Estimation of frequency and duration of future Canadian Armed Forces Operations

Dr. Peter Dobias
DRDC CORA, 60 Moodie Drive
Ottawa, ON, K1A0K2, Canada
peter.dobias@drdc-rddc.gc.ca

Abstract

Government of Canada policy mandates that the Canadian Armed Forces (CAF) be able to conduct multiple missions (in this context viewed as a broader category of operations, such as "provide humanitarian assistance") concurrently. Force Mix Structure Design (FMSD) is an analytical initiative that looks at the capacity of a particular force structure to deliver on this demand. In order to model the demand, it is first necessary to establish the likelihood of occurrence and duration of a particular mission type. It was proposed to use the record of past operations to inform the estimates of the likelihood and duration distributions. This data is held within a database, called 5W, which contains information about approximately 300 named CAF operations between 1990 and 2015. Since the decisions to commence or terminate a military mission are driven by political decisions of the day, the relevance of historic data to predict future demand under different security and political environment is not clear-cut, especially if used in isolation. However, if used in conjunction with professional military and strategic judgement, and within the constraint of the defence policy and other foresight documents, it can still provide valuable insights either as an initial input, or as a challenge to any purely judgement-based estimates. For the purposes of the analysis, the 5W database records were mapped against a set of scenarios developed specifically for the FMSD. The frequency and duration distributions of the operations categorized by scenario, size, and environment were then derived using data fitting through maximum likelihood method; in order to validate the estimates the duration distribution likelihood was then estimated using bootstrapping. The implications, as well as their relevance of the results for the current security environment, are discussed.

Key words: frequency estimate, future operations, bootstrapping, maximum likelihood

Introduction

Government of Canada defence policy *Strong, Secure, Engaged* (SSE) [1] demands that in order to meet its domestic and international obligations, and to contribute to global peace and prosperity, the Canadian Armed Forces (CAF) should be able to conduct multiple operations concurrently. Force Mix Structure Design (FMSD) is an analytical initiative that looks at the capacity of a particular force structure to deliver on this demand. In its very nature, FMSD is a supply-demand analysis, where demand takes the form of expected mission types and corresponding force elements (e.g., squadrons of fighters, or infantry platoons) required as a part of the military response, and supply are the force elements available for employment. There are many intangibles entering the future demand for military forces: changes in the geopolitical and security environment, technological shifts, political and economic pressures, etc. This makes forecasting future requirements difficult if not impossible. The standard approach in the military planning is to use strategic estimates and other foresight documents to develop a set of planning scenarios (i.e., fictitious operational situations that provide background information against which military planners can devise specific plans of employing desired capability sets); these scenarios can then be used to estimate future force demand. However, while the scenarios in isolation may provide insights about which capabilities will be required, additional information is required to estimate how much of each capability will be needed. The latter depends not only on the capability demand for a particular scenario, but also on the concurrent demand from other scenario (and can also change with the duration of a mission). To estimate the overall demand, it is necessary to be able to determine the likelihood of occurrence of each scenario, as well as their probable durations. Historical CAF operations were used to inform these frequency-duration estimates. Given that the analysis of the historical data in the present form did not consider the political and strategic considerations that led to the decisions to undertake some missions and to forego others, the results were only intended to provide input, or a challenge function, to the professional military judgement further informed by the strategic foresight documents (e.g., [2]).

Estimation of frequency of discrete events is common across many different research areas, such as the insurance industry (e.g., estimate of the likelihood of accidents and other claims), hydrology (forecasting severity and frequency of floods), or seismology (risk estimates for earthquakes) [3]. Two commonly used frequency estimation methods based on historical data are maximum likelihood [4,5] and Bayesian methods [6]. For the parametric estimates, the events are usually assumed to originate from the Poisson distribution with rate Λ that can in general vary across the analysed population. According to reference [4] the estimates were traditionally done using negative binomial model (NBM) or condensed negative binomial model; alternatively using the generalization of NBM proposed by [7]. Further generalization, considering time variable Λ is possible using Bayesian approach [8].

In case of floods and earthquakes, the historic analysis typically serves to estimate distribution of both frequencies and severity of the events. These estimates are important for example for design of flood- or earthquake resistant infrastructure or for insurance purposes [9,10]. One of the challenges of forecasting extreme events is that the underlying distributions are frequently un-truncated (with infinite variance) [11]. In many applications (earthquakes, stock markets) the distributions are power law rather than Poisson [12]. For such processes Keim [11] proposed a method analyzing inter-arrival times between events; this method works if the length of individual events is short relative to these intervals. For the purposes of this analysis the duration of a particular mission could be possibly interpreted as severity; consequently, the methods based on the inter-arrival times would not work. The only exceptions would be short-duration humanitarian aid and search and rescue missions. Hence, for this analysis, maximum likelihood methods [4] were used.

A non-parametric alternative, applicable even if the assumptions of the parametric methods are not satisfied, is bootstrap [13]. The bootstrap methods assume that the sample cumulative distribution function (CDF) is equal to the population CDF; this assumption is violated for heavy-tailed distributions and as a consequence, the sample mean does not converge to the population mean [14]. The bootstrap methods were later expanded to include Bayesian bootstrap [15].

The application of the frequency-duration analysis to estimate future likelihood of CAF Operations from historical data presents a particular challenge because of the very limited sample size. It was done previously for maritime operations [16,17]. The frequency confidence interval estimate employed maximum likelihood approach using Johnson-Neyman method [18]; a similar approach is employed in this paper. For the duration estimates Greenwood [17] employed a triangular distribution fitting using, again, maximum likelihood method. Johnson-Neyman maximum likelihood method [18] was used in this paper as well, just with different distributions (exponential and/or uniform). In order to validate the maximum likelihood approach for this analysis, a bootstrap method was used to assess the frequency and duration likelihoods in addition to the parametric estimates.

It can be argued that many military planning scenarios represent extreme and in some cases rare events. While some of the mission types (e.g., the humanitarian assistance) will be more common, others, such as major war fighting are quite rare. For instance, Canada has not been involved in a major peer conflict since the Korean War. However, any prudent military planning must include these extreme events, particularly because the required capabilities cannot be acquired at a moment's notice. This presents a challenge, because the historic data for the more demanding mission types will likely be sparse [19].

Frequency and Duration of CAF Operations

For the analysis purposes, a historic database of CAF operations dating back to 1990 was developed in 2004 by Mason [20], with the stated aim to catalogue the "Who-What-Where-

When-Why of Canadian Forces Deployments Since 1990". The database was updated in 2012 [21], and again in 2015 [22]; during the last upstate it was converted from a spreadsheet to a relational database. The database captures names, dates, usually some information about types of forces employed, and any other pertinent information that was publicly available. Only named operations reported to public were included in the original database. Low-profile, routine operations were excluded, as were major exercises unless they happened in the context of a named operation [20].

The use of historical data for the estimate of the future brings several challenges:

1. Availability of historical data: there is no database of record documenting historical CAF operations with sufficient richness of captured information.
2. Quality of historical data: historical data is somewhat incomplete, with some of the missions broken down across multiple operations, and often lacking supplementary information about personnel and employed capabilities.
3. Subjectivity of the mapping of historical operations to planning scenarios.

As a consequence, the results should be interpreted with caution. Ideally, they should function either as input to professional judgement, or as a challenge to subjective estimates of the likelihood of occurrence and duration of particular types of missions. Furthermore, as was mentioned above, the decision to take on (or not) a particular mission is driven by a variety of geopolitical and strategic considerations that are normally not possible to capture in a historical analysis. In particular, since the comprehensive data on the potential operations and conflicts in which Canada chose not to participate (such the invasion of Iraq in 2003) is not available, it was not possible to consider the patterns in omitted missions. [23]

To estimate the frequency and duration likelihoods for the proposed planning scenarios, each operation in 5W was mapped to one of the scenarios developed specifically for the FMSD analysis (in order to avoid potential sensitivities, an adjusted list of scenarios aggregating some of the related scenarios was used in this paper; in addition, the scenarios are simply referred to by their numbers rather than titles). Since the mapping of the historical operations to the FMSD scenarios was by necessity subjective, best available subject matter expertise was employed. Each operation was mapped to one scenario. In addition, it was deemed useful to consider which was the primary element (army, navy, air force) employed during the operation; although this information was not used in the presented analysis. Furthermore, the Canadian defence policy categorized mission size based on the number of personnel. This categorization was also added to 5W where the information was available; major operations would include more than 500 personnel, while minor operations would have less than 500 personnel [1].

The structural patterns in the historical operations were beyond the scope of this analysis. Therefore, it was assumed that the occurrence of a particular mission type was independent of the preceding and concurrent missions. In addition, seasonality and geographical patterns were not considered either. While these aspects of the historical trends could be of interest, they could not be accommodated in the existing demand model and therefore were not considered.

The initial work used very simplistic frequency estimate in order to accommodate tight client deadline. The mean frequency $< F_i >$ of an *i*-th scenario was estimated as

$$< F_i > = N_i \,/\, T$$

where N_i is the number of operations corresponding to scenario *i*, and *T* is the length of the period covered by the database (25 years or 300 months). Figure 1 shows the numbers of major and minor operations corresponding to 20 selected scenarios. However, due to their limited count, for the frequency analysis major and minor operations were considered together.

In order to estimate a confidence interval, it was then assumed that the operations would follow a Poisson distribution. Since the number of observations is generally low, Johnson-Neyman method using inverse χ^2 function [18] provides best estimate for the lower (L) and upper (U) bounds for the frequency estimate in events per year:

$$L = \frac{1}{2y}(\chi^2)^{-1}\left(\frac{1+c}{2}, n\right) \text{ and } U = \frac{1}{2y}(\chi^2)^{-1}\left(\frac{1-c}{2}, n\right)$$

where $(\chi^2)^{-1}$ denotes inverse χ^2 function, y denotes the length of the considered interval (25 years), *c* is the desired confidence level, and *n* is the number of occurrences of the scenario. For this analysis, confidence interval of 95% was used.

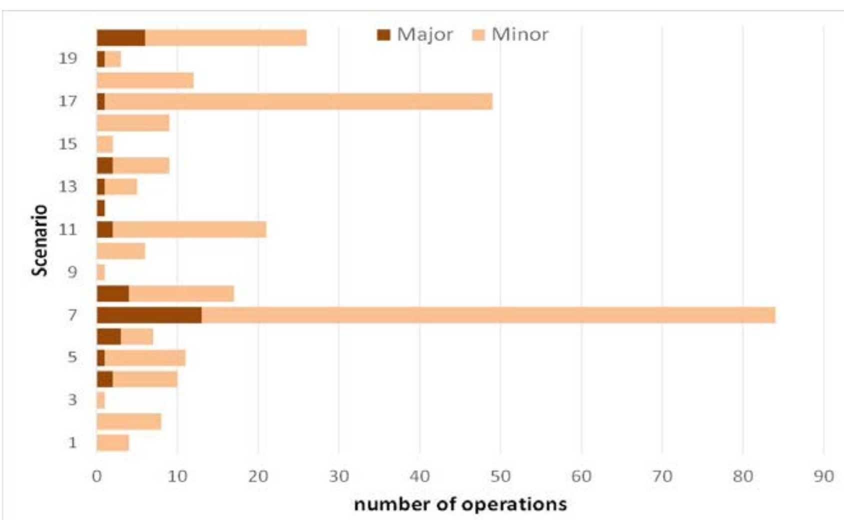

Figure 1: Number of operations assigned to each scenario.

Figure 2 shows the estimated frequencies of various scenarios including upper and lower estimates at 95% confidence level. For example, Scenario 17 occurs with a mean frequency of 1.96 (i.e., a new operation commencing approximately twice a year). With

95% confidence, the frequency is between 1.26 and 2.81; in other words, a new operation starts between every 4 and 9 months.

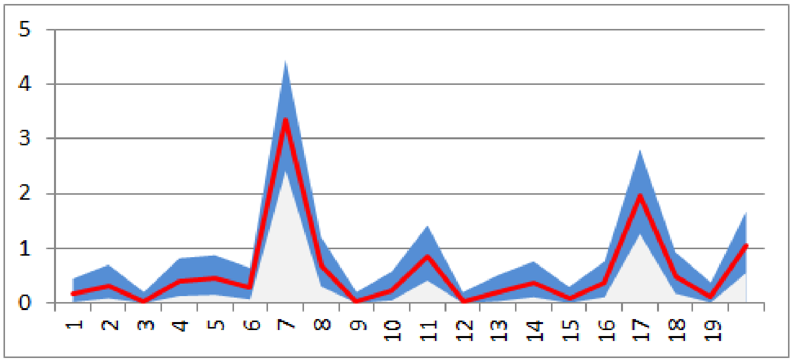

Figure 2: Annual frequency estimate for planning scenarios.

Two types of distributions were prevalent in the duration distributions. For the majority of the scenarios most of the operations were short term; for these the exponential distribution

$$p(x) = \Lambda \; exp\,(-\Lambda(x - X_{min})),$$

with X_{min} as the lower bound of the distribution, and Λ is the scaling parameter, was used. The lower bound was set to be zero (there can be no operations with negative duration). Two of the scenarios exhibited somewhat bimodal behaviour; uniform distribution

$$p(x) = \frac{x - X_{min}}{(X_{max} - X_{min})}$$

fitted these few scenarios best. In addition, four of the scenarios did not have a sufficient number of corresponding operations to fit them to distributions; these were eliminated from the analysis.

Table 1: Duration distribution for the modeled scenarios

Scenario	Distribution function	Median	Min	Max	95% confidence	
					Lower estimate	Upper estimate
1	Exp	35	6	93	1	188
2	Unif	161	0	321	8	313
3	none	42				
4	Exp	2	0	8	0	11
5	Exp	28	0	348	1	147
6	Exp	49	27	184	2	263
7	Exp	2	0	49	0	11
8	Exp	5	0	120	0	28
9	none	6				
10	Exp	53	0	307	2	284
11	Exp	34	0	207	1	182
12	none	36				
13	Exp	6	0	37	0	30
14	Exp	106	0	744	4	563
15	none	36				
16	Exp	7	0	79	0	36
17	Exp	49	1	778	1	142
18	Exp	31	2	161	1	164
19	Unif	19	1	37	1	36
20	Exp	14	1	107	1	73

Maximum likelihood estimate [24], as implemented in Python's *Numpy*[1] distribution was used to fit the distributions to the data. The results are in Table 1.

The summary duration across all vignettes, broken down by size only, matched an exponential distribution. In general, 50% of major operations lasted 16 or fewer months, while minor operations tend to be longer, with 50% under 30 months. With high confidence (97.5%), major operations last below 85 months, while minor could be as long as 161 months.

[1] www.numpy.org (accessed 7 July 2019)

In order to validate the parametric estimate of the durations, it was decided to use a non-parametric method, i.e., bootstrap. This was done particularly since a) the data sets were quite limited, and b) the fitted functions were selected because they looked reasonable, but there was no *a priori* reason that the data would correspond to a particular distribution function.

The implementation was in Python, using the *Numpy.Random*[2] library. Twenty thousand samples with replacement were generated for each of the scenarios; for each of these the mean scenario duration was obtained. Since some of the scenarios had a very small number of corresponding operations, the bootstrap was done only for those with 4 or more operations; for the ones with fewer operations the estimated duration was simply an arithmetic mean of the historical durations. The results are in Figure 3. The resulting estimates were shorter; for some scenarios by as much as a factor of two, than the parametric estimates. This suggests that the parametric estimates using infinite distribution function might not have sufficiently reflected the data.

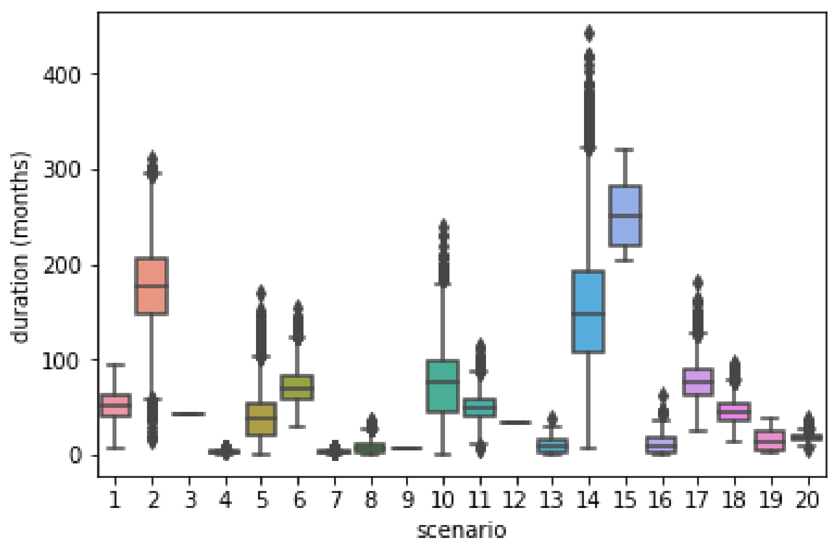

Figure 3: Bootstrap results for mean scenario duration.

Then the estimates for minor and major operations across all scenarios were generated. For the major operations, it was estimated that the median of the means obtained for 20000 resamplings was 22.5 months, with the 95% confidence interval being between 11 and 37 months. For minor operations the median of the means was 43.5 months, with the 95% confidence interval being between 32 and 57 months. The results are in Figure 4. Thus the estimated duration of major operations was slightly longer, while that of minor operations much shorter than was obtained from parametric estimates.

[2] https://docs.scipy.org/doc/numpy/reference/routines.random.html (accessed 7 July 2019)

Since the parametric estimates are somewhat less robust, and more dependent on the variety of assumptions about the data, it is likely that the bootstrap method provided more robust estimates of the durations.

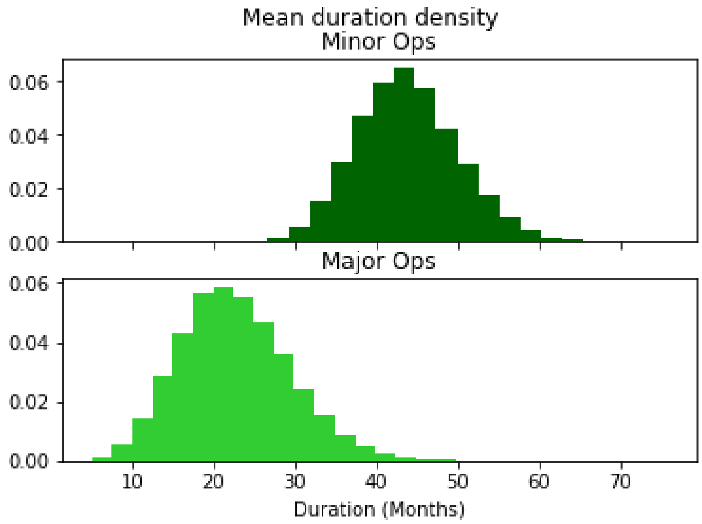

Figure 4: Duration of major and minor operations across all scenarios.

Conclusions

The presented analysis shows the estimate of the frequency and durations for a set of scenarios developed for the force structure capacity analysis. A maximum likelihood approach was used to estimate the frequencies. For the duration estimates, first maximum likelihood was used; however, it was determined that because of the irregularities in the distribution functions, this would be validated using bootstrap methods.

Generally, the estimates obtained from bootstrap were somewhat shorter than those obtained through parametric methods. This result suggests that the actual shape of the distributions might be somewhat different from the estimated infinite (exponential) distributions. If the parametric estimates are to be used in the future, a modified function might be necessary. Due to potential complexities, this makes non-parametric methods more attractive. However, since the parametric distributions are generally easier to implement in the demand analysis, a hybrid approach, possibly including parametrization of bootstrap-obtained distributions, might be desirable.

The obtained probability distributions from the historical analysis then informed the subject matter experts in deciding final frequency and duration parameters for the demand analysis. This last steps considered both current policy and foresight documents that were ignored by the historical analysis itself. The resulting scenario frequencies and durations (and corresponding future mission demand) were then assessed by senior military experts. Because the FMSD process only looks at long-term force structure implications, the in-year variations would be of limited value. Consequently, no additional analysis of patterns and trends in the data, such as potential seasonality of certain types of operations (e.g., search and rescue incidents [25]) was performed. However, further in-depth analysis of the data is currently planned. This may include considering the sequencing/concurrence limitations and changes in government policies. Possibly it may lead to the development of forecast models that could be used to explore implications of possible different futures on the demand.

References:

1. Strong, Secure, Engaged: Canada's Defence Policy, Department of National Defence, D2-386/2017E, June 2017.
2. Chief of Force Development, The Future Strategic Environment 2013-2040, 2014.
3. D.G. Morrison, D.C. Schmittlein, Predicting future random events based on past performance, Management Science, Vol. 27, No 9, 1006-1023, 1981.
4. R.A. Fisher, Theory of statistical estimation, Proc. Cambridge Philosophical Soc. Vol. 22, 700-725, 1925
5. J. Aldrich, R.A. Fisher and the making of maximum likelihood 1912 - 1922, Statistical Science, Vol 12, No 3, 162-176, 1987
6. P.M. Lee, "Bayesian Statistics: An Introduction," 4th edition. Wiley. ISBN 978-1-118-33257-3, 2012.
7. H. Robbins, Prediction and estimation for the compound Poisson distribution, Proc. National Acad. Sci., USA, Vol. 74, 2670-2671, 1977
8. D. Rosenfield, A model for predicting frequencies of random events, Management Science, Vol. 33, No. 8, 947-954, 1987
9. K. Engeland, D. Wilson, P. Borsanyi, L. Roald, E. Holmqvist, Use of historic data in flood frequency analysis: a case study for four catchments in Norway, Hydrology Research, Vol. 49, No. 2, 466-486, 2018
10. S.B. Mathewson, S Earthquake frequency estimation - ratemaking for unusual events, https://www.casact.org/pubs/forum/99wforum/wf99133.pdf (Accessed 2 Apr. 2019)
11. B.D. Keim, J.F. Cruise, A technique to measure trends in the frequency of random discrete events, Journal of Climate, Vol 11, 848-855, 1998
12. D.D. Sornette, Multiplicative processes and power laws, Phys. Rev E, Vol 57, 4811-4813, 1998
13. B. Efron, Bootstrap methods: Another look at the jackknife, The Annals of Statistics, Vol 7, No 1, 1-26, 1979
14. K.B. Athreya, Bootstrap of the mean in the infinite variance case, The annals of Statistics, Vol 15, No. 2, 724-731, 1987
15. D.B. Rubin, The Bayesian bootstrap, The Annals of Statistics, Vol. 9, No 1, 130-134, 1981

16. Eisler, S. Force Structure Evaluation for Capability-Based Planning: An Illustrative Methodology, Defence Research and Development Canada, Scientific Report, DRDC-RDDC-2015-R026, March 2015.
17. Greenwood, N., JY Forcier, H. Fitzpatrick, Maritime Vignette Analysis – Rationalization and Response Development, DRDC-RDDC-2018-C153, 2018
18. P.O. Johnson, J. Neyman, Tests of certain linear hypotheses and their applications to some educational problems. Statistical Research Memoirs, 1:57–93, 1936.
19. N.N. Taleb, "The Black Swan: The Impact of the Highly Improbable", Random House, ISBN 978-1400063512, 2007.
20. D.W. Mason, The Canadian 'Five-W' Database ('Canadian Forces Including Vehicles and Equipment Worldwide'): The Who-What-Where-When-Why of Canadian Forces Deployments since, 1990, DOR(MLA) Research Note 2004/10, 2004.
21. K. Grant, International Canadian Forces Operations from 1990-2011, Supplement to the FIVE-W Database, DRDC CORA CR 2012-270, 2012.
22. M. Stevens, Canadian Armed Forces Operations from 1990–2015: Update 2015 to the FIVE-W Database, Defence Research and Development Canada ,Contract Report, DRDC-RDDC-2015-C350, Valcom Consulting Group Inc., November 2015.
23. P. Dobias, Analysis of likely frequency and duration for force mix structure design based on historical Canadian Armed Forces operations, DRDC-RDDC-2019-L053, 2019
24. H. Cramér, Mathematical methods of statistics, Princeton Univ. Press, 1946
25. P. Dobias, Review of search and rescue (SAR) demand: Baseline trends for 2008-2013 in the Pacific SAR region, DRDC-RDDC-2014-L162, 2014