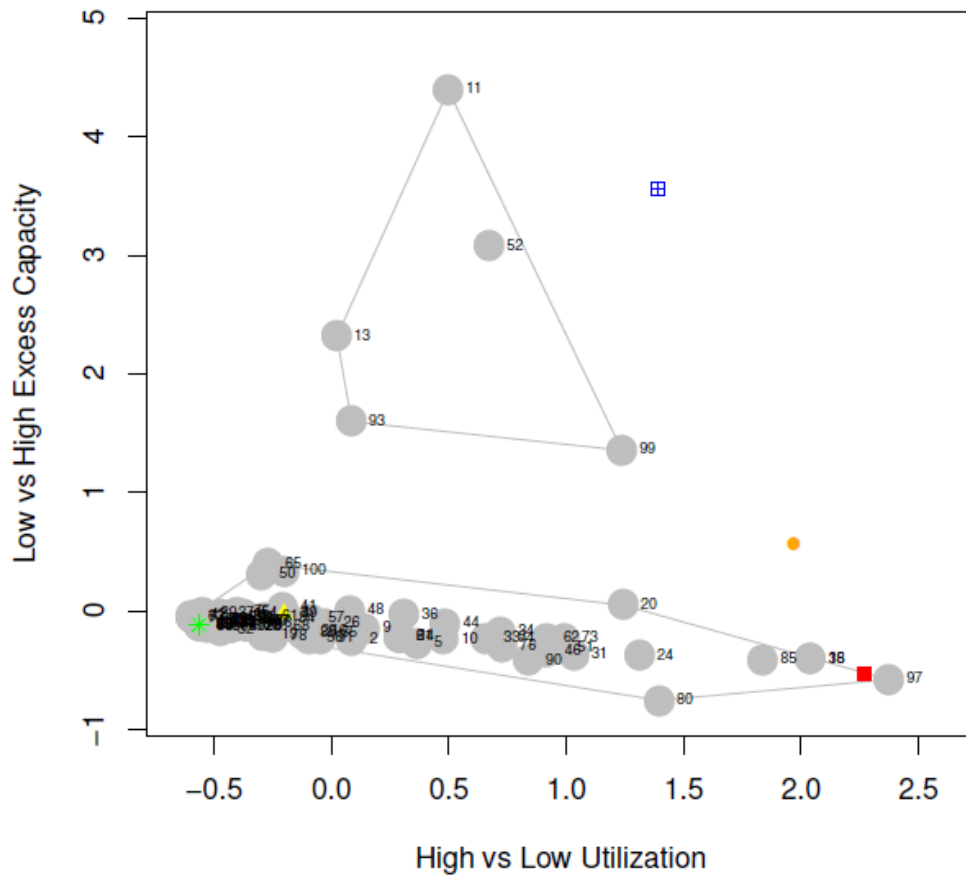
Figure 2: Capability importance and capacity visualizations – excess capacity

- **Low Capacity Capabilities:** Capabilities in the upper region of Figure 3 and 4 are more likely to be assessed as `No Capacity' or `Ad Hoc'. These capabilities may be require investment to ensure the ability to achieve success in the planning scenarios; and

- **Excess Capacity Capabilities:** Capabilities in the upper region of Figure 2 and the right region of Figure 4 are more likely to be assessed as `Excess'. These capabilities may be candidates for divestment.

Suppose a decision--maker is interested in capabilities that may require investment, such as Capability 80. Figure 3 shows that this capability has a low capacity and Figure 1 shows that it is an enabling capability with relatively high importance (between the 'Essential'(▲) marker and the 'Critical' (*). Therefore, this capability may be a good candidate for investment − an

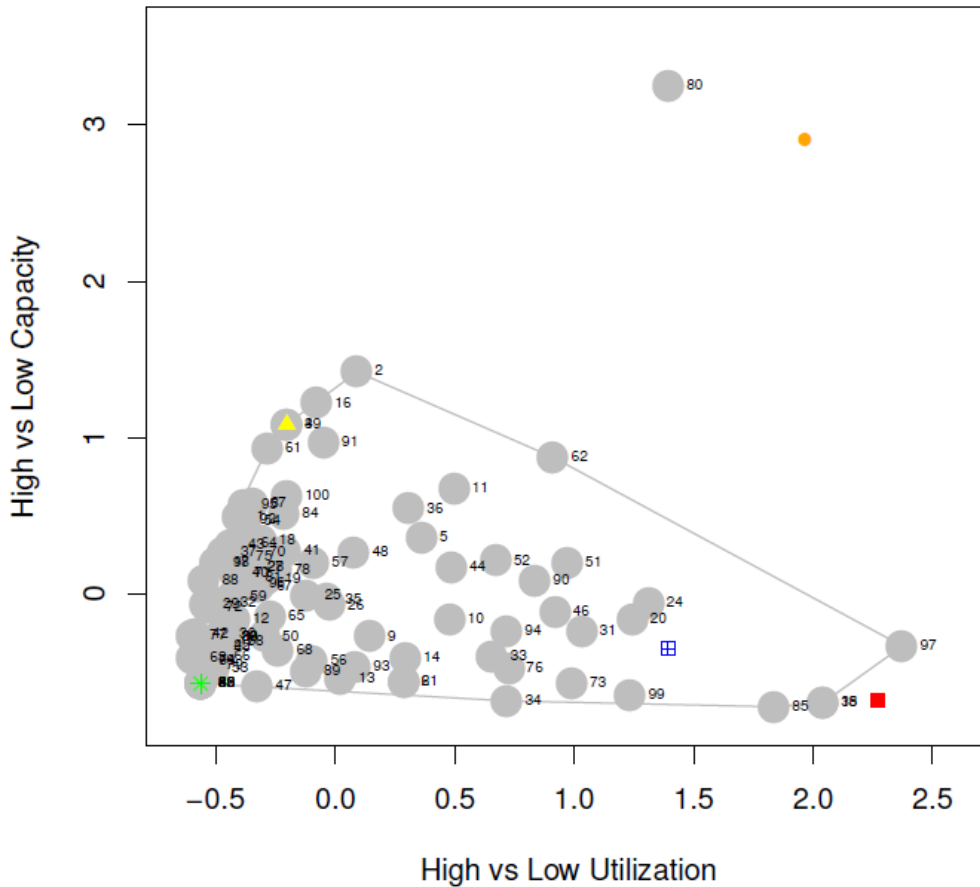enabler with relatively high importance and low capacity.   Likewise, suppose a decision−maker is

Figure 3: Capability importance and capacity visualizations – low capacity.

interested in capabilities that are candidates for divestment, such as Capability 99. Figure 2 shows that this capability has relatively high excess capacity and Figure 1 shows that it has a low importance. This capability may be a good candidate for divestment − relatively low importance and excess capacity.

When determining if a capability requires investment or is a candidate for divestment, the following must be considered:

- the capability assessments are only valid for the scenarios in the data set;
- the capability assessments are only valid for the planned force structure and given time horizon; and
- the concept of scenario concurrency is not captured.

These are a result of the data collected. Most importantly, these limit the conclusions that can be drawn about investments and divestments. If a capability is shown to have excess capacity,
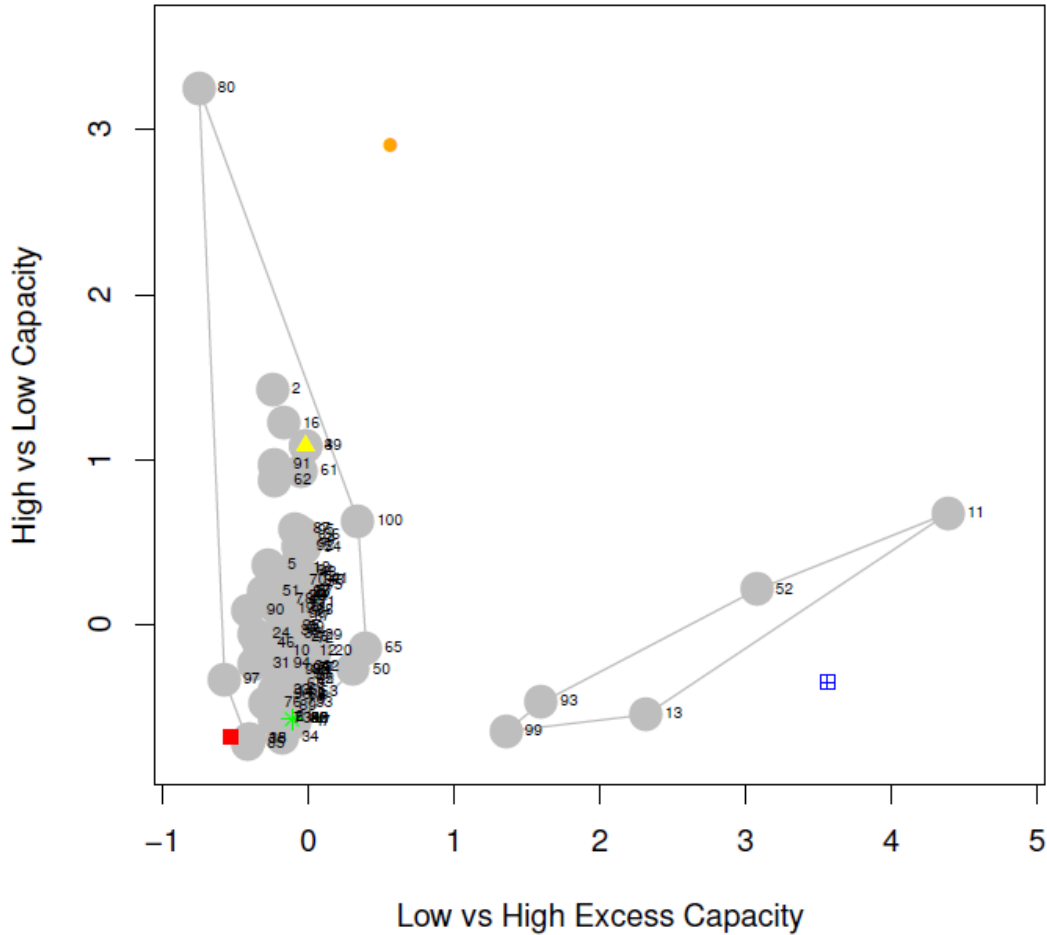

Figure 4: Capability importance and capacity visualizations – low vs excess capacity.

this means that it has excess capacity in individual scenarios. It can not be concluded that the capability has excess capacity if multiple scenarios occur concurrently. This must be verified in a follow−on analysis.

Even with these considerations, there are benefits to using this approach versus the ranking approach: (1) it gives decision−makers a simple visualization of the capabilities' importance and the planned force structure's capacity; (2) it immediately highlights those capabilities that may require investment and those that may be candidates for divestment; and (3) it provides a view of those capabilities that are enabling and those that are special purpose. Thus, this approach provides decision−makers with insights that supports their investment and divestment decisions.

5.   Conclusion

In this paper we described how the CAF use dimensionality reduction and partition clustering in their CBP process to create visualizations that show the importance of capabilities and the capacity of a planned force structure.   The resulting visualizations allow decision−makers to quickly determine those capabilities that may require investment, those capabilities that may be candidates for divestment, and separates capabilities into those that are enablers and those that are special purpose. Thus, these visualizations and the insights gained directly support decision−makers in their investment and divestment decisions.

References

[1] Bezdek, J. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms.* Kluwer Academic Publishers, Norwell, MA, USA.

[2] Chief of Force Development (2009). *The Future Security Environment 2008-2030 Part 1 Current and Emerging Trends.* Canadian Armed Forces, Ottawa.

[3] Christopher, G., Blakeney, D., Petryk, R., Taylor, B., Kerzner, L., Fong, V., Ball, M., and Beard, A. (2009). Strategic capability roadmap version 1.0 analytic framework. Technical Report DRDC CORA TR 2009–013, Defence Research and Development Canada.

[4] Commonwealth of Australia (2013). *Defence White Paper 2013.*

[5] Davis, P. (2002). Analytic arhitecture for capabilities–based planning, mission–system analysis, and transformation. Technical report, RAND.

[6] Dunn, J. (1974). Well separated clusters and optimally fuzzy partitions. *Journal of Cybernetics*, **4**, 95-104.

[7] Economist (2010). A special report on managing information. *The Economist*, **394**(**8671**), 3-18.

[8] Emond, E. and Mason, D. (2002). A new rank correlation coefficient with application to the consensus ranking problem. *Journal of Multi-Criteria Decision Analysis*, **11**, 17-28.

[9] Government of Canada (2008). Canada First Defence Strategy.

[10] Halkidi, M., Batistakis, Y., and Vazirgiannis, M. (2001). On cluster validation techniques. *Journal of Intelligent Information Systems*, **17**(**2/3**), 107-145.

[11] Handl, J., Knowles, J., and Kell, D. B. (2005). Computation cluster validation in post– genomic data analysis. *Bioinformatics*, **21**(**15**), 3201-3212.

[12] Hastie, T., Tibshirani, R., and Friedman, J. (2003). *The Elements of Statistical Learning.* Springer Series in Statistics. Springer, New York, Second edition.

[13] Hristov, N., Radulov, I., Iliev, P., and Andreeva, P. (2010). Prioritization methodology for development of required operational capabilities. *In NATO RTO SAS-081 Analytical Support to Defence Transformation*. NATO Research and Technology Organization.

[14] Jain, A., Murty, M., and Flynn, P. (1999). Data clustering: A review. *ACM Comput. Surv*., **31**(**3**), 264-323.

[15] Joint Systems and Analysis Group Technical Panel 3 (2005). Guide to Capability–Based Planning. Technical report, The Technical Cooperation Program.

[16] Jolliffe, I. (2002). Principal Component Analysis. Springer Series in Statistics. Springer, New York, Second edition.

[17] NATO Research and Technology Organization (2003). Handbook on long term defence planning. Technical Report RTO TR 069, NATO Research and Technology Organization.

[18] Nenadic, O. and Greenacre, M. (2005). Computation of multiple correspondence analysis, with code in R. Working Paper Series 887, Universitat Pompeu Fabra.

[19] New Zealand Government (2010). *Defence White Paper* 2010. Ministry of Defence.

[20] Pelletier, E. (2013). Capability ranking and clustering for multiple scenario analyses with weights: A comparative study. Technical Report DRDC CORA TM 2013–031, Defence Research and Development Canada.

[21] Rempel, M. (2010). Advancing the capability based planning process for strategic decisionmaking within the Canadian Forces. In 41st *Decision Sciences Institute Annual Meeting*, pages 1441-1452.

[22] Rempel, M. (2012). Visualizing capability requirements in planning scenarios using principal component analysis. In *Proc. IEEE Symp. Computational Intelligence for Security and Defence Applications (CISDA)*, pages 1-8.

[23] Rousseeuw, P. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, **20**(**0**), 53-65.

[24] Saaty, T. L. (2001). Decision Making for Leaders: *The Analytic Hierarchy Process for Decisions in a Complex World: 1999/2000 Edition*, volume 2. RWS publications.

[25] Salkind, N. J., editor (2007a). *Encyclopedia of Measurement and Statistics*, volume 1. SAGE Publications.

[26] Salkind, N. J., editor (2007b). *Encyclopedia of Measurement and Statistics*, volume 2. SAGE Publications.

[27] Skillicorn, D. (2007). *Understanding Complex Datasets: Data Mining with Matrix Decom-positions*. Data Mining and Knowledge Discovery Series. Chapman & Hall / CRC, Boca Raton.

[28] Theodoridis, S. and Koutroumbas, K. (2006). *Pattern Recognition*. Academic Press, Inc., Orlando, FL, USA, Third edition.

Mark Rempel
Centre for Operational Research and Analysis
Defence Research and Development Canada
Ottawa, Ontario K1A 0K2, CANADA
mark.rempel@drdc-rddc.gc.ca