Finding a Needle in the Haystack: Predicting the Location of Lost People Using Agent-Based Modeling and Behavioral Inertia

John Nguyen, Caroline Joseph, Bailey Richardson, Roy Hayes, Ricardo Pakula, and Robert Koester

Abstract— **Around 100,000 persons go missing annually in the US. Many factors go into predicting the location of a lost person, like geography, climate, age, health status, gender, disabilities, walking speed, and more. Technology and machine learning can advance the success and speed of search and rescue (SAR) missions. Agent-based modeling is a popular method for predicting the location of a lost person. A recently published paper by Hashimoto et al. demonstrated the ability to tune an agent-based model's parameters so that its emergent behavior statistically matches the lost hiker data in International Search and Rescue Database.**

Hashimoto et al.'s work assumed that a lost person randomly selects a reorienting behavior at every time step. We build upon the work performed by Hashimoto et al. by adding the concept of inertia as a parameter. We hypothesize that a lost person will likely continue their reorienting behavior for some time before changing. We used International Search and Rescue Incident Database SAR incidents with geolocation and Sava et al.'s scoring methodology to compare our inertia-enabled agent-based model with Hashimoto et al.'s model and validate our results. We find our model outperforms Hashimoto et al. within our testing set.

I. INTRODUCTION

When one becomes lost in the wilderness, survivability often relies on the possibility of being found by Search and Rescue (SAR) teams, especially in the earliest few hours. However, SAR operations are dangerous and resource intensive. Between 2004 and 2014 SAR operations within the U.S. National parks cost $51.4 billion [1].

Despite the time sensitive nature and expense of SAR operations, the tactics used in SAR operations have not evolved to take advantage of 21st-century machine learning tools. Currently, when establishing a search area, SAR commands use a variation of the following tactics:

- Last known position and expanding radius – where the radius expands based on the rate of movement in the terrain,
- Point of interest – based on previous SAR operations in the area,
- The Mattson Method – where individuals develop their own search area individually and combine them with the groups to create an expert weighted search grid.

The current SAR tactics results in large search areas but with limited resources and time, the SAR team can only cover a small area [2]. Therefore, several researchers have examined the use of machine learning to improve SAR operations. A challenge is that each lost person (LP) is unique, comprised of a variety of features, such as age, health status, disability, and experience, which affects how that person behaves.

Dr. Robert Koester, the creator of the International Search and Rescue Incident Database (ISRID), a database of thousands of SAR incidents containing specific details of the incident and the demographics of the people involved [3], has catalog the features into a taxonomy, enabling researchers to develop models for specific geographic areas and classes of people, such as adult hikers in mountainous, temperate terrain.

When a person is lost, defined as not knowing their current location and how to navigate to their desired destination, they will apply reorientation techniques, such as backtracking or seeking higher elevation to enhance their view [4]. Our research expands the recent work of Hashimoto et al., which used agent-based models to predict the location of a lost person (LP). Our model extends Hashimoto et al. by:

- Adding inertia to the agent's decision, defined as the agent's reluctance to change their reorientation behavior until some time has passed,
- Exploring new reorientation techniques that we named following-the-natural-drift and moving-towards-civilization,
- Adjust the movement speed of agents based on elevation and land cover.

The next section presents a literature review of previous work, followed by a methodology section. Our paper concludes with an experiment, result, and a conclusion section that presents ideas for future work.

II. LITERATURE REVIEW

In 2012, Jared Doke used data in Yosemite National Park to identify 'hotspots,' locations where people were more likely to become lost [5]. Additionally, he found that the distribution of people lost in Yosemite behaves differently than the ISRID dataset would predict, such as LPs moving downhill at greater rates. Doke's finding indicates that local terrain affects a lost person's behavior and that you cannot apply the aggregate statistics of the ISRID dataset without considering the terrain.

Dr. Koester also agrees, stating, “Search planners tend to prefer local data… They also worry that data from different locales may not be appropriate if the terrain and climate differ from their own…The critical question was how to classify land areas in a way that considered these features. Fortunately, a well-established system already existed that classified different types of terrain, vegetation, and climate [4].” Dr. Koester and Dokes illustrate the importance of considering local terrain in developing a prediction model for an LP.

In 2016, Sava et al. developed a method for directly comparing different machine learning prediction models [6]. They bound the search area and then divided it into equal grids. A prediction model then gives each grid cell a probability that the LP is within it, with p being the probability the model assigned to the cell in which the LP was found. The equations below provide a score that compares each model. The upper-case N is the total number of grid cells, and the lower-case n is the total number of grid cells that the model assigned a higher probability than p (i.e., the model believed it was more likely for the LP to be found in a different cell). The lower-case m is the number of cells with probability equal to p. R is the model score, with -1 being the worst and +1 being the best.

$$r = \frac{n+m/2}{N}, \tag{1}$$

$$R = \frac{0.5-r}{0.5}. \tag{2}$$

The score compares how many grid cells the SAR team would need to search before they found the individual. Sava et al. used their method to compare a model that only considered distance, one that only considered resistance to crossing a watershed, and one that considered both distance and resistance to crossing a watershed. The combined distance and watershed model performed best over 355 test cases.

Alanis et al., in 2019, used a Decision Tree Algorithm (DTA) to simulate a lost person’s behavior and predict their likely path [7]. Their DTA model uses a person’s physical and mental energy as two input variables, with both variables decreasing as time passes and the agent moves through the terrain. The simulated agent is given a goal destination to reach and is rewarded for moving closer to the goal. Their algorithm attempts to balance between moving towards the goal and minimizing the energy exerted.

Alanis et al. model can be thought of as an agent-based model, where an agent is given a set of characteristics (energy) and simple rules (reorientation strategies). The agent interacts with its environment based on these characteristics and rules, and its path emerges from these interactions. While Alanis et al. simulation matched their expectations, they did not test it on real-world incidents.

In a research paper published in 2021, Verdaguer applied machine learning in SAR operations to predict if a person would be found unharmed, injured, or dead [8]. He included features about the LP, such as their age, health status, and the activity they were doing when they became lost (i.e., hiking, biking, etc.). He trained and tested a decision tree classifier on previously collected data from ISRID and a local SAR team. He found that he could correctly predict the outcome of 60% of the cases. Additionally, he discovered that an LP’s characteristics, such as mental health, affect whether they will be found alive or dead. Verdaguer’s findings indicate that LP prediction models must account for individual characteristics.

In 2022 Hashimoto et al. published a paper demonstrating the ability to tune an agent-based model’s parameters so that its emergent behavior statistically matches the lost hiker data in ISRID. Hashimoto et al. first use ArcGIS, a geographic information system provider, to collect geospecific data around SAR incidents within the ISRID dataset. They translate the geographic information into a 2-dimensional grid, defining passable and inaccessible areas based on elevation changes and large bodies of water. The 2-dimensional grid represents the environment in their agent-based model.

Hashimoto et al. provide their agent with six reorientation strategies, initially found in ‘Lost Person Behavior’ [4] and used in Alanis et al. simulation [7]. Our implementation of the reorientation strategies is listed below. We limited the Random Walk strategy to altering the agent’s heading to $\pm 90°$, which our SAR experts believed was more realistic than Hashimoto et al.’s interpretation.

- Random Walk – the agent adds a number between $[-90°, 90°]$ to its existing heading, thereby altering its heading. It then applies its current velocity and moves to the next location.
- Route Traveling – The agent determines if a neighboring grid cell in the direction of its current heading (within 120°) contains a route (e.g., road or trail). If multiple grid cells contain a route, then the agent randomly selects a cell to move towards, and updates its heading. If no neighboring grid cell in the direction of its current heading contains a route, then the agent performs the Random Walk strategy.
- Direction Traveling – The agent maintains its current heading.
- Staying Put – The agent does not move during this time step.
- View Enhancing – The agent examines the eight neighboring cells, selects the cell with the highest elevation to move towards, and updates its heading.

If no neighboring cell has a higher elevation, then the agent performs the Staying Put strategy.

- Backtracking – The agent adds 180° to its current heading, causing the agent to move in the opposite direction.

An agent randomly chooses a reorientation strategy at each time step based on a probability mass function (PMF). Hashimoto et al. provide the following example, "an LP with a probability of [RW, RT, DT, SP, VE, BT] = [1/2, 0, 1/6, 1/3, 0, 0] has a 50% chance of random walking, a 17% chance of direction traveling, and a 33% chance of staying put at a given time step. Independent realizations of this distribution of behaviors are generated at each time step and the agent's position is updated by the randomly selected strategy."

Hashimoto et al. calibrate their agent-based model to predict hikers' location in 65 incidents. Using a brute-force technique, they iterate over all possible PMF, incrementing the PMF by 1/6 for each iteration. They determined that the average hiker spends 56% of their time performing direction traveling, 38% of their time performing route traveling, and 5% performing random walking. The remainder of their time is spent doing view-enhancing and staying put.

Our work extends the literature in several significant ways. Like Hashimoto et al., we develop an agent-based model and calibrate the reorientation strategy to match the ISRID dataset. However, we further restrict our model to adult hikers in mountainous, temperate environments. Dr. Koester and Doke [4, 5] demonstrated that to accurately predict an LP's location, the model must consider the LP's characteristics (e.g., adult hiker) and the environment. We hypothesize that we can develop a more accurate model by further restricting the dataset to a specific climate and terrain.

In our agent-based model, we also reduce the movement speed of an agent based on elevation changes and land cover factors. Additionally, we introduce a concept called inertia and examine two new reorientation strategies.

Lastly, rather than brute forcing all possible combinations, our team uses a swarm algorithm to search the input space, reducing the number of iterations required to converge to a solution. We compare our model using the Map Score technique developed by Sava et al. The following sections provide a detailed overview of our approach.

III. APPROACH

A. Lost Person Agent-Based Model

The research method presented in the paper builds off Hashimoto et al. Using ArcGIS, we retrieve land cover, roads, trails, and elevation data for incidents in the ISRID dataset involving an adult hiker lost in temperate, mountainous regions. The search area is limited to a 25 square kilometer around the incident, with the last known position of the hiker at the center of the box. We selected the search area to align with Sava et al.

The data is transformed into a 2-dimensional grid, and the agent is placed in the center of the grid, with an initial heading. The exact number of grid cells varies slightly, based on the exact location the incident occurred in the world, but each grid cell is approximately 10 square meters. The simulated agent is also given an initial velocity, which is also the maximum velocity of that agent. The initial velocity is calculated by equation 3 below, where one m/s is the maximum velocity of any agent, and the speed percentage is an input into the model. The simulation timestep is set so that the agent, moving at its initial velocity speed, can transition from one cell to another within the timestep; see equation 4 below.

$$Inital\ Velocity\ = 1 * speedPercentage \qquad (3)$$

$$Time\ Step = \frac{Cell\ Width\ (M)}{Inital\ Velocity\ (M/S)}. \qquad (4)$$

An agent can only exist in one grid cell at a time but unlike Hashimoto et al., we do not limit the agent to the center of the grid. At each time step the agent selects a reorientation strategy, based on its PMF. Our team examines the previously mentioned reorientation strategies.

Like Hashimoto et al. we use land cover, roads, and trails retrieved from ArcGIS [9]. Land cover is categorized into the nine categories listed below. Clouds are given as a land cover when clouds obscure the satellite imagery. The cloud land cover has no affect in our simulation.

Water
Crops
Snow/Ice
Trees
Built Area
Clouds
Flooded Vegetation
Bare Ground
Rangeland

The speed of the agent is affected based on land cover. The agent is not allowed to traverse water. If the next move of the agent places it in water, then the agent performs the ‘staying put’ strategy. Additionally, if an agent traverses a cell designated with the tree land cover, the agent’s speed is reduced by 30%. The 30% metric was determined through consultations with search and rescue experts, and future research will need to validate this assumption.

Our model also considers elevation change when determining how far an agent moves during each timestep. If the agent moves to a location with a higher elevation than its current location, the elevation change is considered when determining the agent’s new position. However, suppose the agent is moving over flat ground or moving downhill. In that case, the elevation change is not considered, with the assumption that the agent is getting assisted by gravity to move faster. One limitation of our work is that we do not create impassable areas based on steep terrains, such as cliff faces. Future work will consider how significant elevation changes affect the movement of an LP.

B. Agent-Based Model - Advancements

Hashimoto et al.'s model assumes that an LP randomly selects a reorientation strategy at each timestep. We hypothesize that most people will not select a new reorientation strategy approximately every 4.2 seconds [1]. However, there may exist a subset of the population, such as dementia patients, where that assumption is valid. Therefore, we develop the concept of inertia, which represents the reluctance of the LP to select a new reorientation strategy.

By adjusting inertia, we can simulate an LP that rapidly switches between reorientation strategies, such as a lost child or a mentally ill person. However, we can also simulate knowledgeable survivalists, such as expert hikers, who may be more deliberate in selecting their reorientation strategies and, therefore, less likely to switch their strategy.

We simulate the concept of inertia by assigning an agent a probability between 0 and 1 of reassessing their current strategy. At each timestep, the agent selects a random number, and if that number is below their inertia value, they will choose a reorientation strategy based on their PMF. For simplicity, we keep the inertia probability constant. Future research can examine applying a time-weighted discount factor to the inertia probability, where the agent becomes more likely to reassess their reorientation strategy over time.

Previous researchers find that 'route traveling' is the most common strategy employed by an LP [5]. Therefore, we hypothesize that it is unlikely for an LP to leave a route once they find one. To test this theory, we create a route interruption variable. Like inertia, an agent is assigned a probability between 0 to 1. Given that an agent is currently on a route and has passed its inertia check, it will select a random number. The agent can choose a reorientation strategy based on its PMF if that number is below the route interruption variable. The pseudo-code below illustrates the inertia and route interruption variable.

Pseudo Code: Reorientation Strategy Selection

```
If random(0,1) < Inertia:
    If Agent-On-Route == True:
        If random(0,1) < Route-Interrupt:
            Select-New-Strategy()
        Else: Continue-Prev-Strategy()
    Else: Select-New-Strategy()
Else: Continue-Prev-Strategy()
```

Our team also extends the reorientation strategies that are examined. Previous researchers found that many LPs are discovered at lower elevations than their initial starting point [3, 5], indicating that people follow the natural elevation contours downhill. We examine this hypothesis by developing the reorientation strategy called 'following-the-natural-drift.'

Our discussions with search and rescue experts indicated that an expert hiker might not know where they are but can often orient themselves in a cardinal direction, such as using the sun to determine east. Additionally, experts can often recall general directions to major landmarks or roads, such as knowing the highway is south. Therefore, the person is lost because they do not know where they are or how to navigate to their destination, but they can orient themselves in a direction they are most likely to find help. We defined this reorientation strategy as 'moving-towards-civilization.'

- Following-Natural-Drift – The agent examines the eight neighboring cells, selects the cell with the lowest elevation to move towards, and updates its heading. If no neighboring cell has a lower elevation, then the agent performs the Staying Put strategy.
- Moving-Towards-Civilization – The agent changes their heading and moves toward the closest manmade structure (e.g., road, built area).

C. Machine Learning Approach

Hashimoto et al. discovered the PMF that best matched the ISRID dataset by exploring all possible combinations of PMF. However, their approach does not scale, as additional reorientation strategies and other considerations like inertia area added into the simulation. The brute force method grows in exponential time complexity.

We implemented a particle swarm algorithm using the Python class Pyswarm, in Python 3.7. A swarm algorithm is a method for exploring a search space with the objective of identifying the parameter settings that maximizes or minimizes an objective function [10]. Our algorithm finds the parameters that minimizes the average distance between the simulated lost person position at the end of the simulation and where the real person was found.

IV. EXPERIMENT

The experiment aims to assess if considering climate and terrain improves the accuracy of agent-based modeling. Additionally, we examine if our concept of route interruption, inertia, and new reorientation strategies, improves the model's accuracy.

Our study replicates Hashimoto et al. We limit our dataset to ISRID incidents involving hikers that are lost in temperate, mountainous environments. Additionally, we required that the ISRID incidents have the GPS coordinates of the initial planning point (i.e., last known location of the person), where the person was found (i.e., find location), and how long the search was conducted. Lastly, we require that the LP in each incident be found between 1 and 12 KM from the initial planning point.

The ISRID dataset contains 171 incidents that match the previously mentioned criteria, but due to computational resource constraints, we further limit our dataset to a random 50 incidents, similar to Hashimoto et al.'s 65 incidents. Our computation resource constraints also led to an unorthodox decision to train on 10 incidents (20% of the dataset) and test on the remaining 40 incidents. Since our model has a long runtime, approximately 24 hours, we were interested in its applicability to a more extensive set of incidents.

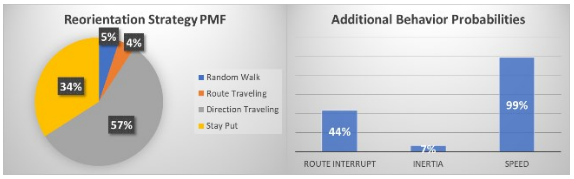
As mentioned previously, the search area is limited to 25 square kilometer around the incident, with the last known position of the hiker at the center of the box. The data is transformed into a 2-dimensional grid, and the agent is placed in the center of the grid, with an initial heading. The exact number of grid cells varies slightly, based on the exact location the incident occurred, but each grid cell is approximately 10 square meters.

We implemented a particle swarm algorithm in PySwarm. The swarm algorithm uses a global best optimization with 5 particles, 100 iterations, $C_1 = 0.5$, $C_2 = 0.3$, and $w = 0.9$. The variable C_1 gives the confidence a particle has in itself and C_2 provides the confidence that a particle has in its neighbor. The variable w is an inertia value. The reader can find a more detailed review of particle swarm optimization in [10]. The objective function of the swarm algorithm minimizes the average Euclidean distance error of the predicted find location and the actual find location across the 10 training incidents.

For each incident, we simulate 100 LPs that apply reorientation strategies, inertia, route interruption, and movement speed based on swarm algorithms input. The maximum movement speed of the simulated LPs is 1 m/s. The simulation ends once the incident's search time has expired.

V. RESULTS

Figure 1. LP Behavior Model - Known Direction Model



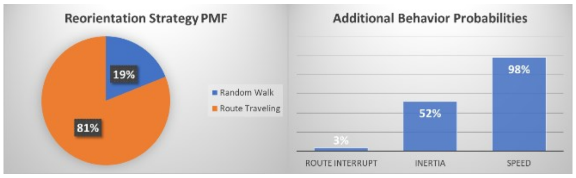
Our team first trained a model, using the original reorientation strategies from Hashimoto et al. During the training all simulated LPs are initialized with a heading facing that incident's find location. Therefore, assuming no rivers or lakes in the path, an agent that performs only direction traveling will eventually reach the find location. We call this model Known Direction, Fig 1. illustrates the weights that the swarm model converged to over 100 iterations.

We find like Hashimoto et al. that hikers do not perform 'backtracking' or 'view enhancing' strategies. Additionally, we find that our agent performs the 'staying put' strategy significantly more and the 'random walk' strategy significantly less than Hashimoto et al. However, we do converge to approximately the same 'direction traveling' percentage. We also find the probability of changing a strategy (i.e., inertia) is 7%, indicating a reluctance to frequent strategy changes.

Based on Fig 1.'s results we replaced the 'backtracking' and 'view enhancing' strategies because they were not used with our 'move-towards-civilization' and 'following-natural-drift' strategies. However, we converged to the same results as seen in Fig 1.

We trained a second model, where the simulated LPs were initialized with a random heading, portraying a scenario where the SAR team does not know the hiker's desired destination, or the hiker is completely disoriented. We called this model the Unknown Direction, Fig 2. illustrates the weight convergence.

Figure 2. LP Behavior Model - Unknown Direction Model



The model converged to weights that only used the 'route traveling' and 'random walk' strategy. Like before, we replaced 'backtracking' and 'view enhancing' with our hypothesized strategies and retrained the model. The model converged to the same result as Fig 2. Our findings indicate that hikers wander randomly until they find a road or trail and then travels down that route. Additionally, once on the route the hiker is unlikely to leave the route.

We next applied our two models (Known Direction and Unknown Direction) to 40 test cases. Using the Map Score described above, we compared our two models with Hashimoto et al.'s model, which we encoded into our simulator. For each test scenario, we simulate 100 LPs, initialized with random headings (i.e., hiker does not know the heading to reach their desired location).

To apply the Map Score methodology, we initialize each grid cell with a numeric count of zero. At the end of each simulated incident, we iterate through the final location of each simulated LP and increment the numeric count of each grid cells within a 1 KM radius of that LP by one. Afterward, we divided the numeric count in each grid cell by the sum of all the cells, creating a probability density map. The 1 KM radius is derived from [11], which recommends 1 KM as the maximum search grid size. Fig 3. and Fig 4. illustrates the results.

Figure 3 Map Score 95% Confidence Interval of Mean

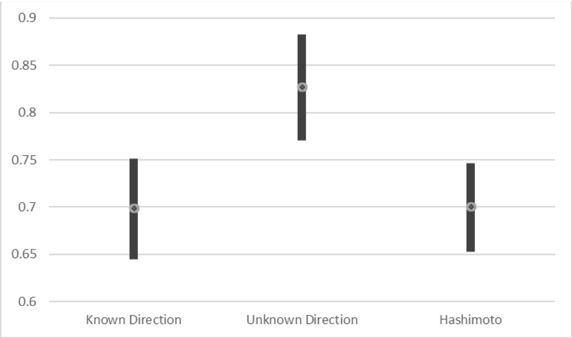


Figure 4: ISRID Incident and Prediction Heatmap

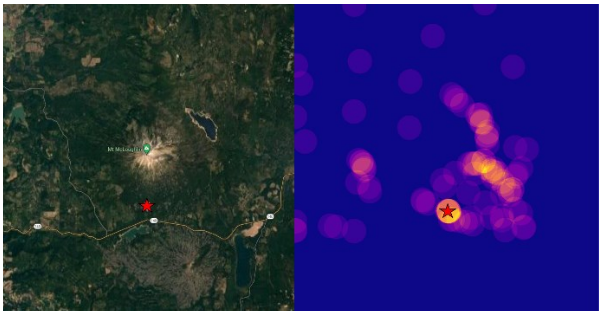


Fig 3. Shows our Unknown Direction model outperforms Hashimoto, achieving achieved a mean map score of 0.82 (95% CI: 0.77 – 0.88). Fig 4. Demonstrates an example of our Unknown Direction model prediction. The red star in both images indicate where the person was found. Our model correctly predicted that the lost person was likely located along the highway.

VI. Conclusion

Our Unknown Direction model is equivalent to Sava et al.'s published results and outperforms Hashimoto et al. in our test cases. Our study finds that hikers solely use the random walk and route traveling reorientation strategies when completely disoriented (i.e., hiker does not know the heading to reach their desired location). Furthermore, the route interruption, inertia, and PMF probability of our Unknown Direction model result in a 0.03% chance of leaving a route the hiker is traveling on, indicating that lost hikers are unlikely to deviate from a known path.

Our Known Direction model, which was trained with simulated LPs initialized facing the find location, matches Hashimoto et al. direction traveling rate. However, our agent moved significantly slower, with a maximum speed 0.5 m/s slower than Hashimoto et al., while staying stationary 37% of the time compared to their <1%. Additionally, our agent's inertia was 7%, indicating that the agent has a resistance to changing strategies.

We hypothesize the difference in findings arises because we limit our study to hikers in temperate, mountainous environments, where Hashimoto et al. examined incidents across the United States. Our Map Score performance and speed discrepancy indicate that local climate and terrain affect the accuracy of a model.

Our study illustrates that using agent-based models and swarm algorithms is a viable approach to train an LP location predictor. Our model is trained using 10 incidents and tested on 40 holdout cases. We achieved a Map Score higher than Hashimoto et al. and equivalent to the published results of Sava et al. Future work will decrease the simulation runtime by implementing just-in-time compilation. Improving our simulation efficiency will enable us to perform machine learning over a larger dataset and potentially achieve greater accuracy.

Future work will also seek to improve accuracy by incorporating additional features such as watershed information. We will also integrate the behavior models into LandSAR, an Air Force Research Lab Search and Rescue tool.

VII. REFERNCES

[1] A. Hashimoto, L. Heintzman, R. Koester, and N. Abaid, "An agent-based model reveals lost person behavior based on data from wilderness search and rescue," *Scientific reports*, vol. 12, no.1, p. 5873, April 2022.

[2] L. Lin and M. Goodrich, "A Bayesian approach to modeling lost person behaviors based on terrain features in Wilderness Search and Rescue," *Computational Mathematical Organization Theory*, vol. 16, p. 300–323, 2010. https://doi.org/10.1007/s10588-010-9066-2

[3] R. Koester, "International Search & Rescue Incident Database (ISRID)," *SAR Research*. Retrieved February 10, 2023, from https://www.dbs-sar.com/SAR_Research/ISRID.htm

[4] R. Koester, *Lost Person Behavior: A search and rescue guide*. dbS Productions, 2008.

[5] J. Doke, "Analysis of search incidents and lost person behavior in Yosemite National Park," PhD dissertation, University of Kansas, 2012.

[6] E. Sava, C. Twardy, R. Koester, and M. Sonwalkar, "Evaluating lost person behavior models," *Transactions in GIS*, P. 38, 2016.

[7] J. Alanis, M. Brown, J. Kitchens, J. Magana, C. Velastegui, M. Thakura, et. al, "Topography and behavior based movement modeling for missing hikers in land-wilderness settings," *Mathematical Theoretical Biology Institute,* p. 1-53, 2019.

[8] E. Verdaguer, "Applications of machine learning in decision analysis for rescuing lost people," PhD dissertation, ETSI_Informatica, 2021.

[9] K. Krishna, C. Kontgis, Z. Statman-Weil, J. Mazzariello, M. Mathis, and S. Brumby, "Global land use/land cover with Sentinel 2 and deep learning," In *2021 IEEE international geoscience and remote sensing symposium IGARSS*, pp. 4704-4707. IEEE, 2021.

[10] J. Blondin, Particle swarm optimization: A tutorial. *Availaible from: http://cs. armstrong. edu/saad/csci8100/pso tutorial. pdf, 34*. 2009.

[11] Emergency ManagementBC, "Ground Search & Rescue", British Columbia Emergency Management, 1999.