Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnline CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 56-2 (2023) 1466–1471

GIS Data Driven Probability Map Generation for Search and Rescue Using Agents

Jan-Hendrik Ewers * **David Anderson** * **Douglas Thomson** *

* *University of Glasgow, Glasgow, Scotland, G12 8QQ*

Abstract: Predicting the final resting location of a missing person is critical for search and rescue operations with limited resources. To improve the accuracy and speed of these predictions, simulated agents can be created to replicate the behavior of the missing person. In this paper, we introduce an agent-based model, to simulate various psychological profiles, that move over a physical landscape incorporating real-world data in their decision-making without relying on per-location training. The resultant probability density map of the missing person's location was the result of a combination of Monte Carlo simulations and mobility-time-based sampling. General trends in the data were comparable to historical data sets available. This work presents a flexible agent that can be employed by search and rescue that easily extends to various locations.

Copyright © 2023 The Authors. This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/)

Keywords: search and rescue, probability map generation, agent-based, data-driven, prediction

1. INTRODUCTION

Search and Rescue (SAR) of vulnerable missing persons is unfortunately a common task for the Police and other emergency services. Organizations like the Centre for Search and Rescue (Perkins et al. (2011)) and the Grampian Police (Grampian Police (2007)) carry out research and training in areas related to SAR and while their training and published papers offer a valuable resource to the people responsible for finding lost or missing persons, they focus only on land-based search i.e. directing teams of individuals. This is a slow and methodical process that would undoubtedly benefit from the assistance of an airborne surveillance platform. As a result of rapid advances in the drone sector over the last decade, multi-rotors have become cheaper and more accessible than ever before. Consequently, several concept evaluation trials for measuring the efficacy of incorporating drones into the search process have recently been undertaken in Scotland (Skeleton and Smith (2020)).

The mission profile for UAVs in a SAR setting is solely for search. This requires finding the lost person (LP) as quickly as possible. The search further breaks down into prediction, flying, and sensing. Flying and sensing are done in real-time (Brown and Anderson (2020)), but the prediction of where the LP could be can be done en route to the scene. After generating the Probability Distribution Map (PDM), a trajectory can be created for a UAV to search the problem space whilst accumulating a maximum probability over the trajectory (Ewers et al. (2022)). This implies increasing the probability of finding an LP faster and therefore saving lives.

In order to adapt the PDM to the LP, historical data has been ordered into significant categories by the likes of Koester (2008) and Perkins et al. (2011). Both agree that significant behavioral profiles exist in collected data from historical SAR cases. This means that a solo hiker behaves differently from an elderly dementia patient when it comes down to where they were found. By using this a priori information along with other location data, these PDMs can be highly customized on a per-location and per-person basis resulting in better PDMs (Ewers et al. (2022))

To generate the PDM, Heintzman et al. (2021) implements a 6-strategy model, where each strategy emulates a self-rescue technique. These are then randomly selected at every time step based on a learned weighting vector. This was then further developed by Hashimoto et al. (2022) using a leave-one-out analysis to further improve the accuracy of the resultant behavior vector compared to the real-world data from Koester (2008). However, this vector has to be learned for every new location as results differed substantially between trials.

In this paper, our research hypothesis is that by using the same data as Hashimoto et al., as well as the location found data, a more generalized model can be created that doesn't need training. We use the find location data from Perkins et al. (2011) to drive four different behaviors, each with the intent of arriving at a unique target location. We then compare the artificially generated data to historical data to analyze the results.

This paper's methodology is discussed in section Sect. 2, results are presented and discussed in section Sect. 3. Finally, section Sect. 4 concludes this paper and discusses future work.

2. METHOD

The PDM is generated by running an LP surrogate in a Monte Carlo manner to generate large amounts of path data. These paths are then used to infer the resting location of the LP using historical data. However, at its

2405-8963 Copyright © 2023 The Authors. This is an open access article under the CC BY-NC-ND license.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2023.10.1834

core, every simulation run is a behavior traversing the landscape from a starting point until it reaches $D_{\max}$ steps. The starting points are sampled from the multivariate gaussian probability density function

$$PDF_{start}(\boldsymbol{p}) = \frac{\exp\left[-\frac{1}{2}(\boldsymbol{p}-\boldsymbol{\mu})^T\boldsymbol{\sigma}^{-1}(\boldsymbol{p}-\boldsymbol{\mu})\right]}{\sqrt{(2\pi)^k \det \boldsymbol{\sigma}}} \quad (1)$$

centred around a set location, $\boldsymbol{\mu} \in \mathbb{N}^2$, values of which are defined later in Sect. 3, with covariance $\boldsymbol{\sigma}$ to simulate the uncertainty of the PLS.

2.1 Agent Dynamics

The LP is modeled as a 2D cellular automaton on a $x-y$ plane with rectangular grids of shape $n\text{m} \times m\text{m}$. Its viewpoint sits at a constant 1.6m above the surface, allowing it to see over smaller obstacles like rocks or long grass. Furthermore, the agent does not tire during path generation but rather attempts to finish as close as possible to the given maximum distance $D_{\max}$. At each time step, the total distance traveled is incremented by either D (immediately adjacent cells) or $\sqrt{2}D$ (corner cells) depending on which direction is chosen by the behavior (i.e. the control loop). The simulation loop finishes when $D > D_{\max}$, and $D_{\max}$ is typically a larger number, such as $10000D$. As outlined below in Sect. 2.3, the path generated by the agent is then sampled via a distribution scaled by the average preferred walking speed of a hiker over rough terrain; $3.87\text{km}\,\text{h}^{-1}$ Gast et al. (2019).

2.2 Lost Person Behaviour

In order to model different physiological profiles, location-found and travel-time data sets were used. By changing which profile's data is applied throughout the simulation the various parameters (age, activity, etc.) can be modeled. This data will be primarily sourced from Perkins et al. (2011) with missing data being taken from Koester (2008). One of the largest datasets reported by both is *Hiker (solo)* which is the one used in this paper.

In their work, Perkins et al. found that when analyzing the location found there was no statistical significance between genders but still reported them separately. As this study concerns itself with *Hiker (solo)* as a whole, the genders will be merged for this dataset accordingly. Tbl. 1 shows the location found by significant groupings with a sample size of $n = 130$.

Table 1. Location found data for a *Hiker (solo)* from Perkins et al. (2011)

Location found	n	%
Open Ground	53	40.8
Travel Aid	33	25.4
Building	30	23.1
Linear Feature	9	6.9
Trees	4	3.1
Water	1	0.8

From the data in Tbl. 1, an appropriate behavior can be selected $P\%$ of the time that attempts to seek out the desired final location. *Building*, *Trees* and *Water* were given individual behaviors, but *linear feature* and *travel aid* were merged together. Furthermore, Hyp. 2.1 (with later empirical evaluation) was employed to generalize the *open-ground* find data.

Hypothesis 2.1. The open-ground find location data naturally results from an LP trying to navigate to other possible locations.

As the datasets available are highly generalized and don't provide nuances in the LP's behavior, it is also assumed that extraordinary circumstances are probabilistically insignificant. Scenarios such as the agent accidentally falling into the river whilst following it are ignored. Adding such complexity to the model is outside of the scope of this paper, and will be explored in future work.

Head To Water The logic behind each behavior is also highly dependent on the available GIS data at hand. Through data from Morris and Flavin (1990), the *water* behavior has a lot of unique information to work with. Therefore, a simple vector field follower using the water outflow direction may be sufficient. To increase the fidelity of the model, the cumulative catchment area and the water surface type can further be used to force the agent to navigate around large enough bodies of water. This was done by evaluating the next step of the agent, and if it was of types *lake*, *sea*, or *river* then the agent evaluated a scaled probability given by

$$p(v) = \begin{cases} a, & v \le a \\ (v-a)\frac{d-c}{b-a} + c, & else \\ b, & v \ge b \end{cases} \quad (2)$$

where a and b are the lower- and upper-bounds of the input respectively, and $c = 0, d = 0$ are the lower- and upper-bounds of the output respectively. v is the cumulative catchment area value in the next time step, and thus $p(v)$ is the scaled percentage chance for the agent to step into the body of water. If the check failed, then the agent would walk around the obstacle. In a real-world scenario, this equates to a hiker stepping over a small stream or encountering a lake and walking along its edges.

Head to Buildings and Head to Trees Similarly, using the land cover ID and digital elevation maps, the agent could *see* the landscape and act accordingly. Using a viewshed algorithm means that expensive ray-casting can be mitigated. Viewshed algorithms are popular in GIS, for applications such as radio tower positioning. The resultant map is a mask that outlines which areas within a given radius are visible to the observer. From this mask, any given map can be easily analyzed using an element-wise AND operation. The visible cells can then be analyzed further. To create a behavior around this masked area, every possible value is given a weight and the maximum weights are the only ones considered in the map which allows multiple map values to be treated as equals. Then, the mean angle to each visible cell, $\theta(\bar{p})$ is selected as the command direction as seen below in Eqn. 3.

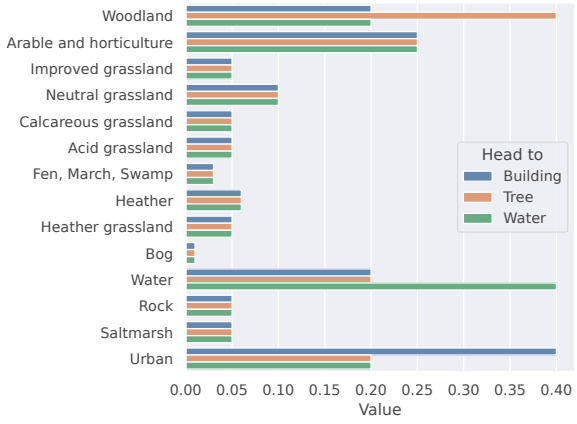


Fig. 1. Viewshed behavior weights used in J1 for the Building and Tree behavior with data from GeoPackage geospatial data (2021). Water was included to show how this behavior could be done without the data from Morris and Flavin (1990). Note how favorable destinations (like Urban) were always left high to encourage the agent to head even if the most desirable weight was not visible.

$$\theta(\bar{p}) = \frac{1}{N}\sum_{i=0}^{N} \tan^{-1}\left(\frac{p_y - c_{i,y}}{p_x - c_{i,x}}\right) \tag{3}$$

where p is the current position, c is the cell's position, and N is the number of visible cells.

The weights from Fig. 1 for each map were chosen by hand, with logic dictating that an LP will avoid a bog, whilst slightly preferring to walk over neutral grassland than heather.

Head to Roads The final behavior to consider is *Road* (which includes paths, trails, etc.) which was split into two behaviors. The first uses a vector field (like the water behavior) which was created to show the direction to the nearest road, from Ordnance Survey (2021), at every pixel k such that

$$\cos\theta_k = \frac{\boldsymbol{p}_{E,k,min} \cdot \boldsymbol{p}_k}{|\boldsymbol{p}_{E,k,min}||\boldsymbol{p}_k|} \tag{4}$$

where $\boldsymbol{p}_{E,k,min}$ is the closest point on any road. Once the agent encounters a road, it switches to the next behavior which treats the road network as a graph, where the roads are edges and junctions are nodes. From there it randomly traverses the network which encourages backtracking.

2.3 Generation of PDM from Path

The simulation length was capped at $10,000$ steps of 5m which is equivalent to 12.82h and accounts for around 95% of LP scenarios.

To fit the mobility time distribution, multiple other functions were evaluated: exponential, log-gamma, and normal. However, by using the symmetrical Kullback-Leibler divergence function (Kullback and Leibler (1951))

$$\begin{aligned} KL(A||B) &= \sum_{\gamma\in\chi} A(\gamma)\log\left(\frac{A(\gamma)}{B(\gamma)}\right) \\ SKL(A||B) &= KL(A||B) + KL(A||B) \end{aligned} \tag{5}$$

as a metric to compare the similarity between A and B over the sample space χ, it was found that the log-normal function was the best fitting. Where the log-normal is defined as

$$\begin{aligned} f(y) &= \frac{1}{s\cdot y\cdot\lambda\sqrt{2\pi}}\exp\left(-\frac{\log^2(y)}{2s^2}\right) \\ y &= \frac{x-\mu}{\lambda} \end{aligned} \tag{6}$$

s is the shape parameter, μ is the mean, and λ is the scaling parameter.

Ideally, every final location would be the result of a dedicated simulation (i.e. one run for every point). However, with run times per path exceeding 30s, this isn't feasible. Therefore to transform the N generated paths from Sect. 2.2 to find locations, Eqn. 6 was treated as a cumulative distribution function and sampled M times. Every sample $T_{S,i}$ represents the time an LP has been traveling for, and using a constant $4.87\text{h}\,\text{km}^{-1}$ (Gast et al. (2019)), this can be converted to distance. This is then used to calculate a point along the paths resulting in $N\cdot M$ locations.

To select a minimum sample size M of statistically significant, Eqn. 5 was used to determine the performance of $M \in \mathbb{N}$ for the resultant PDM. Fitting a logarithmic regression to the output of evaluating M_i and M_{i-1} gave a best fit of

$$\begin{aligned} f_{score}(M) &= -584.28\log(M) + 3582.43 \\ f'_{score}(M) &= \frac{-584.288}{M} \end{aligned} \tag{7}$$

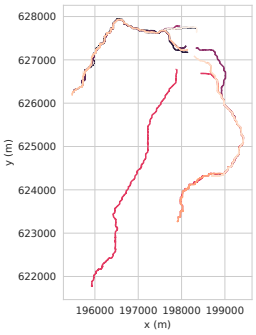
Solving for $f'_{score}(M) < 1$ (a 1×10^{-4} difference per cell) of the discrete PDM gives $M > 584$. Throughout the results, this condition was met with a final M of 3402 and a theoretical $f'_{score} = 0.17$.

3. RESULTS AND DISCUSSION

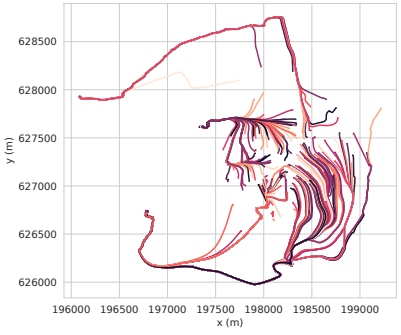
A key element in deciding which location to analyze was the availability of complete digital elevation map coverage. The LiDAR survey undertaken by ASC geospatial data (2014) has 100% of the Isle of Arran, Scotland, and therefore this location was chosen. An added benefit of using a rural island as the test area is the natural boundary that the sea provides, preventing the agent from going out of bounds. Furthermore, the two locations selected in Tbl. 2 gave a varied environment for the agent to be tested in. Tighvein is on top of a hill at around 500m and situated close to human presence whereas Glen Iorsa has a limited line-of-sight as it sits within a valley.

3.1 Paths

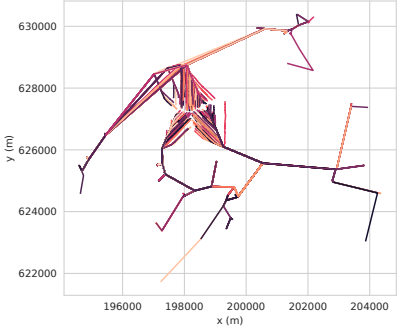
Fig. 2 shows a random sample of the simulation runs for the Glen Iorsa location. This figure clearly shows



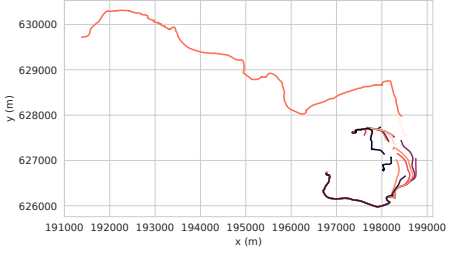
(a) Head to water



(b) Head to buildings



(c) Head to roads



(d) Head to trees

Fig. 2. Paths traversed by agents using the 4 different behaviors starting at Glen Iorsa with a bivariate gaussian covariance of $\boldsymbol{\sigma} = [100000, 100000], \rho = 0$. Color was used for contrast between paths.

Table 2. Location 1 is in the south of the Isle of Arran closer to woodland, and agriculture, whilst location 2 is in the north of the island deep in the hills and glens.

#	Location Name	x	y
1	Tighvein	198235	627232
2	Glen Iorsa	194435	642918

the different behaviors and the varied starting locations. Head to water (Fig. 2a), shows a set of paths without much variance. This is to be expected as the watershed of the area is followed which, by definition, follows the slope downhill and merges into larger streams of water. Ultimately, the agent would reach the coast if the number of steps had been sufficiently high. The head-to-roads behavior, seen in Fig. 2c, shows similar behavior in its initial stage with the agent being guided towards the nearest road. However, this behavior has a clear problem in ignoring the line-of-sight and treating the area as a flat plane where everything is perfectly visible. The next stage of head-to-roads is also very constricted, as it perfectly follows the roads in the network. Head to buildings and head to trees (Fig. 2b and Fig. 2d respectively) show a problem with the viewshed-based behaviors, in that once the agent arrives at a desired location it is very unlikely to further explore the terrain. This is more clearly seen in Fig. 3, where both the aforementioned behaviors' paths are, in general, much shorter than from the non-viewshed algorithms. An interesting result of the viewshed weights shown in Fig. 1, is that even though the head-to-trees behavior will prioritize trees if none are visible it will still head to the next most highly weighted locations which is the case for paths finishing around $(196700, 626650)$ (Fig. 2b and Fig. 2a).

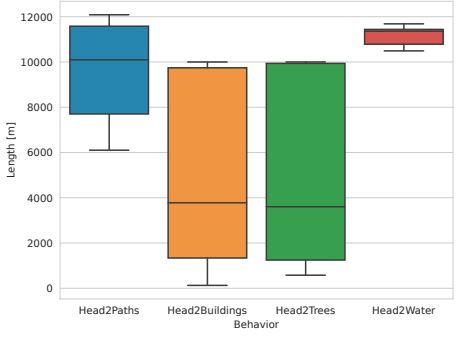
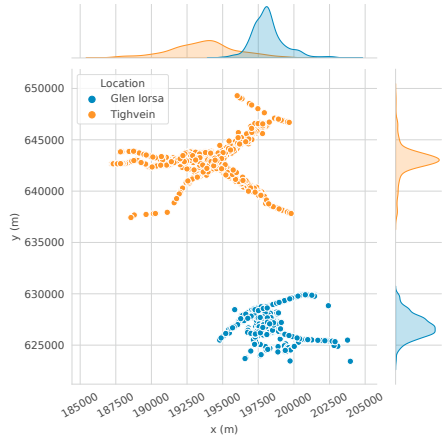


Fig. 3. Both viewshed-based algorithms are seen to have a low median around, 4.00km, whilst the vector map algorithms are closer to 11.00km

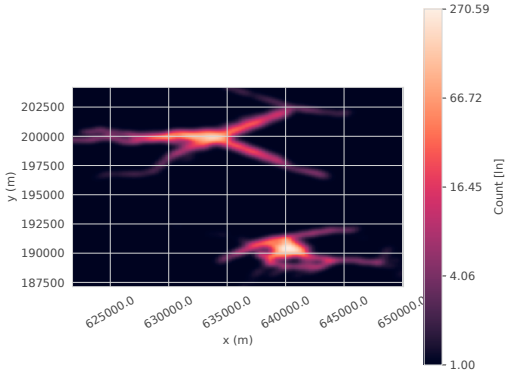
Fig. 4 shows the sampled points for Glen Iorsa and Tighvein. The distribution of the sampled points along the y and x axis can be seen to center around the PLS as expected in Fig. 4a. However, Fig. 4b shows that location 2 has several paths *burnt* into the heatmap, with the same problem being much less severe for location 2.

3.2 Location Found

The resultant sampled locations of the agents is the most important metric to gauge the effectiveness of this model.



(a) Locations found by simulation run (2 for each location), showing the distribution of points in the x-y axis



(b) Logarithmic PDM of all locations clearly showing the start points for both locations, around which the highest probability of finding an LP is seen

Fig. 4. Paths traversed by agents using the different behaviors with $\boldsymbol{\sigma} = [100000, 100000], \rho = 0$

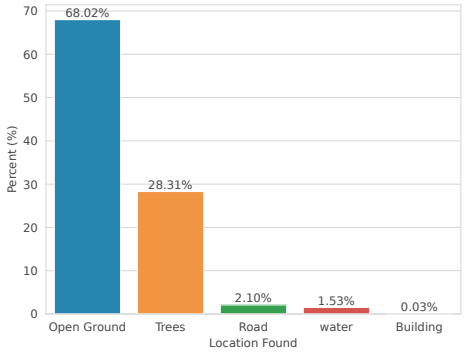


Fig. 5. Classified sample points by location found using similar metrics to Perkins et al. (2011). The road network was given a 5m buffer which was subtracted from the land cover ID vector for the other categories

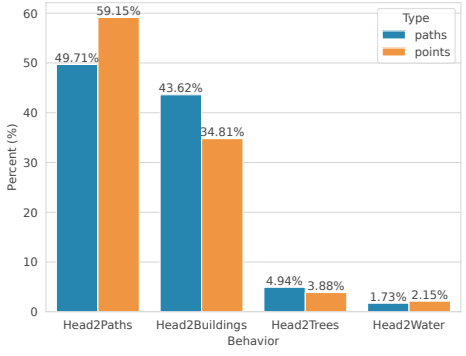


Fig. 6. Behaviors distribution of original paths and sampled points showing the filtering in action

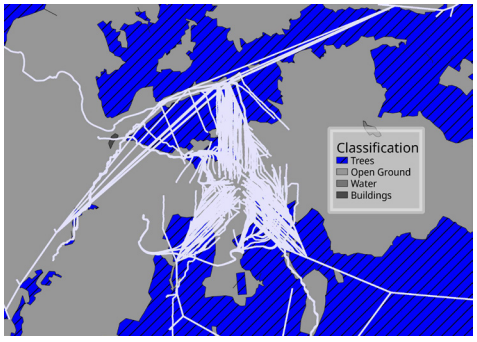


Fig. 7. Paths passing through trees classification area

As can be seen from Fig. 5, the *open ground* category is 26.4% larger, when compared to Tbl. 1, than desired at the cost of the other classifications. Even though this is a significant difference from the original data, this does empirically prove Hyp. 2.1 in that no dedicated *open ground* behavior is required for LPs to naturally end up in that classification.

A further large discrepancy is in the *building* with 0.03% rather than the expected 23.1%. Conversely, *trees* is 26.03% over target even though the main behavior (Head2Tree) is also purely viewshed based. One source of this error could be originating from the manner that the sampling is being done. Fig. 6 shows that even though 43.5% of paths are due to the Head2Buildings behavior, only 34.6% of samples were as a result of the aforementioned paths. This is due to Head2Buildings having a low mean 6.09km length compared to the largest of 11.20km as is seen in Fig. 3.

Another reason for this discrepancy is that 25.11% of the Isle of Arran is covered in trees which results in the other behaviors also passing through trees and as such, samples are being classified as finishing within a *trees* area. This is more clearly seen in Fig. 7.

3.3 Sampling Effectiveness

The short path distances from early termination of behaviors is a problem, as touched on in Sect. 3.2, resulting in samples not being taken at distances longer than the path in question. Data from Perkins et al. (2011) states that 90% of hikers traveled for a further 2.45h, with all coming to rest within 131h. This final 10% means that a substantial amount of samples were ignored.

Fig. 8 shows the times at which a point was sampled is skewed heavily to the left, with the original distribution being more spread out along the time axis. The sampled

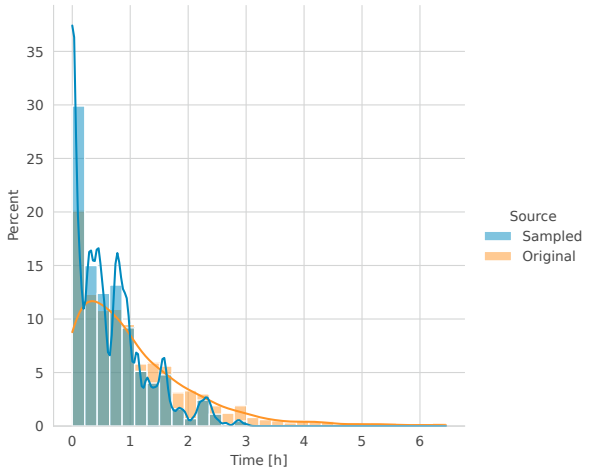


Fig. 8. Time distribution of the time used for the sampled points (blue) and the original distribution (orange) showing the filtering in action

points have a mean time of 0.64h and a standard deviation of 0.61, conversely, the original distribution has a mean of 1.04h with a standard deviation of 1.02. Another indicator of this is that 72.89% of points sampled from the mobility distribution were ultimately used to sample from a path.

4. CONCLUSION

This paper explored the creation of a model emulating the movement of an LP over a landscape based on their physiological profile. By characterizing the profile into four distinct behaviors (head-to-water, head-to-trees, head-to-roads, and head-to-buildings), the model can be adjusted to match the location-found description from local datasets. These distinct, simpler, behaviors are essentially their model and traverse the landscape as if they were an LP with a single goal in mind. It was hypothesized that running each behavior a percentage amount of times, from the dataset, would produce a resulting distribution of find locations that match the aforementioned dataset. However, in practice, this was not the case as certain find location classifications were much more abundant. For example, more than a quarter of the simulated LPs were found in an area marked as trees even though only 3.82% of simulations were run with the *Head2Trees* behavior. This is a direct result of the assumption that the location found and behavior are directly related, whereas it is more realistic that the find location is related to all behaviors used. This is primarily due to paths crossing over areas that are not the goal (such as the agent aiming for buildings but having to cross open ground). Similarly, the hypothesis that the open-ground find location from the datasets is a direct result of LPs traversing the landscape is true. However, future work should also consider that other find locations will also experience this phenomenon.

Whilst the classified find locations are not the same as that from Perkins et al., the general trend of more LPs being found in open-ground than the others is correct. Likewise, the trend of the water find locations being the smallest is also accurate. However, the purely viewshed-based behavior (head-to-buildings) is incorrect by orders of magnitude and the approach to the algorithm needs reworking. It is possible that running this algorithm with more start locations would smooth out the data.

Future work will further explore extending the capabilities of the various behaviors. Following this, more GIS data will be incorporated to further establish the agent's roots in the real world. Replacing behaviors with machine learning models is also something that will need to be explored. Furthermore, the usage of PDMs to create UAV search trajectories will be explored (Ewers et al. (2022)).

REFERENCES

ASC geospatial data (2014). Lidar Composite Digital Terrain Model Scotland (Phase2) 1m resolution.

Brown, A. and Anderson, D. (2020). Trajectory optimization for high-altitude long-endurance UAV maritime radar surveillance. *IEEE Transactions on Aerospace and Electronic Systems*, 56(3), 2406–2421. doi:10.1109/TAES.2019.2949384.

Ewers, J.H., Anderson, D., and Thomson, D. (2022). Optimal Path Planning using Psychological Profiling in Drone-Assisted Missing Person Search (in press). *Advanced Control for Applications: Engineering and Industrial Systems*.

Gast, K., Kram, R., and Riemer, R. (2019). Preferred walking speed on rough terrain; is it all about energetics? *Journal of Experimental Biology*, jeb.185447. doi: 10.1242/jeb.185447.

GeoPackage geospatial data (2021). Land Cover Map 2020.

Grampian Police (2007). Missing persons: Understanding, planning, responding. Technical report, Grampian Police, 4 Burnside St, Rothes, Aberlour AB38 7AA.

Hashimoto, A., Heintzman, L., Koester, R., and Abaid, N. (2022). An agent-based model reveals lost person behavior based on data from wilderness search and rescue. *Scientific Reports*, 12(1), 5873. doi:10.1038/s41598-022-09502-4.

Heintzman, L., Hashimoto, A., Abaid, N., and Williams, R.K. (2021). Anticipatory Planning and Dynamic Lost Person Models for Human-Robot Search and Rescue. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 8252–8258. doi: 10.1109/ICRA48506.2021.9562070.

Koester, R.J. (2008). *Lost Person Behavior: A Search and Rescue Guide on Where to Look for Land, Air, and Water*. dbS Productions, Charlottesville, VA.

Kullback, S. and Leibler, R.A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79–86. doi:10.1214/aoms/1177729694.

Morris, DG. and Flavin, RW. (1990). A digital terrain model for hydrology. In *Proc 4th International Symposium on Spatial Data Handling*, volume 1, 250–262. Zurich.

Ordnance Survey (2021). OS MasterMap Highways Network®.

Perkins, D., Roberts, P., Feeney, G., and Mrt, P. (2011). The U.K. Missing Person Behaviour Study. *CSR*.

Skeleton, L. and Smith, C. (2020). Remotely Piloted Aircraft System (RPAS) Evaluation Report. Technical report, Scottish Police Authority, SPA Policing Performance Committee.