18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009
http://mssanz.org.au/modsim09

Detecting general patterns in fish movement from the analysis of fish tagging data

Daphney Dagneaux, Fabio Boschetti, Russ Babcock and Mat Vanderklift

Abstract: We describe a method to extract patterns of local fish movement from acoustic tagging data in small scale surveys covering distances in the order of tens of kms. The purpose of our analysis is two-fold: first, to provide a general insight into fish movement which allows us to approximately evaluate the efficiency of current local protected areas; second, to provide a statistical description of fish movement as input to ecological numerical models which can be used to evaluate alternative design for sanctuary zones.

We analysed 3 months of data from 41 Spangled Emperors (Lethrinus nebulosus) in the Ningaloo Marine Park, in Western Australia. The data was collected via 38 acoustic receivers spread over an area of approx 180 km2 covering roughly 5 different habitats typical of shallow water, reef environments.

The frequency of fish detection from the acoustic receivers has a strong seasonal and daily pattern with detections decreasing at day-time and from December to February. Our initial analysis suggests that this is not the result of biological factors, rather of environmental factors (mostly waves due to sea breeze) and that the pattern of fish movement is mostly uniform during the sampled period.

Using wind data from a nearby station we devised an approximate rule relating the effective variability in the receivers' detection range with wind speed, which in turns is transformed into uncertainty on the fish position as a function of time. This has been further processed to generate a) the statistics of fish transitioning from one receiver position to another and b) a probability distribution of likely fish movement in un-sampled areas (outside the detection ranges)

With this information we produced 2 types of maps. One map represents the probability of occurrence of Spangled Emperors at a given location in the studied area given only the measurements and their relative uncertainty. This can be considered as the information most constrained by the data. A second map represents the probability of occurrence assuming all fish use movement patterns extracted from the data, that is they act as agents trained on the measured data set. This can be considered as data extrapolation and is least constrained by the data. 'Real' fish movement, given the data at hand, may be expected to lie somewhere in between these two representations.

By assigning receivers' location to different habitats, from these two maps we extracted two probability distributions of fish movement from habitat to habitat. First results suggest that Spangled Emperors are mostly sedentary, which should facilitate the design of protected areas.

Finally, this analysis has highlighted several possible improvements in data collection, receiver positioning, data analysis and algorithm design which are also discussed and which may lead to improved future surveys.

Keywords: *Agent based modeling, animal behavior, statistics, animal tagging*

1. ANALYSIS OF DETECTION RECORDS

In this section we analyse the spatial and temporal distribution of the tag detections with the purpose of establishing a rough approximation of the uncertainty in fish position as a function of time. The survey area is shown in Figure 1**,** together with the delineation of different habitats.

In Figure 2 we can see the number of tag detections per hour over the survey period; we can notice that:

- the average number of detections decrease from December to February;
- hourly detections seem to follow an approximate quasi-periodic daily pattern.

The first question we want to address is whether these two trends are due to environmental or biological factors, that is, whether they are the result of specific features in the fish behaviour or are due to external factors.

We first note that the data in Figure 2 refers only to the fish which have been detected over the entire survey period, that is, the decrease in detections going from December to February is not due to loss of individual fish. Next, in Figure 3a, we show the histogram of the total number of tag detections at different hour of the day, over the overall survey period. We can clearly see that most tag detections occur at night and that the minimum of detection occurs in the afternoon.

An obvious question arising from these two figures is whether certain fish may behave differently in the afternoon compared to night time, for example certain individual fish may prefer to hide or migrate outside the studied area in the hottest hours of the day. To verify this we checked the number of individual fish which are detected at different hours of the day, over the overall survey period. This information is plotted in Figure 3b; this is similar to Figure 3a, except that for each hour we plot the total number of individual fish detected rather that the total number of tag detections. As we can see, Figure 3b shows much less variability than Figure 3a, that is the decrease in tag detections does not seem to be due to the decrease in the number of individual fish detected. This suggests (although does not prove) that the variability in tag detections during the day may be related to environmental factors rather than fish behaviour.

A likely candidate may be the sea breeze: the resulting water turbulence can generate background noise as well as air bubbles (Pincock, 2006), which may affect the detection range of the acoustic sensors. This hypothesis may be further corroborated by Figure 4; it shows the number of tag detections per hour at different locations in the survey area: the sharp decrease in detections is evident everywhere except that in the channel area, that is in the area characterised by least amount of reef obstacles and deeper water, which may result in less turbulence and air bubbles.

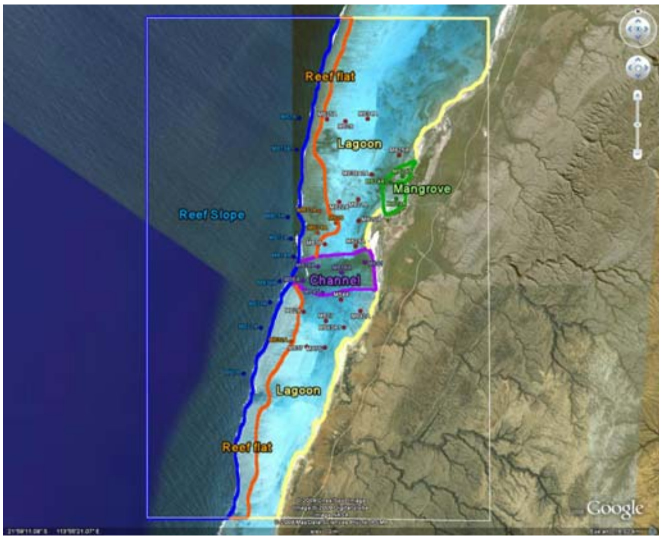


Figure 1. Visualization of the delimitation of the habitats. Blue: Reef slope; Orange: Reef flat; Green: Mangrove area; Purple: Channel; Pale yellow: Lagoon.

This inverse indirect relation between air temperature and tag detections (via sea breeze and water turbulence) may explain not only the short term daily variability but also the longer term decrease in detections between December and February. Other factors may also be at play however, including the growth of algae on the receiver sensor, or the batteries losing power.

The relation between wind speed from a near-by station and tag detections has been carried out, suggesting a correlation between the frequencies of the two signals (not shown). However, it is clear that the relation between the two signals is complex, that other factors (wind direction, tidal waves for example) may play a role and that sophisticated statistical techniques are needed to establish a more reliable relation between the two variables. In the absence of further information, in this work we assume that wind speed is the main factor controlling tag detections.

Background noise and air bubbles affect the number of tag detections by changing the detection range of the acoustic sensors, that is the size of the area around the sensor in which it is likely that a signal emitted by a tag is received and decoded correctly. Usually the detection range is known only under ideal, low background noise conditions and in our study can be assumed to be 300 m. Since we have a record of the average number of detections per hour, we can estimate an approximate relation between the wind speed and the expected detection range as follows:

$$DetectionRange(h)^2 = \frac{MaxDetectionRange^2 \cdot DetectionNumber(h)}{MaxDetectionNumber} \qquad \text{Eq 1}$$

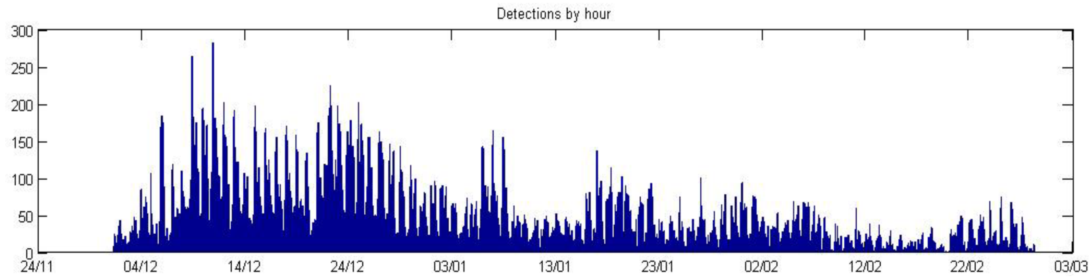
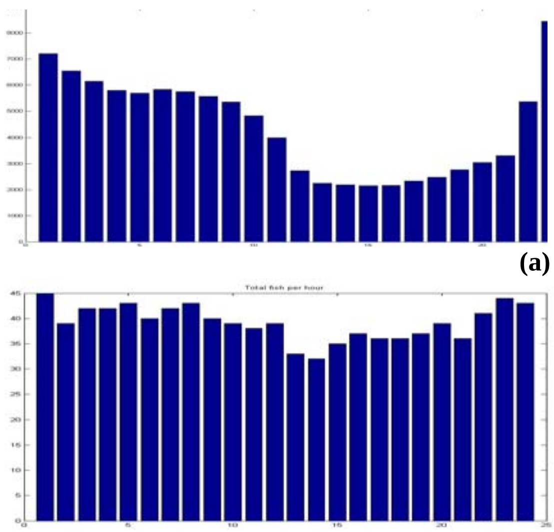


Figure 2. Number of tag detection per hour during the survey period. We notice the average decrease in detection number in February compared to December as well as a strong quasi-period daily variability

where, DetectionRange(h) is the expected detection range at time h, MaxDetectionRange is 300 m in our study, MaxDetectionNumber is the maximum number of hourly tag detections we measured (which we assumed happened under the ideal conditions of 300m detection range) and DetectionNumber(h) is the number of tag detections at time h. Needless to say, this relation can be considered only as an approximation, but it has the benefit to circumvent the need to account for the actual wind speed in the studied area, which may not be available; it should be considered merely as a statistical relation which assumes that the number of expected tag detections is only a function of the detection range around the acoustic sensor.

The purpose of Equation 1 is to provide an approximate bound of the uncertainty in fish position: when a tag is detected by an acoustic sensor, the fish may be anywhere within the detection range of the sensor. Under ideal conditions, a tag is more likely to be detected, but the area where the fish may be located is larger; under noisy conditions, it is less likely we detect a tag, but if we do, the fish position is known more precisely since the detection range is smaller. This information will be used to model the actual fish position and movement in the following sections.



(a)

(b)

Figure 3. (top) Total number of detections for each hour of a day, for the overall period by all receivers. (bottom) Total number of fish detected by all receivers, for each hour of a day, during the overall period .

2. DATA DRIVEN FISH MOVEMENT STATISTICS

In the previous section we obtained a general overview of the data at hand, their trends, the likely factors affecting their quality and quantity and their likely uncertainty. In this section we extract some simple statistics of fish movement. The aim is twofold: first, we would like to estimate the likely position of each tagged fish at any given time; second, we would like to use the information on fish movement to train 'virtual fish' and simulate their behaviour in the studied area.

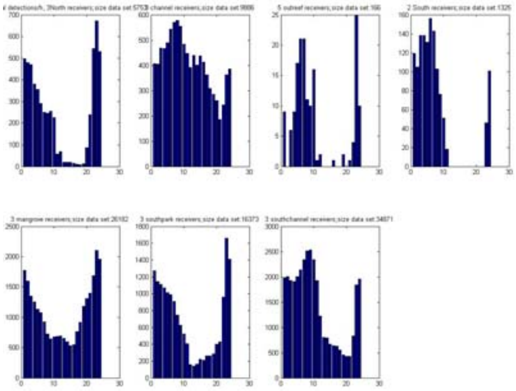


Figure 4. Number of detections by some receivers for each hour of a day, at 7 different locations, for the overall survey period.

As we have seen the frequency of tag detection varies considerably during the survey period. Consequently we decide to discretise the analysis in time over one-hour intervals; this is justified by the fact that the average interval between tag detections over the entire survey period is roughly 30 minutes. In particular, at each hour, a fish position is assigned to the acoustic sensors where the fish has been most often detected; had the fish tag never been detected at that specific hour, the fish position is taken as unknown and assigned stochastically based on the previous and next hourly detection (see below).

Once the data has been discretised, we generated three kinds of statistics for each fish:

1. the transition probability from the detection range of one acoustic sensor to the detection range of another sensor; we store all hours at which a fish tag has been detected at a particular sensor (sensor A, say) and for which we also have a detection at the next hour (that is the position of the fish at the next hour is known). We then count how at many times the tag has been detected in the next hour at each surveyed sensor; by normalising this count we obtain a transition probability, for each fish, of going from sensor A to any other sensor. We then repeat this procedure for each sensor.

2. The probability of 'losing track' a fish after seeing it at one acoustic sensor; for each fish, this is given by the counts of how many times a fish tag is not detected at any sensor, after seeing the fish at a specific sensor (sensor A, say) the previous hour. By dividing this number by the number of times a fish tag is seen at sensor A we obtain the probability of 'losing track' of a fish at sensor A. This procedure is also then repeated for each sensor.

3. The likely movement of a fish when it is not within any sensor's range. Obviously we have no direct measurement of this behaviour. We thus sample all detected movements (that is all transitions sampled at point 1), both in terms of length and direction of movement and we generate a probability distribution. A fish movement outside any sensor's range is then picked randomly within this distribution. This distribution inherently contains information about the likely fish speed. In order to make the movement statistics habitat dependent, a different probability distribution is generated for each habitat so that we allow fish to move differently depending on where they are within the studied area.

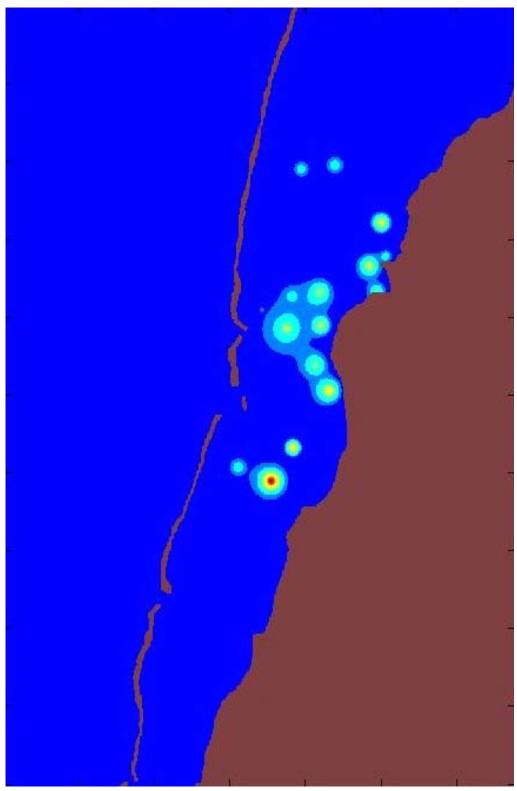


Figure 5. Probability distribution of fish position arising from tag detections and accounting for location uncertainty and missing detections.

3. DATA ANALYSIS

In this section we analyse the distribution of likely fish positions which we obtain from the tag detections. There are two reasons why this can not be obtained by a straightforward plot of the detection data: first, once a fish is detected by a sensor, the fish can be anywhere within the sensor's detection range; second; a fish may not be detected by any sensor for a long interval of time, often spanning several hours, especially during the hottest hours of the day.

The first problem is addressed by spreading the probability of finding a fish within a sensor's detection range. In our case we used a 2D Fermi's function (Reif, 1965) which reaches its maximum around the sensor location and decreases smoothly to zero at the border of the detention range; by suitable choice of parameters the decrease to zeros can be made as smooth as desired. The 2D integral of this function is then normalised to one to be converted to a probability distribution which effectively represents the constraints we have on the fish position, given a detection of its tag. The width if this probability distribution is given by the detection range of the acoustic sensor, which changes as a function of time as described above, to respect the uncertainty on the fish position due to the effect of the sea breeze.

Table **1**. Likely time spent by a fish within the sanctuary it has been released in

Alternative Sanctuary design	Permanence
1	.52
2	.35
3	.51
4	.35
5	.64
6	.40
7	.63
8	.46

When a fish tag is not detected at a given hour, the fish position is constrained by a) when and where the fish was seen last, b) when and where the fish is seen next and c) the fish speed. Constraints a) and b) are addressed as described in the previous paragraph. Constraint c) tells us how far a fish could possibly be by both a) and b) at a given time.

The resulting likely position of a fish at a given hour is then obtained by suitable iterative convolution of the likely position of the fish at the previous and next hour via a kernel represented by a Fermi function with width equal to the likely fish speed. The results can be seen in Figure 5.

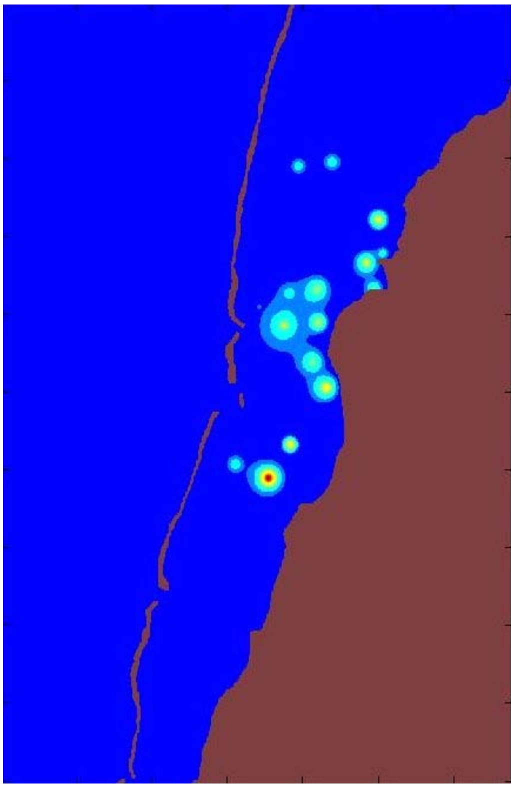


Figure 6. Probability distribution of fish position arising from tag detections and accounting for location uncertainty and missing detections.

4. AGENT-BASED SIMULATION OF VIRTUAL FISH

In the previous section we presented a probability distribution of fish presence at each location in the surveyed area which we obtained from the tag detections accounting for uncertainty in the fish position. This distribution is strongly affected by the location of the acoustic sensors: a fish could be seen at a location far from a sensor only if it could have transitioned through it in its path from one sensor to another. In principle, there is no reason why a fish may not be equally likely to be present at a location far from an acoustic sensor.

The purpose of the simulation we describe in this section is to account for this. Rather than using the tag detections as we did in the previous section, here we use the statistics described in Section 3 in order to 'train' the movement of a set of virtual fish. We then 'release' the virtual fish in the studied area and let them swim for a period of time equivalent to the survey period and store the locations they occupy at each hour. This is then turned into a probability distribution similar to the one presented in Figure 5. In particular the following steps are followed:

1. a fish is released at a random location within the studied area;

2. movements occur hourly;
3. if the fish happens to be within a receiver range, its next move is determined by the transition probability from the detection range of one acoustic sensor to the detection range of another sensor, as described at steps 1 and 2 in Section 3;
4. if the fish happens to be outside the detection range of any sensor then a random movement is chosen according to the probability distribution of movement described at step 3 in Section 3; the movement depends on the habitat the fish is in, as described above.
5. in order to account for uncertainty in the exact position of the fish, the actual location of the fish is convolved with the Fermi filter described above;
6. at the end of the simulation the location occupied by the fish are normalised in order to produced a probability distribution.

Table **2**. Probability of fish moving between habitats

Habitat	Lagoon	Reef slope	Mangrove Bay	Reef flat	Channel
Lagoon	0.8221	0.0285	0.0641	0.0545	0.0307
Reef slope	0.3664	0.4796	0.0300	0.0813	0.0163
Mangrove Bay	0.7369	0.0036	0.1197	0.0374	0.0135
Reef flat	0.9276	0.0499	0.0902	0.1450	0.0343
Channel	0.9246	0.0804	0.1075	0.1106	0.0578

An example of simulation output, for 15000 fish, can be seen in Figure 6. Fish are now allowed to occupy locations where no acoustic sensor is located as it can be seen by the image. This image should be interpreted as an extrapolation of the sampled data to un-sampled locations, constrained by the data statistics.

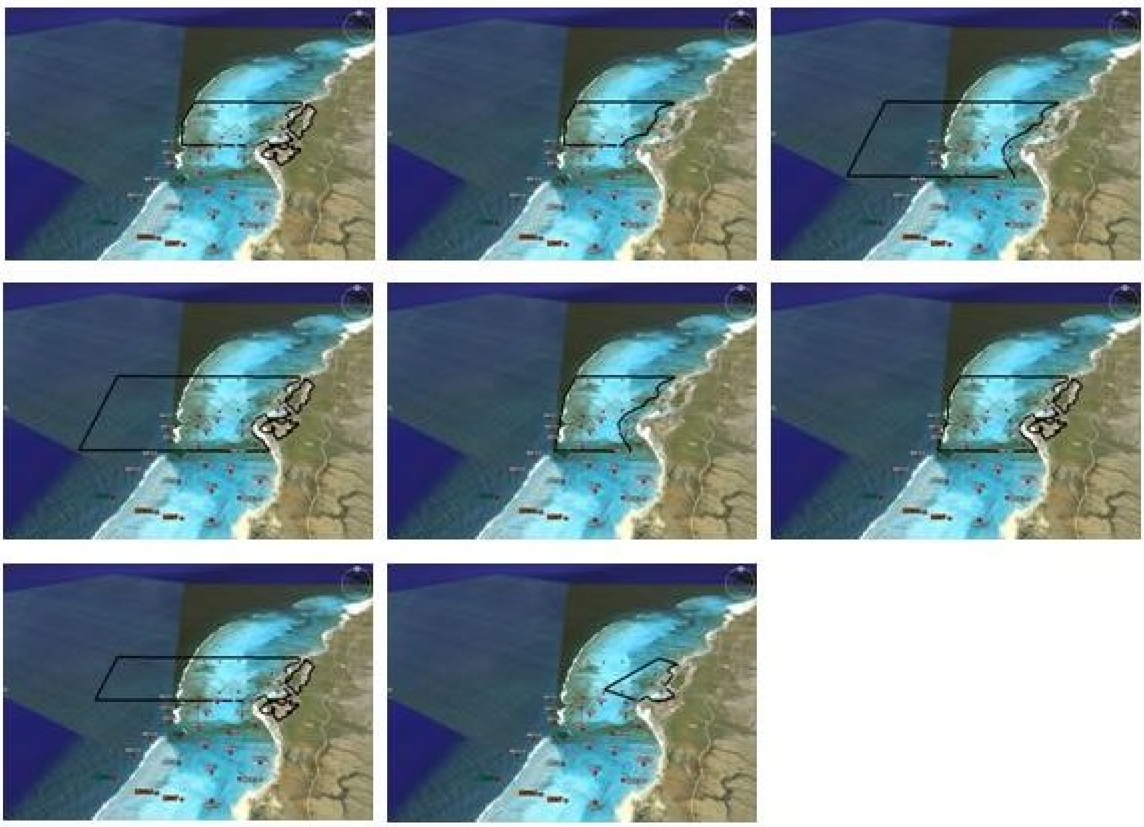


Figure 7. Alternative sanctuary designs analysed by evaluating the probability of fish released in the sanctuary to swim outside of it in a given time.

5. SUITABILITY OF SANCTUARY ZONE DESIGN AND TRANSITION BETWEEN HABITATS

The simulation technique described in the previous section offers a natural way to evaluate the suitability of alternative sanctuary zone designs. Given a sanctuary zone, we 'release' a number of 'virtual' fish within its boundaries and let them 'swim' for a given time. At the end of the simulation we calculate the amount of time these fish have spent inside and outside the sanctuary zone. A good sanctuary zone design is one which, everything else being the same, maximises the permanence of fish within it, consequently protecting the fish from fishing which occurs outside.

In Figure 7, nine alternative sanctuary zone designs are shown. For each we calculated the amount of time a fish released in a sanctuary will likely spend inside the same sanctuary in a 3 month period and the results are presented in Table 1.

Similarly, we can calculate the transition probability between different habitats, as shown in Figure 1, by releasing 'virtual' fish in one habitat and measuring the amount of time they spend in each different habitat in a given period. This is presented in Table 2.

6. DISCUSSION AND CONCLUSIONS

We have described an attempt to employ acoustic tagging data to extract information about fish movement. Unlike other studies, in this work the survey area was very small and specific information about short range movement was sought. The crucial components of our approach include a) estimating the detection range of the acoustic sensors as a function of environmental factors like the sea breeze, b) associating the time-varying detection range with the uncertainty on fish position, c) constraining the fish position when detections are missing by its likely swimming speed and the previous and next location of the detection, d) constraining the fish movement by storing occurrences of different possible movements and e) simulating the likely movement of 'virtual' fish by allowing it to move according to the sampled statistic in a Markov-like mode. Inevitably, this approach employs a number of algorithms and several heuristics, each of which could be changed affecting the outcome and more testing will be necessary to evaluate its effectiveness.

We are aware of several limitations in our analysis. The most important is that we would expect the probability distribution of 'virtual' fish position in Figure 6 to be somehow smoother and less centered on the actual location of the sensors. We suspect the main reason for this result lies in an incorrect estimation of the likely detection range of the acoustic sensors and we plan to use additional data, as well as specifically designed sampling recently carried out, to improve on this crucial variable. Also, the results presented in this work relate to 3 months of data. Additional data has recently been collected and will soon be included in the analysis in order to further evaluate the method.

REFERENCES

Pincock, D. (2006). "Understanding the Performance of Vemco 69 kHz Single Frequency Acoustic Telemetry," *http://www.vemco.com/education/range.php*, ISSN.

Reif, F. (1965). "Fundamentals of Statistical and Thermal Physics." McGraw-Hill, p. 341.