

1 Modeling

We will represent the fish’s location by Cartesian coordinates in the plane of the surface of the water, with the origin set to some latitude / longitude coordinate O . Latitude / longitude coordinates of the fish measured during the track are mapped to this plane. Let $r[t]$ denote the x -coordinate (distance east of O) and y -coordinate (distance north of O) of the fish at time $t > 0$.

1.1 Interpolation and Weighting

Since the fish’s location cannot be determined at regular intervals, the times for which $r[t]$ is defined also fail to occur at regular intervals. Such an uneven sampling is inconvenient for modeling the fish’s location. Interpolation is used to generate an approximation $R[t]$ of the fish’s location, defined for any $t = kT_{\text{sample}}$, where k is a nonnegative integer. For convenience, we may use the notation $R[k] = R[kT_{\text{sample}}]$ to index the fish’s location by integer k .

This summer, we used the MATLAB function `interp1` to generate $R[t]$ by linear interpolation. The real fish data we used had a median interval between samples of almost exactly two minutes, so we chose $T_{\text{sample}} = 2$ mins as our constant.

1.2 Modeling Individuals vs. Population

After interpolation, there are two possible paths for further analysis. Both paths have the primary goal of a transition model that will place robots as close to the target fish as possible. The *individual* approach creates a transition model in which the probability of the transition $i \rightarrow j$ at time t approximates the probability of a real fish making the transition $i \rightarrow j$ at time t , based on analysis of the data $R[k]$ and $R[k + 1]$. However, the *population* approach creates a transition model in which the density of robots at node i and time t approximates the density of fish at node i and time t , based on analysis of the same data. The individual approach makes individual robots behave as individual fish, while the population approach makes the population of robots behave as the population of fish.

1.3 Individual Model

Once the fish’s location $R[k]$ is discretized to the N nodes of the graph, we may represent the fish’s movement between each time step as $d_i[k]$, defined recursively as follows:

$$d_i[k] = \begin{cases} R[k + 1] & R[k] = i \\ d_i[k - 1] & R[k] \neq i \text{ and } k > 0 \\ i & R[k] \neq i \text{ and } k = 0 \end{cases}$$

Informally, $d_i[k]$ represents the node that the fish will occupy at time step $k + 1$ if it’s location at time step k is node i . Of course, we only observe the fish’s transition from one source node per time step. If $R[k] = i$ and $R[k + 1] = j$, then we may easily choose $d_i[k] = j$; however, we are left to guess where the fish *would* have gone from some other node $x \neq i$ to choose the value of $d_x[k]$. The above definition makes the guess that, if $R[n + 1] = y$ for the largest $n < k$ at which $R[n] = x$, then $d_x[k] = y$. An alternative definition of $d_i[k]$ may make the guess that $d_i[k] = R[k + 1]$ for any i , though such a definition may overemphasize the current location of the fish, at the expense of losing information about the fish’s path.

The next step is to convert $d_i[k]$ into a function $T[k]$, which yields the $N \times N$ transition matrix for time step k . This summer, we used a “windowed transition counting model” to extract transition probabilities from $d_i[k]$. Under this model, we define $T[k]_{i,j}$ as follows:

$$T[k]_{i,j} = \sum_m^L f(m|k, \sigma/T_{\text{sample}}) \langle d_j[m] = i \rangle,$$

where L is the largest value such that $d_i[L]$ is defined, $f(x|\mu, \sigma)$ is a discrete normal distribution (defined for integers 0 to $L - 1$ with a total sum of 1), and $\langle P \rangle$ is 1 if the proposition P is true and 0 otherwise. Note that, by this definition, every column in $T[k]$ has a sum of 1, and every element of $T[k]$ is in the range $[0, 1]$.

1.4 Population Models

Let $\vec{R}[k] = (x_k, y_k)$ be a vector denoting the fish's location at time step k , and let $\vec{V}_i = (x_i, y_i)$ denote the location of node i . From these quantities we may define an *attenuation vector* $\mathbf{a}[k]$ as follows:

$$\mathbf{a}_i[k] = e^{-\|\vec{R}[k] - \vec{V}_i[k]\|^2 / (2\sigma_{\text{atten}}^2)}$$

where σ_{atten} is a free parameter. Informally, $\mathbf{a}[k]$ describes the weight of a particular node in the fish's location. If the fish's location at some k precisely equals the location of node i , then $\mathbf{a}_i[k] = 1$; but as the fish moves away from i , then $\mathbf{a}_i[k] \rightarrow 0$. The parameter σ_{atten} sets the “attenuation” of the a node's assigned weight as the fish moves away from it.

We then define a density vector $\mathbf{p}[k]$ as a weighted sum of the attenuation vector:

$$\mathbf{p}[k] = \sum_m^L f(m|k, \sigma/T_{\text{sample}}) \mathbf{a}[m].$$

Ideally, we would like to construct a transition matrix $T[k]$ such that $\mathbf{p}[k+1] = T[k]\mathbf{p}[k]$. In some cases, however, such a construction may not be possible. Instead, we use Matlab's `lsqlin` function to find a $T[k]$ that minimizes $\|\mathbf{p}[k+1] - T[k]\mathbf{p}[k]\|^2$, subject to the constraint that $T[k]$ is a stochastic matrix.