Title: Estimating individual animal movement from observation networks

Authors: Martin W. Pedersen[1,2], Kevin C. Weng[1]

[1] Pelagic Fisheries Research Program, Department of Oceanography, University of Hawaii at Manoa

[2] Corresponding author: martinwp@hawaii.edu (alt: map@aqua.dtu.dk)

Running title: Estimating movement from observation networks

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/2041-210X.12086

Summary (Abstract)

1. Observation network data comprise animal presences detected by observer stations at fixed spatial locations. Statistical analysis of these data is complicated by spatial bias in sampling and temporal variability in detection conditions. Advanced methods for analysis of these data are required but are currently underdeveloped.

2. We propose a state-space model for observation network data to estimate detailed movements of individual animals. The underlying movement model is an Ornstein-Uhlenbeck (OU) process, which is stationary and therefore has an inherent mechanism that models home range behaviour. An integral part of the approach is the detection function, which models the probability of logging animal presences. The detection function is also used to provide absence information when animals are undetected. Since the ability to detect an animal often depends on time varying external factors such as environmental conditions we use covariate information about detection efficiency as control variables.

3. Via simulation we found that movement estimation error scales log-linearly with network sparsity. This result can be used to indicate the number of stations necessary to achieve a desired upper bound on estimation error. Furthermore we found that the state-space model outperforms existing techniques in terms of estimating detailed movements, and that estimates are robust toward misspecification of the detection function. We also tested the importance of accounting for time varying detection conditions and found that the probability of making wrong conclusions decreases substantially when covariate infor-

mation is exploited.

4. The model is used to estimate movements and home range of a humphead wrasse (*Cheilinus undulatus*) at Palmyra Atoll in the central Pacific Ocean. Here, detection conditions have a strong diel component, which is controlled for using detection efficiency information from a reference device.

5. The presented approach enhances the toolbox for analysis of observation network data as collected by acoustic telemetry or potentially other aspiring methods such as camera trapping and mobile phone tagging. By explicitly modelling movement and observation processes the model integrates all sources of uncertainty and provides a sound statistical basis for making well informed management decisions from imperfect information.

Key words: State-space model, detection probability, acoustic telemetry, Ornstein-Uhlenbeck process

1 Introduction

Observation networks monitoring animal movements are used throughout ecology (McConnell *et al.*, 2004; Heupel *et al.*, 2006; Rowcliffe *et al.*, 2011). Data produced by such networks comprise animal presences detected by automated observer stations positioned at fixed spatial locations. These data are akin to manually collected presence-absence data from e.g. capture-recapture (Jolly, 1965), mark-resight (Hestbeck *et al.*, 1991), or similar experiments, however with the difference that observation network data can be gathered continuously over extended periods of time with comparatively minor effort. Traditional presence-absence studies focusing on movement have estimated dispersal as quantified by rates of movement between geographical areas (states) in multistate models (Lebreton & Pradel, 2002). Since the number of observations per individual often is limited owing to manual data collection, these studies aggregate observations from many individuals (Hestbeck *et al.*, 1991). In contrast, observation network data are sampled at a higher rate, and therefore have potential for individual-level modelling (Simpfendorfer *et al.*, 2002).

With presence-absence data a common modelling concept is the animal detection probability (Jolly, 1965; Ovaskainen, 2004), i.e. the probability of detecting an animal present at the surveyed location. For manually collected data the detection probability is typically modelled as a constant over the survey area (Gimenez *et al.*, 2007). Similarly, for automated data collection by observation networks, a simplistic model for animal detectability would assume that observer stations have a detection zone within which all animal presences are logged and outside of which presences cannot be detected (binary

detection). More realistic models substitute binary detection with a general function relating the probability of animal detection to its spatial location relative to the station (How & de Lestang, 2012). Such a detection function can take into account both spatial and temporal variation in detection probability.

In addition to location, a number of other factors have been documented to influence detection probability (Heupel *et al.*, 2006). Studies have, for example, reported on factors such as time of day (Payne *et al.*, 2010), physical environment (Bergé *et al.*, 2012), anthropogenic noise (Thorstad *et al.*, 2002), to affect detection efficiency. A common feature of these factors is that their impact can be very site specific and therefore difficult to quantify in general.

The established approach to extract detailed individual movement from observation network data is the weighted-mean method (Simpfendorfer *et al.*, 2002). Recently, alternative methods relying on local polynomial regression have also been considered (Hedger *et al.*, 2008). These are non-mechanistic methods, which do not account for uncertainties induced by the movement and observation processes. Uncertainties arise because the continuous movement process is observed non-continuously and indirectly via some proxy at distinct points in time. For observation network data this proxy is the detection of an animal near an observer station. In the ecological literature on animal movement, state-space models (SSMs) have become a popular statistical approach to handle indirect and auto-correlated movement observations (Patterson *et al.*, 2008). Previously, SSM analyses of presence-absence data have been used to estimate demographic parameters such as sur-

vival and population dynamics (Gimenez *et al.*, 2007; Royle, 2008), however SSMs have yet to be applied to high frequent data from observation networks.

Movement of species exhibiting site fidelity, which are often studied using observation networks, are spatially limited to regions within their home range. This behaviour is typically modelled by incorporating a bias in movement toward a fixed point, which can be interpreted as the home range centre. An example of such a model is the Ornstein-Uhlenbeck (OU) process (Blackwell, 1997), which is stationary and therefore has an inherent mechanism that mimics movement behaviour of animals with home ranges. Diffusion without bias is a sub-model of the OU process (when the drift term is set to zero), enabling the process to also model movements of species without a home range.

This paper develops an SSM for observation network data to estimate animal movement using the OU process. The model assumes that the detection function can be estimated by independent ranging experiments. Using simulation, we check the robustness of this assumption by gauging the performance of the model when the detection function is misspecified. The performance of the SSM relative to two established approaches (Simpfendorfer *et al.*, 2002; Hedger *et al.*, 2008) is also assessed. Some studies may have detection functions that extend farther than others; study animals with different home range or movement scales; or have observation networks that extend over different areas. In order to compare studies with differing spatial or temporal scales, we develop dimensionless performance metrics functioning as universally comparable characteristics of observation network studies. Via the detection function the presented approach is able

to control for time-varying detection conditions by incorporating covariate information from a reference device. We evaluate the importance of this feature by estimating models exploiting either full, partial (realistic), or no covariate information. Finally, we apply the modelling framework to fixed acoustic receiver data from a humphead wrasse (*Cheilinus undulatus*) at Palmyra atoll in the central Pacific Ocean.

2 Materials and Methods

2.1 State-space model for observation network movement data

A state-space model (SSM) is composed of two linked sub-models: the movement model and the observation model. Since the main focus of this paper is to model the data collection mechanisms of the observation network we begin with the observation model. Subsequently, we present the movement model, which is a bivariate Ornstein-Uhlenbeck (OU) process (Blackwell, 1997).

2.1.1 Observation model

The observation process for observation network data is intuitively simple: an area is monitored continuously by n_o observer stations and present individuals are detected with some probability. Since acoustic monitoring of animals carrying electronic tags (Heupel *et al.*, 2006) is the most common observation network type we develop the model with this framework in mind. Electronic tags are, for simplicity, assumed to emit signals at known and constant rates given by the sample interval dt. For a given station, the probability

of detecting a signal conditional on the animal's location is described by the detection probability $p = P(\text{detection}|\text{location})$. For now we assume that detection conditions do not change i.e. that p is constant in time. The distance, d, from the observer station to the animal is an important variable for explaining the spatial distribution of p (How & de Lestang, 2012). In heterogeneous environments the animal's angle to the observer station, θ, may also influence p. Thus, for a given station, the function relating detection probability to spatial location is

$$p = f(d, \theta). \tag{1}$$

It is assumed that f can be estimated from ranging experiments using spatially fixed reference devices prior to network deployment (How & de Lestang, 2012). The modelling framework presented here is generic, i.e. no assumptions are made for the functional form of f.

Station i records information, $I_{i,t}$, at time t, which can be either a presence ($I_{i,t} = 1$) or an absence ($I_{i,t} = 0$). If $I_{i,t} = 1$, f can be interpreted as a likelihood of the animal's location, $L(\text{location}|\text{detection})$, meaning that locations close to the station are more likely than distant locations (Fig. 1). Conversely, in the case of an absence ($I_{i,t} = 0$) the location likelihood is $1 - f$ meaning that locations farther from the station are more likely than nearby locations (Fig. 1).

Owing to the dichotomous nature of the data and the constant sample interval, absence or presence information is available for each of the n_o stations in the observation network at a temporal resolution given by dt. Information from multiple stations is combined by multiplying their likelihood functions, thus identifying their spatial overlap as the most likely location of the animal at the given time point (Fig. 2). The resulting function, $L(t, x, y)$, describing the overlap at time t does generally not have a closed form expression (such as a Gaussian function) so we resolve it non-parametrically on a two-dimensional horizontal grid with $n_x \times n_y$ grid cells. Mathematically, the location likelihood is calculated for each spatial position (x, y) and each time point (t) as

$$L(t, x, y) = \prod_{i=1}^{n_o} I_{i,t} f(d_i, \theta_i) + (1 - I_{i,t})[1 - f(d_i, \theta_i)], \tag{2}$$

where d_i and θ_i are distance and angle from station i to (x, y) respectively. Different examples of $L(t, x, y)$, including the information derived from absence data, are illustrated in Fig. 2.

2.1.2 Movement model

The grid $L(t, x, y)$ only utilises data collected at time t to indicate location. However, data from other points in time also hold information about the location at t. This is because data are temporally auto-correlated, which occurs when the sample rate is high relative to the animal movement rate. In other words, data are auto-correlated if, given a known

current location, the range of possible future locations is limited by the animal's maximum movement speed. This biological limitation on movement is incorporated in SSMs via the movement model, which enables the SSM to also use information from other time points to obtain a more accurate location estimate at time t.

The modelling framework we outline here can in principle accommodate any movement model, be it random walk variants (Codling *et al.*, 2008) or continuous-time processes (Preisler *et al.*, 2004). Observation networks are often used for studying species with home ranges, which restrict their movements to a limited region. We therefore use a bivariate Ornstein-Uhlenbeck (OU) process (Blackwell, 1997) as the movement component of the SSM. The OU process is stationary with a fixed point of attraction, (μ_x, μ_y), which can be interpreted as the home range centre. The following stochastic differential equation describes OU-based movements

$$\begin{pmatrix} dX_t \\ dY_t \end{pmatrix} = B \begin{pmatrix} \mu_x - X_t \\ \mu_y - Y_t \end{pmatrix} dt + \sigma \begin{pmatrix} dW_x \\ dW_y \end{pmatrix} \tag{3}$$

where X_t and Y_t are coordinates in two-dimensional space at time t, the matrix

$$B = \begin{pmatrix} b_x & b_{xy} \\ b_{xy} & b_y \end{pmatrix} \tag{4}$$

determines the strength of attraction toward (μ_x, μ_y), σ is the movement rate, and dW_x and dW_y are Brownian motions. The home range of the animal can be estimated by calculating the stationary distribution of the OU process, which is bivariate Gaussian

with mean (μ_x, μ_y) and covariance $0.5\sigma^2 B^{-1}$, which determines the shape and extent of the home range. Diffusion without attraction is a sub-model of the OU process (if $B = 0$), which is useful for species that do not exhibit site fidelity.

2.1.3 Numerical implementation

The SSM is implemented and fitted as a spatial hidden Markov model (Pedersen *et al.*, 2011). Details pertaining to the implementation of the OU process are given in Appendix S1. Model parameters were estimated with maximum likelihood in Automatic Differentiation Model Builder (Fournier *et al.*, 2012). Model code is found in Appendix S2. To summarise the model fit we calculate, in addition to the estimated model parameters, the expected animal location with associated confidence region at each time t from the smoothed probability distributions of the state as returned by the SSM (Pedersen *et al.*, 2011).

2.1.4 Time varying detection conditions

Several time varying factors influence the detection probability in observation networks (Heupel *et al.*, 2006). It is therefore unreasonable to assume that p is constant in time. By letting detection function parameters explicitly depend on factors such as season or environmental conditions, some time varying effects can be accounted for (Rowcliffe *et al.*, 2011; How & de Lestang, 2012). This approach, however, requires the relation between the covariate and the detection function to be explicitly modelled, which is not necessarily

trivial. For example, in order to account for changing weather conditions synoptic data must be acquired and often temporally and spatially interpolated to match the location and scale of interest. Estimation errors associated with the integration of synoptic data products can be difficult to quantify and may lead to biased results.

A simple alternative is to use supplementary reference data collected at the study site during the experiment, to provide site-specific information about detection conditions at a high temporal resolution. To gather reference data a reference device, which is able to trigger presence observations, should be placed at a known location within the detection zone of an observer station. The reference device should be easily detectable by the observer station in good detection conditions, and less detectable in poorer conditions, such that shifting conditions can be identified. The resulting reference data serve as a proxy for detection efficiency in that a large proportion of the expected number of signals will be logged under good detection conditions and vice versa. As an alternative to external and often large scale data products the use of reference data is appealing because it provides information about detection conditions at the exact spatial location of interest and on a fine temporal scale.

The detection efficiency, E_t, indicates the detection conditions at time t and is based on the proportion of expected signals that were in fact recorded over a time interval around t. Specifically, we obtained a continuous detection efficiency curve by a weighted average (or smoothing) of the reference data using a Gaussian kernel (Hastie *et al.*, 2001). The detection efficiency curve (E_t) was normalised such that $E_t = 1$ under optimal detection

conditions (upper bound) and $E_t = 0$ when detection is impossible (lower bound). Determination of the kernel bandwidth, i.e. the time interval over which to smooth, is a modelling decision with larger values leading to a smoother detection efficiency curve at the cost of losing fine scale variations. Previously the probability of detection (f) was a function of d and θ. Now, to incorporate temporal variability, the detection function (h) is also a function of detection efficiency

$$p_t = h(d, \theta, E_t). \tag{5}$$

The relationship between p_t and E_t can be study specific. It is clear, however, that a 50% reduction in detection efficiency implies a 50% reduction in p_t at the reference distance (or equivalently when $E_t = 0.5$, then $h(d_{ref}, \theta, 0.5) = 0.5f(d_{ref}, \theta)$, where d_{ref} is the distance between reference device and observer station). It is, however, unclear how the detection function shape changes at distances other than d_{ref}. It is tempting to use $h(d, \theta, E_t) = E_t f(d, \theta)$ (blue curve in Fig. 3), which implies that the detection probability decreases evenly over all distances (shifts downward on the p-axis). However, this is a different behaviour to what we have observed in our ranging experiments (Fig. 1 in Appendix S3). Instead, we use a model where the detection function shifts along the d-axis (red curve Fig. 3). Assuming axis-symmetry (omitting θ) the detection function becomes

$$p_t = h(d, E_t) = f\left(d\frac{d_{noise}}{d_{ref}}\right), \tag{6}$$

where $d_{noise} = f^{-1}(E_t f[d_{ref}])$, which is the distance the reference device would have to be from the station to get the observed probability of detection, assuming optimal detection conditions. If $E_t < 1$ then $d_{noise} > d_{ref}$ resulting in a p which is lower than it would be under optimal conditions. Here, f^{-1} is the inverse function of f, i.e., the equation is solved for d such that $d = f^{-1}(p)$. With the model in eqn. 6 the detection probability scales with E_t at d_{ref}, which is consistent with changes in the reference data under changing conditions. In addition, the detection probability at $d = 0$ is constant regardless of the value of E_t, meaning that locations closer to the observer station are less affected by poor detection conditions than locations farther away (Fig. 3).

2.2 Simulation study

Observation network data were simulated to address: 1) how SSM estimation performance varies with movement rate and network sparsity 2) the performance of existing movement estimation approaches and of SSMs with misspecified detection functions, and finally 3) the importance of accounting for variation in detection conditions.

2.2.1 Temporal and spatial Scales

As studies using observation networks are conducted over different spatial and temporal scales we introduce the following scale independent descriptors of network sparsity and sampling properties. Temporally, the main factors are the sampling interval, dt, and the

movement rate, σ. To represent the effective movement capacity of the animal in relation to the detection range we use the ratio of the root mean squared (rms) displacement ($\sigma\sqrt{dt}$) within dt to the detection range d_r

$$\phi = \frac{\sigma\sqrt{dt}}{d_r}. \tag{7}$$

Specifically, d_r is the distance at which the detection function has reached $f(d_r) = 0.05$. The quantities d_r and dt are often known prior to data analysis making ϕ proportional to σ and therefore a dimensionless indicator of movement capacity. For example, $\phi = 0.1$ can be interpreted as an animal with a capacity to move 10% of the detection range within a sample interval, dt. The effective movement capacity (ϕ) can also be viewed as a signal to noise ratio with the important property that increasing d_r reduces ϕ. Hence, it is more difficult to extract fine-scale movements when the detection range is large.

Spatially, we calculate an absolute measure of station closeness, a, as the median of $\{a_1, \ldots, a_{n_o}\}$, where a_i is the distance from station i to its nearest neighbouring station. Network sparsity is then defined as

$$\delta = \frac{a}{2d_r}. \tag{8}$$

which is a dimensionless quantity enabling scale independent comparison of observation networks. If $\delta < 1$ the network mostly has detection functions that overlap, whereas $\delta > 1$ implies a sparser network with mostly non-overlapping detection functions. Another spatial scale of interest is the extent of the network relative to the animal's movement

range. Treatment of this scale is outside the scope of this study. Instead we assume that the chance of the animal leaving the observation network is negligible, which seems reasonable at least for animals with a home range.

2.2.2 Detection function

When simulating observation network data we omit, for simplicity, the detection function's dependence on the observation angle, thereby mimicking data collection in a homogeneous flat terrain. We adopt a sigmoidal relationship between detection probability and distance, which has the form

$$p = f(d) = \frac{p_{max}}{1 + \exp(\log[19][d - D_{50}]/[D_{95} - D_{50}])}, \quad p = \min(f[d], 1), \tag{9}$$

where D_{50} and D_{95} are distances where f has declined by 50% and 95% of its maximum value (p_{max}) respectively. The constant $\log(19)$ is necessary to ensure that $f(D_{95}) = 0.05 \times p_{max}$. The sigmoidal function has been shown as a flexible descriptor of the relationship between detection probability and distance in different observation network environments (How & de Lestang, 2012). An example of this detection function is shown in Fig. 1. Table 2 contains detection function parameter values used in the simulation study. When analysing field data, detection function parameters should be estimated from site-specific ranging experiments.

2.2.3 Simulation scenario 1

Observer stations were placed in an equilateral triangle pattern at distances of a from each other. Isotropic attraction to (μ_x, μ_y) was assumed by setting $B = bI$, where b is the attraction parameter and I is a 2×2 identity matrix. This results in four model parameters (μ_x, μ_y, b, σ) to be estimated.

Observation network data were simulated by first generating random movement trajectories from eqn. (3) using three different movement rates (σ) with fixed d_r (Table 1). Values of b were adjusted such that the home range size remained constant at different σ. Then, using the simulated trajectory, artificial signals were emitted at a fixed rate, $dt = 180$ s, mimicking the mechanism underlying acoustic telemetry data. A total of 4801 signal were emitted corresponding to ten days of data. The emitted signals were detected at each station with a probability given by eqn. (9) using parameter set A (Table 2) leading to three different data sets; one for each movement rate. Each data set was analysed with the described SSM approach ($n_x \times n_y = 51 \times 51$) to check how estimation error varied with animal speed. To obtain scale-independent results, we use a dimensionless error metric akin to the coefficient of variation

$$c_e = \frac{e_{rms}}{\sigma\sqrt{dt}}, \tag{10}$$

where e_{rms} is the rms error between estimated and true locations. The whole procedure was carried out repeatedly for network sparsities $\delta \in [0.5, 3.25]$ to also ascertain how estimation error scales with decreasing detection coverage. The range for δ represents the

interval where state-space analysis is possible and estimation of location by triangulation is not possible. A log-linear model for c_e as a function of δ was used to statistically test for differences in model performance. Data for the log-linear model were down-sampled to remove temporal auto-correlation leading to the sample sizes in Table 1.

2.2.4 Simulation scenario 2

Similar to scenario 1, but focusing only on the low speed case (Table 1). Generated data were analysed with a weighted-mean method (Simpfendorfer *et al.*, 2002) and a local polynomial regression method (Hedger *et al.*, 2008). Data were also analysed with the described SSM in three separate cases using either parameter sets A, B, or C (Table 2), thereby making slight (set B) and substantial (set C) misspecification of the detection function. Log-linear regressions of differences in rms location error relative to the rms location error of the data generating model (SSM A) were used as model performance metric. By taking the exponential of the log rms location error we calculated the percentage difference in model performance.

2.2.5 Simulation scenario 3

To evaluate the effect of integrating reference data as covariate information three models were estimated using data generated with time varying detection conditions: an SSM

assuming constant conditions (SSM D), an SSM employing reference data to estimate and control for varying conditions (SSM E), and an SSM using the true detection conditions used to generate the data (SSM F). See Appendix S4 for full simulation details.

2.3 Field data

For illustrative purposes we used the SSM to analyse a data from a humphead wrasse (*Cheilinus undulatus*) collected by a submerged acoustic observation network in a shallow water tropical reef setting located at Palmyra Atoll in the central Pacific Ocean (data set is provided in Appendix S2 along with model code). The network consisted of 14 observer stations. The fish was tagged with a Vemco V13-1H acoustic transmitter with an average transmission interval of 120 s. To minimise signal interference between transmitters, each signal is sent with a random delay of ±60 s. We therefore binned data into $dt = 240$ s intervals for analysis to ensure that at least one signal was sent per time interval. As a consequence, intervals lacking signal receptions at a station indicate an absence. To avoid messy visualisation of results we extracted a subset of 57 hours of movement data for analysis containing $n = 342$ detections.

Prior to network installation, ranging experiments were carried out. Ranging data were unevenly sampled at different distances from the station and it was therefore necessary to impute missing data to obtain equal sample sizes. Specifically, we used multiple imputation for estimation of the detection function parameters as described by Nakagawa & Freckleton (2008) (see Appendix S3 for full details on detection function estimation). Parameter

values of the sigmoidal detection function (eqn. 9) under good detection conditions were estimated to: $p_{max} = 1.86$, $D50 = 7.22$, and $D95 = 149.2$. A reference device with a transmission interval of $dt_{ref} = 600$ s was installed at a distance of $d_{ref} = 60$ m to a station to collect information about detection conditions. Detection efficiency was calculated using a kernel bandwidth of $6 \times dt_{ref}$. Using the median detection efficiency (median$[E_t] = 0.257$) the effective detection range was $d_r = 79.2$ m leading to an an effective network sparsity of $\delta = 0.63$.

In approximating the spatial area of interest, $n_x \times n_y = 77 \times 33$ grid cells were used, with a grid cell size of 10×10 m. Model fitting was carried out in projected Cartesian coordinates. Movement results are presented in terms of latitude and longitude by inverse projection.

3 Results

3.1 Simulation scenario 1

The ratio (c_e) of rms location error to rms displacement increased with increasing network sparsity for all values of ϕ (Z-test for slopes: $Z_{\phi=0.1} = 7.29$, $Z_{\phi=0.2} = 4.30$, $Z_{\phi=0.4} = 60.3$, all $P < 0.001$, Fig. 4). Relative location error increased with decreasing values of movement speed (Z-test for means; $\widehat{c}_{e,\phi=0.1} > \widehat{c}_{e,\phi=0.2}$: $Z = 4.43$, $P < 0.001$; $\widehat{c}_{e,\phi=0.2} > \widehat{c}_{e,\phi=0.4}$: $Z = 15.8$, $P < 0.001$). For a low speed animal log-linear regression indicated a 41% decrease in c_e if network sparsity (δ) is reduced by a value of 0.5. For moderate

and high speed animals the corresponding decreases were 32% and 28% respectively. This confirms the intuition that lower speed animals require low sparsity (i.e. high density) networks to resolve detailed movement.

3.2 Simulation scenario 2

Slight misspecification of the detection function (SSM B) had no detectable influence on location estimation (mean: $Z = 1.36$, $P = 0.176$; slope: $Z = -0.168$, $P = 0.867$; Fig. 5 pane a) relative to the optimal model SSM A. Substantial misspecification of the detection function (SSM C) resulted in significantly (mean: $Z = 5.56$, $P < 0.001$) increased location error. For dense networks ($\delta = 0.5$) the error was increased by 45% (Fig. 5 pane b). This difference in location error converged to zero for increasingly sparse networks (Fig. 5 pane b). Location error of the local polynomial regression was on average 33% larger than that for SSM A (mean: $Z = 10.1$, $P < 0.001$, Fig. 5 pane c). This increased error was constant for all network sparsities (slope: $Z = -1.07$, $P = 0.284$). The weighted-mean method showed a similar pattern with a 47% average increased error (mean: $Z = 11.9$, $P < 0.001$; slope: $Z = -1.49$, $P = 0.136$; Fig. 5 pane d).

3.3 Simulation scenario 3

When the true detection conditions were assumed known (SSM F), the 95% confidence region around the estimated spatial location provided by the SSM contained the true location 94.8% of the time (mean different from 0.95: $Z = -0.988$, $P = 0.326$). In the more realistic situation where only reference data are available (SSM E), the 95% confidence regions contained the true location 94.8% of the time on average (mean different from 0.95: $Z = -0.745$, $P = 0.458$). If no information about detection conditions were used in the estimation (SSM D) the confidence regions became unreliable with the true location contained 64.9% of the time (mean different from 0.95: $Z = -94.9$, $P < 0.001$). Similarly, for estimation of movement rate (σ) and attraction to home range centre (b), a significant bias was observed for SSM D (Fig. 2 in Appendix S4, top row), however estimation of the home range centre itself was unbiased. All parameter estimates using approximate or true covariate information were unbiased (Fig. 2 in Appendix S4, middle and bottom rows respectively).

3.4 Field data

During day-light hours locations of the fish were estimated with a 95% confidence region radius of 50 m when most accurate (Fig. 6, pane a). Absence information during good detection conditions lead to confidence regions that "avoided" observer stations resulting in oddly shaped distributions (Fig. 6, pane b). At night no observations of animal presence were made within the network. A possible reason for this is the general reduction in

detection efficiency between sunset and sunrise (inferred from reference data). Such a decrease is consistent with other similar studies, which have found acoustic activity of nocturnal creatures to interfere with data collection (Payne *et al.*, 2010). Another possible explanation for the lack of detected animal presence is that the fish either left the network or sheltered itself in an unobservable location, as many reef fishes are known to enter holes and caves at night thereby hindering detection. The SSM responded to the lack of presence data by increasing the confidence regions (Fig. 6, pane c), thereby implying that the specific location of the fish was unknown, but that it was expected to remain within its home range.

Estimated model parameter are shown in Table 3. The diagonal elements of B indicated a stronger attraction in the latitudinal direction relative to the longitudinal ($b_x < b_y$) leading to an elongated home range shape (Fig. 6). Furthermore, the off-diagonal element, b_{xy}, of the matrix B was different from zero (Table 3) resulting in a home range ellipse, which has angled major axes relative to the spatial coordinate system (Fig. 6). The estimate of σ lead to a rms displacement of 26.6 m (95% CI: 23.0; 30.7 m) for $dt = 240$ s, or equivalently an effective movement capacity of $\phi = 0.336$ (95% CI: 0.291; 0.387).

Plotting the confidence regions (as in Fig. 6) in temporal succession reveals fine scale details of the estimated movement (animation is included in Appendix S5). In the es-

timated movement some bias toward observer stations was evident. The SSM, however, also captures movement between stations. This happens if the animal is detected on multiple stations simultaneously in which case the model combines information using the overlapping detection functions (Fig. 2). Estimation of movement in regions outside the detection range relies on a combination of absence information and the OU movement model. This information is weak relative to presence information and therefore leads to increased uncertainty of location estimates in these regions.

4 Discussion

This study presented an SSM for movement data collected by observation networks, which incorporates knowledge about underlying biological and observational processes. Shifting from non-mechanistic methods (kernel smoothing, local polynomial regression) to an SSM approach provides uncertainty quantification and parameter estimation in a flexible statistical framework.

Our simulation study showed that the error of estimated locations relative to the animal rms displacement decreased log-linearly with an increase in network density. This confirms the intuition that estimation error can be reduced by installation of additional observer stations. Results also indicated that faster moving animals have lower relative error when compared with slower animals. This is an expected result since slower moving animals can cover less space in a fixed time interval and are therefore exposed to fewer observer stations than faster animals would be. While slower moving animals are associated with larger

relative error our results indicate that they also benefit from a larger relative improvement in error when reducing network sparsity. This can be viewed as slower animals having a higher information gain per added station relative to faster animals.

In terms of network design the estimated log-linear relationship can be used to indicate the network density necessary to achieve a certain upper bound on estimation error. This only requires that the detection function and animal rms displacement are known or can be estimated prior to network deployment. Naturally, since this upper bound is calculated from an idealised albeit realistic simulation scenario it should only be used as a general guideline for expected estimation error.

The SSM outperformed existing non-mechanistic methods (Simpfendorfer *et al.*, 2002; Hedger *et al.*, 2008) when estimating location using simulated data. This is unsurprising since the SSM utilises additional information, the detection function, in its estimation; however SSM performance was superior even with misspecified detection function (Fig. 5). Non-mechanistic approaches are simple and therefore able to provide quick but rough estimates of movement with minor implementation effort, however studies aiming for higher estimation precision and improved ecological understanding of the data can benefit substantially from the presented framework.

It is well documented that temporal variation in detection conditions occur (Heupel *et al.*, 2006; Payne *et al.*, 2010). Our simulation results showed that ignoring this variability can lead to large biases in parameter estimates (Fig. 2 in S4) and misleading confidence regions for location. Variations in detection probability may lead to misinterpretation of

animal behavior. For instance if a diel change in detection conditions is present but unaccounted for, data may be misinterpreted to indicate that animals leave the monitored area at night. We found that when using supplementary data collected by a reference device, variability in detection conditions can be approximated resulting in unbiased estimates and reduced risk of misinterpretation (Fig. 2 in S4). We therefore echo conclusions by other studies (e.g. How & de Lestang, 2012) that reference devices should be considered an important component in modern observation networks.

Owing to manufacturing variability internal clocks at observer stations may not be synchronised. In severe cases (time difference larger than the sampling interval) this increases the risk of biased model results. However, clock drift is typically linear and can be corrected (cf. VUE software manual, www.vemco.com/pdf/vue_manual.pdf, accessed 23 April 2013) thereby eliminating this risk. Since our state-space approach does not rely on triangulation to calculate location, as is the case with the accurate Vemco VR2W Positioning System, sub-second differences between station clocks will have negligible impact on results.

An integral component of the presented SSM is the detection function, which describes the probability of detecting an animal as a function of its location in space relative to an observer station. This is a well established concept within studies of occupancy (Rowcliffe *et al.*, 2011) and dispersal (Ovaskainen, 2004), which, with the presented method, also has found a use within individual movement estimation. Importantly, the SSM can accommodate any functional form of the detection function, and is therefore not specific to the

sigmoidal relationship used here. The detection function is useful not only when animals are present within the detection range, but also when absences are logged. When animals are undetected our model uses the detection function to exclude regions near observer stations from the range of possible locations. Absence information cannot be exploited by traditional non-mechanistic methods, but as shown here it can make a substantial difference in the shape of the estimated confidence region (Fig. 6) leading to reduced estimation error (Fig. 5).

We introduced two characteristic numbers that describe data collection by observation networks. The first, δ, is the network sparsity, which relates average distance between stations to the detection range. This number can be calculated for any network and is an objective way of comparing spatial properties of different studies independent of scales. In the case of a very dense network ($\delta < 0.5$) most regions should be covered by at least three observer stations and therefore make location via triangulation possible. On the other hand, when $\delta > 3$ estimated locations as provided by the SSM converges to the error obtained using a naïve estimator (simple mean of the station locations). The use of the SSM is therefore only relevant in the range $0.5 < \delta < 3$. The second characteristic number, ϕ, is the effective movement capacity relating animal speed to the detection range. The effective movement capacity can also be interpreted as a signal to noise ratio, which is proportional to the model's ability to estimate detailed movements (Fig. 4).

The biological component of the SSM is the movement model. In this study we have used the OU process as model for movement. The OU process is particularly suited for

animals with a home range and therefore appropriate for many species studied with observation networks. The main limitation of the OU process is its representation of home range as elliptical and unimodal. Blackwell (1997), however, presented extensions to the basic OU process using mixtures of OU processes with different home range centres thereby mimicking multimodality. Another extension of the movement model could be to incorporate random effects and individual covariate information in model parameters when analysing data from multiple individuals. This would improve estimation of data limited individuals by information sharing, while providing estimates of demographic parameters. Another interesting model extension would be to let movement parameters vary dynamically as functions of the ambient environment and internal physical condition. This could aid in understanding the fundamental ecological mechanisms underlying home range based movement behaviour (Börger *et al.*, 2008).

In this paper we have demonstrated that the potential of observation network data has so far been underutilised with respect to the estimation of movement and home range. We believe that the presented model is a first step toward an integrated framework, which, in addition to a sound statistical analysis of observation network data, also has the potential to become a valuable guide to designing observation network studies via the proposed dimensionless descriptors of spatio-temporal scales. Today, automated data collection using observation networks is commonly achieved via acoustic receivers and animal borne electronic transmitters. However, other branches of ecology are also experiencing technological advancements e.g. enabling automated logging of animal presences using camera

traps (Rowcliffe *et al.*, 2011) or mobile phone technology (McConnell *et al.*, 2004). These trends indicate a future with increasing access to large observation network data sets and therefore an increasing need for advances in statistical methods.

5 Acknowledgements

We thank Luca Börger and two anonymous reviewers for comments that greatly enhanced this paper. MWP and KCW were funded by the Pelagic Fisheries Research Program (PFRP) under Cooperative Agreement NA17RJ1230/NA09OAR4320075 between the Joint Institute for Marine and Atmospheric Research (JIMAR) and the National Oceanic and Atmospheric Administration (NOAA). Research at Palmyra Atoll was funded by the NOAA Coral Reef Conservation Program NA05OAR4301108. The views expressed herein are those of the authors and do not necessarily reflect the views of NOAA or any of its subdivisions. We thank Anders Nielsen for providing adjoint ADMB code. We thank Dodie Lau, Johnoel Ancheta, and John Sibert for support of PFRP. We thank Andrew Gray, Jeff Muir, Christina Comfort, Leilani Itano, Andrew Purves, Amanda Meyer and Kydd Pollock for assistance in the field, Kim Hum and Brenda Santos for logistical support. Work at Palmyra Atoll was conducted under permit from the US Fish and Wildlife Service and the University of Hawaii Institutional Animal Care and use Committee.

References

Bergé, J., Capra, H., Pella, H., Steig, T., Ovidio, M., Bultel, E. & Lamouroux, N. (2012) Probability of detection and positioning error of a hydro acoustic telemetry system in a fast-flowing river: Intrinsic and environmental determinants. *Fisheries Research*, **125**, 1–13.

Blackwell, P. (1997) Random diffusion models for animal movement. *Ecological Modelling*, **100**(1-3), 87–102.

Börger, L., Dalziel, B.D. & Fryxell, J.M. (2008) Are there general mechanisms of animal home range behaviour? A review and prospects for future research. *Ecology letters*, **11**(6), 637–650.

Codling, E.A., Plank, M.J. & Benhamou, S. (2008) Random walk models in biology. *Journal of the Royal Society Interface*, **5**(25), 813–834.

Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A. & Sibert, J. (2012) AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software*, **27**(2), 233–249, doi:10.1080/10556788.2011.597854.

Gimenez, O., Rossi, V., Choquet, R., Dehais, C., Doris, B., Varella, H., Vila, J.P. & Pradel, R. (2007) State-space modelling of data on marked individuals. *Ecological Modelling*, **206**(3), 431–438.

Hastie, T., Tibshirani, R. & Friedman, J.J.H. (2001) *The elements of statistical learning*, vol. 1. Springer New York.

Hedger, R., Martin, F., Dodson, J., Hatin, D., Caron, F. & Whoriskey, F. (2008) The optimized interpolation of fish positions and speeds in an array of fixed acoustic receivers. *ICES Journal of Marine Science: Journal du Conseil*, **65**(7), 1248–1259.

Hestbeck, J., Nichols, J. & Malecki, R. (1991) Estimates of movement and site fidelity using mark-resight data of wintering Canada geese. *Ecology*, **72**(2), 523–533.

Heupel, M., Semmens, J. & Hobday, A. (2006) Automated acoustic tracking of aquatic animals: scales, design and deployment of listening station arrays. *Marine and Freshwater Research*, **57**(1), 1–13.

How, J. & de Lestang, S. (2012) Acoustic tracking: issues affecting design, analysis and interpretation of data from movement studies. *Marine and Freshwater Research*, **63**(4), 312–324.

Jolly, G. (1965) Explicit estimates from capture-recapture data with both death and immigration-stochastic model. *Biometrika*, **52**(1/2), 225–247.

Lebreton, J.D. & Pradel, R. (2002) Multistate recapture models: modelling incomplete individual histories. *Journal of Applied Statistics*, **29**, 353–369.

McConnell, B., Beaton, R., Bryant, E., Hunter, C., Lovell, P. & Hall, A. (2004) Phoning

home – A new GSM mobile phone telemetry system to collect mark-recapture data. *Marine Mammal Science*, **20**(2), 274–283.

Nakagawa, S. & Freckleton, R.P. (2008) Missing inaction: the dangers of ignoring missing data. *Trends in Ecology & Evolution*, **23**(11), 592–596.

Ovaskainen, O. (2004) Habitat-specific movement parameters estimated using mark-recapture data and a diffusion model. *Ecology*, **85**(1), 242–257.

Patterson, T.A., Thomas, L., Wilcox, C., Ovaskainen, O. & Matthiopoulos, J. (2008) State-space models of individual animal movement. *Trends in Ecology & Evolution*, **23**(2), 87–94.

Payne, N., Gillanders, B., Webber, D. & Semmens, J. (2010) Interpreting diel activity patterns from acoustic telemetry: the need for controls. *Marine Ecology Progress Series*, **419**(Nov), 295–301.

Pedersen, M.W., Patterson, T.A., Thygesen, U.H. & Madsen, H. (2011) Estimating animal behavior and residency from movement data. *Oikos*, **120**(9), 1281–1290.

Preisler, H.K., Ager, A.A., Johnson, B.K. & Kie, J.G. (2004) Modeling animal movements using stochastic differential equations. *Environmetrics*, **15**(7), 643–657.

Rowcliffe, J., Carbone, C., Jansen, P., Kays, R. & Kranstauber, B. (2011) Quantifying the sensitivity of camera traps: an adapted distance sampling approach. *Methods in Ecology and Evolution*, **2**(5), 464–476.

Royle, J.A. (2008) Modeling individual effects in the Cormack–Jolly–Seber model: a state–space formulation. *Biometrics*, **64**(2), 364–370.

Simpfendorfer, C., Heupel, M. & Hueter, R. (2002) Estimation of short-term centers of activity from an array of omnidirectional hydrophones and its use in studying animal movements. *Canadian Journal of Fisheries and Aquatic Sciences*, **59**(1), 23–32.

Thorstad, E., Økland, F., Rowsell, D. & McKinley, R. (2002) A system for automatic recording of fish tagged with coded acoustic transmitters. *Fisheries Management and Ecology*, **7**(4), 281–294.

List of Figures

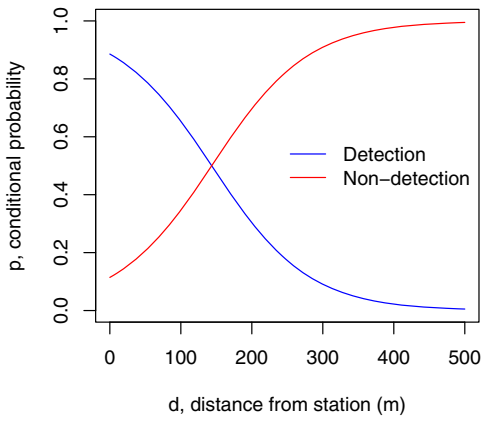

Figure 1: Detection and non-detection probabilities conditional on distance d from animal to station. Because of the dichotomous nature of the observations (either detection or non-detection) the two functions always sum to one for a given value of d. These curves can also be interpreted as location likelihood functions in the case of observing either a presence ($I_{i,t} = 1$, blue) or an absence ($I_{i,t} = 0$, red), see Fig. 2 and main text for more details. The detection function is sigmoidal with Set A parameter values (Table 2).

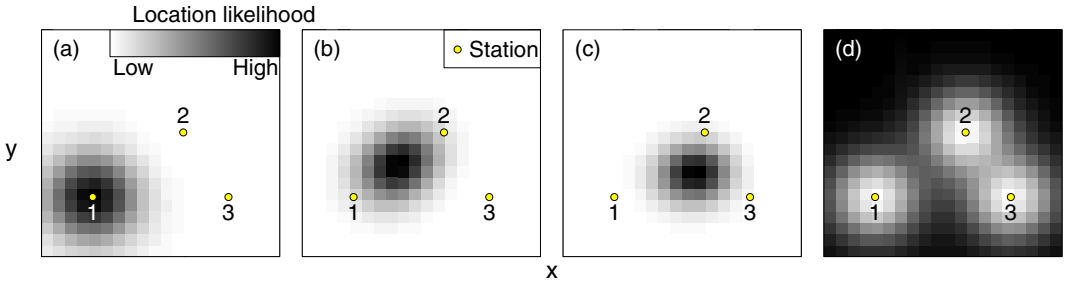

Figure 2: Examples of location likelihoods $L(x, y, t)$ in a three station setup where detection probability only depends on distance to station. (a) Detection at station 1 only. (b) Detection at station 1 and 2. (c) Detection at station 1, 2, and 3. (d) No detection. It is clear from (d) that absence data can provide useful information about the location of the animal.

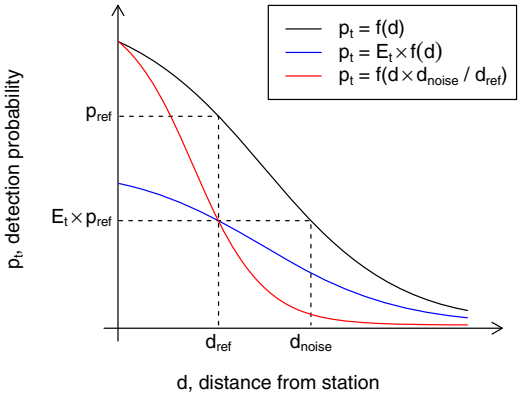

Figure 3: Effect of detection efficiency on detection function. Black: detection function under optimal conditions. Blue: scaling p-axis with E_t. Red: scaling d-axis with d_{noise}/d_{ref} as in $h(d, E_t)$.

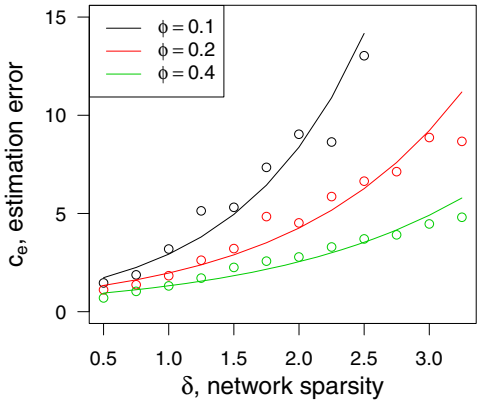

Figure 4: Simulation scenario 1: location estimation error (c_e,the ratio of rms location error to rms displacement, $\sigma\sqrt{dt}$) as a function of network sparsity and effective movement capacity (ϕ). Location estimation errors increase as a function of sparsity with increasing rate for decreasing values of movement speed (Z-test for slopes; $\phi = 0.1$ vs. $\phi = 0.2$: $Z = 4.79$, $P < 0.001$; $\phi = 0.2$ vs. $\phi = 0.4$: $Z = -16.6$, $P < 0.001$). For a slow moving animal (black line) estimation error converged to that obtained by a naïve estimator (movements estimated by a simple mean location) at $\delta = 2.5$ above which advanced estimation techniques seem unnecessary.

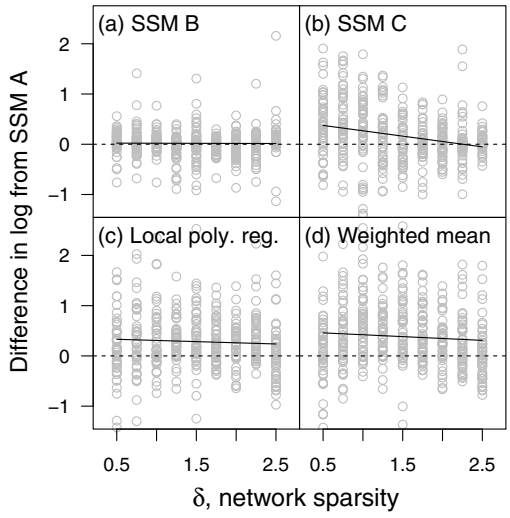

Figure 5: Simulation scenario 2: model estimation performance at slow speed (Table 1), solid line (represented by log rms location error minus log rms location error obtained with SSM A) as a function of network sparsity (δ). (a) SSM with slight misspecification of the detection function. (b) SSM with substantial misspecification of the detection function. (c) Local polynomial regression (Hedger *et al.*, 2008). (d) Weighted-mean method (Simpfendorfer *et al.*, 2002).

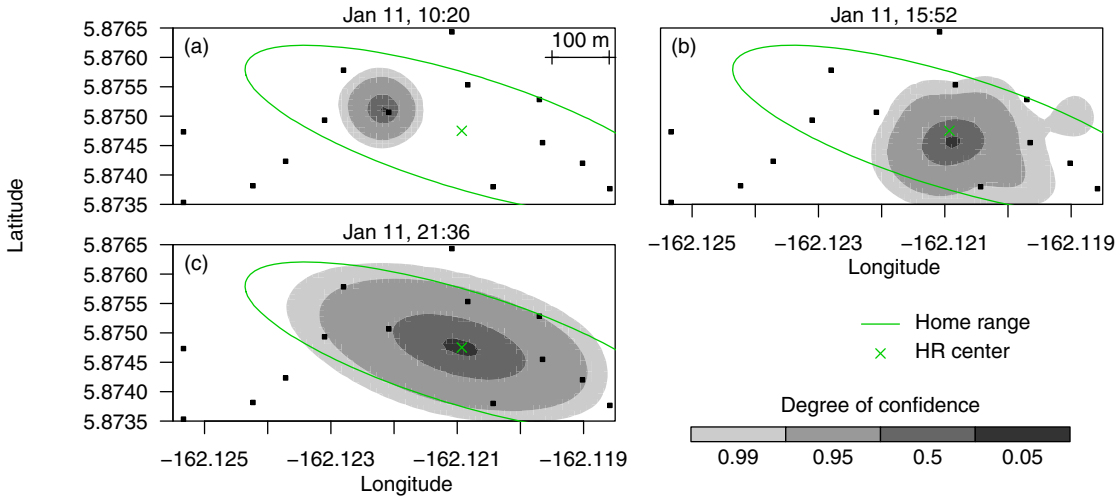

Figure 6: Confidence regions (grey contours) for the location of a humphead wrasse at Palmyra Atoll along with home range (HR, green line) estimated by the SSM. Dark points are observer stations. (a) a circular confidence region typical for day-light hours, here with a radius of approximately 50 m. (b) at times with no detections the fish is less likely to be located near stations, which can result in oddly shaped probability distributions. (c) after sunset the fish is believed to hide on the reef making it difficult to locate as indicated by the wide confidence region. Note that the distribution does not "avoid" observer stations as in (b). This is because detection efficiency is reduced on the reef after sunset (inferred from reference data). Animated confidence regions with point estimates of location showing detailed movement are found in Appendix S5

List of Tables

Speed	ϕ	n
low	0.1	441
moderate	0.2	1155
high	0.4	5772

Table 1: In simulations d_r and dt are fixed making ϕ proportional to movement rate. Lower speeds with a fixed sample interval leads to higher temporal auto-correlation and therefore smaller uncorrelated sample sizes (n) for log-linear model fitting.

Set	p_{max}	D_{50}	D_{95}
A	0.99	145	344
B	1.10	133	421
C	0.68	361	442

Table 2: Detection function parameter sets used in simulation studies. The values are based on published ranging experiment results (Table 2 in How & de Lestang, 2012).

Parameter	Estimate	95% CI
μ_x (°)	−162.1209	±69.2 m
μ_y (°)	5.8748	±25.5 m
$b_x \times 10^{-4}$ (s^{-1})	1.23	0.64; 2.36
$b_y \times 10^{-4}$ (s^{-1})	6.91	4.52; 10.6
$b_{xy} \times 10^{-4}$ (s^{-1})	2.10	0.21; 3.44
σ ($m{\cdot}s^{-0.5}$)	1.72	1.49; 1.98

Table 3: Parameter values estimated from field data. Confidence bounds for the home range centre (μ_x, μ_y) are given in metres to ease interpretation. Other confidence intervals (CIs) are given as upper; lower.