



Calhoun: The NPS Institutional Archive
DSpace Repository

NPS Scholarship

Publications

2018-02

Optimal location of Navy recruiters

Salmeron, Javier; Buttrey, Samuel E.

Monterey, California. Naval Postgraduate School

<https://hdl.handle.net/10945/57354>

This publication is a work of the U.S. Government as defined in Title 17, United States Code, Section 101. Copyright protection is not available for this work in the United States.

Downloaded from NPS Archive: Calhoun



Calhoun is the Naval Postgraduate School's public access digital repository for research materials and institutional publications created by the NPS community. Calhoun is named for Professor of Mathematics Guy K. Calhoun, NPS's first appointed -- and published -- scholarly author.

Dudley Knox Library / Naval Postgraduate School
411 Dyer Road / 1 University Circle
Monterey, California USA 93943

<http://www.nps.edu/library>

NPS-OR-18-003



NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

OPTIMAL LOCATION OF NAVY RECRUITERS

by

Associate Prof. Javier Salmeron and Associate Prof. Samuel E. Buttrey

February 2018

Prepared for: Navy Recruiting Command

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			<i>Form Approved</i> OMB No. 0704-0188		
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) 31-12-2017		2. REPORT TYPE Technical Report –Project Deliverable		3. DATES COVERED (From-To) 1-Oct-2016 - 31-Dec-2017	
4. TITLE AND SUBTITLE OPTIMAL LOCATION OF NAVY RECRUITERS				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Associate Professor Javier Salmeron and Associate Professor Samuel Buttrey				5d. PROJECT NUMBER SERIAL NRS: W3201-W3300	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) AND ADDRESS(ES) Operations Research Dept., Naval Postgraduate School 1411 Cunningham Rd. Monterey, CA, 934943				8. PERFORMING ORGANIZATION REPORT NUMBER NPS-OR-18-003	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Naval Research Program Research & Sponsored Programs Office 699 Dyer Road, Bldg 234 Naval Postgraduate School Monterey, CA, 93943				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited					
13. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S. Government.					
14. ABSTRACT This research has developed and computationally implemented the “Navy Recruiter Prediction and Optimization Model” (NRPOM). NRPOM can assist Navy Recruiting Command (NRC) with the assignment of recruiters to geographical areas across the U.S. Under given assumptions, NRPOM optimizes: (a) the allocation of a limited number of recruiters to candidate recruiting stations in a region; (b) the assignment of Zip codes to recruiting stations; and (c) the (fraction of) time recruiters should spend at each Zip code. The research has also developed a predictive tool that produces input data for the optimization. Experiments conducted on realistically-sized problems demonstrate that these tools can be used to guide NRC’s decisions. However, NRPOM has only been tested with notional data from the state of California, and for this case some of the required inputs have not been provided by NRC; instead, the authors have used estimations that have no guarantee of reflecting actual data. Thus, we believe that NRPOM is a starting point by which to approximate a truly optimal solution to the problem; however its development is not finalized yet.					
15. SUBJECT TERMS recruiting; forecasting; optimization; facility location; assignment.					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 35	19a. NAME OF RESPONSIBLE PERSON Javier Salmeron
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			
			19b. TELEPHONE NUMBER (include area code) (831) 656 2779		

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

THIS PAGE INTENTIONALLY LEFT BLANK

**NAVAL POSTGRADUATE SCHOOL
Monterey, California 93943-5000**

Ronald A. Route
President

Steven R. Lerman
Provost

The report entitled "OPTIMAL LOCATION OF NAVY RECRUITERS" was prepared for Navy Recruiting Command and funded by Naval Research Program.

Further distribution of all or part of this report is authorized.

This report was prepared by:

Javier Salmeron
Associate Professor

Samuel Buttrey
Associate Professor

Reviewed by:

Johannes Royset
Associate Chairman for Research
Department of Operations Research

Released by:

Patricia Jacobs, Chair
Department of Operations Research

Jeffrey Paduan
Dean of Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

This research has developed and computationally implemented the “Navy Recruiter Prediction and Optimization Model” (NRPOM). NRPOM can assist Navy Recruiting Command (NRC) with the assignment of recruiters to geographical areas across the U.S. Under given assumptions, NRPOM optimizes: (a) the allocation of a limited number of recruiters to candidate recruiting stations in a region; (b) the assignment of Zip codes to recruiting stations; and (c) the (fraction of) time recruiters should spend at each Zip code. The research has also developed a predictive tool that produces input data for the optimization. Experiments conducted on realistically-sized problems demonstrate that these tools can be used to guide NRC’s decisions. However, NRPOM has only been tested with notional data from the state of California, and for this case some of the required inputs have not been provided by NRC; instead, the authors have used estimations that have no guarantee of reflecting actual data. Thus, we believe that NRPOM is a starting point by which to approximate a truly optimal solution to the problem; however its development is not finalized yet.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

I.	INTRODUCTION.....	11
A.	RESEARCH CONTEXT	11
B.	SCOPE AND BENEFITS.....	11
II.	THE DATA.....	13
A.	RECRUITING’S CHICKEN-AND-EGG PROBLEM	13
1.	The Fundamental “Problem” of Predicting Recruiting	13
B.	ELIGIBILITY AND PROPENSITY.....	13
1.	Definitions.....	13
C.	COUNTY-LEVEL AND ZIP CODE-LEVEL DATA	14
1.	Data and Sources.....	14
2.	ZCTA and County-level Data	15
D.	THE PROBLEM OF DIMINISHING MARGINAL RETURNS	15
1.	Defining Marginal Returns	15
2.	A Very Simple Model for the Number of Recruits per Recruiter	17
III.	THE OPTIMIZATION MODEL	21
A.	APPROXIMATION OF RECRUITING EFFORT IN OPTIMIZATION ...	21
B.	MATHEMATICAL MODEL	22
C.	INPUT FILES FOR OPTIMIZATION	25
1.	Miscellaneous Input File.....	25
2.	Station Input Files.....	26
3.	Zip Input Files	27
D.	EXAMPLE OF OPTIMIZATION RESULTS.....	29
1.	California Scenario Setup	29
2.	Scenario Results	31

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

Figure 1: Examples of Diminishing Returns Curves for $A=40$ and Different Values of B	16
Figure 2: The Gamma Distribution with Shape = 8 and Scale = $\frac{1}{4}$	19
Figure 3: Examples of Two Piece-wise Linear Approximations of a Recruiting Effort Curve.....	21
Figure 4: Example of Miscellaneous (<i>Misc.csv</i>) input file	26
Figure 5: Example of Station and Station Date input files: <i>S.csv</i> (left) and <i>S_Data.csv</i> (right)	27
Figure 6: Example of Zip and Station-Zip input files: <i>Z.csv</i> (left) and <i>SZ_Dist.csv</i> (right)	28
Figure 7: Example of Zip Production input file, <i>Z_Production.csv</i>	28
Figure 8: Example of Zip Fit input file, <i>Z_Fit.csv</i>	29
Figure 9: Summary of Test Cases and Results	31
Figure 10: Station Assignment Excerpt from Scenario 500-80	32

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. RESEARCH CONTEXT

This document, along with a series of computer programs and tools described herein, is the final deliverable for the project “Optimizing Location of Navy Recruiters” developed by the authors of the document as part of the Naval Research Studies Program at the Naval Postgraduate School. The project has been sponsored by the Chief of Naval Operations and executed between October 2016 and December 2017, with the U.S. Navy Recruiting Command (NRC) as the final customer.

The NRC’s mission is to “leverage an inspirational culture to inform, attract, influence and hire the highest quality candidates from America's diverse talent pool to allow America's Navy to assure mission success and establish the foundation for Sailors to thrive in a life-changing experience.”¹

This research has developed and computationally implemented “Navy Recruiter Prediction and Optimization Model” (NRPOM). NRPOM can assist Navy Recruiting Command (NRC) with the assignment of recruiters to geographical areas across the U.S. Under given assumptions, NRPOM optimizes: (a) the allocation of a limited number of recruiters to candidate recruiting stations in a region; (b) the assignment of Zip codes to recruiting stations; and (c) the (fraction of) time recruiters should spend at each Zip code. The research has also developed a predictive tool that produces input data for the optimization.

B. SCOPE AND BENEFITS

NRPOM has been developed in the Windows 7² operating system. The optimization requires (as additional software) the General Algebraic Modeling System (GAMS³) optimization environment with the GAMS/CPLEX⁴ solving engine. It also requires that the user can write and read comma-separated value (CSV) files for data input and

¹ United States Navy Recruiting Command (2017). Online: <http://www.cnrc.navy.mil/about.htm>

² Microsoft Corporation (2017). Online: <http://www.microsoft.com/en-us/default.aspx>

³ GAMS (2017). Online: www.gams.com

⁴ GAMS/CPLEX (2017). Online: https://www.gams.com/latest/docs/S_MAIN.html

resulting output, respectively. The predictor module requires no additional software although some data preparation is required.

Experiments conducted on realistically-sized problems demonstrate that the tool can be used to guide NRC's decisions. However, NRPOM has only been tested with notional data from the state of California, and for this case some of the required inputs have not been provided by NRC; instead, the authors have used estimations that have no guarantee to reflect the actual data. Thus, we believe NRPOM is a starting point to approximate a truly optimal solution to the problem, but its development is not finalized yet.

II. THE DATA

A. RECRUITING'S CHICKEN-AND-EGG PROBLEM

1. The Fundamental “Problem” of Predicting Recruiting

The fundamental problem we face in predicting recruiting is that recruiters almost always produce the number of recruits they are told to produce. This can come at great cost to their quality of living, and it might even in some cases lead to lower-quality recruits. Still, the costs of failure are very high. Conversely, the benefit to producing more recruits than assigned are quite small, so recruiters have an incentive to “game” the system – to delay the induction of some recruits until a later month, for example. So in areas where there are few available recruits, recruiters work harder, and in areas where there are plenty of recruits, recruiters can work less hard.

We put “problem” in quotation marks here because in some ways this is not really a problem. When the Navy is meeting its recruitment quotas, accounting for different types of recruits and timing, then the Navy has no recruiting problem at all. But the fact that recruiters almost always reach their goal leads to problems for analysts trying to make recruiting more efficient.

If all recruiters reach their goals, then there is no obvious way to determine the relative efficacy of recruiters. In such a world, all recruiters seem equally capable, and all regions seem equally productive (in that they produce exactly as many sailors as they are assigned). This has an effect on the Navy’s missioning strategy, too, since that strategy arises, at least in part, from the historical levels of recruiting achieved in each area.

B. ELIGIBILITY AND PROPENSITY

1. Definitions

To refine the idea of measuring the number of possible recruits in an area, let us use the terms **eligibility** and **propensity**. A recruit is “eligible” to join the Navy if he or she is not disqualified. Overall, about 70% of U.S. youth aged 17 to 24 are ineligible for one or

more reasons like health and fitness, test scores, certain tattoos, or criminal history⁵. Of course the proportion of eligible youth can be expected to vary from place to place. “Propensity” describes the interest a youth shows in joining the military. This is obviously a personal decision for each person, but in aggregate we expect youth in different areas to have different average propensities to join, based on factors like access to university education, local economic conditions, exposure to military through nearby military bases or concentrations of veterans living nearby, and so on. Only about 1% of young people are both “eligible and inclined to have a conversation with” the military about possible service

C. COUNTY-LEVEL AND ZIP CODE-LEVEL DATA

1. Data and Sources

In order to determine the best locations for recruiters we need to know where potential recruits can be found. Thus we need to know about the numbers of young people, and also, to the extent possible, about their eligibility and propensity rates. There have been a number of attempts in the literature to model the number of available recruits by Zip code. Among these the recent thesis of Fulton⁶ examines a number of publicly available data sets that can be brought to bear. These include community health status indicators from the Centers for Disease Control, Zip Code-level income tax data from the Internal Revenue Service, and locations of universities from the National Center for Education Statistics. The Census Bureau maintains a number of potentially valuable databases, including Economic Census and County Business Patterns data, as well as information on the number of veterans residing in each community. Other data that has been brought to bear in other research includes FBI crime data used, for example, in Intrater et. al ⁷. The data used in this project is taken from the thesis of Fulton.

⁵ Feeney, 2014: “Pentagon: 7 in 10 Youths Would Fail to Qualify for Military Service,” <http://time.com/2938158/youth-fail-to-qualify-military-service/>, retrieved 21 Feb 2018

⁶ Fulton, B.: “Determining market categorization of United States ZIP codes for purposes of Army recruiting,” Master’s Thesis, Naval Postgraduate School, Department of Operations Research, 2016

⁷Intrater, B., Alt, J., Buttrey, S., House, J., and Evans, M.: “Understanding the Impact of Socio-Economic Factors on Navy Accessions,” *Military Operations Research Journal*, to appear 2018.

2. ZCTA and County-level Data

Inevitably a number of issues arise. First, Zip codes are not quite sufficient; they describe mail-delivery routes, rather than polygonal areas, and change from time to time as the Postal Service's requirements change. It is common to use Zip Code Tabulation Areas; these generalized areas, maintained by the Census Bureau, are more stable, and essentially every US household sits in exactly one ZCTA. Second, a lot of data is available only at the county level. For these data sets we distribute county values to ZCTA codes in proportion to the population in each ZCTA (correcting for the ZCTAs that cross county lines where applicable). This has the effect of adding noise to the data and making neighboring areas look more similar than they should.

D. THE PROBLEM OF DIMINISHING MARGINAL RETURNS

1. Defining Marginal Returns

It is important that the model impose a diminishing rate of return as additional recruiters are added to an area. Otherwise, the optimal approach would be to assign all recruiters to the one area with the largest recruitable population – perhaps New York City. We assume that the number of recruits in any area is finite and exhaustible, and that as additional recruiters are added, the number of recruits they are able to secure will eventually decrease.

As an example, suppose that a recruiter's mission in a particular area is twelve recruits per year and she is able to reach that goal. A pair of recruiters in that area might be able to procure a total of 24 recruits – twelve each – per year, but if a third were to be added, the three recruiters collectively might be able to procure only, say, 30, not 36, recruits, since so few recruitable young people would remain.

There is little evidence as to the diminishing return associated with adding recruiters. Dertouzos and Garber⁸ build a simulation model for recruiting and, by modifying the numbers of recruiters in the simulation, conclude that "...the estimated elasticity of

⁸ Dertouzos and Garber, "Performance Evaluation and Army Recruiting," Rand Corporation Tech Report RAND_MG562.pdf, 2008, p. 54-55

contracts in each category is approximately 0.5. In other words, contracts increase only by about 5 percent when recruiters are increased by 10 percent.” We expect that the real relationship will be much more complicated. As in our example, adding one more recruiter might approximately double the number of recruits – depending on local factors like, for example, economic conditions. On the other hand, after there are too many recruiters in an area, adding yet one more might produce little additional benefit.

We want to be able to model the effect of a partial recruiter, as well, so that we can split recruiters across areas – assigning a recruiter to spend half his time in one area and half in another, for example. We implemented a simple formula to describe the diminishing rate of return, as shown here:

$$R(r) = A(1 - e^{-Br}),$$

where r is the number of recruiters, $R(r)$ is the number of recruits expressed as a function of r , and A and B are two numbers (called “parameters”) that will be determined separately for each area. Figure 1 depicts $R(r)$ for several values of A and B .

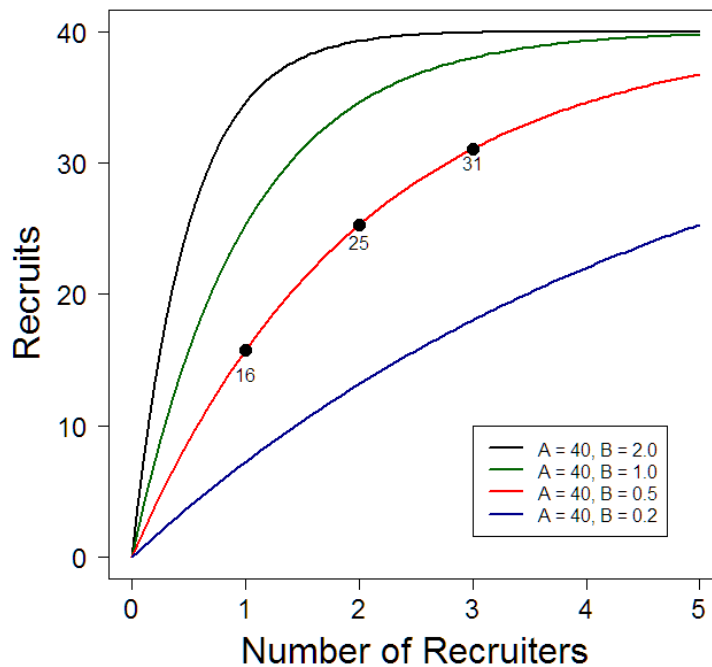


Figure 1: Examples of Diminishing Returns Curves for A=40 and Different Values of B

The parameter A describes the maximum number of recruits available in an area, while B describes the rate at which the return decreases as more recruiters are added. A little calculus shows that the slope of the line for a particular value of r is given by ABe^{-Br} . When $r = 0$ the slope is the product of A and B ; as r becomes large, the slope tends to 0.

2. A Very Simple Model for the Number of Recruits per Recruiter

For this work we have constructed a very simple model for the number of recruits available in an area, as a function of the number of recruiters placed there. The model is intended only as a discussion aid and starting point for further research.

In the model, each young person in an area is given a number called a “recruitability” score. This number describes the level of effort a recruiter would require to get that young person signed up for the military. We scale the recruitability scores from 0 (meaning that the young person volunteers for the service unasked, with no recruiter effort) to infinity (meaning there are no circumstances under which this young person would join the service, regardless of recruiter effort).

Each recruiter has 12 units of recruiting “power” which he or she can apply to the young people in his or her area. We chose 12 here to match the common requirement that each recruiter brings in one recruit per month. We envision recruiters taking a sample of young people and selecting those with the smallest recruitability numbers, then using their recruiting power to persuade those people to join the service. Specifically, the model proceeds like in this “Recruiter Calculator Algorithm” (RCA):

Recruiter Calculator Algorithm (RCA)

1. Examine the Qualified Military Available (QMA) number for each Zip code. Recruitability scores are generated for that many young people, using a random number generator based on a distribution we describe below. To start, with every young person in the QMA for the population is available to be recruited.
2. Assume recruiters operate sequentially, in a loop, starting with recruiter #1.
3. The current recruiter samples a set of the available recruitability scores. We expect that young people who are easier to recruit – that is, who have lower recruitability numbers – will be more represented in the set of potential recruits encountered by this recruiter. So a young person with score r is given a probability of entering the sample that is proportional to $.01 + \max(r) - r$. Young people whose scores are among the very largest have the smallest probability of entering the sample. (Note: The added .01 is there largely for technical reasons.)
4. Each recruiter samples 50 young people (or all of the young people, if there are fewer than 50). The scores are sorted from lowest to highest and the 12 smallest scores, corresponding to the young people who are easiest to recruit, are selected.
5. If the 12 scores for Step 4 add up to a number smaller than 12, all 12 of those young people are recruited. Otherwise we take as many of the 12 as possible, such that the sum of those scores is at most 12. In either case, the recruited individuals are marked as no longer available.
6. If there are fewer than 12 young people left, we stop. If the last recruiter was the sixth, we stop as well, because we have set six as the maximum number of recruiters in one area. Otherwise the next recruiter operates, starting from Step 3.

When the above RCA completes, we have a number of recruits for each recruiter in the area. Because of the randomness this might not be non-increasing: we might have the first recruiter getting 11 recruits and the second, 12, for example. Thus, we sort the numbers from largest to smallest and report that as our final estimate of the numbers of recruits for each recruiter. This simple approach produces numbers that are not inconsistent with the idea of diminishing returns.

One critical element of this simulation model involves the distribution of recruitability scores. We have selected the so-called “gamma” distribution with shape parameter 8 and scale parameter 1/4 to generate these scores. The resulting distribution has mean value 2,

representing our belief that a large segment of the population is difficult to recruit. On the other hand about 5% of this population has values smaller than 1.

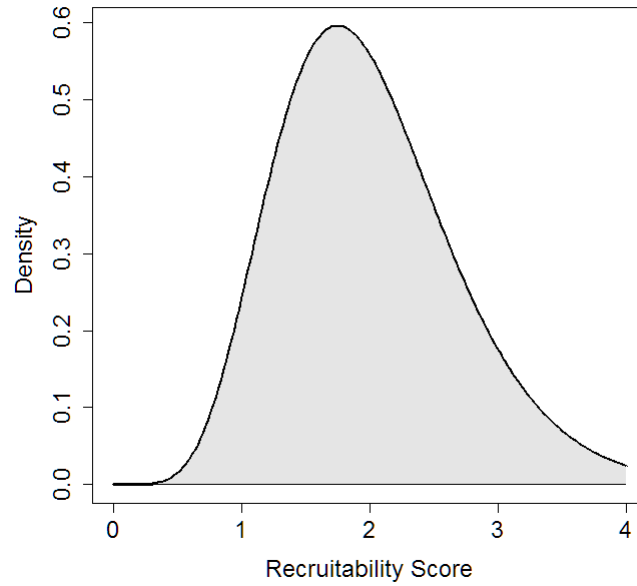


Figure 2: The Gamma Distribution with Shape = 8 and Scale = $\frac{1}{4}$

The relevant gamma distribution is shown in Figure 2. We could imagine, with additional information, tailoring the distribution to the area, taking into account other factors such as propensity to enlist, local economic and educational opportunities, and so on.

In addition to the shape of the recruitability dimension, a number of elements of this set-up have been chosen heuristically. These include the size of the sample that each recruiter takes and the way the probabilities of selection are computed.

THIS PAGE INTENTIONALLY LEFT BLANK

III. THE OPTIMIZATION MODEL

A. APPROXIMATION OF RECRUITING EFFORT IN OPTIMIZATION

The approximation of recruiting effort given by $R(r) = A(1 - e^{-Br})$ and depicted in Figure 1 needs to be further linearized to become usable in NRPOM, which uses a mixed-integer, linear optimization model. (Note: In this section NRPOM refers to the optimization model, even though elsewhere NRPOM refers to the predictor tool too.)

We accomplish this by approximating $R(r)$ between two consecutive integers (e.g., between $r = 0$ and $r = 1$, or between $r = 1$ and $r = 2$) by a piece-wise linear function, as shown in Figure 3 for $A = 40.0$ and $B = 2.0$. The linearization assumes the marginal recruiting effort is constant in-between break points.

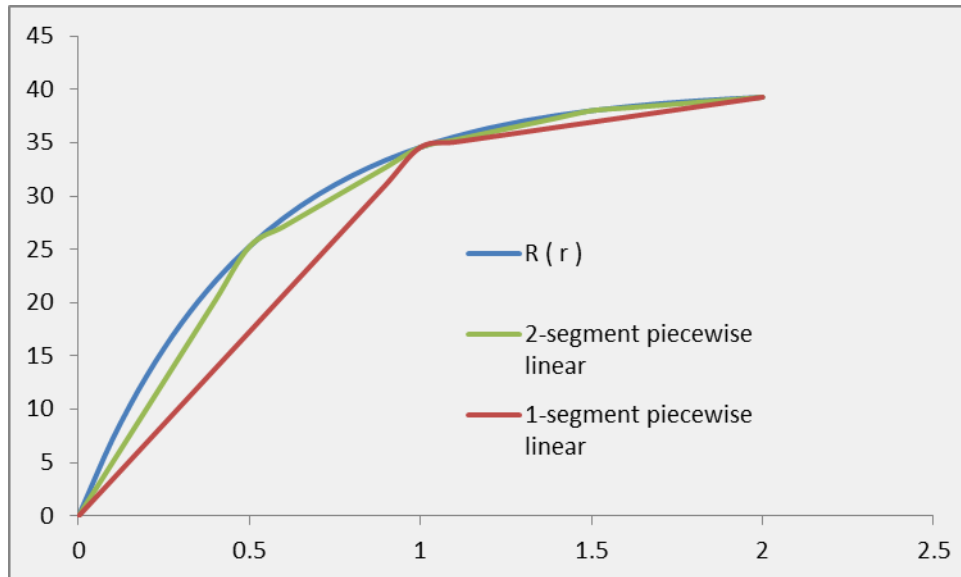


Figure 3: Examples of Two Piece-wise Linear Approximations of a Recruiting Effort Curve

Obviously, the more segments per recruiter, the better the approximation of the intended recruiting function $R(r)$ will be. For example, in Figure 3, if we use a one-segment approximation, only $r = 0, 1$ and 2 recruiter effort is estimated exactly, with some estimations being notably imprecise: for example, the one-segment piecewise

linearization is 17.25 recruits at $r=0.5$, when the actual value is $R(0.5)=25.28$ (the worst value of this approximation occurs at $r=0.419$ where the actual function is $R(r)=22.707$ and the approximation is 8.205). On the other hand, with two segments, our estimation improves notably: it is exact at $r=0, 0.5, 1, 1.5, \text{ and } 2$, and has a worst case at $r=0.229$ where the actual function is $R(r)=14.715$ and the approximation is 7.932).

B. MATHEMATICAL MODEL

For a given region of interest, NRPOM seeks to optimally determine: (a) what candidate stations must be operated; (b) how many recruiters must be assigned to each operated station; and (c) what recruiter effort should be assigned from each operated station to each Zip code.

Notation and full mathematical formulation for NRPOM follows:

Indices and sets:

$s \in S$	set of candidate stations where recruiters can be assigned;
$z \in Z$	set of Zip codes;
$k \in K$	set of ordered segments for the production approximating function, $K = \{1, 2, \dots K\}$.

Data [units]:

nr	total number of recruiters available [recruiters];
ns	number of stations that can be placed [stations];
$dmax$	maximum distance desired between any station and a Zip code assigned to the station [miles or hours];
$wdmeps$	weight that distance to military entrance processing station (MEPS) has in the penalty function [fraction in interval $[0,1)$];
$mineff$	minimum effort that can be assigned to any Zip code that receives recruiters [fraction in interval $[0,1)$];
$maxr_s$	maximum number of recruiters that can be assigned to station s [recruiters];
$dmeps_s$	distance from station s to the closest MEPS [miles or hours];
d_{sz}	distance from station s to central point in Zip code z [miles or hours];
r_{zk}	nominal production rate of segment k for Zip code z [recruits/recruiter];
m_{zk}	maximum number of recruiters within segment k of production function in Zip code z [recruiters].

Derived Data:

$K_z \subset K$	subset of segments applicable to Zip code z given the production rates: $K_z = \{k \in K \mid r_{sz} \geq 0\}$;
$S_z \subset S$	subset of stations that can recruit Zip code z given the maximum distance: $S_z = \{s \in S \mid d_{sz} \leq dmax\}$;
$prod_{szk}$	modified production rate of segment k if Zip code z is assigned to station s [recruits/recruiter]: $prod_{szk} = r_{zk} \left(1 - \frac{d_{sz}}{dmax}\right) \left(1 - wdmeprs\right)^{\frac{dmeprs}{dmax}} . \quad (1)$

Decision variables [units]:

X_s	1 if candidate station s is used, and 0 otherwise;
Y_{sz}	1 if Zip code z is assigned to station s , and 0 otherwise;
NS_s	number of recruiters allocated to station s [recruiters];
NSZ_{sz}	recruiter effort (fraction allowed) in Zip code z from station s [recruiters];
NZK_{zk}	recruiter effort (fraction allowed) in Zip code z , segment k [recruiters];
$Zobj$	the value of objection function [recruits].

Auxiliary variables:

U_{szk}	product of NZK_{zk} and Y_{sz} with the actual number of recruits produced from station s in Zip code z and segment k [recruiters].
-----------	---

Formulation:

$$\text{Maximize } Zobj = \sum_z \sum_{s \in S_z} \sum_k prod_{szk} U_{szk} \quad (2)$$

subject to:

$$\sum_{s \in S} X_s \leq ns \quad (3)$$

$$NS_s \leq maxr_s X_s \quad \forall s \in S \quad (4)$$

$$NS_s \geq 2X_s \quad \forall s \in S \quad (5)$$

$$\sum_{s \in S} NS_s \leq nr \quad (6)$$

$$\sum_{s \in S_z} Y_{sz} = 1 \quad \forall z \in Z \quad (7)$$

$$Y_{sz} \leq X_s \quad \forall s \in S_z, z \in Z \quad (8)$$

$$0 \leq NZK_{zk} \leq m_{zk} \quad \forall k \in K_z, z \in Z \quad (9)$$

$$0 \leq U_{szk} \leq NZK_{zk} \quad \forall s \in S_z, k \in K_z, z \in Z \quad (10)$$

$$U_{szk} \geq NZK_{zk} - nr(1 - Y_{sz}) \quad \forall s \in S_z, k \in K_z, z \in Z \quad (11)$$

$$NZK_{zk} \leq \sum_{s \in S_z} \max_r Y_{sz} \quad \forall k \in K_z, z \in Z \quad (12)$$

$$NSZ_{sz} = \sum_{k \in K_z} U_{szk} \quad \forall s \in S_z, z \in Z \quad (13)$$

$$NSZ_{sz} \leq \max_r Y_{sz} \quad \forall s \in S_z, z \in Z \quad (14)$$

$$NS_s = \sum_{z \in Z} \sum_{k \in K_z} U_{szk} \quad \forall s \in S \quad (15)$$

$$\text{mineff } Y_{sz} \leq NSZ_{sz} \quad \forall s \in S_z, z \in Z \quad (16)$$

$$X_s \in \{0,1\} \quad \forall s \in S \quad (17)$$

$$Y_{sz} \in \{0,1\} \quad \forall s \in S, z \in Z \quad (18)$$

A brief explanation of the formulation follows:

- Equation (1) expresses a reduced production rate from the nominal rate r_{zk} due to (a) relative distance from the recruiting station to the Zip code compared to a maximum distance desired, and (b) relative weight of the distance from the recruiting station to its closest MEPS.
- Equation (2) is the objective function of optimization model: total number of recruits produced.
- Constraint set (3) ensures that the number of stations does not exceed the requirement on the number of stations that can be opened.
- Constraint sets (4) and (5) ensure that the number of recruiters assigned to a station does not exceed the maximum that the station can have, and a minimum of two, which must be enforced if the station is open.
- Constraint set (6) ensures that exactly the total number of recruiters is assigned to some station.
- Constraint sets (7) and (8) ensure that Zip codes are assigned to available stations.
- Constraint set (9) limits the recruiter effort in each segment and Zip. It also establishes variable non-negativity domain.

- Constraint sets (10) and (11) enable the calculation of the recruiting effort by station, Zip code and segment, via a linearization of the product of two variables. They also establish variable non-negativity domain.
- Constraint set (12) is not needed but could help achieving a solution faster by limiting the recruiting effort by station, Zip code and segment from a different angle.
- Constraint set (13) calculates recruiter effort from each station to each Zip code.
- Constraint set (14) ensures recruiter effort from each station to any Zip code can only occur if the station has been assigned to that Zip code.
- Constraint set (15) calculates number of recruiters allocated to each station.
- Constraint set (16) may not be needed but could help achieving a faster solution by limiting recruiting effort by station and Zip code from a different angle.
- Constraint sets (17) and (18) establish additional binary variable domain.

C. INPUT FILES FOR OPTIMIZATION

NRPOM uses seven input files, all of which must be in .csv format. The files are described below:

1. Miscellaneous Input File

This file, that must be named *Misc.csv*, contains the following miscellaneous inputs:

- *nr*: number of recruiters available (positive integer);
- *maxns*: maximum number of stations to operate (positive integer);
- *Dmax*: maximum distance desired between a station and a Zip code assigned to the station (units may be distance or time, but consistent with other inputs);
- *weight_dmeps*: weight that distance from Zip to MEPS has in the penalty function (fraction in interval [0,1]);
- *min_effort*: minimum effort that can be assigned to any Zip code that receives recruiters [fraction in interval [0,1)];
- *effort_breaks*: number of segments to distinguish different rates of recruiting effort for a recruiter (positive integer, recommended between 2 and 6);
- *regression_option*: regression option between two integers (use 1 for linear, or 2 for exponential approximated as piece-wise linear of *effort_breaks* segments);

- *meanErr_override*: mean squared error required to override a Zip code's exponential approximation of recruiting effort by the piece-wise linear approximation (squared number of recruits);
- *maxTimeMinutes*: maximum time given to the optimization solver to find an optimal solution (minutes). If not provable optimal at the time limit, the solver will stop and return the best solution found.

nr	40
maxns	6
Dmax	200
weight_dmeps	0.3
min_effort	0.1
effort_breaks	2
regression_option	2
meanErr_override	0.10
maxTimeMinutes	60

Figure 4: Example of Miscellaneous (*Misc.csv*) input file

2. Station Input Files

There are two station input files. These must be named *S.csv* and *S_data.csv*, and examples are displayed in Figure 5.

The first file, *S.csv*, simply contains a list of station names.

The second file, *S_data.csv* contains that same list in the first column (except for the first row, which is blank). The other columns are as follows:

- *d_MEPS*: distance from the station to the closest MEPS (distance or time units);
- *mr*: maximum number of recruiters that can be assigned to the station (positive integer);
- *cost*: fixed cost of operating the station (i.e., of assigning recruiters to that station). Note: cost is not included in the current model formulation because we do not have any budget information, but it can be added in future versions.

s94015	s94015	d_MEPS	mr	cost
s94109	s94109	29.18	10	2.5
s94402	s94402	42.77	10	4.6
s94501	s94501	16.9	10	3.2
s94513	s94513	34.64	10	2.5
s94519	s94519	43	10	4.6
s94538	s94538	39	10	3.2
s94550	s94550	10.4	10	2.5
s94578	s94578	26	10	4.6
s94939	s94939	20.7	10	3.2
		44.4	10	4

Figure 5: Example of Station and Station Date input files: *S.csv* (left) and *S_Data.csv* (right)

3. Zip Input Files

There are three Zip input files and one additional Zip-station input file.

The first Zip input file is *Z.csv*, and it simply contains a list of Zip codes.

The second file we describe is the Zip-station distance input file. This file must be named *SZ_Dist.csv* and it contains the Zip list in the first column (except for the first row, which is blank), and the station codes in the first row (except the first column, which is blank). That is, *SZ_Dist.csv* has the structure of a two-entry table (or matrix) with Zip codes down and stations across. The data by Zip code and station is the “distance” between the station and the Zip code (distance here refers to driving distance or time for a recruiter to travel from the station to the Zip code).

Samples of both files *Z.csv* and *SZ_Dist.csv* are shown in Figure 6.

90001	s91055	s92105	s91427
90002	421.6	841.6	570.4
90003	421.2	841.2	570
90004	420.8	840.8	569.6
...

Figure 6: Example of Zip and Station-Zip input files: *Z.csv* (left) and *SZ_Dist.csv* (right)

The next Zip input file is the “production” file, *Z_Product.csv*. This file contains the Zip list in the first column (except for the first row, which is blank), and column headers in the first row labeled “*Rec0*”, “*Rec1*”, ..., “*Rec6*”. Each of this refers to a recruiting effort of 0, 1, ..., 6 recruiters, respectively. (The limit of 6 exists in the current version but can be easily increased.) An example of this file is given in Figure 7.

The data by Zip code recruiting effort is the number of recruits that can be recruited by the exact recruiting effort shown in the corresponding column. The first column, “*Rec0*” can be omitted, as the number of recruits is assumed to be zero. The numbers given by row must be non-decreasing, but if no strictly positive increase happens, a blank can be given (instead of repeating a previous value). For example, in Figure 7, the row for Zip code *90001* could have blanks for *Rec5* and *Rec6*.

	Rec0	Rec1	Rec2	Rec3	Rec4	Rec5	Rec6
90001	0	15	25	29	31	31	31
90002	0	7	11	13	13	13	13
90003	0	25	40	47	49	49	49
90004	0	3	5	5	5	5	5
...

Figure 7: Example of Zip Production input file, *Z_Production.csv*

The last Zip input file is the “fit” file, *Z_Fit.csv* (see Figure 8). This file is mostly filled by the optimization, with the exception of the headers: “*a*”, “*b*”, “*meanSqErr*” and “*meanErr1*”. That is, even the first time that a particular scenario is run, NRPOM needs the file to exist with at least those headers in columns 2, 3, 4 and 5, respectively.

The file contains the parameters and statistics for a least-squares regression fit for each Zip code, assuming the $R(r) = A(1 - e^{-Br})$ model. NRPOM uses the inputs in the *Z_Product.dat* to calculate the fit. However, if the file already contains data for a Zip code (e.g., from a previous run), then NRPOM will save substantial time by not recreating the fit and using the existing information on the file.

If *Z_Product.dat* has changed for an existing Zip code, then the user must manually delete the row in *Z_Fit.csv* corresponding to that Zip code, so as to indicate NRPOM that a fit for that Zip code must be recalculated.

If a new Zip code is added to *Z_Product.dat*, the user needs not change *Z_Fit.csv* (the lack of that record will make NRPOM calculate it and record it for the next time).

	a	b	meanSqErr	meanErr1
91055	78.17	0.5	0.64	0.03
92105	25.51	0.69	0.04	0.02
91427	59.76	0.71	0.11	0.01
92449	13.27	1.01	0.09	0.04
...

Figure 8: Example of Zip Fit input file, *Z_Fit.csv*

D. EXAMPLE OF OPTIMIZATION RESULTS

We have tested Predictor and NRPOM with notional data to emulate a California-like scenario, but some data has not been drawn from reliable NRC sources.

1. California Scenario Setup

The California scenario draws from a subset of available predictors produced in the Fulton thesis references above. These include population size (census and QMA), and density, poverty levels, proportions of residents by age and by race, proportions without a high school diploma, obesity rate and unemployment rates, and indicators which divide all Zips into between 4 and 17 clusters based on variables that describe demographics,

health, education, military presence, and economic strength. Overall, the CA scenario includes 2,156 Zip codes of which 121 are candidates to host a recruiting station.

We do not have enough information to differentiate among the candidate stations in terms of size (how large or small those stations could be) and cost (set up, annual overhead and/or staffing). Thus, we simply assume that each station can host anywhere from 2 to 20 recruiters.

We use actual Euclidean distances between Zip code centroids. For example, the distance between Zip codes 90001 (a neighborhood in Los Angeles County) and 92121 (Sorrento Valley in San Diego County) is estimated as 96.37 miles.

Nominal recruiting functions are based on the Recruiter Calculator Algorithm above and a least-square error fit for a recruiting function of the form $R(r) = A(1 - e^{-Br})$, as described in Section II.D. For example, for Zip code 90001 we estimate

$R(r) = 353.7(1 - e^{-0.033845r})$, which gives, approximately 0, 11.8, 23.2, 34.2, 44.8, 55.1 and 65.0 recruits for effort levels $r=0, 1, 2, 3, 4, 5$ and 6 recruiters, respectively. In contrast, Zip code 92517 (a downtown neighborhood in Riverside County near San Bernardino) has a nominal recruiting function $R(r) = 12.34(1 - e^{-1.02693r})$, which produces 0, 7.9, 10.8, 11.8, 12.1, 12.3 and 12.3 recruits for the same recruiting efforts as above, respectively.

The above nominal recruiting functions are first approximated with four-segment piecewise linear functions for any two consecutive integers (number of recruits). That creates nominal rates $r_{z,k}$ for recruiting a certain Zip code z in a given segment effort k .

This recruiting rate is further altered into final production level $prod_{szk}$ based on the specific station s assigned to the Zip code (see Equation (1)). The relevant parameters used here are a maximum desired distance $dmax=200$ miles, and a relative weight of the distance to closest MEPS $wdmeps=0.2$.

We run several scenarios that differ in the maximum number of stations that can be opened. Specifically we explore a maximum of $maxns=30, 40, 50$ and 80 stations,

respectively. We also vary the number of recruiters available to assign to those stations: for $maxns=30, 40, 50$ we assign $nr=150$ recruiters; and, for $maxns=80$ stations we use $nr=500$ recruiters. For brevity we refer to a specific instance as “Scenario $nr-maxns$,” such as Scenario 150-40.

2. Scenario Results

We list our summary results in Figure 9. As expected, run time (in seconds) increases with the complexity of the scenario (more stations and/or recruiters to assign). The most complex case (Scenario 500-80) takes approximately 1.5h. This time assumes the all regression fits have been performed in advance (this process may take an additional 30 minutes but it needs to be done only once). All scenarios are solved with the GAMS/CPLEX optimization engine to a relative error under 1%.

Scenario	Run time	Recruits	Original	Reduction	Open
150-30	3,397	2,049.2	2,061.4	239	30
150-40	2,087	2,057.2	2,076.6	235	40
150-50	3,297	2,064.6	2,295.1	230	50
500-80	5,359	5,929.1	6,310.6	382	80

Figure 9: Summary of Test Cases and Results

The number of recruits (Recruits column) is our final goal. We observe no significant improvement by allowing more than 30 stations if the number of total recruiters remains the same (150).

The column labeled Original shows the recruits as assessed by nominal rates r_{zk} approximation. This is just a function of the recruiting efforts (i.e., it disregards the associated recruiting stations). When distance and MEPS weight are factored in, we estimate a reduction (shown in the next column) that produces the final recruit figure. The last column simply displays the number of stations open in the final configuration. In all of our scenarios, it matches the maximum available to open.

In Figure 10 we can see an excerpt from the detailed output for Scenario 500-80. Specifically, we see two of the Zip codes that NRPOM recommends for a recruiting station: 92121, with six recruiters, and 92802, with 11 recruiters. These recruiters will share their time among several Zip codes in fractions multiple of 25%. We will have 1.25 recruiting effort of station 92121 in its own Zip code (this can be achieved, for example, by one recruiter 100% and another recruiter 25% of their time). But we will have only 0.5 recruiting effort from station 92121 in Zip code 92007. The last two columns are identical to the ones described for Figure 9, but specific to a recruiting effort in question. For example, we can expect 13.9 and 5.4 recruits from the aforementioned efforts, respectively.

station	ZIP	nRec	Obj	Orig	Reduc
92121		6			
	92121	1.25	13.9	14.1	0.3
	92007	0.5	5.4	5.7	0.3
	92014	1	11.1	11.6	0.5
	92075	0.75	8.1	8.5	0.4
	92129	1.5	16.4	16.9	0.5
	92130	1	10.8	11.1	0.3
92802		11			
	92802	0.75	8.2	8.5	0.3
	90621	2	22.6	23.9	1.3
	90680	0.5	5.4	5.7	0.3
	92801	2.75	31	32.3	1.3
	92806	1.25	14	14.7	0.7
	92833	1.25	14	14.7	0.7
	92835	0.25	2.7	2.9	0.2
	92841	2	22.9	24	1.1
	92843	0.25	2.7	2.9	0.1

Figure 10: Station Assignment Excerpt from Scenario 500-80

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
Ft. Belvoir, Virginia
2. Dudley Knox Library
Naval Postgraduate School
Monterey, California
3. Research Sponsored Programs Office, Code 41
Naval Postgraduate School
Monterey, CA 93943
4. Mr. Gary C. Peterson
CNRC N00B, Navy Recruiting Command
gary.c.peterson@navy.mil
5. Mr. Daniel Ammons-Moreno
N37, Operations Analysis Division, Navy Recruiting Command
daniel.ammonsmoreno1@navy.mil