A quantitative approach to static sensor network design

Journal:	*Methods in Ecology and Evolution*
Manuscript ID:	MEE-14-06-320.R1
Manuscript Type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Pedersen, Martin; Technical University of Denmark, National Institute of Aquatic resources Burgess, Greg Weng, Kevin; University of Hawai'i, Oceanography
Keywords:	Spatial or time-series < Statistics, Monitoring < Population Ecology, Modelling < Population Ecology
Abstract:	1. Static sensor networks to observe animals are widely used in ecological, management and conservation research, but quantitative methods for designing these networks are underdeveloped. 2. In the context of aquatic systems, we present a method for quasi-optimal network design, which accounts for blocking of detections by obstacles, horizontal and vertical movement behaviour of the target animals, and type of research question (is the network intended for estimation of detailed movement or home range?). Optimal design is defined as the sensor configuration that maximises the expected number of unique animal detections. As finding the global optimum is generally time consuming we use a greedy algorithm instead, which places sensors optimally relative to already placed sensors. The design method requires access to topographic data of the study site and knowledge of the sensor detection range. 3. We illustrate the method with real topographic data from a rugose coral reef where network performance is highly influenced by detection shadowing. Network performance is visualised by a coverage map indicating the probability of detection at any location in the study area. The reported unique recovery rate summarises the expected ability of the network to collect data given the design constraints. Because sensors are placed sequentially the information gain per sensor can be evaluated and used as a proxy for sensor value. 4. The presented method formalises important considerations, when designing sensor networks, that were previously often based on heuristics and intuition. The method provides a guide to maximising the information potential of future monitoring studies as well as a means to improve existing networks. The method is available as an R package and can be tested via an online web tool.

Title: A quantitative approach to static sensor network design

Authors: Martin W. Pedersen[1,2,4], Greg Burgess, Kevin C. Weng[2,3]

[1] National Institute of Aquatic Resources, Technical University of Denmark, Charlottenlund Slot, Jægersborg Allé 1, 2920 Copenhagen, Denmark

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

[3] Present address: Department of Fisheries Science, Virginia Institute of Marine Science, Gloucester Point, Virginia 23062-1346 USA.

[4] Corresponding author: map@aqua.dtu.dk

Running title: Quantitative static sensor network design

Summary (Abstract)

1. Static sensor networks to observe animals are widely used in ecological, management and conservation research, but quantitative methods for designing these networks are underdeveloped.

2. In the context of aquatic systems, we present a method for quasi-optimal network design, which accounts for blocking of detections by obstacles, horizontal and vertical movement behaviour of the target animals, and type of research question (is the network intended for estimation of detailed movement or home range?). Optimal design is defined as the sensor configuration that maximises the expected number of unique animal detections. As finding the global optimum is generally time consuming we use a greedy algorithm instead, which places sensors optimally relative to already placed sensors. The design method requires access to topographic data of the study site and knowledge of the sensor detection range.

3. We illustrate the method with real topographic data from a rugose coral reef where network performance is highly influenced by detection shadowing. Network performance is visualised by a coverage map indicating the probability of detection at any location in the study area. The reported unique recovery rate summarises the expected ability of the network to collect data given the design constraints. Because sensors are placed sequentially the information gain per sensor can be evaluated and used as a proxy for sensor value.

4. The presented method formalises important considerations, when designing sensor networks, that were previously often based on heuristics and intuition. The method provides a guide to maximising the information potential of future monitoring studies

as well as a means to improve existing networks. The method is available as an R package and can be tested via an online web tool.

Key words: optimal design, detection function, home range, animal movement, bioacoustics, camera traps

1 Introduction

Static sensor networks are a wide-spread tool to continuously monitor presences of free-ranging animals enabling researchers to address fundamental ecological topics such as movement behaviour, home range, survival, and habitat utilisation. Within terrestrial ecology common static sensor network types include animal-borne radio-transmitters detected by fixed receivers (Mennill *et al.*, 2012), microphone arrays that record the sounds of untagged animals (Blumstein *et al.*, 2011), and camera traps that capture photographs of animals upon detection (Rowcliffe *et al.*, 2011). Within aquatic systems, networks of static sensors that acoustically detect tagged animals are among the only feasible means to gather ecological information. This technology has been applied to a wide range of species and environments: diadromous fishes (Solomon & Potter, 1988; Klimley *et al.*, 2013), reef fishes (Meyer *et al.*, 2010), deep water fishes (Weng, 2013), sharks (Yeiser *et al.*, 2008), turtles (Seminoff & Jones, 2006), and highly migratory pelagic fishes (Honda *et al.*, 2010). Generally, the use of static sensor networks is growing rapidly within ecology (Kessel *et al.*, 2013; Marques *et al.*, 2013) primarily as a means to monitor animals in environments where continuous human observation is difficult.

In light of the popularity of static sensor networks and the many factors influencing network performance it is surprising that quantitative guidelines for designing these networks are limited. Some studies have addressed technology-specific design considerations such as the choice of microphone type for animal sound recordings (Rempel *et al.*, 2013), the effect of camera trap spacing and total survey area on the number of species recorded (Tobler *et al.*, 2008), and the influence of obstacles and suspended matter on detection probability of aquatic acoustic sensors (Welsh *et al.*, 2012). However, a general framework for network design that integrates underlying assumptions and network performance metrics is lacking.

A fundamental consideration when designing static sensor networks is whether the network will recover sufficient and appropriate data to test the research question of interest. Different research questions require different network design. For example, if the research objective is to estimate detailed movements, overlapping detection zones within the preferred habitat of the species are important to accurately pinpoint location (Biesinger *et al.*, 2013). On the other hand, if home range is the focus of the study, increased spatial extent of the network to capture movement over different spatial scales is more important than detection overlap (Jackson, 2011). Biological aspects of the target animals such as known habitat preferences and site fidelity also influence network performance and should therefore be factored into design considerations. Many animals live in complex environments where detection is inhibited by physical obstacles. The data recovery rate of a static sensor network installed in such environments will therefore depend strongly on the placement of sensors relative to potential detection shadows.

This paper presents, in an aquatic context, a quantitative method for finding a quasi-optimal static sensor network design that is adapted to the research objective (estimating detailed movement or home range), assumed movement behaviour or spatial distribution of the target animals, and spatial environment in the form of shadowing effects and detection conditions. These aspects are important because they dictate the ability of static sensor networks to provide sufficient high-quality data to address a given scientific question. A number of performance metrics are furthermore introduced to facilitate a more standardised understanding and comparison of performance between competing designs and among separate networks.

2 Methods

The design method was developed focusing on static acoustic sensor networks for aquatic systems. The presentation of the method is therefore placed in this context. However, because many of the presented concepts apply broadly to static sensor networks such as camera traps, acoustics, or radio-based technology used in terrestrial systems (see discussion), a terminology with broader appeal, i.e. "sensor network", is maintained over "receiver array", which is the common term in underwater acoustics.

An overview of the mathematical symbols used to describe the method is provided in Table 1.

2.1 Optimality criterion

Optimising a network design involves weighing the value of data quantity and data quality. Data from static sensor networks comprise animal detections. Here, a detection is defined as an encounter of an animal by a sensor. Naïvely aiming to maximise the number of logged animal detections seems inappropriate as the resulting network would consist of sensors that are all positioned at the location where most animals are expected to be present. While this indeed is the optimal design in terms of data quantity, most detections will be duplicates (i.e. simultaneous detections of the same individual by different sensors) and therefore have low average information content. It seems more appropriate to maximise the number of unique detections, i.e. the number of non-duplicate detections. This can be viewed as maximising the data information content under the assumption that all detections have equal importance. However, in home range studies that require data from different spatial scales, detections may differ in information content depending on their location. For now we ignore this complexity and assume equal information content per detection. We therefore define optimality as the design that provides the maximum number of unique detections or, in other words, maximises the unique recovery rate. We will subsequently explore the case of detections with different information content.

Sensor placement is optimised in the horizontal plane only. Three dimensional optimal placement is theoretically possible, but in an open water environment without potential vertical obstacles the solution is trivial in that sensors are optimally placed as high off the bottom as possible. However, vertical placement close to the surface is often impratical because of boat traffic, currents, waves etc. Instead we assume that sensors are located at a constant predefined offset relative to the bottom.

2.2 Design prerequisites

In determining the optimal horizontal sensor configuration several design parameters are required. These design parameters constrain the optimisation problem. The spatial extents and topographic properties of the study site must be specified. This information is typically available on a spatial topographical grid (T) with a known grid cell size (dx). Another fundamental design parameter is the number of available sensors (n) for which to calculate the optimal configuration. Sensors are assumed to have omnidirectional detection capabilities in the horizontal plane.

An additional prerequisite to optimising design is knowledge about detection conditions at the study site. This can be quantified by conducting ranging experiments to approximate the average horizontal detection range (d_r) of a sensor at the study site. In this work the detection range is defined as the horizontal distance from the sensor where the probability of detecting an animal has decreased to 5%.

The possible target animals of static sensor networks range from a few tagged individuals over sub-populations to communities of possibly multiple species. Prior to network deployment some knowledge pertaining to the general type of movement behaviour or horizontal distribution of the target animals may be available possibly from pilot field studies or from the literature. This information can aid in optimising the sensor placement to target these animals. Specifically, prior knowledge of the horizontal animal distribution can be expressed in a spatial grid (U_0). For aquatic species a preferred vertical habitat range also provides information about the horizontal species distribution because regions with depths outside the preferred vertical habitat range can be excluded from the array of possible horizontal sensor locations.

[Table 1 about here.]

2.3 Optimising sensor placement

Sensors are placed sequentially using a so called greedy algorithm (Leiserson *et al.*, 2001). Placement of sensor $s \in \{1, \ldots, n\}$ requires information stored in a number of grids (matrices): the detection grid (D_s), the grid of undetected animals (U_s), and the goodness grid (G_s). All grids have dimensions identical to the topographical grid (T). For reasons of clarity we first give an overview of the algorithm, and then define the grids in the following sections.

The greedy algorithm places sensor s by following the steps:

1. Use the grid of undetected animals (U_{s-1}) and the topographical grid (T) to calculate the goodness grid (G_s) in which the value of each cell represents the ability of sensor s to detect animals in nearby cells if placed in that cell.

2. Place sensor s in grid cell (k_s, l_s), where $G_s(k_s, l_s)$ is the maximum value of G_s.

3. Calculate U_s by down-weighing values of U_{s-1} in proximity to sensor s using the detection grid (D_s).

Following this procedure for all n sensors results in the quasi-optimal sensor configuration (Fig. 1).

[Figure 1 about here.]

2.3.1 Detection grid

A cell in the detection grid, $D_s(i, j)$, contains the probability, $p(i, j, k_s, l_s)$, that sensor s placed in cell (k_s, l_s) detects animals that are located in cell (i, j). The detection probability $D_s(i, j)$ has two components: the first is the proportion $(\alpha[i, j])$ of the target animals' vertical distribution in (i, j) that is visible to sensor s; the second is the horizontal detection probability $(\beta[i, j])$ as described by the detection function (f). The detection function describes the decline in animal detection probability as a function of distance from the sensor (Kessel *et al.*, 2013). Since our focus is to optimise the horizontal sensor configuration, we only apply the detection function to horizontal distances.

The visible vertical range can be calculated using a three-dimensional line of sight algorithm similar to Akbarzadeh *et al.* (2012). A known vertical distribution of the target animals is modelled as a normal distribution around a preferred mean position in the water column. For example, the preferred position of a demersal species could be defined as a constant elevation off the bottom. By adjusting the standard deviation of this distribution the degree of confidence in the depth preference can be controlled. Thus, if no prior knowledge of the depth preference is available a large standard deviation can be specified. Given a mean and a standard deviation of the vertical species distribution, the visible proportion $(\alpha_s[i, j])$ of this distribution in cell (i, j) can be calculated as the area under the normal probability density function within the visible range of a sensor placed in (k_s, l_s) (Fig. 2).

[Figure 2 about here.]

The horizontal detection probability $(\beta_s[i, j])$ is calculated by evaluating the detection function with detection range d_r at the horizontal distance $d(i, j, k_s, l_s)$ from sensor s

to (i, j), i.e. $\beta_s(i, j) = f(d[i, j, k_s, l_s]; d_r)$. The combined detection probability in cell (i, j) is then calculated by assuming independence of the two contributions $p(i, j, k_s, l_s) = \alpha_s(i, j)\beta_s(i, j)$, thus

$$D_s(i, j) = p(i, j, k_s, l_s) \tag{1}$$

2.3.2 Grid of undetected animals

The grid U_s contains values in the interval $[0, 1]$ representing the available data potential (undetected animals) remaining within the region of interest after s sensors have been placed (Fig. 1). For $s = 0$, i.e. before any sensors are placed, U_0 contains the expected horizontal distribution of the target animals (Fig. 1b). In this context the target animals represent the focal species, population, or individuals of the study. The Eulerian (i.e. population-based) interpretation of $U_0(i, j)$ is as the proportion of the target population located within cell (i, j) at a given time. The alternative Lagrangian (i.e. individual-based) interpretation of $U_0(i, j)$ is as the probability of a moving individual being located within cell (i, j) at a given time. In both cases the sum of all values in U_0 is one. Thus, U_0 can be regarded as a probability distribution of the location of the target animals. With monitoring systems that rely on tagged individuals the Lagrangian interpretation is the most intuitive, whereas when studying untagged populations, e.g. visually or via sound recordings, the Eularian seems the most meaningful. The implementation of U_0 is unaffected by the interpretation.

As assessing the horizontal distribution of the target animals may be the very objective of the monitoring experiment, U_0 is generally unknown. However, knowledge of the target animals' movement behaviour or habitat preference may be available from the literature or pilot studies and can aid in making informed assumptions about general shape of U_0. For instance, a study species known to inhabit denser vegetation should have higher values of U_0 in these regions. As a result, the design algorithm will bias sensor placement toward densely vegetated regions. In essence, U_0 is an instrument the designing researcher can use to focus the sensor network in regions of interest. Thus, U_0 does therefore not necessarily need to reflect high animal density, but could also highlight protected areas or areas of ecological importance with low occupation (e.g. migration routes).

A complete lack of prior knowledge about the horizontal distribution can be modelled by a uniform distribution in U_0. In the Lagrangian interpretation, a uniform distribution arises when an individual follows a random walk movement model. For species known to prefer certain vertical habitats the values of cells in U_0 that do not accommodate the preferred vertical range are zero. The target population may be known to have a preferred horizontal location (μ_x, μ_y) within the study site, which can be reflected in U_0 via a bivariate normal with mean equal to (μ_x, μ_y) and covariance Σ (Fig. 1b). These parameters can be interpreted as a home range center and home range extents respectively. In Lagrangian terms, the individual movement model leading to this distribution is the Ornstein-Uhlenbeck (OU) process (Pedersen & Weng, 2013), which has a bivariate normal as long-term average (i.e. stationary) distribution. Thus, cell (i, j) in U_0 is then given by

$$U_0(i,j) = \frac{1}{K} N([i,j], [\mu_x, \mu_y], \Sigma), \tag{2}$$

where $N(\cdot)$ is a bivariate normal density function evaluated at (i, j) with mean (μ_x, μ_y) and covariance Σ. The constant $K = \sum_{i,j} N([i,j], [\mu_x, \mu_y], \Sigma)$ normalises U_0 such that all cells of U_0 sum to one. As a parallel to Bayesian statistics, U_0 can be viewed as reflecting the *a priori* information about the horizontal distribution of the target animals. Less certain prior knowledge is indicated by large values in Σ leading to a wider and less informative distribution. Extending this OU-based model to multiple regions of interest is trivial and can be represented via a mixture of normal density functions.

When sensor s is placed at (k_s, l_s) the value of cell (i, j) in U_s is calculated as

$$U_s(i,j) = U_{s-1}(i,j)(1 - D_s[i,j]), \tag{3}$$

recall here that $D_s(i,j) = p(i,j,k_s,l_s)$. Thus, U_s is similar to U_{s-1}, however with values of cells in proximity to sensor s down-weighed because the probability of finding undetected animals (unique detections) is reduced in those cells.

2.3.3 Goodness grid

The goodness grid (G_s) holds values $G_s(i,j)$ representing the goodness of placing sensor s in cell (i,j) after $s-1$ sensors have been placed (Fig. 1c and 1e). The goodness value in a cell is defined as the ability of a sensor located in that cell to detect animals. The goodness therefore depends on the detection function, the topography surrounding the cell, the vertical placement of the sensor in the cell, and on the horizontal and vertical distribution of undetected animals in the surrounding cells. For instance, a sensor placed in close proximity to large objects is likely to have a reduced goodness because of detection shadows (obscured line of sight between animal and sensor). In contrast, a sensor placed in an open environment or above potential obstacles is likely to result in an improved overall detection capability and therefore increased goodness. Generally, the key for a sensor to obtain high goodness is to have an unobscured line of sight to the areas and depths that the target animals are most likely to inhabit.

The goodness value of a sensor placement in cell (k,l) is

$$G_s(k,l) = \sum_{i,j} U_{s-1}(i,j)p(i,j,k,l), \tag{4}$$

where $p(i,j,k,l)$ is the detection probability in cell (i,j) of a hypothetical sensor placed in cell (k,l). The whole goodness grid is calculated by repeating the procedure in Eqn. (4)

for all valid (k, l). The range of locations valid for sensor placement may be reduced by excluding regions that fall outside topographical constraints (e.g. shallower or deeper than certain limits). Invalid locations will therefore have zero goodness.

2.4 Network metrics

A number of metrics can be calculated to assess the quality of a generated network design.

2.4.1 Goodness grid

The initial goodness grid before sensors are placed (G_1) highlights locations that are most suited for sensor placement. This information could in theory be used to manually place sensors. The initial goodness grid also aids in understanding the optimal sensor placements reported by the greedy algorithm.

2.4.2 Coverage grid

The value of a grid cell in the coverage grid (C_s) after s sensors are placed is the probability of at least one sensor detecting a signal emitted in that grid cell. For grid cell (i, j) the coverage is

$$C_s(i, j) = 1 - \prod_{r=1}^{s}(1 - D_r[i, j]). \tag{5}$$

The coverage grid when all sensors are placed (C_n) is a useful tool for spatially visualising the aggregate detection zone of the network.

2.4.3 Unique recovery rate

The unique recovery rate (ρ_s) is an indicator of the overall performance of a network with s sensors. In short, ρ_s is calculated as the expected proportion of the assumed horizontal distribution (U_0) that is covered by the network. For individuals with attached transmitters ρ_s can be interpreted as the expected proportion of all emitted signals that would be detected by at least one sensor. The unique recovery rate is bounded between zero and one ranging from no coverage at all to perfect coverage. The unique recovery rate is

$$\rho_s = \sum_{i,j} U_0(i,j) C_s(i,j), \tag{6}$$

which is basically a sum of the overlap between the initial behaviour grid (U_0) and the coverage grid (C_s). Limited spatial overlap between U_0 and C_s will result in a low ρ_s and vice versa.

Owing to the high cost of sensors the spatial extent of the animal distribution as represented by U_0 is likely to exceed the detection coverage (C_n). Therefore, even with good individual sensor performance, the overall performance of the array in terms of ρ_s might be relatively low. Adding more sensors (increasing n) will intuitively improve the unique recovery rate. With sensor networks that utilise transmitters ρ_s may be improved by investing in more powerful transmitters, which often increases detection range and therefore results in an improved coverage.

2.4.4 Sparsity

The network sparsity (Pedersen & Weng, 2013) is defined as

$$\delta = \frac{a}{2d_r}. \tag{7}$$

Here a is an absolute measure of sensor closeness calculated as the median of $\{a_1, \ldots, a_n\}$, where a_s is the distance from sensor s to its nearest neighbouring sensor. If $\delta < 1$, the network mostly has detection functions that overlap, whereas $\delta > 1$ implies a sparser network with mostly non-overlapping detection functions. Thus, for $\delta < 1$ the spatial density of sensors is high, which will make the network suited for estimating detailed animal movement. Sparser arrays ($\delta > 1$) will result in higher uncertainty of location estimates, but will in turn have larger spatial extents for a fixed number of sensors thus improving the ability of the network to resolve the entire movement range of the target animals.

2.4.5 Absolute recovery rate

The absolute recovery rate (ω_s) is the expected total number of detections relative to emitted signals of a network with s sensors. The equation for calculating ω_s is

$$\omega_s = \sum_{r=1}^{s} \sum_{i,j} U(i,j) D_r(i,j). \tag{8}$$

If a large number of multiple detections of the same animal is expected ω_s can, in contrast to ρ_s, exceed 100%. This will occur for networks with a high density of sensors and good detection coverage, i.e. substantial overlap between U_0 and C_n. Thus, the ratio of unique to absolute recovery rate is also an indicator of network sparsity. If $\omega_s = \rho_s$ the network has non-overlapping detection function, whereas $\omega_s > \rho_s$ indicates overlap of sensors' detection functions.

While the absolute recovery rate is a useful metric it is important to note that it can be maximised by simply placing all sensors in the cell containing the maximum value of the goodness grid. Biologically, however, such an array configuration seems sub-optimal since the information gain from detecting a signal multiple times decreases rapidly per extra detection.

2.4.6 Sensor value

Adding sensors sequentially following a greedy algorithm will gradually improve the unique recovery rate of the network. Plotting ρ_s as a function of number of sensors s enables monitoring of the convergence of ρ_s toward the 100% asymptote. Another useful metric is the value of sensor s, which is defined as

$$\upsilon_s = \rho_s - \rho_{s-1}.$$

Thus, the value of sensor s is the increase in unique recovery rate when the sth sensor is added. Since $\rho_s \geq \rho_{s-1}$ the sensor value is always non-negative.

The sensor value can be interpreted as the added information provided by placing sensor s. This enables the designing researcher to evaluate whether the information gain of deploying additional sensors outweighs the associated cost. The sensor value could also be used as a tool to distribute a larger number of sensors among different research projects by maximising the total sensor value of all projects. Comparing sensor values between projects may be complicated, however, because of potential differences in economic and scientific premise.

2.5 Adapting design to research question

The horizontal extent and number of sensors of a static sensor network are determining for the range of research questions to be addressed. Designing networks that maximise the unique recovery rate typically results in static sensor networks with low sparsity ($\delta < 1$), which are suited for estimation of detailed movements. For estimating home range of mobile species where the expected movement range is large relative to the sensor detection range, the detection coverage and unique recovery rate will be poor regardless of sensor placement. In this case, increasing the network sparsity to distribute sensors more evenly throughout the study region would improve the network's ability to resolve the movement range of the target animals and provide an improved basis for home range estimation.

The mechanism that controls the spacing between sensors is the down-weighing, or suppression, of the goodness grid by the detection grid (D_s) as described by Eqn. (3) and

(4). To obtain increased spacing between sensors, we calculate a suppression grid (W_s) defined as

$$W_s(i, j) = f(d[i, j, k_s, l_s]; qd_r), \tag{9}$$

which is similar to the definition of $\beta_s(i, j)$, however with the detection range replaced by a suppression range specified as a product of d_r and the suppression range factor q. Thus, the sparsity of the resulting network is proportional to the chosen value of q. Using W_s the algorithm for placing sensor s is simplified to two steps:

1. Place sensor s in grid cell (k_s, l_s), where $G_s(k_s, l_s)$ is the maximum value of G_s.

2. Calculate G_{s+1} by down-weighing values of G_s in proximity to sensor s using the suppression grid (W_s). Specifically do $G_{s+1}(i, j) = G_s(i, j)(1 - W_s[i, j])$ for all (i, j).

The iteration is initialised with G_1 calculated by Eqn. (4). If $q = 1$, this suppression scheme is an approximation to that of Eqn. (3) and Eqn. (4), with the difference that obstacles blocking transmissions are not accounted for. The degree of approximation when disregarding obstacles is reduced for $q > 2$ because sensor detection zones do not overlap. In general for $q > 1$, goodness values at locations close to placed sensors are down-weighed more than if $q = 1$. This relaxes the assumption of equal data information content because the suppression range exceeds the detection range making locations distant from sensors become more attractive to new sensor placements. Thus, the resulting network design is optimal in terms of the unique recovery rate under the constraint of a required minimum sparsity given by the suppression factor q.

Factors influencing the choice of q are site and study specific and will depend on available prior knowledge of the general movement range and preferred habitat of the target animals, and on the research question. For instance, estimating the home range of a species that is expected to exhibit large scale movements would benefit from a high value of q resulting in widely distributed sensors that are more likely to also capture extreme movements. Depending on the amount and quality of prior information, network designs for a range of q values can be calculated and evaluated with respect to unique recovery rate, feasibility of implementation etc. Thus, the specific choice of q relies on informed judgment by the designing researcher. Devising rigorous methods for translating prior knowledge to an optimal choice of q is outside the scope of this work, but is a potential avenue for future research.

[Figure 3 about here.]

2.6 Numerical implementation

The outlined design algorithm is available via the R package "acoustic" (Supplementary material). Furthermore, a web-based graphical user-interface is available for download and local install via https://github.com/gregorylburgess/acoustic-deploy. For demonstration purposes, the web interface can be accessed with some restrictions at http://www.soest.hawaii.edu/PFRP/acoustic/pages/.

2.7 Example

The design algorithm is illustrated using two scenarios both utilising a 5 meter resolution bathymetry (http://www.soest.hawaii.edu/pibhmc/pibhmc_pria_pal_bathy.htm, accessed 21 March 2014) from a shallow water tropical reef setting at Palmyra Atoll located in the central tropical Pacific Ocean. We assume that $d_r = 120$ meters as determined via ranging experiments. Our hypothetical study species is attracted to a complex reef structure, the Rubble Pile, located in the western part of the atoll (Fig. 4a and 4b). The study species is a demersal reef-associated fish so we assume a mean elevation off the bottom of 0.5 m and a standard deviation of 1.5 m. Six sensors are available for the experiment ($n = 6$). In scenario 1) the network is designed to monitor detailed movements near the Rubble

Pile by optimising the unique recovery rate. In scenario 2) the network is designed to determine the home range of the study species. To achieve a wider distribution of sensors the suppression range factor is set to $q = 3$. Code to replicate the results of these scenarios is an example in the "acoustic" R package (Supplementary material).

3 Results

The initial goodness grid before any sensor was placed (G_0, Fig. 5) generally has larger values near the Rubble Pile where the animal concentration was assumed to be high. It appears, however, that because of detection shadows only few areas atop the Rubble Pile itself are suited for sensor placement. The areas around the Rubble Pile generally have higher goodness, but with substantial variation in goodness as a result of the rugose environment.

In scenario 1 the optimal sensor configuration (Fig. 4c) resulted in a sparsity of $\delta = 0.68$, a unique recovery rate of $\rho_6 = 0.131$ (Fig. 6a), and an absolute recovery rate of $\omega_6 = 0.139$. The greedy algorithm placed the sensors in the region around the Rubble Pile structure where the animal concentration is expected to be high (Fig. 4b). The suppression range factor of $q = 1$ resulted in network coverage with mostly overlapping detection functions (Fig. 4c). This is reflected in the sparsity parameter ($\delta < 1$) and also by the fact that $\omega_s \geq \rho_s$, which occurs when detection functions overlap.

In scenario 2 the optimal sensor configuration (Fig. 4d) resulted in a sparsity of $\delta = 1.30$, a unique recovery rate of $\rho_6 = 0.096$ (Fig. 6a), and an absolute recovery rate of $\omega_6 = 0.096$. The first sensor was placed at the same location as in scenario 1. As a result of the increased suppression range factor ($q = 3$) the subsequent sensors were distributed wider within the expected movement range of the study species (Fig. 4d) resulting in $\delta > 1$, which is better suited for home range estimation.

Optimising the network design with estimation of detailed movement in mind (scenario 1) resulted in a higher unique recovery rate than the design optimised for home range estimation (scenario 2). Furthermore, the sensor value v_s were consistently lower in scenario 2 than in scenario 1 except for the first sensor, which was the same (Fig. 6b). By construction, the greedy algorithm will always provide the highest ρ_n when maximising the unique recovery rate, however in studies focusing on home range estimation detection of extreme movements is more valuable than collecting many detections near the home range center.

[Figure 4 about here.]

[Figure 5 about here.]

[Figure 6 about here.]

4 Discussion

We have presented a method for quasi-optimal design of static sensor networks, which can account for signal blocking, type of research question, and species distribution and habitat preference. The resulting design is quasi-optimal because each sensor placement is optimal relative to already placed sensors, but not relative to subsequently placed sensors. Still, the method formalises important design decisions that were previously often based on heuristics and intuition and therefore enables researchers to base sensor network design on quantitative results.

The presented method has a number of limitations that are important to keep in mind. The quality of the resulting network design depends on the resolution of the available topography grid. The method is therefore only able to account for fine scale topographic features if a high resolution topography is provided. Furthermore, in the network installation phase, local conditions may inhibit the exact placement of sensors resulting in network properties that are likely to deviate from those calculated theoretically. The method optimises horizontal sensor placement and does therefore not account for obstacles in the vertical direction such as would occur in caves. Theoretically, extending the algorithm to work in three dimensions is trivial, however this would require access to vertically resolved topography data. The optimal sensor configuration will, for these reasons, be an approximation to the network installed in practice and should therefore be used as a theoretical guideline for network design rather than an absolute protocol for implementation.

The method uses a temporally constant detection range, which may be an unrealistic assumption in many environments because of variation in biological, physical or anthropogenic activity affecting detection conditions, e.g. diel variation in background noise influencing the efficiency of acoustic networks (Kessel *et al.*, 2013). It is therefore advisable to optimise network design using different detection ranges to evaluate potential implications for network performance.

While the main advantage of the method is in the design phase of a new static sensor network the presented network metrics can also be used to assess the theoretical performance of existing networks. Visualising the coverage of an existing network with known sensor locations may aid in interpreting already collected data e.g. by highlighting potential shadow zones. Similarly, the performance of hypothetical manually designed networks can be evaluated using the network metrics. The unique recovery rate, sparsity, and coverage metrics indicate whether the network is particularly suited to address certain research questions and may provide guidance when expanding existing networks.

The greedy design algorithm places sensors sequentially resulting in a quasi-optimal configuration. Simulated annealing (Van Laarhoven & Aarts, 1987) and evolutionary methods (Deb *et al.*, 2001) are examples of algorithms that search for the global optimum, i.e. the configuration resulting in the maximum possible unique recovery rate. These algorithms, however, often require hand-tuning of optimisation parameters and are not guaranteed to find the global optimum (Van Laarhoven & Aarts, 1987; Hansen & Ostermeier, 2001). In contrast, the greedy algorithm has low computational complexity and it always converges to an optimum without requiring a deep knowledge of the effect of tuning parameters. Additionally, because the greedy algorithm is deterministic it is straightforward to expand presented method by implementing alternative optimisation criteria such as to place sensors until a target unique recovery rate is achieved or until the sensor value drops below a certain threshold.

The design method has been presented in the context of static acoustic sensor networks for aquatic systems. While this is the most immediate application of the method owing to the widespread use of this technology, the presented framework also has potential use within other branches of ecology. Regardless of the specific application the method has two main requirements: 1) an estimate of sensor detection range; 2) high resolution topographical information of the study site. Analogous to aquatic systems, terrestrial ecologists use animal-borne transmitters together with static receivers for monitoring species (Mennill *et al.*, 2012). The detection range for this technology can be quantified by range tests, however obtaining accurate information about topography and obstacles could pose a challenge in particular in dense forest environments where canopy may obscure satellite-based data collection. An alternative technology is microphone arrays that enable terrestrial ecologists to monitor wild untagged animals by recording their sounds (Blumstein *et al.*, 2011). In this context, reliable estimation of detection range may prove challenging as it is a function of both microphone type and sound characteristics (i.e. higher frequency bird song may have longer range than lower frequency mammalian grunts). Camera traps that take photographs of animals upon detection classify as a static visual-based sensor network (Rowcliffe *et al.*, 2011). A notable difference from acoustic data collection is that the sensors (cameras) are directional. As methods for quantifying cameras' detection zone are well-developed (Rowcliffe *et al.*, 2011), the main challenge of using the presented design approach therefore lies in optimising not only sensor location but also direction. Theoretically, adding sensor angle to the parameter space is trivial (Akbarzadeh *et al.*, 2012), however a significant associated increase in computation time must be expected. While sound may penetrate low-density obstacles such as light vegetation, the quality of camera trap data depend critically on an unobscured line-of-sight. Thus, in addition to topographical information, the performance of the presented method in the context of camera traps will also depend on the availability of detailed information about vegetation, which, as a further complication, may vary in time. Overall, while adapting the presented method to non-aquatic systems may require additional research efforts, it is clear that the method has potential to also become an important resource in the design and analysis of terrestrial static sensor networks.

Static sensor networks are generally costly to acquire, install and maintain, particularly in unpredictable aquatic environments. The presented method enables objective comparison of manual and quasi-optimal network designs providing a basis for assessing the resources needed to address conservation and management objectives. In addition the method provides a theoretical platform for optimising the data potential of available resources and aids researchers in making well-informed decisions when designing future static sensor networks.

5 Acknowledgements

KCW and MWP were supported by the Pelagic Fisheries Research Program under Cooperative Agreements NA17RJ1230 and NA09OAR4320075 between the Joint Institute for Marine and Atmospheric Research and NOAA. Research at Palmyra Atoll was funded by a grant to KCW from NOAA's Coral Reef Conservation Program NA05OAR4301108 and

NA09OAR4300219. We thank Stephen Scherrer, Johnnoel Ancheta and Dodie Lau.

References

Akbarzadeh, V., Gagné, C., Parizeau, M., Argany, M. & Mostafavi, M. A. (2012). Probabilistic Sensing Model for Sensor Placement Optimization Based on Line-of-Sight Coverage. *IEEE Transactions on Instrumentation and Measurement*, **62**, 293–303.

Biesinger, Z., Bolker, B. M., Marcinek, D., Grothues, T. M., Dobarro, J. A. & Lindberg, W. J. (2013). Testing an autonomous acoustic telemetry positioning system for fine-scale space use in marine animals. *Journal of Experimental Marine Biology and Ecology*, **448**, 46–56.

Blumstein, D. T., Mennill, D. J., Clemins, P., Girod, L., Yao, K., Patricelli, G., Deppe, J. L., Krakauer, A. H., Clark, C., Cortopassi, K. A. *et al.* (2011). Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus. *Journal of Applied Ecology*, **48**, 758–767.

Deb, K. *et al.* (2001). *Multi-objective optimization using evolutionary algorithms*, *Volume 2012*. John Wiley & Sons Chichester.

Hansen, N. & Ostermeier, A. (2001). Completely derandomized self-adaptation in evolution strategies. *Evolutionary computation*, **9**, 159–195.

Honda, K., Hobday, A. J., Kawabe, R., Tojo, N., Fujioka, K., Takao, Y. & Miyashita, K. (2010). Age-dependent distribution of juvenile southern bluefin tuna (Thunnus maccoyii) on the continental shelf off southwest Australia determined by acoustic monitoring. *Fisheries Oceanography*, **19**, 151–158.

Jackson, G. D. (2011). The Development of the Pacific Ocean Shelf Tracking Project within the Decade Long Census of Marine Life. *PloS one*, **6**, e18999.

Kessel, S., Cooke, S., Heupel, M., Hussey, N., Simpfendorfer, C., Vagle, S. & Fisk, A. (2013). A review of detection range testing in aquatic passive acoustic telemetry studies. *Reviews in Fish Biology and Fisheries*, pp. 1–20.

Klimley, A. P., MacFarlane, R. B., Sandstrom, P. T. & Lindley, S. T. (2013). A summary of the use of electronic tagging to provide insights into salmon migration and survival. *Environmental Biology of Fishes*, **96**, 419–428.

Leiserson, C. E., Rivest, R. L., Stein, C. & Cormen, T. H. (2001). *Introduction to algorithms*. The MIT press.

Marques, T. A., Thomas, L., Martin, S. W., Mellinger, D. K., Ward, J. A., Moretti, D. J., Harris, D. & Tyack, P. L. (2013). Estimating animal population density using passive acoustics. *Biological Reviews*, **88**, 287–309.

Mennill, D. J., Doucet, S. M., Ward, K.-A. A., Maynard, D. F., Otis, B. & Burt, J. M. (2012). A novel digital telemetry system for tracking wild animals: a field test for studying mate choice in a lekking tropical bird. *Methods in Ecology and Evolution*, **3**, 663–672.

Meyer, C. G., Papastamatiou, Y. P. & Clark, T. B. (2010). Differential movement patterns and site fidelity among trophic groups of reef fishes in a Hawaiian marine protected area. *Marine Biology*, **157**, 1499–1511.

Pedersen, M. W. & Weng, K. C. (2013). Estimating individual animal movement from observation networks. *Methods in Ecology and Evolution*.

Rempel, R. S., Francis, C. M., Robinson, J. N. & Campbell, M. (2013). Comparison of audio recording system performance for detecting and monitoring songbirds. *Journal of Field Ornithology*, **84**, 86–97.

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**, 464–476.

Seminoff, J. A. & Jones, T. T. (2006). Diel movements and activity ranges of Green turtle (Chelonia mydas) at a temperate foraging area in the Gulf of California, Mexico. *Herpetological Conservation and Biology*, **1**, 81–86.

Solomon, D. & Potter, E. (1988). First results with a new estuarine fish tracking system. *Journal of Fish Biology*, **33**, 127–132.

Tobler, M., Carrillo-Percastegui, S., Leite Pitman, R., Mares, R. & Powell, G. (2008). An evaluation of camera traps for inventorying large-and medium-sized terrestrial rainforest mammals. *Animal Conservation*, **11**, 169–178.

Van Laarhoven, P. J. & Aarts, E. H. (1987). *Simulated annealing*. Springer.

Welsh, J., Fox, R., Webber, D. & Bellwood, D. (2012). Performance of remote acoustic receivers within a coral reef habitat: implications for array design. *Coral reefs*, **31**, 693–702.

Weng, K. C. (2013). A pilot study of deepwater fish movement with respect to marine reserves. *Animal Biotelemetry*, **1**, 1–9.

Yeiser, B., Heupel, M. & Simpfendorfer, C. (2008). Occurrence, home range and movement patterns of juvenile bull (Carcharhinus leucas) and lemon (Negaprion brevirostris) sharks within a Florida estuary. *Marine and Freshwater Research*, **59**, 489–501.

List of Figures

6 a) Unique recovery rate and b) the sensor value as a function of number of sensors placed in the two example scenarios. Because the greedy algorithm places sensors sequentially starting with the best, the sensor value will always be a decreasing function. Scenario 1 aims to maximise the unique recovery rate while scenario 2 is designed for home range estimation ($q = 3$) and therefore has a lower recovery rate but a higher network sparsity (Fig. 4). 35

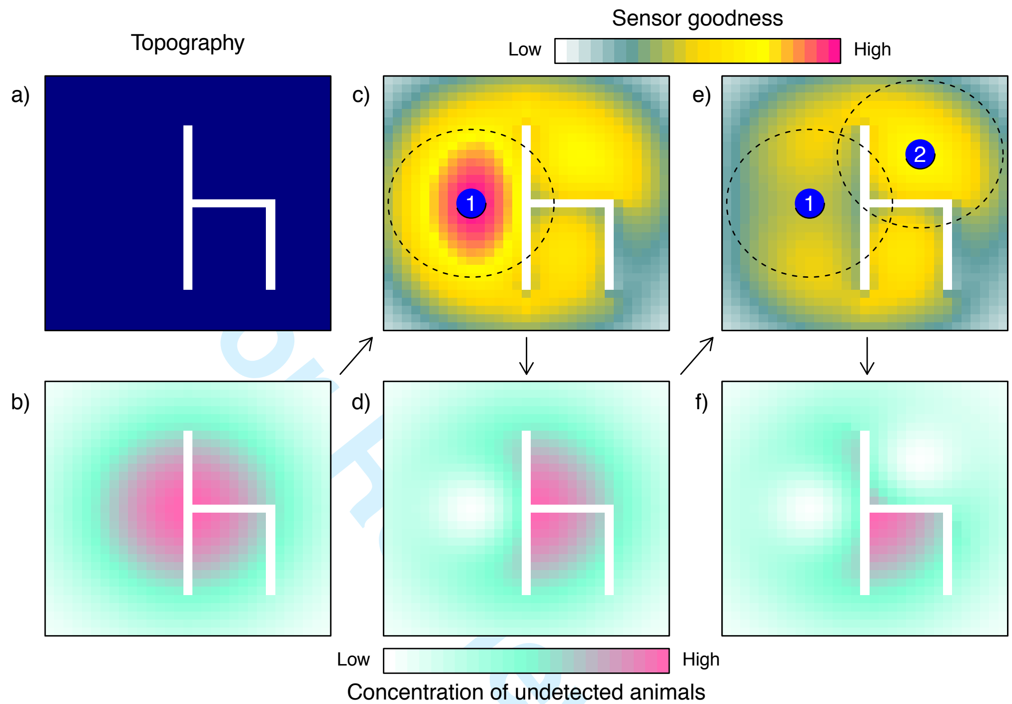

Figure 1: Overview of greedy algorithm using a simple artificial topography with arrows indicating the algorithmic flow. Solid blue circles indicate sensor locations with number indicating order of placement. Dashed lines indicate sensor detection range d_r. a) Topographical grid (T), primarily flat (dark blue) with an h-shaped island acting as potential obstacle (white). b,d,f) Grid of undetected animals; b) shows an initial distribution of animals (U_0) generated from a Ornstein-Uhlenbeck (OU) process with attraction toward the center of the study area; d) shows U_1, which has a lower concentration of undetected animals in the left part of the region as an effect of sensor 1; f) shows U_2 as affected by sensors 1 and 2. c,e) Goodness grids; c) shows G_1, indicating that sensor 1 should be placed in the grid cell with the maximum value of G_1; e) shows G_2 with goodness values resulting from placement of sensor 1, and indicating that sensor 2 should be placed in the upper right corner as a result of fewer detection shadows as compared to the lower right corner.

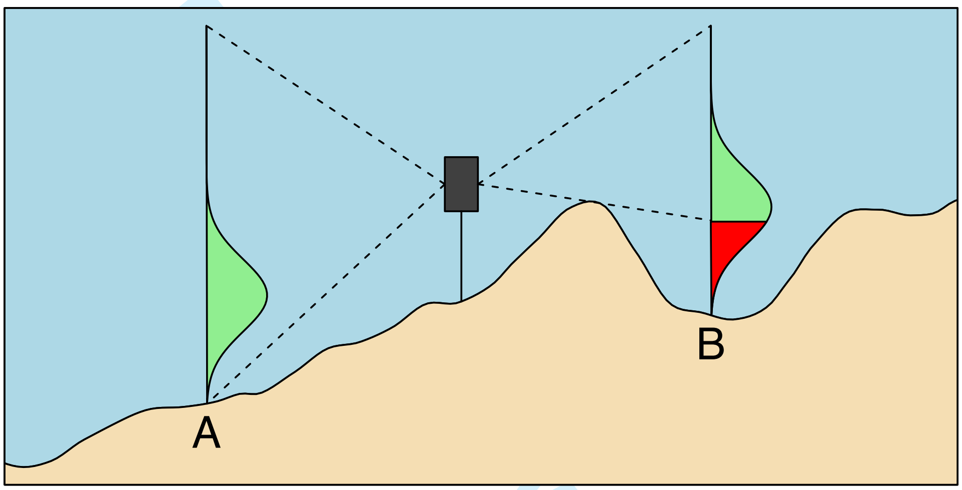

Figure 2: Sketch showing how the visible vertical range is calculated. At point A 100% of the vertical distribution of the species is visible to the sensor whereas at point B only 66% of the vertical distribution is visible. The green area of the normal curve indicates the visible part of the species' vertical distribution at a point, and red indicates the obscured part. The visible proportion of the vertical species distribution is defined as the area of the visible (green) part of the vertical distribution.

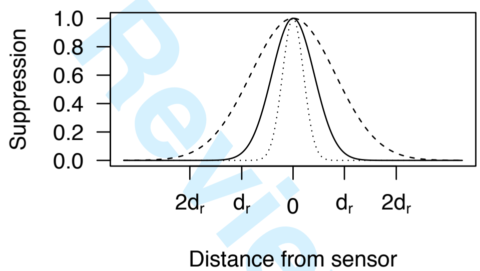

Figure 3: Example of a custom suppression function with $q = 2$ (dashed line) and $q = 0.5$ (dotted line) relative to using the horizontal detection function (solid line) for suppression ($q = 1$). The detection range (d_r) is the horizontal distance from the sensor where the probability of detecting an animal has decreased to 5% of its maximum.

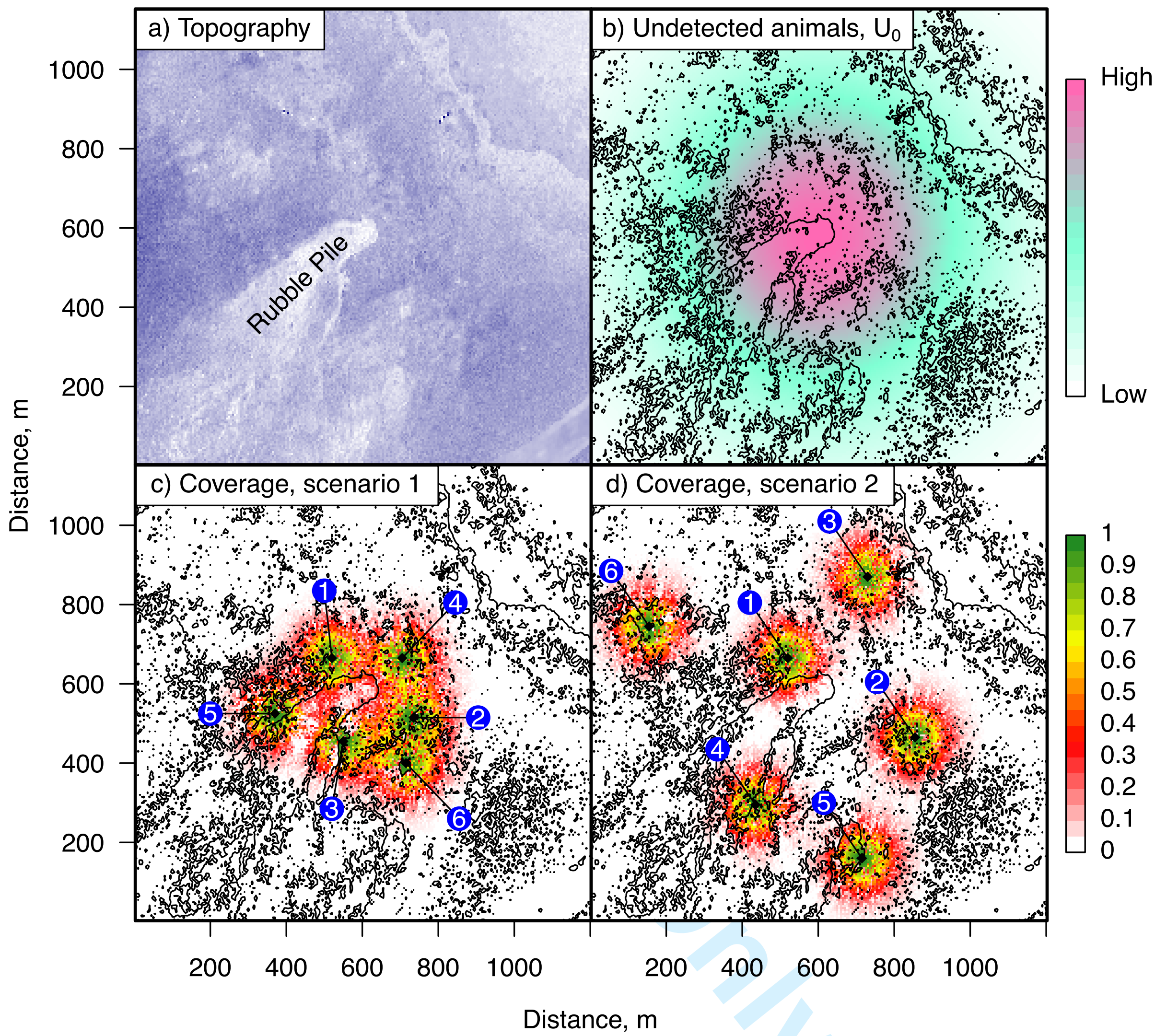

Figure 4: Design result of scenarios 1 and 2. a) depth range is $[-25, 0]$ meters with lighter colors indicating shallower depths. b) the reef has a particularly shallow feature (Rubble Pile) around which the fish are located. Contour lines indicate topography. c) sensor locations and resulting acoustic coverage when network design focuses on detailed movement (scenario 1). Numbers indicate order of sensor placement. The shallowness of the Rubble Pile forces sensors to be placed around this feature. d) sensor locations and coverage when network design focuses on home range (scenario 2) resulting in a sparser network.

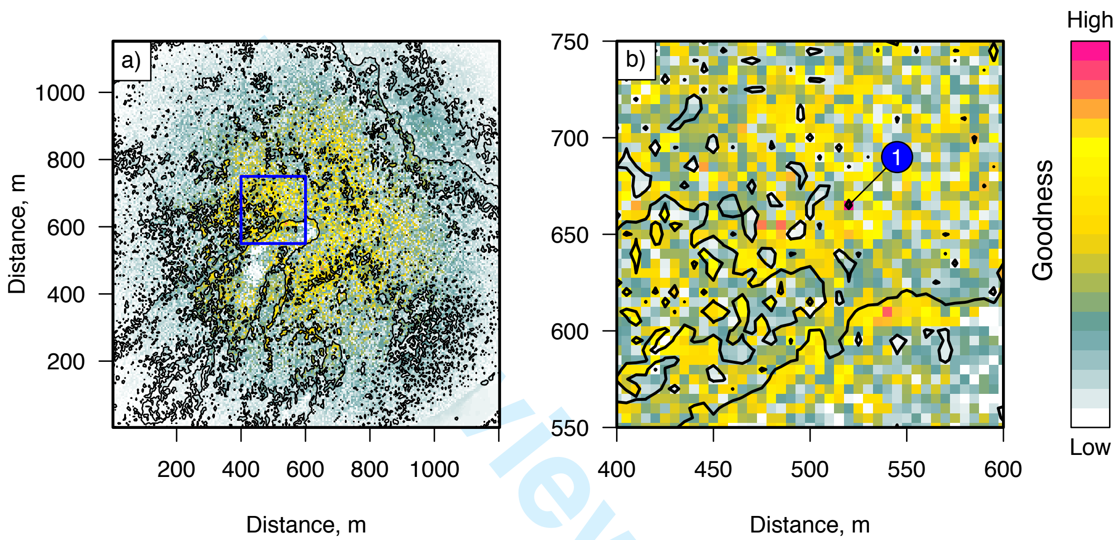

Figure 5: Initial goodness grid (G_1) of scenario 1 and 2. Prior to the placement of sensors, the initial goodness grid indicates favourable and unfavourable locations for sensor placement given an assumed distribution of the target species (U_0, Fig. 4b) and while accounting for signal shadowing by the topography. a) shows the whole study area. b) shows a zoom (indicated by the blue line in a) of the location of maximum goodness and therefore optimal placement of sensor 1 (numbered circle). Holes (e.g. western part of b) and shallow parts of the Rubble Pile appear unsuited for sensor placement, while elevated areas surrounded by flatter topography have higher goodness.

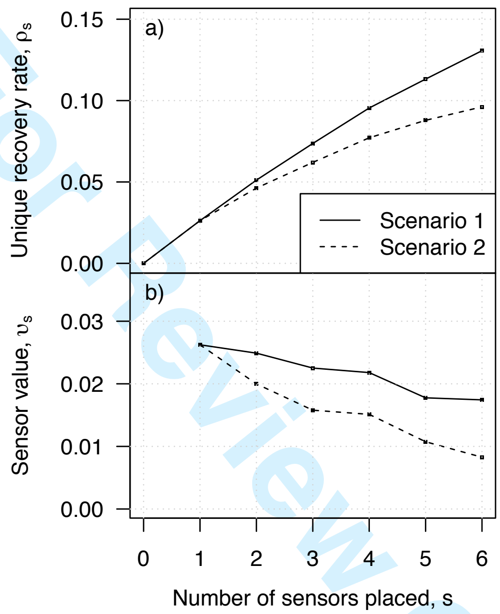

Figure 6: a) Unique recovery rate and b) the sensor value as a function of number of sensors placed in the two example scenarios. Because the greedy algorithm places sensors sequentially starting with the best, the sensor value will always be a decreasing function. Scenario 1 aims to maximise the unique recovery rate while scenario 2 is designed for home range estimation ($q = 3$) and therefore has a lower recovery rate but a higher network sparsity (Fig. 4).

List of Tables

Symbol	Meaning
a	Distance from sensor to nearest other sensor.
α	Visible portion of vertical distribution of animal.
β	Horizontal detection probability.
C_s	Coverage grid.
D_s	Detection grid.
$d(i, j, l, k)$	Distance between cell (i, j) and (k, l).
d_r	Detection range.
dx	Spatial grid cell size.
δ	Network sparsity.
$f(d; d_r)$	Detection function with detection range d_r evaluated at d.
G_s	Goodness grid.
K	Normalizing constant for OU process.
i	Longitudinal position of animal in the grid.
j	Latitudinal position of animal in the grid.
k	Longitudinal position of sensor.
l	Latitudinal position of sensor.
μ_x	Mean longitude of OU process.
μ_y	Mean latitude of OU process.
N	Bivariate normal distribution.
n	Number of sensors to use in the design.
ω_s	Absolute recovery rate.
$p(i, j, k, l)$	The probability of a sensor in (k, l) detecting an animal in (i, j).
q	Suppression range factor.
ρ_s	Unique recovery rate.
Σ	Covariance of OU process.
s	Sensor number.
T	Topographical grid.
U_s	Grid of undetected animals.
U_0	Expected distribution of animals.
υ_s	Value of sensor s.
W_s	Suppression grid.

Table 1: Nomenclature.