Estimating the distribution of demersal fishing effort from VMS data using hidden Markov models

D.L. Borchers*

Centre for Research into Ecological and Environmental Modelling, The Observatory, Buchanan Gardens, University of St Andrews, Fife, KY16 9LZ, Scotland

and

D.G. Reid

FRS Marine Laboratory Aberdeen, PO Box 101, Victoria Rd. Aberdeen, AB11 9DB, Scotland

October 22, 2008

Summary.

Keywords: Vessel Monitoring System, hidden Markov model, Fishing effort, Mixture model

1. Introduction

This paper sets out to utilise the Vessel Monitoring System (VMS) data that has recently been made available to the Fisheries Research Services. The advantage of VMS is that it covers most of the fleet (above 12.5m), and

*email: dlb@mcs.st-and.ac.uk
throughout their fishing activity, and is available since 2003. The research question we address here is how to predict vessel state (e.g. steaming, scouting and fishing - the last being of most interest) from VMS data. The ultimate purpose is to partition catch and effort at higher spatial resolution than has been possible to date. Standard landings reports provide landings, days at sea and occasionally effort (hours fished). Location of fishing is at the level of ICES rectangle, catch will often cover more than one fishing operation. As a result, logbooks can only give a broad brush impression of catch and effort distribution. VMS provides position at relatively high frequency (every two hours, sometimes every hour), along with speed and direction. Combined with survey data from the North Sea bottom trawl surveys the VMS data should be able to provide effort and catch distributions (maps) at high spatial resolution. These maps will allow us to evaluate the effectiveness of real time closures and other measures and to understand where the effort and catch are displaced to from such closed areas.

In this paper we develop hidden Markov model (HMM) methods to predict where vessels were fishing from VMS data. HMMs are useful for drawing inferences about the state of a system in discrete time when (a) there is Markov dependence in the evolution of the state over time (i.e. the probability of being in any given state now depends on the state in the previous time period) and (b) the state of interest cannot be observed itself but some random variable whose distribution differs between states can be observed. The random variable can be thought of as a "noisy" manifestation of the state, contaminated by "noise" from other states.

The primary state of interest in the context of fisheries management using
VMS data is the state “fishing”. The VMS data do not contain data on this state itself but they do contain observations of variables which are related to whether or not the vessel is fishing. Vessel speed is the primary such variable: demersal fishing is associated with slower vessel speeds than steaming to and from fishing grounds. We might therefore expect that vessel speed would be a good predictor of fishing/non-fishing state.

2. The Data and exploratory analyses

The vessel monitoring system (VMS) data consist of vessel position (latitude and longitude), speed, and course at approximately hourly intervals (about 40% of the observations) or two-hourly intervals (about 60% of the observations) for a selection of 37 vessel-trips in 2007. The sample sizes for each vessel are shown in Figure 1.

[Figure 1 about here.]

The vessel tracks while collecting VMS data are shown in Figure 2

[Figure 2 about here.]

We hope to use vessel speed (and possibly course) to draw inferences about vessel state. One can consider the observed distribution of vessel speeds to have arisen from a mixture of probability distributions, each arising from a specific vessel state. For speed to be informative about fishing state without additional information on state, it must be possible to decompose the mixture distribution into its components. This will be difficult or impossible if different states generate the same (or very similar) speed distributions. We
would therefore like the distribution of vessel speeds to show evidence of being a mixture of two or more distributions. Fortunately this is the case.

It can be seen from Figure 3 that the distribution of vessel speeds has more than one mode. The left-hand plot has a mode in the first speed interval. The majority of these are speeds of zero occurring at the start of vessels’ VMS data sequences and likely corresponding to times when the VMS was in operation but the vessel was effectively inoperative (in port, for example). If speeds are left-truncated to remove these observations, the distribution has two clear modes.

Mixtures of two, three and four normal probability densities were fitted to the left-truncated speed data by maximum likelihood. The likelihood for a normal mixture can be written as

$$L(\theta) = \prod_{i=1}^{n} \sum_{m=1}^{M} \phi_m f_m(x_i)$$

where $f_m(x_i)$ is a normal probability density function (pdf) with mean $\mu_m$ and variance $\sigma_m^2$ ($m = 1, \ldots, M$) (in our case this pdf is left-truncated at 0.2), $M$ is the number of components in the mixture, $\phi_1, \ldots, \phi_M$ are the mixture weights, constrained so that $\sum_{m=1}^{M} \phi_m = 1$, and the vector of parameters to be estimated is $\theta = (\phi_1, \ldots, \phi_{M-1}, \mu_1, \ldots, \mu_M, \sigma_1^2, \ldots, \sigma_M^2)$.

Convergence problems occurred with the 4-component mixture model. The 3-component model provided a better fit and lower Akaike Information Criterion (AIC) value than the 2-component model (12,033 vs 12,426). This
suggests that a model with three underlying (unobserved) states might be appropriate for these data. We expand on this in the next section. Fits of the 2- and 3-component models are shown in Figures 4 and 5.

[Figure 4 about here.]

[Figure 5 about here.]

It proved impossible to fit mixture models separately to each vessel-trip because of the small sample size of some vessel-trips and because not all vessel-trips contained both fast and slow vessel speeds.¹

The distribution of vessel speeds in space suggests that speeds in the vicinity of the mean of the second mixture component (2.7 knots) in the case of the 3-component mixture, or the mean of the first component (2.5 knots) in the case of the two-component mixture, correspond to a state of fishing. (Although the state itself is not observable, segments of vessel trackline spent in transit clearly have higher speeds and segments spent intensively in a smaller area correspond to slower speeds.) See Figure 6.

[Figure 6 about here.]

It is possible, or even likely, that other observable data also contain information about whether or not the vessel was fishing at any given time. The

¹Until January 2009, histograms of individual vessel-trip speeds overlaid with the mixture model obtained from fitting to all vessels can be found in the zipped file “HMMpredictions.zip” at http://www.creem.st-and.ac.uk/dlb/temp/temp.html. Use the password “supplementary” to unzip the files.
other bit of data we have from the VMS (aside from location) is the vessel’s course. While it is most unlikely that the course itself is informative, it seems possible that vessels may change course more frequently than otherwise while fishing. However, exploratory analysis of the course data suggest that they contain no information about fishing not available in the speed data. In particular, while it seems clear that certain speeds are associated with fishing, course change seems independent of speed. This can be see from Figure 7, which contains scatterplots of differences in course at lags of between 1 and 9 time intervals plotted against speed. Course change was not considered further as a predictor of vessel fishing state.

[Figure 7 about here.]

Inferences about fishing state are obviously improved if some data with known fishing states are available. While a sample of known-state data are available they have not yet been used because there are some data validation issues which remain to be resolved. The problem is illustrated in Figure 8, which shows the locations of hauls from a sample of vessels with discards observers on board. The fact that some hauls occur on land indicates that at least some of these records are in error. These data will be used once the validation issues have been resolved. (Meanwhile, methods for drawing inference about fishing locations are developed below.)

[Figure 8 about here.]

3. The hidden Markov model

Vessel state is modelled by allowing vessels to be in one of $M$ states; the state at time $t$ is denoted $S_t$. Given this state, the probability density function
(pdf) of an observable random variable (speed in our case) which we denote \( X_t \), is \( f_{X_t|S_t}(X_t|S_t) \). Here \( f_{X_t|S_t}(X_t|S_t) \) is assumed to be a normal probability density function with mean \( \mu_{S_t} \) and variance \( \sigma^2_{S_t} \), left-truncated at 0.2. The (unobserved) vessel state \( S_t \) takes on values \( 1, \ldots, M \) according to a Markov process with \( M \times M \) transition matrix

\[
\Gamma = \begin{pmatrix}
\gamma_{1,1} & \cdots & \gamma_{1,M} \\
\vdots & \ddots & \vdots \\
\gamma_{M,1} & \cdots & \gamma_{M,M}
\end{pmatrix}
\]  

(2)

where \( \gamma_{j,k} \) is the transition probability from state \( j \) to state \( k \) \((j, k \in \{1, \ldots, M\})\).

Exploratory analyses described above suggest that \( M = 2 \) or \( M = 3 \) would be appropriate for our VMS data. The key difference between the simple normal mixture models described above and a HMM with normal \( f_{X_t|S_t}(X_t|S_t) \) is that while both comprise mixtures of normal pdfs, the HMM has an underlying Markov model for the evolution of vessel states that imposes temporal correlation on states and accommodates “runs” of each state in a way which a simple mixture model which assumes independence of the \( X_t \)s cannot.

Writing the stationary distribution of the Markov chain defined by \( \Gamma \) as \( \pi(\gamma) = (\pi_1(\gamma), \ldots, \pi_M(\gamma)) \) and the vector of the \( M \times (M - 1) \) transition probabilities in \( \Gamma \) to be estimated\(^2\) as \( \gamma \), the likelihood for \( \gamma \) and \( \theta \) for a

\(^2\)The constraint that \( \sum_{k=1}^M \gamma_{j,k} = 1 \) means that one \( \gamma_{j,k} \) in each row of \( \Gamma \) is determined by the other \( (M - 1) \) \( \gamma_{j,k} \)s.
single vessel can be written as follows (see MacDonald and Zucchini (1997, p79)):

\[
L(\gamma, \theta|X) = \pi(\gamma) \left( \prod_{t=1}^{T} B_t(\gamma, \theta) \right) 1'
\]  

(3)

where \(1'\) is a column vector of \(M\) 1s, \(B_t(\gamma, \theta) = \Gamma f(X_t)\) and \(f(X_t) = \text{diag}(f_{X|1}(X_t|S_t = 1), \ldots, f_{X|M}(X_t|S_t = M))\). (Note that \(\Gamma\) depends on \(\gamma\) and \(f(X_t)\) depends on \(\theta\) although for brevity we do not show these dependencies explicitly.)

This likelihood is appropriate when the starting state of the system is unknown. When the system is known to start in state \(m\), \(\pi\) should be replaced with a vector with a 1 in position \(m\) and zeros elsewhere, which we denote \(u_m\).

Given \(\gamma\) and \(\theta\), the conditional probability mass function for states \(S = (S_1, \ldots, S_T)\), having observed \(X = (X_1, \ldots, X_T)\) can be written as

\[
P(S|X) = \frac{P(X|S)P(S)}{\sum(S) P(X|S)P(S)}
\]  

(4)

where \(\sum(S)\) indicates the sum over all possible \(S\). We obtain our prediction of the states of a vessel at each of the \(T\) time points by maximixing Equation 4 with respect to \(S\). This in turn gives our prediction of the times at which the vessel is in the fishing state. Since each time is associated with a known vessel location, we have predictions of the locations at which fishing took place.

In fact Equation 4 gives us more than a means of predicting where fishing occurred - it also gives us a means of estimating the uncertainty associated
with this prediction. Conditional on $\gamma$ and $\theta$, this uncertainty can be evaluated by drawing samples of $S$ from Equation 4.

Notice that because the denominator of Equation 4 is constant with respect to $S$, we need only evaluate the numerator in order to find the maximum of $P(S|X)$ with respect to $S$, and we need only evaluate the numerator to draw samples from $P(S|X)$. Evaluating the numerator is relatively straightforward as it is

$$P(X|S)P(S) = \prod_{t=1}^{T} f_{X|S}(X_t|S_t) \times \pi_{S_1} \prod_{t=2}^{T} \gamma_{S_{t-1},S_t} \tag{5}$$

(Here $\pi_{S_1}$ should be replaced with $u_m$ if the system was known to be in state $m$ at time 1.) Uncertainty about $\gamma$ and $\theta$ can be incorporated by sampling from an estimate of their joint distribution obtained in maximising the likelihood of Equation 3.

3.1 Estimation with known-state data

Consider the case in which we have observations of both $S$ and $X$. In this case the likelihood function for $\gamma$ and $\theta$, given $S$ and $X$ is simply Equation 5 considered to be a function of $\gamma$ and $\theta$:

$$L(\gamma, \theta|X, S) = P(X|S)P(S) \tag{6}$$

and we can obtain maximum likelihood estimates of $\gamma$ and $\theta$ by maximising this function with respect to $\gamma$ and $\theta$. 
3.2 Using both known-state and unknown-state data

While we can estimate the model parameters $\gamma$ and $\theta$ when only $X$ is observed, we have no reliable way of knowing how any of the unobserved states in the model relates to the state of interest (i.e. fishing). Suppose however, that for some vessels we observe both $X$ and $S$ (these are the known-state data) and for others we observe only $X$ (these are the unknown-state data) and that one of the $M$ observed states is fishing. The known-state data relate the observations $X$ directly to the state of interest (fishing) as well as providing direct observations of the evolution of states over time. They provide the means of linking the unknown-state data to the state “fishing”.

The joint likelihood for the known-state and unknown-state data is

$$ L(\gamma, \theta) = \{L(\gamma, \theta|X)\} \times \{L(\gamma, \theta|X, S)\} \tag{7} $$

where $\{L(\gamma, \theta|X)\}$ indicates the product of the $L(\gamma, \theta|X)$s for each vessel with unknown-state data and $\{L(\gamma, \theta|X, S)\}$ indicates the product of the $L(\gamma, \theta|X, S)$s for each vessel with known-state data.

With both known-state and unknown-state data, a variety of estimation options is available. These include (i) parameter estimation by maximisation of Equation 7 or (ii) by maximisation of $P(S)$ alone using only known-state data. In both cases the estimated model can be used to predict states from known-state and unknown-state data. For either (i) or (ii) one could use only one HMM state or more than one HMM states for non-fishing. Estimation with more than one HMM state for non-fishing will require some method development.
Having a sample of known-state data allows one to rely much less heavily on circumstantial evidence of fishing to infer fishing state and having these data is therefore important for reliable inference about the times and locations at which fishing occurred.

4. Results

Models were fitted using a modified version of the R library HiddenMarkov.

While a three-state model provided a better fit to the data, it is difficult to interpret the states without some known-state data. (It is not clear whether the first two normal pdfs in Figure 5 correspond to fishing, or only one of them, and if one, which one.) We therefore used a two-state HMM for inferences in the interim while the known-state data validity is investigated.

Figure 10 shows the estimated distribution of fishing effort from the HMM, overlaid with vessel tracks colour-coded by speed. Figure 11 shows the level of agreement between the HMM predictions of fishing locations and that predicted from a simple rule which treats all speeds less than some cutoff speed (shown on the horizontal axis) as corresponding to fishing.

[Figure 10 about here.]

[Figure 11 about here.]

Figure 11 indicates that there is good agreement between the HMM method and the simple rule with cutoff speed of about 4.5 knots. Without knowing where fishing actually occurred, it is impossible to say which of the methods performs best.
5. Discussion

While we can’t at this stage say which method predicts fishing best, it is clear that estimates from a simple rule are somewhat sensitive to which cut-off speed is used. The HMM method is considerably more complicated than a simple rule, but it does have advantages. Primary among these is the fact that it is an objective, probability-based method and it consequently provides a measure of uncertainty associated with model predictions (although this has not yet been implemented) and does not require any subjective decisions regarding the relationship between speed and fishing state. By drawing samples from the pdf given in Equation 4 one can evaluate (for example) the probability that there were at least \( h \) hours of fishing within some nominally closed area of the sea. The HMM is also flexible enough to allow inclusion of explanatory variables (vessel size and/or bottom depth, for example).

In conclusion, the HMM method shows promise and seems to have worked well when applied to the 37 vessel-trip VMS sample data. We expect it to provide a useful method for inferring the distribution of fishing effort in space, particularly when known-state data are incorporated.

References

**Figure 1.** Vessel monitoring system (VMS) sample sizes (number of VMS records). Each bar corresponds to a different one of the 37 vessel-trips in the dataset.
Figure 2. Tracks of locations of 37 vessels from VMS data.
Figure 3. Distribution of vessel speeds from VMS data. Vessel speeds of 0.2 knots or less have been removed from the right plot.
Figure 4. Fit of the 2-component mixture model (dark black line) to the vessel speed histogram. Mixture components are shown using dashed coloured lines.
Figure 5. Fit of the 3-component mixture model (dark black line) to the vessel speed histogram. Mixture components are shown using dashed coloured lines.
Figure 6. VMS vessel tracks, color coded by speed.
Figure 7. Scatterplots of course difference against vessel speed. Course difference is the difference between current course and course $l$ time units previously, where $l$ is the lag, which varies from 1 to 9 time units.
Figure 8. Nominal locations of hauls from discard observer data from a sample of vessels.
Figure 9. Schematic representation of a hidden Markov model (HMM). The underlying states $S = S_1, \ldots, S_{t-1}, S_t, S_{t+1}, \ldots$ are not observed. But a random variable $X = X_1, \ldots, X_{t-1}, X_t, X_{t+1}, \ldots$, whose distribution depends on $S$ is observed. Here $f_{X|S}(X_t|S_t)$ ($t = 1, 2, \ldots$) is the probability density function of $X_t$, given that the system is in state $S_t$. The probability of making a transition to state $S_{t+1}$ in the next time step, given that the system is in state $S_t$ at time step $t$ is $\gamma_{S_tS_{t+1}}$ ($t = 1, 2, \ldots$).
Figure 10. Intensity of estimated fishing effort (in VMS time units) of all 37 fishing vessels in one-degree by half-degree rectangles. The tracks of the vessels are also shown, color-coded by vessel speed.
Figure 11. Agreement between HMM predictions of fishing locations and predictions from a simple rule which assumes all points at which the vessel was going less than some cutoff speed to have occurred during fishing. The cutoff speed is shown along the horizontal axis. Plot (a) shows the percentage agreement between the HMM and cutoff methods; the dashed horizontal line is at 100%. Plot (b) shows the number of points which are in agreement; the dashed horizontal line is at the number of fishing points predicted by the HMM.