A probability-based lost person simulation for wilderness search and rescue operations

Max Metcalfe
Department of Geography, Boston University

Advisor: Luis Carvalho
Department of Mathematics and Statistics, Boston University

April 4th, 2012

Chapter 1

Introduction

Search and Rescue (SAR) operations require fast, organized, and knowledgable search teams in order to safely locate the lost subject and remove the subject from harm's way. Each search case is unique, requiring meticulous investigation and preparation, and search teams should be aware of the current search strategies and be able to apply these strategies when they are applicable. Research related to SAR began after World War II, primarily for locating enemy submarines (Koopman, 1956, 1957) . Later research modified these early search methods for use in land-based SAR scenarios (Mattson, R. J., 1975; Stone, 1975; Cooper, Frost, and Robe, 2003; O'Connor, 2004). Land-based scenarios are more complex than ocean and air cases because both the geography of the terrain and the behavior of the lost subject directly influence the subject's possible locations. Modern search methods involve the analysis of lost person behaviors (Syrotuck, 1977; Koester, 2008) as well as a complex system of probability assignments (Bownds et al, 1991). More recently, data from historical SAR cases have become more available and this has motivated further research (Koester, 2008).

While a portion of the current and past research focuses on search team organization (sweep methods, effort allocation, etc.), our research aims specifically at determining which areas are most likely to contain the lost subject and constructing probability areas for use by search teams. Since search resources are limited and efficiency is critical, determining these areas is crucial in the search planning process (Stone, 1975). Our methods use a lost person simulation, which is described in Chapter 2.

Chapter 2

Lost Person Simulation

The behavior of a lost subject is dictated by a range of inter-related factors; both the terrain of the region and the phycological status of the subject (inferred from witnesses, family members, and physical evidence) have powerful and varying effects on the true path of the subject. In order to account for these factors, we have created a probability-based lost person simulation. The foundation of the simulation is within a series of decision points, during which the subject chooses a move through an eight-cell sampling process (see Figure 1). A walk is defined as sequence of moves, with each move beginning with a decision and ending with the subject moving to the chosen destination. At each decision point, the subject can either rest (no motion) or move (to one of the eight neighboring cells). Within each move decision, each potential move will be referred to as a *move option (i).*

The lost person simulation was implemented using GIS (Geographic Information Systems) Python scripting. The visual capabilities and spatial processing functions of GIS make it a useful tool for SAR applications. This version of the model includes a sequential set of functions which contribute values to the final move sampling. In describing the model, we will refer directly to these functions.

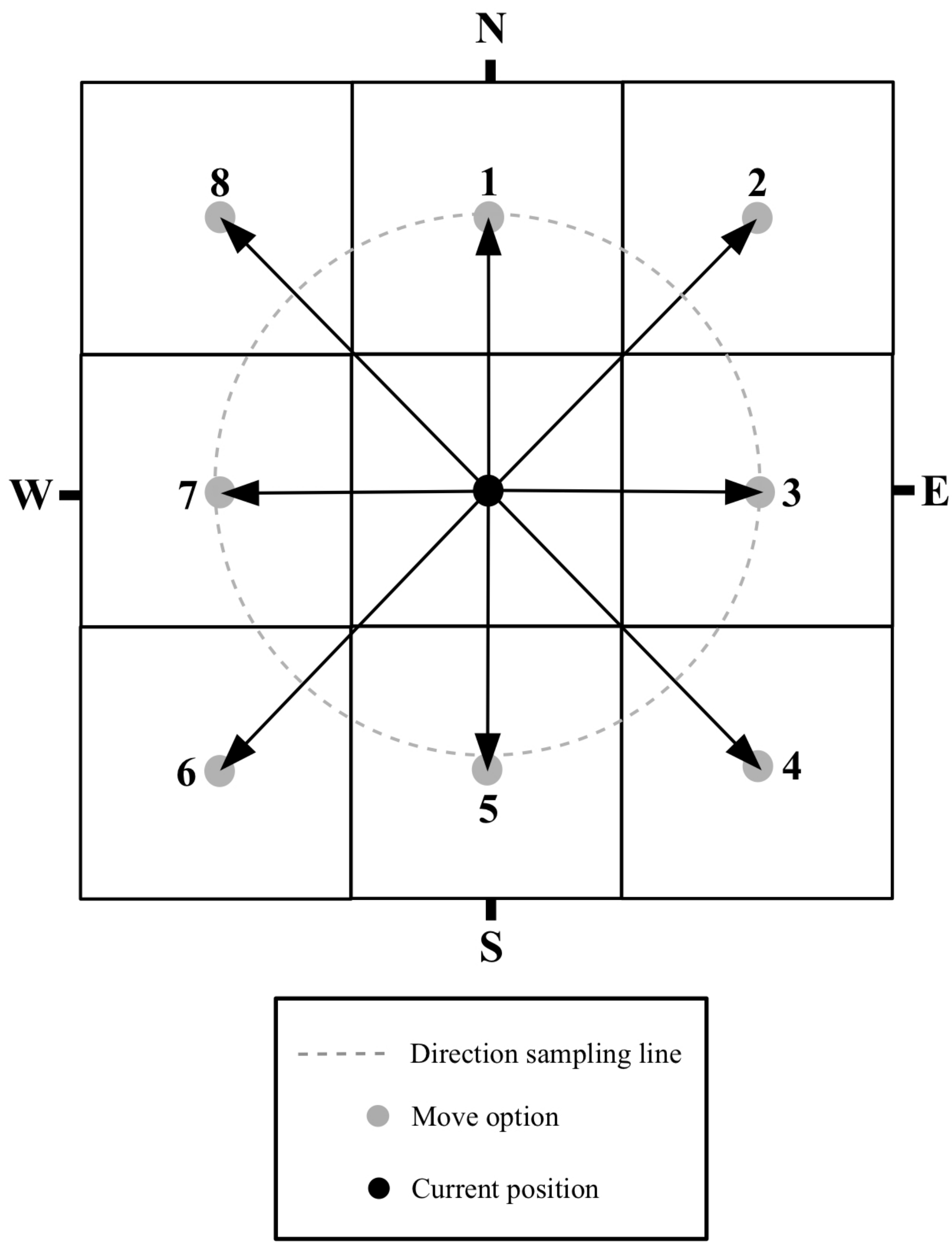

Figure 1: Eight-cell move diagram. This diagram represents an example decision point, showing the eight possible move options.

2.1 Point Last Seen

Every search begins with a Point Last Seen (PLS). The PLS is the most recent confirmed location of the subject. Witnesses, a note from the subject, or other ancillary information may contribute to deciding upon a PLS. In the lost person simulation, all walks begin from the PLS.

2.2 Point of Interest

The point of interest (POI) models a travel tendency of the subject towards a goal location. The POI can either be a physical location towards which the subject wants to travel, or a phycological tendency (e.g. through a valley, towards the mountain). The difficulty in deciding upon a POI is threefold: (1) The subject may have not provided information regarding a goal location (therefore not requiring a POI). (2) The intentions and travel tendencies may change throughout the search. (3) The subject is likely to not have a correct inclination of the true location of the POI. This third factor is likely considering the subject is lost and disoriented. The reason for including a POI is to model the subject's preference in traveling, and it should only be used if there is evidence of a phycological travel bias (i.e. not directly influenced by the terrain).

The accuracy of the subject's perception of the POI is modeled with a width parameter θ, measured in compass degrees. θ is centered on the true compass bearing from the current location of the subject towards the true location of the POI. The perceived direction is sampled from a uniform distribution within this compass region (see Figure 2). Large values of θ denote large errors in the subject's sense of direction, and small values of θ denote an accurate understanding of the POI's location (see Figure 3).

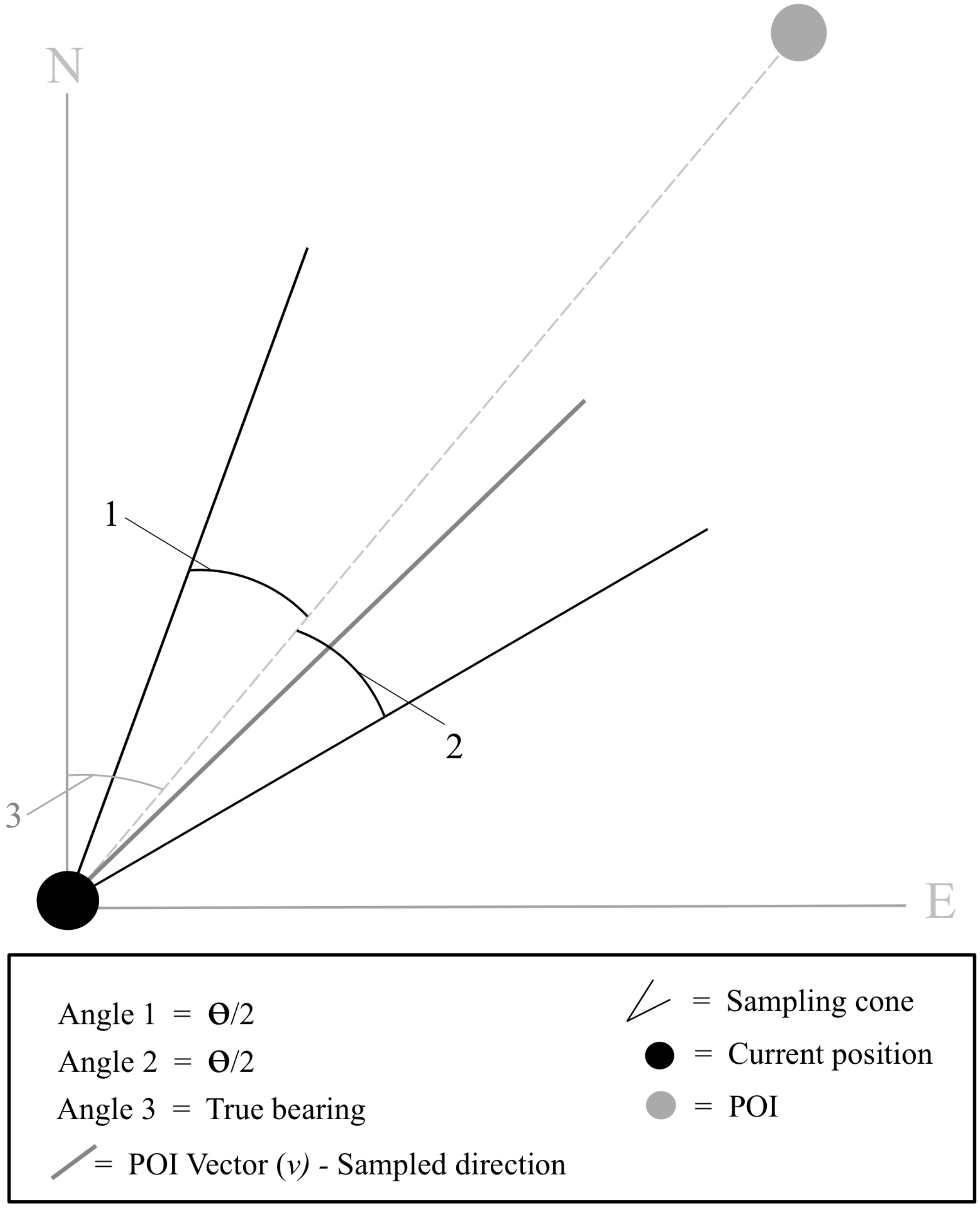

Figure 2: POI direction sampling. Sampling within a cone of width = ϴ produces the subject's perceived direction of the POI (shown here as POI Vector (*v*)).

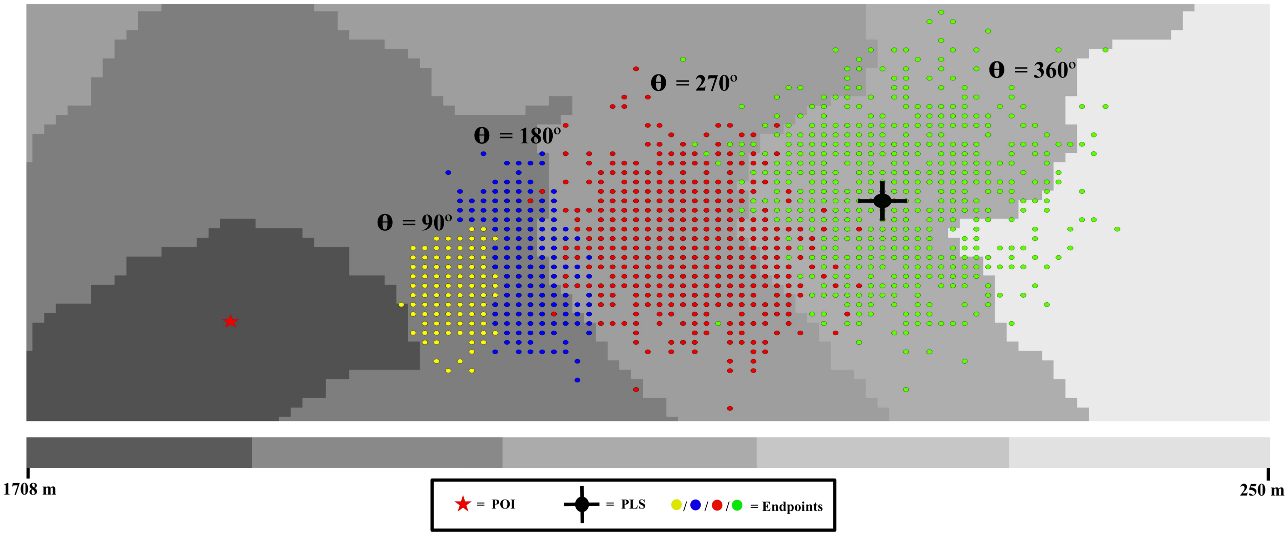

Figure 3: Effect of ϴ on endpoint distribution (n = 1000, S*top Time* = 5000 s (~ 1 hr 23 mins) , *Long* = .00001, *Short* = .0001, $Clock_{r:1/2}$ = 14400 s (4 hrs) , P_0 = .01, $\eta = 0$, $\gamma = 20$, T_{rest} / *Walk Clock = .25,* $V_T / V_F = 2$).

The influence of the POI Vector is quantified through a dot product calculation. For each move option, the dot product is computed between the cell move vector (*u)* and the POI Vector (*v)*. The dot product quantifies the alignment of each move option with the POI Vector (e.g. if *v* is aligned with *u* , the value of the dot product is 1. If *v* is in the opposite direction of the *u*, the value of the dot product is -1). The final probability calculations (described later in Section 2.6) use these dot product values for each move option.

2.3 Coordinates, Distance, and Elevation Gain

The first calculation of every decision point is to get the coordinates and distances of the eight neighboring cells (North, Northeast, East, Southeast, South, Southwest, West, and Northwest). The *Check Around* function gathers the move option coordinates and distances by adding and/or subtracting the cell size from the subject's current location. On any given move, the subject can move from a certain location in the current cell to that same location in any of the eight neighboring cells. The distances of each move is defined as the raster cell size, the raster being the local digital elevation model (DEM).

While a move from the current cell to a diagonal cell (i.e. Northeast, Southeast, Southwest, or

Northwest) is longer than a move directly horizontal or vertical (i.e. North, East, South, or West), the travel distances are all defined as the cell size. Keeping these distances uniform removes bias against the diagonal moves, since the longer distances would lead to longer time values for these moves. This exception is only applicable for the move sampling - the true move coordinates are used when the subject executes the move.

The *Check Gain* function gathers the elevation of each move option and then computes the elevation gains.

2.4 Travel Pace

Using the elevation gain values, the *Check Pace* function returns the fresh travel pace for each move option. The fresh travel paces V_F are listed below for each elevation gain range. We have estimated these paces using Naismith's Rule - allow 1 hour for every 3 miles forward, and an additional 1 hr for every 2000 ft of ascent. We have also simplified this rule and assumed a single value for flat, uphill, and downhill travel (Naismith, 1892).

Gain $>$ 0 m (uphill move): $V_{F\,(Up)} = 0.335$ m/s

Gain $=$ 0 m (flat move): $V_{F\,(Flat)} = 0.670$ m/s

Gain $<$ 0 m (downhill move): $V_{F\,(Down)} = 1.340$ m/s

(1) <u>Travel Aids:</u> Trails, or other features that promote easier travel (i.e. roads, fields, etc.), are defined as travel aids. If the lost subject approaches a trail, it is likely for the subject to travel on the trail rather than continue through the wilderness. A *Check Trail* function checks each move option for a nearby travel aid. If a move option utilizes a travel aid, an “on trail” pace is calculated as a multiple of the fresh pace (i.e. $V_T = 2V_F$). This relationship should be defined by the search planner based upon the terrain and the subject.

(2) Long Fatigue: To model the decreasing pace of the subject over the whole duration of the walk, the *Long Fatigue* function decreases the fresh pace after each rest period. The parameter *Long* controls this relationship.

$$V_{L_i} = V_{F_i} * \exp(-Long * Clock) \qquad (1)$$

$V_{F\,i}$ = Fresh travel pace

$V_{L\,i}$ = Pace after long fatigue

Clock = Duration of walk

Long = Long-term fitness of subject

(3) Short Fatigue: To model the decreasing pace of the subject between rest stops (i.e. within each walk segment), the S*hort Fatigue* function decreases the initial pace of the walk segment ($V_{L\,i}$) as the subject travels away from the rest stop. The parameter *Short* dictates this relationship.

$$V_{S_i} = V_{L_i} * \exp(-Short * Walk\,Clock) \qquad (2)$$

V_{Li} = Travel pace after long fatigue

V_{Si} = Travel pace after short fatigue

Walk Clock = Time since last rest stop

Short = Short-term fitness of subject

We have chosen to use the above fatigue equations based upon the assumed tendencies of a lost person (i.e. the faster the subject travels, the more rapid the subject fatigues). This is applicable in both short and long-term fatigue. This logic is more clearly expressed in the differential equation shown in Equation 3 below; both fatigue functions are solutions to this equation.

$$dv/dt = -\ Short * V \qquad (3)$$

2.5 Time and Probability

The final step before calculating the move probabilities is determining the time values for each move option. Using the travel pace ($V_{S\,i}$) and the terrain distance (calculated using the sampling distance and the elevation gain), the *Check Time* function returns the time values for each move option.

(1) <u>Probability of Move Options:</u> From the time values calculated above and the POI Vector (described in Section 2.2 Point of Interest), the *Check Probabilities* function calculates the probability of each move option. Here, the two parameters γ and η balance the influence of the POI Vector (v) with the time values for each move option (t_i). γ models the desire of the subject to follow his/her perception of the POI, while η models the influence of the time-related factors (terrain, fatigue, travel aids). Figure 4 demonstrates the effects of both η and γ.

$$P_{move_i} = \exp(-\eta t_i + \gamma(v \cdot u_i)) \tag{4}$$

η = Subject's sensitivity to time

γ = Subject's desire to follow POI inclination

u_i = Cell move vector

v = POI Vector

t_i = Time of move option

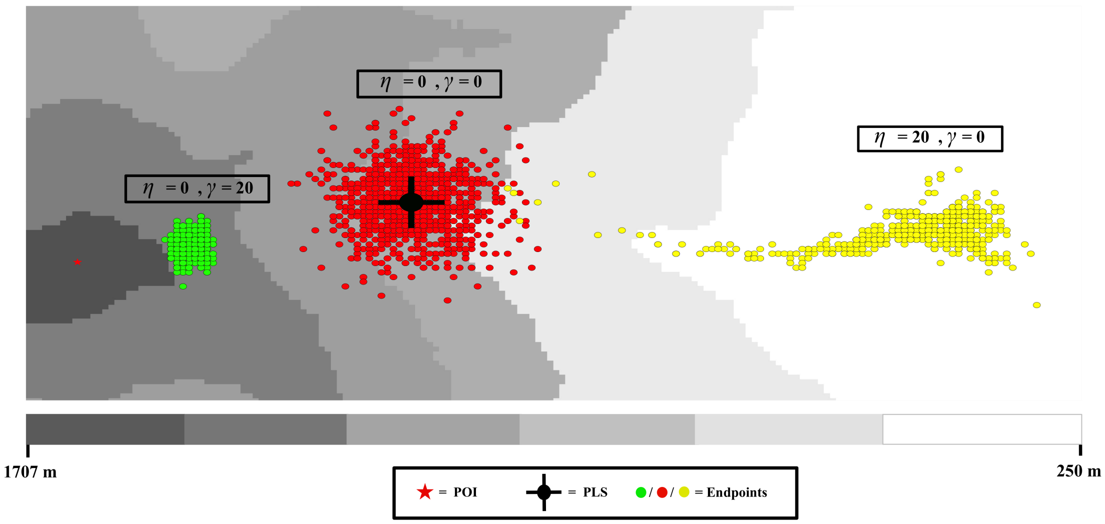

Figure 4: The effect of γ and η on endpoint distribution (n = 1000, *Stop Time* = 5000 s (~ 1 hr 23 mins) , *Long* = .00001, *Short* = .0001, $Clock_{r:1/2}$ = 14400 s (4 hrs) , P_0 = .01, ϴ = 90º, T_{rest} / *Walk Clock* = .25, V_T / V_F = 2)

(2) Probability of Resting: An additional option during each decision point is to rest. It is assumed that as the subject becomes tired, the probability of resting increases. The probability of resting during a decision point is described in Equation 5 below:

$$P_{rest} = \exp((r * Clock) + r_o) \ / \ 1 + \exp((r * Clock) + r_0) \qquad (5)$$

r_0 = Calculated in Equation 6

r = Calculated in Equation 7

Clock = Total walk time

$$r_0 = \log(P_0 \ / \ 1 - P_0) \qquad (6)$$

$$r = -r_o \ / \ Clock_{r:1/2} \qquad (7)$$

Equations 6 and 7 require the parameters P_0 and *r*. P_0 is the probability of resting when *Clock* = 0 (i.e. at the beginning of the walk). This value can be changed according to the fitness and phycological state of the subject, but it could also reflect the immediate health of the subject (e.g. an injury requiring more frequent rests). Additionally, a value of *r* must be estimated by inferring at what point during the walk ($Clock_{r:1/2}$) is P_{rest} = .50. Although this may be a difficult parameter to assess, it is important to estimate a value for each scenario - a determined and athletic runner may avoid resting for extremely long periods of time due both to fitness and attitude, while an un-fit, casual walker may tend to rest after short periods of activity.

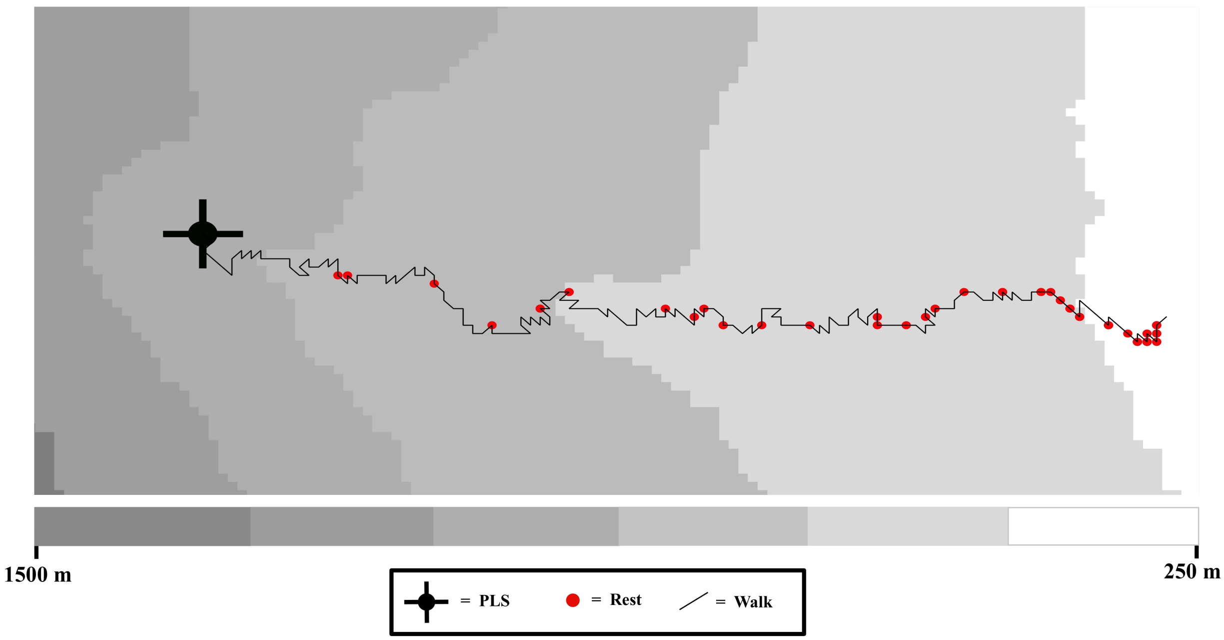

Figure 5: Sample walk showing increased frequency of rest stops over the length of walk (n = 1, S*top Time* = 12000 s (3 hrs 20 mins), *Long* = .00001, *Short* = .0001, $Clock_{r:1/2}$ = 6000 s (1 hr 40 mins) , P_0 = .05, η = 20, γ =*0,* ϴ = 180º, T_{rest} / *Walk Clock = .25,* $V_T / V_F = 2$*).*

Duration of Rest*:* The time spent resting must also be determined for each rest stop. It is assumed that the subject rests for a fraction of the time spent traveling since the pervious rest stop (T_{rest} / *Walk Clock)*. If necessary, search planners can change this proportion to reflect the subject's fitness (similar to the fatigue parameters *Long* and *Short*). A fit and determined subject is likely to spend shorter periods of time resting than an unfit subject.

2.6 Move Sampling

The sampling procedure is divided into two parts: (1) Sampling a move type (rest or move). (2) If the decision is to move, sampling a move option. The probability of resting (P_{rest}, calculated from Equation 5) and the probability of moving ($1 - P_{rest}$) are sampled with these weights. If the result is the choice to rest, the subject rests; if the choice is to move, the eight move options are sampled with their associated probabilities ($P_{move\ i}$ – for each move option as calculated in Equation 4). Prior to sampling, all probabilities are normalized. Once a move is chosen, the subject executes the move and the process is repeated.

2.7 Duration of Walk

Each walk continues until the *Clock* equals the *Stop Time.* The *Stop Time* is the total time of the simulation, and in some cases may be the time associated with the PLS (i.e. estimated time of travel from the PLS). For example, search planners could receive a information that the subject was at a certain trail junction at a specific time. Here, the total time since the subject was confirmed to be at the trail junction could be used as the *Stop Time.* Other situations may merit different uses of the *Stop Time* value, specifically if the PLS does not have an associated time or if multiple scenarios are modeled for a single search case. Additionally, the *Stop Time* helps in determining the probability areas (described later in Chapter 3), especially when multiple sets of areas are constructed throughout the search process. As the search continues, the *Stop Time* increases.

Chapter 3

Applications

SAR-related research is only useful if the results and methods can be applied in the field. The tools used by search teams must be effective and easy to interpret in a fast-paced search environment. GIS has enabled us to present the results of this simulation in simple and clear ways. Referring back to the initial intentions of using GIS, the capabilities of GIS proved to be very valuable in this research.

The goal in viewing the results of this simulation is to identify the areas most likely to contain the lost subject. One possible way to do this is to construct probability areas as convex hulls containing the points closest to the median point away from the PLS (see Figure 6). The probability of each zone is defined by the proportion of points within each convex hull (i.e. the 25% convex hull contains at least $.25 * n$ (the number of iterations) points. These methods have been derived from the probability plots created by William G. Syrotuck in his book *Analysis of Lost Person Behavior: An Aid to Search Planning* (Syrotuck, 1977). However, the results of our simulation and the capabilities of GIS have allowed us to construct these areas over a spatial scale, rather than a linear scale. In Figure 6, we have also collapsed the probability areas into a linear scale to show this type of representation. Whether viewing probability maps or linear distance scales, it is important to interpret these results in relation to the larger search scenario. It is not enough to base the search on these plots alone, but to use them as a tool to understand the relationship between the terrain and the subject's behavior.

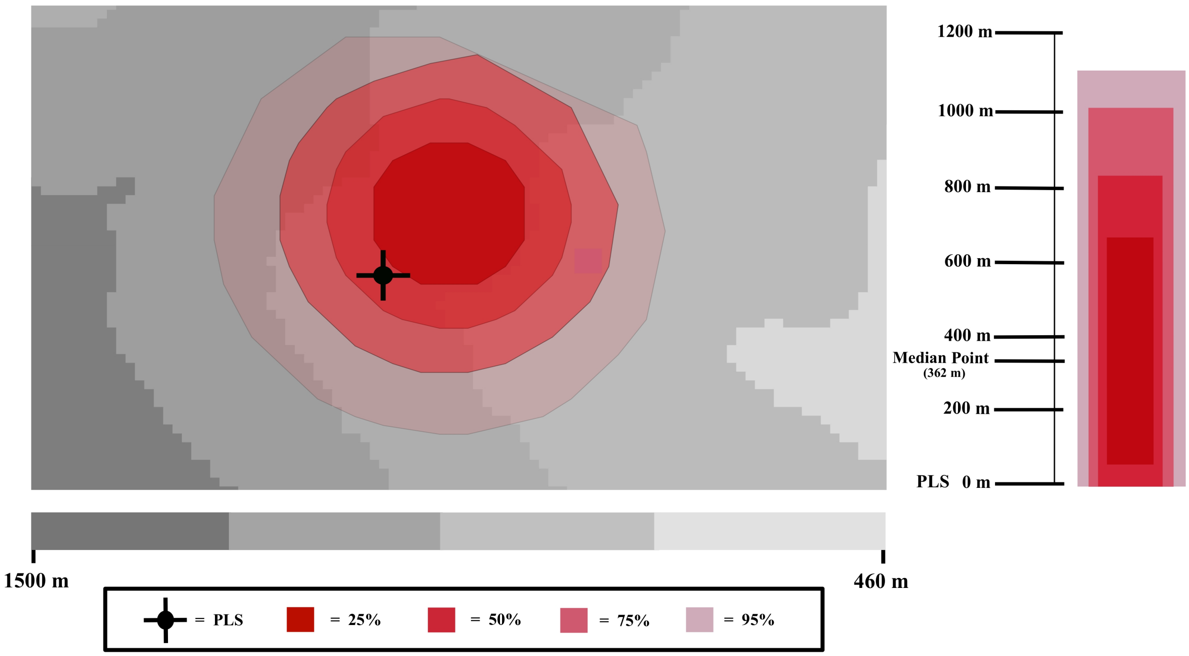

Figure 6: Example probability areas with linear distance from PLS plot (n = 5000, S*top Time* = 8000 (~ 2 hrs 13 mins), *Long* = .00001, *Short* = .0001, *Clock* $_{r:1/2}$ = 14400 s (4 hrs), P_0 = .01, $\eta = 0$, $\gamma = 0$, Θ = 360º, T_{rest} / *Walk Clock = .25,* $V_T / V_F = 2$).

Chapter 4

Discussion

The goal of this research is to design and implement a realistic and simple lost person simulation. Other types of empirical models, such as those which are based purely on SAR data (PLS and find location data points), attempt to model these scenarios without considering the individual travel decisions of the subject. Moreover, many of the current methods in SAR use expert consensus evaluations to determine search areas as well as phycological profiles of lost person behavior. While these methods are useful, they do not model the individual move decisions of the subject. The structure of our model is similar to that of the decision making environment of a lost subject. It is this realism for which we have designed this model. This has also required us to simplify the model and only include those variables which are most relevant in the decision making process. It has been our goal to maintain a balance between realism and simplicity.

This model requires extensive parameterization, considering the large number of parameters. However, the resolution of the data required to parameterize this model cannot be found in the current SAR datasets. Data does not exist for the individual decision points of the lost subject, but rather only for the initial PLS location and the find location. Until we can find this resolution of data, we must use other applicable methods to parameterize the model. One potential method is to use the probability areas as a calibration tool. This would require a series of tests in which the parameters would be changed until the 95% probability area contained the actual find location. This process would then be repeated sequentially for each probability area (75%, 50%, and 25%). Other methods, such as a 10-fold cross validation, could be used for parameterization. It would also be important to create *Subject Profiles,* allowing search teams to quickly set the parameters for each subject. Lastly, it will be necessary to complete a thorough validation of this model. Charles Twardy, a pioneer of SAR research, has designed a lost person simulation validation system and it may be possible to run a complete validation using this tool. This method will enable us to compare the simulation results with a large database of SAR data.

Human behavior is very unpredictable and modeling it is often an impossible task. This research will require many more hours of work - tweaking parameters, validating results, and continually adding and subtracting behavioral factors. However, it is not a single quantifiable validation for which is the goal. The ultimate benefit from this type of modeling is in the continuous testing and investigation which leads to a better understanding of the true nature of lost persons. Potential influences are thrown away and new influences surface, but only those which withstand validation and testing will remain relevant and our overall understanding of these issues will progress.

Appendix A

Parameters

Variable Parameters	
Parameter	**Description**
θ	ɵ determines the width of the POI sampling range in compass degrees. The sampling “cone” is defined as [*POI bearing* – $\theta/2$, *POI bearing* + $\theta/2$]. Large θ values indicate a lower accuracy in the subject's perception of the POI location, and smaller θ values indicate a higher accuracy in the subject's perception of the POI location.
η	η determines the sensitivity of time (i.e. the travel duration of each move option) in the probability calculations.
γ	γ determines the POI sensitivity in the probability calculations. If the subject is determined to follow his/her inclination of the POI (independent of its accuracy), γ should be large. If the subject chooses to ignore the POI, the value of γ should be small. If there is no POI, $\gamma = 0$.
Long	The *Long* variable is a fitness parameter to model the subject's fatigue over the entire duration of the walk. High values of *Long* model low fitness, while low values of *Long* model high fitness.
Short	The *Short* variable is a fitness parameter to model the subject's fatigue between rests. High values of *Short* model low fitness, while low values of *Short* model high fitness.

Fixed Parameters	
Parameter	**Description**
P_0	Probability of resting when *Clock* = 0 (i.e. at the start of the simulation). This value has been set to .005 or .01, depending on the testing scenario.
$Clock_{r:½}$	*Clock* time when the P_0 = .50. This value has been to 1440 seconds (4 hours).
$V_{F\ (Up)}$	The fresh travel pace when gain > 0. Using an estimate from the Naismith's Rule, this value has been set to 0.335 m/s
$V_{F\ (Down)}$	The fresh travel pace when gain < 0. Using an estimate from the Naismith's Rule, this value has been set to 1.34 m/s.
$V_{F\ (Flat)}$	The fresh travel pace when gain < 0. Using an estimate from the Naismith's Rule, this value has been set to 0.67 m/s.
$T_{Rest} / Walk\ Clock$	The fraction of time spent resting after each walk segment. A value of .5 means that the subject rests for one half of the time for which he/she had previously traveled since the last rest point. This value has been set to .25.
$V_T\ /\ V_F$	The travel pace on a travel aid (e.g. trail) compared to the fresh travel pace. This value has been set to 2 (i.e. the subject travels twice as fast via a trail compared to the fresh travel pace).

We have organized the parameters into two categories: fixed and variable. The fixed parameters are very important when running the model, but serve as “backend” environment variables. Efforts should be focused on making reasonable assignments of the variable parameters, before changing the values of the fixed parameters. The variable parameters should be used as the primary parameters when running the model, and reflect both the fitness of the subject and the subject's phycological state. Future parameterization will allow us to have a better understanding of these parameters and of their respective domains.

Bibliography

Bownds, J., Ebersole, M., Lovelock, D., and O'Conner, D. "Reexamining The Search Management Function. Part I," RESPONSE, Winter, 1991, Vol.10, No. 1.

Cooper, D.C., J.R. Frost, and R. Quincy Robe, "Compatibility of Land SAR procedures with search theory," prepared for U.S. Dept. of Homeland Security, 2003

Koester, Robert J. *Lost Person Behavior: A Search and Rescue Guide on Where to Look for Land, Air, and Water.* Charlottesville, VA: DbS Productions, 2008. Print.

Koopman, Bernard O. The Theory of Search. I. Kinematic Bases. *Operations Research.* Vol. 4, No. 3 (Jun., 1956), pp. 324-346

Koopman, Bernard O. The Theory of Search. II. Target Detection. *Operations Research.* Vol. 4, No. 5 (Oct., 1956), pp. 503 - 531

Koopman, Bernard O. The Theory of Search. III. The Optimum Distribution of Searching Effort. *Operations Research.* Vol. 4, No. 5 (Oct., 1957), pp. 613 – 626

Mattson, R. J. "Establishing Search Priorities," Handout, 1975.

Naismith, W.W. Scottish Mountaineering Club Journal 2, 136. 1892

O'Connor, Daniel. Controversial Topics in Inland SAR Planning. NEWSAR, 2004.

Stone, Lawrence D. *Theory of Optimal Search*. Arlington, VA: Military Applications Section, Operations Research Society of America, 1975. Print.

Syrotuck, William G., and Jean Anne. Syrotuck. *Analysis of Lost Person Behavior: An Aid to Search Planning*. Westmoreland, NY: Arner Publications, 1977. Print.