

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/224182030>

# Statistical Interaction Modeling of Bovine Herd Behaviors

Article in IEEE Transactions on Systems Man and Cybernetics Part C (Applications and Reviews) · December 2011

DOI: 10.1109/TSMCC.2010.2073464 · Source: IEEE Xplore

CITATIONS

8

READS

529

8 authors, including:



**Bruce Stephen**

University of Strathclyde

72 PUBLICATIONS 1,255 CITATIONS

[SEE PROFILE](#)



**C. Dwyer**

Scotland's Rural College

8 PUBLICATIONS 116 CITATIONS

[SEE PROFILE](#)



**Jimmy J Hyslop**

Scotland's Rural College

73 PUBLICATIONS 1,436 CITATIONS

[SEE PROFILE](#)



**Matthew Bell**

Hartpury University

63 PUBLICATIONS 1,206 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Automatic Rice Seed Inspection Using Hyper-Spectral Imaging (Newton Fund) [View project](#)



ICT-CMAC: Intelligent Decision Support and Control Technologies for Continuous Manufacturing and Crystallisation of Pharmaceuticals and Fine Chemicals [View project](#)

# Statistical Interaction Modeling of Bovine Herd Behaviors

Bruce Stephen, *Member, IEEE*, Cathy Dwyer, Jimmy Hyslop, Matthew Bell, David Ross, Kae Hsiang Kwong, Craig Michie and Ivan Andonovic, *Senior Member, IEEE*

**Abstract**— While there has been interest in modeling the group behavior of herds or flocks, much of this work has focused on simulating their collective spatial motion patterns which have not accounted for individuality in the herd and instead assume a homogenized role for all members or sub-groups of the herd. Animal behavior experts have noted that domestic animals exhibit behaviors that are indicative of social hierarchy: leader/follower type behaviors are present as well as dominance and subordination, aggression and rank order, and specific social affiliations may also exist. Both wild and domestic cattle are social species, and group behaviors are likely to be influenced by the expression of specific social interactions. In this paper, Global Positioning System coordinate fixes gathered from a herd of beef cows tracked in open fields over several days at a time are utilized to learn a model that focuses on the interactions within the herd as well as its overall movement. Using these data in this way explores the validity of existing group behavior models against actual herding behaviors. Domain knowledge, location geography and human observations, are utilized to explain the causes of these deviations from this idealized behavior.

**Index Terms**—Agriculture, Behavioral science, Global Positioning System, Markov processes

## I. INTRODUCTION

DOMESTIC cattle are known to possess social traits which induce different interactions with the other animals in their herd. Animal behaviorists have long noted the connection between physiology and social activity which emphasizes the importance of increasing understanding of the mechanisms that underlie grouping behavior. Although prior research has produced simulations of herding behavior that approximates collective behaviors this has been in an oversimplified manner that adopts a generic behavior for all animals. The reality is far removed from this homogeneity, with a variety of dominant, subservient and isolationist social traits having been observed by behavioral experts. In a review of cattle and domestic

animal social characteristics, [1] noted that herds exhibited fairly complex social interactions that were indicative of dominance and subordination with clear leaders and followers along with evidence of care and dependency. As an example of how activities of individuals can impact on the welfare of others, [2] noted that groupings of animals diluted their predatory risk and improved their foraging ability confirming that the existence of reactionary, agonistic and learning behaviors will motivate the herd to either respond to or imitate the actions of an individual. Domesticated animals also show differences in social behavior, including both agonistic responses (butting, pushing, avoidance) and affiliative behaviors [3]. While environmental and physiological factors may affect behavior, there remains evidence of a social order within herds; dominant animals are noted to defend territory with age and weight correlating with herd rank [1].

In a management context, understanding and being able to quantify aspects of an animal's behavior may increase the ability to manage the animal for both economic productivity and animal welfare reasons. Modeling GPS based positional datasets on parameters such as distances traveled, spatial distribution and activity levels are an innovative way to gain new insights into both individual animal and herd or group behaviors. Understanding animal behaviors such as grazing patterns, feeding or rumination time budgets and animal to animal social interactions will assist our ability to manage grazing provision, mating strategies and disease epidemiology patterns across and within herds.

Formulating a model of herd behavior that captures all of its complexities has been attempted in several ways with simulations of simple pseudo-autonomous creatures and models derived from geospatial data being at either extreme of the research spectrum; this paper attempts unification between these disparate strands of research by learning a model from observed herd activities and relating its inference to hypothetical behavior constructs used for group simulation. Utilizing a set of GPS transponder data gathered in continuous periods over several weeks from a free ranging herd of beef cows, data driven models are generated. From these models, the original data can be segmented by positional co-ordinates according to the interacting positions and therefore implied behaviors of animals in the herd. It is shown that modeling a herd as a Markov Random Field, through the choice of an appropriate neighborhood system, automatically segments the

Manuscript received November X, 2009. This work was supported in part by ITI Techmedia as part of the Condition Based Monitoring Programme.

Dr. B. Stephen is a Research Fellow in the Advanced Electrical Systems Research Group, Institute of Energy and Environment, University of Strathclyde, Glasgow, G1 1XW (phone: +44 (0)141 548 5864, e-mail: [bstephen@eee.strath.ac.uk](mailto:bstephen@eee.strath.ac.uk))

Dr. C. Dwyer, Dr J. Hyslop, Mr. D. Ross and Mr. M. Bell are with Scottish Agricultural College, King's Buildings, West Mains Road, Edinburgh, EH9 3JG.

herd at each timeframe, labeling animals according to their relationships with others.

This paper is structured as follows: in the next section the existing strategies for modeling herd behavior are reviewed with particular emphasis being placed on the pioneering concepts of steering behaviors [4, 5]. Although the methodology for investigating herding behavior presented here is data driven, the underlying intelligence model and how it relates to what is observed is of primary importance from a ground truthing perspective and drives several of the modeling decisions such as feature selection. In section 3 the model strategy is outlined and justified. In section 4 the collection of herd data from beef cattle is described along with the challenges associated with utilizing GPS data. In section 5, the feature extraction used for deriving the herd models is described and choices justified. Section 6 reviews Markov Random Fields and their derivatives and the specializations required for adapting them to herd modeling are shown. After this, the algorithmic details used to apply the model are presented along with the system implemented to visualize the collective and individual behaviors present in the herd over time. Finally, limitations of the current model and further development required for practical applications are described in the conclusions.

## II. PRIOR GROUP BEHAVIOR MODELING STRATEGIES

Among the earliest attempts to simulate herding animals using only their positions in space, [4] originally proposed the model of group behaviors based on steering behaviors and separations known as ‘boids’. The original ‘boids’ flocking algorithm focused solely on steering behaviors that dictated separation, alignment and cohesion of the group. These assumed uniform behavior throughout the group making no attempt to incorporate individuality, while in reality, some animals will be sociable and others less interested in the rest of their group. An additional level of autonomy was encoded in the spatial models of [6], where a herd of 100 simulated ‘robots’ of point mass was used to demonstrate how simple group behavior could be regardless of the complexity of the individual; like the work of [4], this contribution was important since it attempted to simulate a herd through consideration of motion primitives applied by the individual: namely steady state motion, acceleration, turning and avoidance of obstacles. Unlike the earlier herding models there were more layers of detail: a perception model was incorporated which determined the animals in the herd that were visible within a given radius; a placement algorithm was used to determine the desired relative position; lastly, a simple spring/damper model control system was employed for computing the velocity of the animal within the constraints of the position and velocity of its immediate neighbors. However, individuality was not a consideration, with all animals having the same dynamics.

The next increment in group behavior models, made by [5], continued to build on steering behaviors with ‘autonomous

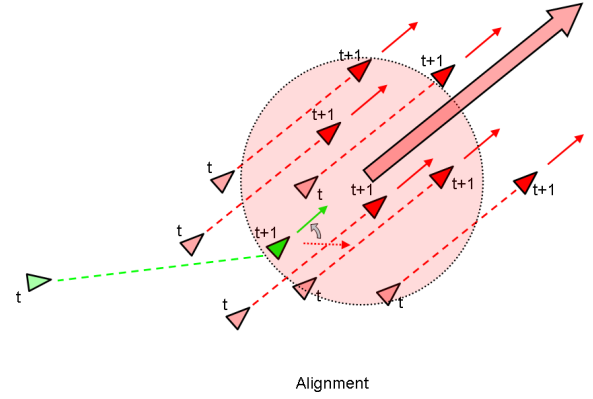


Fig. 1. Alignment steering behavior identified by Reynolds. The individual aligns the directional component of their displacement vector with the aggregate bearing of the group.

characters’, proposing a three layer architecture of action selection, steering and locomotion much like [6]. However, key to this was a set of basic steering behaviors such as ‘seek’, ‘flee’ and ‘evasion’ that dictated a characters placement, velocity and directional bearing based on encounters with other characters – behavior of the individual was no longer homogenous, it could change to reflect circumstance.

Figure 1 shows how herding among groups of animals is modeled using steering behaviors while figure 2, again from [5] shows how a more complex interaction can be described using a finite set of predicates.

Action selection chooses steering then steering chooses locomotion: these control not only the speed and direction the characters were traveling in but also the distance maintained from the rest of the herd. Reynolds gave the example of caribou fleeing from wolves to illustrate the whole concept: sometimes there will be a discrete behavioral switch (action selection goes from grazing to fleeing, there is no mixture) other times there will be mixing of behaviors (avoiding trees while fleeing). The first of these, ‘seek’ and its complement

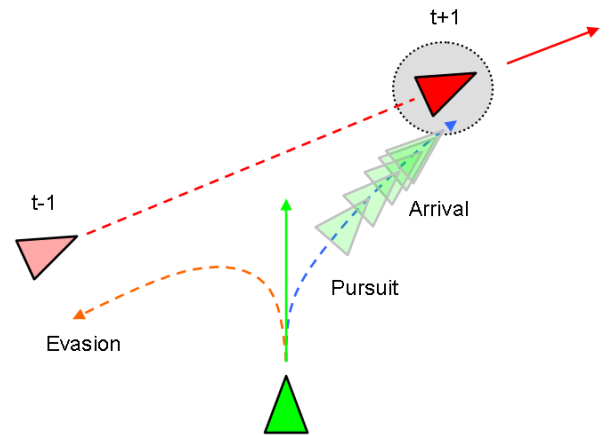


Fig. 2. Pursuit steering behavior identified by Reynolds. The individual aligns the directional component of their displacement vector with the aggregate bearing of the group.

'flee' involve aligning a velocity vector with a target point. In the case of 'flee', the direction of the vector is away from the target. Closing on a stationary target may occur in an agricultural situation when feedstocks are delivered or to a water source. Specializations of this behavior to a moving target are 'pursuit' and 'evasion' both of which can be envisaged as traits of dominance assertion. More complex than 'pursuit' is 'arrival' which still follows the same alignment rules but includes a strategy for linearly reducing the characters velocity on approach to the target so as to avoid overshoot. A final basic steering behavior is 'wander', where the character follows a randomly changing bearing with sufficient trend to give an impression of deliberation.

Combining the steering behaviors led to a secondary set of interaction traits: wandering can form the basis of 'exploration' which manifests itself as an exhaustive coverage of a region of space, while more the specialized 'foraging' combines wandering with resource seeking i.e. a random search then pursuing a stationary target when close enough to register its existence. Additional cues would be required to infer these from spatiotemporal data though. More basic aggregations are 'cohesion', where all members of a herd exhibit a 'seek' behavior on the centroid of the group, converging into a closer ensemble. 'Separation' is where grouped characters repulse, essentially exhibiting the opposite to 'cohesion' and dispersing the herd. 'Leader Follower' combines 'cohesion', to maintain a bearing with another moving entity. Unlike 'seek' or 'pursuit', a 'follower' never crosses the path of a 'leader' or crowds it.

The models of Hodgins and Reynolds were more focused on emulating the behavior of prey animals such as antelope, fish, birds or insects. Lower level limitations of the simple herd/swarm/flocking models prevent the more realistic individual behaviors in a group from being detected or simulated. Spherical field of vision is plausible for insects and fish but for larger, more intelligent animals, this and uniform awareness in all directions is less likely. Energy levels are also unlikely to maintain a consistent level throughout the duration of a simulation. The assumption of a single social standing is also assumed whereas in reality different animals will have differing reactions to peers depending on their size, age and temperament which in turn leads to lack of provision for the existence of cliques or sub-herd groups.

A solution to obtaining a less regimented picture of real herding behavior is to consider a real herd of animals. With recent advances in low cost sensor technology, a number of applications of GPS transponders to herding animals have attempted to capture behavior to previously unknown levels of detail. Such studies have the potential to validate earlier theories of grouping behavior and elaborate on their key features. Using a small herd of Hereford breed beef cattle, [7] trained cluster (specifically, K-Means [8]) models on free-ranging cow behaviors to classify active and inactive states. [7] also used equipment that measured head angle via an inclinometer and magnetometer as well as recording GPS

fixes. [7] then coupled classification with biological observations to infer the meaning of the clusters and found in particular that low speed and high head angle were deemed traits of inactivity while high speed and low head angle were traits of an active animal. Mean cow distance to herd centroid and standard deviation showed that cows are closer during activity than inactivity. Activity levels were shown to mutually influence each other among gregarious animals.

### III. CONSTRUCTION OF A HERD BEHAVIOR MODEL

Construction of a model of group behavior that captures the type of discrete interaction activities that [5] identified but without generalizing this type of behavior to an individual requires three key elements. The first is a set of exemplary behaviors from a real herd to demonstrate how the members of a group behave towards each other and how this changes depending on the presence of others. The second is to identify the features of this set that characterize the behavior of interest. The final element is to capture the particular values of the features or combinations of values that recur on a regular basis and do this in a manner that is robust to noise.

### IV. EXEMPLAR HERD BEHAVIOR

#### A. Data Gathering

Following the example given by [7], the behavior of a group of animals was captured as GPS fixes over time. 14 animals were equipped with collar mounted GPS transponders manufactured by Bluesky Telemetry Ltd of Aberfeldy, Scotland, shown in figure 3, which were configured to record fixes every 3 minutes and periodically transmit this data wirelessly to be ultimately accrued in a relational database.

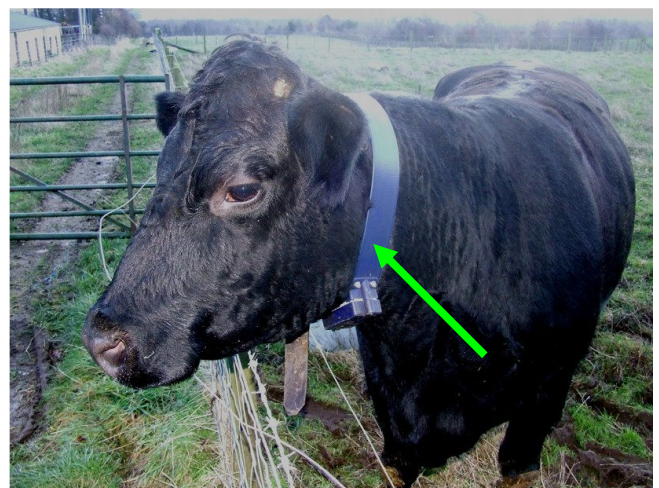


Fig. 3. Beef cow wearing a collar mounted GPS transponder.

The database was populated between June 2006 and September 2006, obtaining around 0.3 million positions during the course of the 85 whole days of monitoring. This gave 946 cow days where the GPS fix was taken every 3 minutes. The database was indexed by fix timestamp and the unique

identifier of the animal to permit individual animal trajectories to be retrieved easily over specified time periods. The animals studied were spring calving beef cows that had all had at least one calf previous to the current year and were grazed as part of a small herd during the months of June – September 2006. Cows and their suckling calves were rotationally grazed in one of two fields of 4.4 and 6.2 hectares (ha) respectively during this period at stocking rates of approximately 1.5 livestock units/ha; typical of commercial farming practice in Scotland. The location was on the eastern mainland of Scotland, UK and the altitude of the fields was approximately 210 m above sea level.

### B. GPS Data Considerations

In order to utilize the GPS data, which arrives in geodetic co-ordinates, a conversion is made into the Cartesian system [9] using the following notation:

- $\psi$  – altitude above the ellipsoid
- $\phi$  - latitude
- $\lambda$  – longitude
- $f$  – flattening
- $R$  – Radius of curvature in prime vertical
- $a$  – semi major earth axis (ellipsoid equatorial radius - OSGB36 is 6377563.396)
- $b$  – semi minor earth axis (ellipsoid polar radius - OSGB36 is 6356256.91)
- $e^2$  eccentricity squared

Where:

$$f = \frac{a-b}{a} \quad (1)$$

$$e^2 = 2f - f^2 \quad (2)$$

$$R(x) = \frac{a}{\sqrt{1 - e^2 \sin^2 x}} \quad (3)$$

To convert the geodetic latitude, longitude and altitude into their Cartesian equivalents at time  $t$ :

$$x_t = (R(\phi_t) + h_t) \cos \lambda_t \cos \phi_t \quad (4)$$

$$y_t = (R(\phi_t) + h_t) \cos \phi_t \sin \lambda_t \quad (5)$$

$$z_t = (R(1 - e^2) + \psi_t) \sin \phi_t \quad (6)$$

Cartesian co-ordinates are easier to work with when computing distances between animals. All GPS estimates of position have some degree of error in them. The extent of this error cannot be accurately estimated but is known to be inversely proportional to the number of satellites used to produce the fix. [10] noted the relationship between the numbers of satellites and the impact this had on position accuracy.

The GPS collars are not synchronized, so in order to maintain a consistent representation of the position of all animals in the herd at a given time, a state machine is used to maintain a synchronized representation of the herd. This receives GPS fixes and converts these into features before updating the features that are common to the entire herd such as the dispersal of animals and the overall movement between fixes.

### V. CHARACTERIZING HERDING FEATURES FROM GEOSPATIAL DATA

To best represent animal behavior with respect to the rest of the herd and its immediate neighbors, more than just the position must be considered. Features derived from GPS data must be able to capture elements of animal to animal social interactions that are significant to the animals welfare and management. These would include individual and total within herd contacts, movements with the herd as a group (or sub-groups within the herd) and the degree of “solitude” or lack of contact. All of these elements may differ widely according to grazing availability, breed type, position within the social herd hierarchy and previous learned experience. Firstly it is assumed that each cow is represented by a point  $c$  in 2-dimensional space at time  $t$  in a herd of size  $H$ . To track the motion of the herd as a group, a useful quantity is the herd centroid  $g$ , which at time  $t$  is given by:

$$g(t) = \frac{1}{H} \sum_{h=1}^H c_h(t) \quad (7)$$

To contextualize the distance between cows, the dispersal  $d$  of the herd is also used – this is just the average distance from the centroid:

$$d(t) = \frac{1}{H} \sum_{h=1}^H \sqrt{(c_h^x(t) - g^x(t))^2 + (c_h^y(t) - g^y(t))^2} \quad (8)$$

As figure 4 shows, these two quantities show where most animals are positioned and how tightly they are assembled. The reason for including this is that smaller fields may force animals to be closer together; to avoid using absolute values, the dispersal is used to normalize the distance between animals. Again, this quantity is calculated at every time step.

For a herd of size  $H$ , there will be a symmetric  $H$  by  $H$  matrix whose cells specify the distance between two cows. A cell in the inter-cow matrix  $B$  measuring the distance between cow  $i$  and cow  $j$  is defined as:



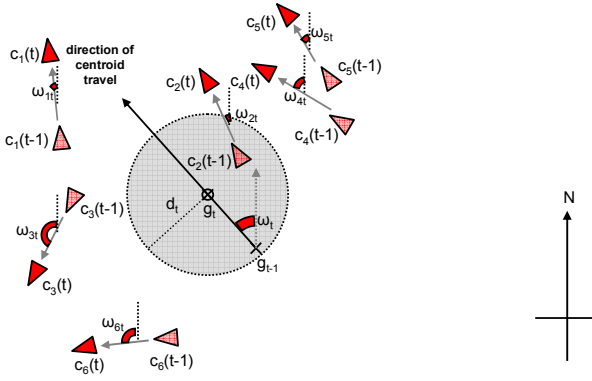


Fig. 4. Illustration of the herd dispersal and centroid bearing between times  $t-1$  and  $t$ .

$$b_{ij}(t) = \sqrt{(c_i^x(t) - c_j^x(t))^2 + (c_i^y(t) - c_j^y(t))^2} \quad (9)$$

The diagonal is obviously zero. This is used to find the animals closest to a given cow normalized using the herd dispersal. Figure 5 illustrates this for a group of 8 animals.

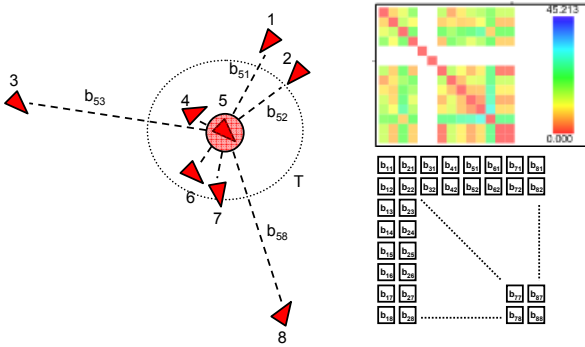


Fig. 5. Illustration of the herd intercow matrix; white cells represent an animal outwith the herd dispersal.

Also shown in figure 5 is how the matrix fills up in practice, with red cells representing close animals and blue cells being those further away. White cells represent those animals that are outwith the herd dispersal radius.

As mentioned, GPS collars are not synchronized so to ensure that there is a representation of all animals at all time steps, a state machine approach to pre-processes the position data extracted from the database is used to calculate the herd dispersal, centroid and inter-cow distance matrix from accumulated fixes. Position and spacing are able to represent static grouping behaviors such as cohesion or solitude, but dynamic behaviors such as pursuit or evasion require temporal and directional observations. Temporal observations are dealt with using the displacement since the previous time step while the directional are given by the bearings of the herd and the individual animal. Bearings  $\omega$  of animal  $h$  at time  $t$  are obtained from:

$$\omega_{ht} = \arctan\left(\frac{c_h^y(t) - c_h^y(t-1)}{c_h^x(t) - c_h^x(t-1)}\right) \quad (10)$$

The difference between the herd bearing and the individual is calculated.

$$\Delta\omega_{ht} = |\omega_{ht} - \omega_t| \quad (11)$$

Where

$$\omega_t = \arctan\left(\frac{g^y(t) - g^y(t-1)}{g^x(t) - g^x(t-1)}\right) \quad (12)$$

The displacement (or distance traveled) of an animal  $h$  at time  $t$  shows how active it has been between fixes and is given by:

$$s_h(t) = \sqrt{(c_h^x(t) - c_h^x(t-1))^2 + (c_h^y(t) - c_h^y(t-1))^2} \quad (13)$$

Figure 6 shows the bearing and displacement for a single animal over a 24 hour period. This illustrates how a single animal can exhibit varying levels of activity through its displacements, and deliberation through its changes in heading.

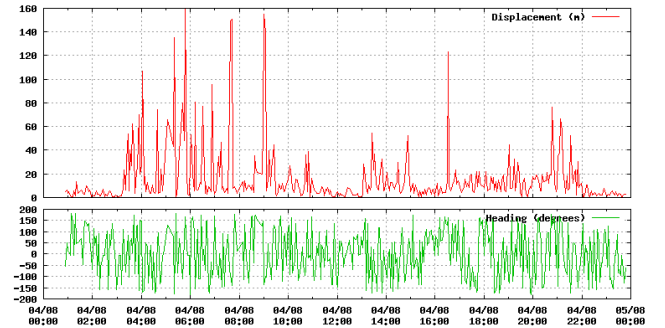


Fig. 6. Illustration of the changes in displacement and bearing for a single animal over 24 hours.

Another useful feature is the displacement of the herd as it contextualizes activity level:

$$S_{H \setminus h}(t) = \frac{1}{H} \sum_{h' \neq h} s_{h'}(t) \quad (14)$$

This is the herd displacement without the contribution of animal  $h$ ; the reason for the omission is that a highly active animal will skew this metric towards high activity and therefore make itself look normal. Aside from providing a context for the level of activity of individuals, on its own the group displacement remains a useful indicator: activity of the

group as a whole may be suppressed by hot weather or excited by the arrival of feedstock.

The cows themselves are assumed to be stateless with no memory of their previous activities. While this is questionable in reality, the temporal independence assumption is replaced by one of dependence on the group.

## VI. MARKOV RANDOM FIELDS

There are two problems inherent in modeling herding behaviors: firstly, the general behavior of individuals needs to be captured and abstracted in such a way that it is robust to imprecise observations; secondly, the interaction between individuals needs a similarly generalized representation to be learned. The preferred means of capturing such dynamics while retaining robustness to noise is through the statistical modeling of the variables in the system. ‘Cluster’ type models such as K-Means [8] or Finite Mixture Models [11] that abstract a multivariate, continuous observation into a single discrete label or state are based on the assumption that there are subpopulations in the observation data generated by processes with differing statistical distributions. However, such models are not immediately suited to capturing complex dependencies between variables, as individual observations are attributed to generator processes that are assumed independent of one another. The works of [12] and later [13] extended the notions of underlying or hidden generator processes that were originally envisaged by [14] for much simpler models. Collections of generators form particular configuration architectures with the configuration being defined by the bonds or interfaces between generators and the state of each generator. [13] saw this model as a formalism of structural or ‘closed’ patterns, a structural familiarity that simplified observations by finding regularities in their constituent parts.

### A. Graphical Models

A convenient way of representing the probabilistic relationships between these processes is by using graphical models [15]. The models feature nodes representing random variables that, if statistically dependent, are interconnected with arcs representing this relationship. In some graphical models, these arcs are directed to imply causality; however there are circumstances where undirected arcs are preferred. It may be unclear whether one variable ‘causes’ another or it may just be the case that causal structure is irrelevant and all that matters is the co-occurrence of variables. The tractability of graphical models is only realizable through the adoption of the Markov assumption which restricts the scope of influence variables have on each other. Like a multidimensional generalization of the order of a Markov chain, a Markov blanket represents the extent of dependence on a particular variable. For spatial data, the Markov blanket extends in 2-dimensions leading to an undirected graph known as a Markov Random Field (MRF). An MRF is typically used to obtain the global effect of local relationships such as the interpretation of images from pixel arrangements or the classification of characters from assemblies of strokes [16]. From the Hammersley Clifford Theorem, an MRF is distinguished from

a Random Field by the fact that the state variables  $q$  in a configuration are distributed according to a Gibbs distribution:

$$P(q) = \frac{1}{Z} \exp\{-U(q)\} \quad (15)$$

Where  $Z$  is the normalizing constant given by summing over all possible observations for  $q$  in the graph:

$$Z = \sum_{q \in Q} \exp\{-U(q)\} \quad (16)$$

$U$  is an energy function that is strictly positive and is chosen according to how configurations are defined in the model application. The energy of a configuration is usually obtained by evaluating all clique potentials  $V$  over all connected sites. The potential function  $V$  is an arbitrary function used to measure the relationship between variables or multiple variables in the clique  $c$ , the definition of this can be application specific. Cliques are subsets of the graph containing connected variables. As an example consider the graph representation of variables in a regular lattice arrangement shown in figure 7. This type of graph is common in image analysis applications where each node represents a pixel. The value each pixel takes may be dependent on at least one of its neighbors, hence the connection structure. The clique size and hence the perceived influence is limited by a function called a neighborhood system,  $N$ , which dictates the variables that can be included in a clique for a site; in the example in figure 7 it is each variable that has an adjacent connected variable in both directions on either side of it. The neighborhood of the site at  $i, j$  contains 4 cliques.

The relationship of all cliques at all sites forms the joint density over all variables in the graph. Cliques come in various orders; an  $n^{\text{th}}$  order clique contains  $n$  connected sites all of which must form part of the neighborhood of the configuration being considered. A scheme used by [16] in character recognition and [17] in image segmentation is to use 1<sup>st</sup> and 2<sup>nd</sup> order clique potentials:

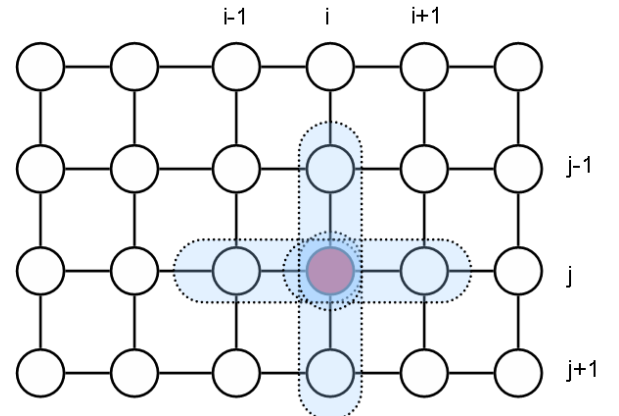


Fig. 7. A lattice shaped random field with a simple neighborhood formed from adjacent nodes. The shaded areas represent cliques of the red site.

$$U(q) = \sum_{i \in I} V_1(q_i) - \sum_{j \in N_i} V_2(q_i, q_j) \quad (17)$$

Although higher order cliques are possible, 1<sup>st</sup> and 2<sup>nd</sup> order capture statistical observation details about the site, how the site is labeled, from the first order potential function and structural information, how the sites interact, from the second order one. This way, the MRF captures localized dependencies without requiring a complex conditional dependency structure for the entire graph - a key part of the ‘recognition by components’ theory of perception [13].

### B. Hidden Markov Random Fields

In situations where the observation is assumed to be noisy and/or must be abstracted into one of a finite number of states that represents a generalization of its value, the MRF can still be used to model the dependency structure of the graph. However, since the state labels are never observed but instead are implied by the observation, a different model called a Hidden Markov Random Field (HMRF) is required. The MRF works as before but an additional layer of inference is required to map the observation or emission field onto the corresponding locations in the graph. For an HMRF the joint distribution of the hidden (Q) and observed (O) variables P(Q, O) is required:

$$P(q, o) = \frac{1}{Z} \exp\{-U(q, o)\} \quad (18)$$

In this case the energy function is of both the observed and hidden variables although the choice of clique potentials in U are made to reflect preferred configurations of site labels implied by observations: likely configurations are still required to have the lowest energy but for an HMRF, the energy function is now a function of both the state and the observation:

$$U(q, o) = \sum_{i \in I} V_1(o_i | q_i) - \sum_{j \in N_i} V_2(q_i, q_j) \quad (19)$$

The observed random variables are assumed to be conditionally independent with the dependency structure captured through the hidden variables. For an observation o, a function is required to relate the observations back to the configuration of the underlying MRF states that generated it.

### C. Herd Model Specialization of HMRF

There are two parts to the HMRF that are domain specific, the neighborhood function and the clique potentials that are used in the energy function. This section discusses how these are formulated for the herd application. While Hidden Markov Random Fields have been extensively employed in pattern analysis tasks, these have mainly focused on imaging data where the regular structure of pixels lends itself to simple

neighborhood functions based on site adjacency like that shown in figure 7. However, there is no concept of spatial adjacency in the herding application as firstly the herd is not regularly aligned in space like lattice data and second, the animals will move causing the neighboring cows for a given site (cow) to change. Instead, this neighborhood will be concerned with the influence of variables rather than their position so the herd neighborhood will be constructed according to the proximity of other cows. The neighborhood function proposed for the herd model is a simple maximization of the inter-cow distance matrix at a given time step. For every cow this forms a neighborhood from the K nearest cows.

If K was set to H then the assumption would be that the entire herd influences each individual which assumes an unrealistic degree of awareness; if K was set to 1 then it would be assumed that a cow was influenced only by its nearest neighbor which would return to the over simplistic herd representation. Neighborhoods consisting of 3 cows and greater will be considered; this lower bound is chosen because building a neighborhood only from pairs would not capture grouping behaviors from cows detached from the herd as these would be part of only one neighborhood. Selecting the ideal neighborhood size is a model selection problem dealt with in section VIIB. In figure 8, the sites (cows) in three neighborhoods are highlighted with a filled shape and the cows in the neighborhood enclosed. Note how only part of the herd is in a particular cow’s neighborhood – this is more realistic and follows the thinking of the real herd models and herd-animals actual awareness of the rest of the herd: there is going to be a limit of the influence of distant animals.

Once the neighborhood function is defined, the remaining specialization of the model is the energy function whose second order clique potential depends on the choice of neighboring sites. The energy function must be defined to prefer recurring combinations of arrangements of cow behavior, that is, cows behaving in a particular way are observed with cows behaving in a (or possibly another)

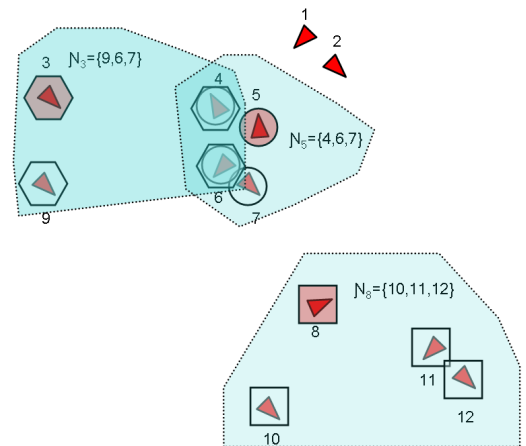


Fig. 8. Neighborhoods formed for animals 3, 5 and 8. Note that animals can belong to multiple neighborhoods. Neighborhoods for the remaining animals are not shown to preserve clarity.



particular way. This is achieved by expressing the first order clique, which captures single animal behavior, as a collection of independent Gaussian distributions, then weighting these with a connectivity matrix which forms the basis of the second order clique and captures behavioral ensembles in the herd. Since the minimization of this function is required for preferred configurations, the log of these is used. Each state observation is assumed to be Gaussian distributed so taking the negative log of this, gives a function that decreases with increasing observation likelihood:

$$V_1(o_i(t)|q_i = k) = -\log \left[ \frac{1}{\sqrt{2\pi}|\Sigma_k|} \exp \left\{ -\frac{1}{2} (o_i(t) - \mu_k)^T \Sigma_k^{-1} (o_i(t) - \mu_k) \right\} \right] \quad (20)$$

Each state has its own mean and elliptic covariance matrix, the latter to relax dependence assumptions between observation features. For the second order clique potential, the co-occurrence of states is used:

$$V_2(q_i, q_j) = -\log a_{ij} \quad (21)$$

The element  $a_{ij}$  of the matrix  $A$  is the probability of the label  $j$  being seen as a neighbor of a site labeled  $i$  – which is analogous to the likelihood of one cow exhibiting a particular type of behavior being seen with a cow exhibiting another particular behavior type. To reiterate, the first order clique sets the most likely state the cow is in with respect to its observed behavior. The second order clique reduces the energy of the model if the cow state is likely within the context of its neighbor's states.

## VII. APPLYING THE MODEL

Table I lists the animals used in the trials which were a mixture of ages, sizes and breeds. These were all either from the Aberdeen Angus (AA) or Limosin (LIM) breeds. While age was not recorded, parity, the number of times a cow has calved provides a measure in proportion to how old the animal is. Similarly, weight gives an approximation of size; both parameters may be useful in explaining interactions such as dominance where size (physical stature) and age may be influencing factors in the roles played within the herd.

### A. Estimation of Model Parameters

The distributions that relate states to observations and the connectivity matrix must be learned from a set of exemplar data. In [16] a Maximum Likelihood implementation of the Expectation Maximization (EM) algorithm [18] was used to learn an HMRF. EM is an iterative procedure that, for maximum likelihood estimation, computes the distribution of the state variables with respect to the observed data then chooses model parameters such that this likelihood is

TABLE I  
ANIMALS USED IN BEHAVIOR TRIAL

| Identifier | Breed | Weight (kg) | Parity† |
|------------|-------|-------------|---------|
| LX1        | LIMx* | 720         | 5       |
| AX1        | AAx   | 748         | 4       |
| LX2        | LIMx  | 568         | 2       |
| LX3        | LIMx  | 568         | 2       |
| LX4        | LIMx  | 686         | 3       |
| AX2        | AAx   | 612         | 2       |
| AX3        | AAx   | 720         | 4       |
| AX4        | AAx   | 642         | 2       |
| AX5        | AAx   | 710         | 8       |
| LX5        | LIMx  | 640         | 6       |
| AX6        | AAx   | 682         | 6       |
| AX7        | AAx   | 706         | 6       |
| LX6        | LIMx  | 798         | 6       |
| AX8        | AAx   | 770         | 5       |

†Parity refers to the number of times the animal has been in calf.

\*Letters denote the following breeds LIMx = Limousin cross-bred, AAx = Aberdeen Angus cross-bred.

maximized. The algorithm continues iterating through the two steps until the log likelihood of the data converges.

[8] provided this general view of the EM algorithm used to estimate a parameter  $\theta$  using a set of observed data  $X$  with unobserved data  $Q$ :

1. Estimate  $\theta$  as  $\hat{\theta}$
2. On the Expectation step calculate  $P(Q|X, \hat{\theta})$
3. Calculate:

$$Z(\theta, \hat{\theta}) = \sum_Q P(q|X, \hat{\theta}) \log P(X, q|\hat{\theta})$$

From which the maximization of  $Z$  gives the parameter update:  $\theta = \arg \max_{\theta} Z(\theta, \hat{\theta})$

4. If  $Z$  has not converged, set  $\hat{\theta} = \theta$  and repeat from the Expectation step (2).

For the Gaussian observation models used in the first order cliques, the K-Means [8] clustering algorithm is used to first find approximations of feature means and variances. Using these, the connectivity matrix is filled in with counts of co-occurring clusters and then normalized.

The beef cows used were kept outside during the duration of the trial and for days chosen for analysis, consequently had no predictable routine outside of a daily visual check by farm and technical staff. Given this limit of distractions, data over a single 24 hour period is used to train the model resulting in 5325 fixes or 408 frames. The model typically converges after around 20 iterations of EM.

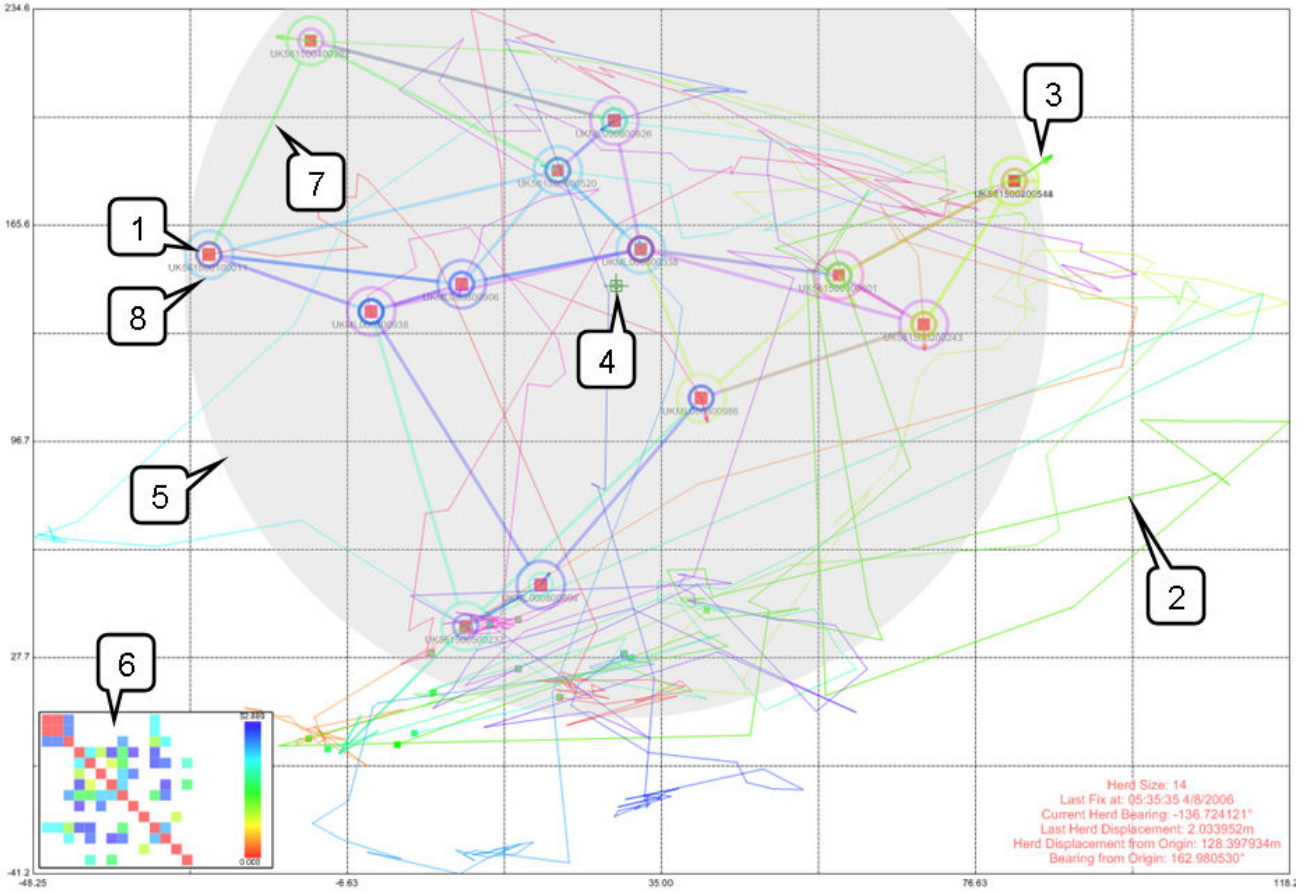


Fig. 9. The herd behavior visualization software showing model inference and salient behavioral features, specifically : 1 – cow, 2 – cow position trace, 3 – bearing with magnitude denoting displacement on last fix, 4 – herd centroid, 5 – herd dispersal, 6 – inter-cow distance matrix, 7 – neighborhood connector arc, 8 – animal state (outer ring color).

### B. Social behaviors as state labels

The cardinality of  $Q$  is another variable quantity in the model: as this effectively abstracts recurring feature vector values, the states can be construed as particular behaviors. The number of behaviors assumed present in the data must be specified prior to starting training.

Many MRF models have a binary state variable, betraying their origins in the Ising model from Statistical Physics which allowed a  $\pm 1$  value for each site to represent magnetically coupled atoms. Multivariate states are known in image processing and texture analysis applications of HMRFs to allow different underlying abstractions to be represented. The same philosophy prompts the choice of multiple states in this model; firstly it allows a more complex emission distribution to be learned and in doing so allows a range of implied group behavior patterns to be revealed. This application assumes that states and configurations of an MRF will correspond to particular group behavior which can be interpreted through the parameters of the emission distribution associated with each state.

To select the optimal number of behavior states, the model most likely to have generated the data is used. This maximum likelihood approach to model selection has the disadvantage that it finds every possible behavior at the expense of loss of generality of the model. Several model selection criteria such

as Bayesian Information Criteria, Normalized Entropy Criterion, Automatic Relevance Determination and Akaike Information Criterion [8, 11] could be employed to lessen this effect although with an increase in computational overhead.

### C. Labeling Animals by Interactions

Once an MRF has been trained it can be used to find the optimal label structure of the herd given a set of observations, much in the same way the Viterbi algorithm is used to find the implied state sequence in a Hidden Markov Model (HMM). Unlike the trellis structure of possible sequence likelihoods in an HMM, the state of the herd is retained in a frame generated at each time step by the herd state machine. This ‘frame’ is an irregular structure that holds the observation feature vector for each animal as well as its state label and pointers to the animals captured by its neighborhood function. Another problem is that unlike the trellis, the graph structure of the MRF model does not permit closed form inference. A common choice of algorithm for this task is Iterated Conditional Modes (ICM) [19] which implicitly minimizes the energy function by maximizing the following posterior density energy function sequentially at each site through the choice of site labels:

$$\max_k P(q_i = k | o_i(t)) \quad (22)$$

The probability at each site is maximized with respect to the observation by piecewise re-labeling of the neighboring sites. Obviously, changing the labeling at one site will impact on the others which share it in their clique, so this optimization must be performed globally by traversing the entire graph until the global energy converges. In an image processing application, this algorithm would traverse the data set in a raster fashion, however, as already noted, this procedure is further complicated by the herd fixes not having this regularized structure.

#### D. Visualizing Herd Behavior

A graphical front end to the relational database was developed to allow animated playback of the herd positions over a chosen time period and is shown in figure 9. This display not only represented the separation of the animals but also their bearings, their past positions and the magnitudes of their displacements or distance traveled. Overlaid on the graphical display were the state labels and the neighborhoods at each time step for each animal as well as the centroid and radius of dispersal for the entire herd. This facilitates interpretation of animal states and through the formation of neighborhoods allows the cliques within the herd to be easily seen. Also displayed is the inter-cow distance matrix which clearly shows the separations between animals.

### VIII. MODELS IN USE

The features used in the test model were the log of the angle between the animal and herd bearings, included to capture alignment. The log of the displacement since the last fix, to capture activity levels, the log of the distance to the nearest animal and the log of the distance to the herd centroid. The last two features are used to capture solitude and cohesion respectively. The log scale is used to capture orders of magnitude rather than absolute values. Based on 24 hours of training data from August 2nd 2006 (midnight to midnight) the model preferred by maximum likelihood had 6 states and a neighborhood of 3. The model preferences for various other parameterizations are shown in figure 10.

Simpler models that focused on a single variable could be used to isolate particular herd traits of interest such as directional alignment or cohesion on their own rather than the approach taken here to find combinations of multiple features, which may result in less clear separation.

#### A. Behavioral interpretation of state labels

The mean parameters for the states of the optimal model are shown in table II and offer some degree of explanation as to what the states represent in terms of behavior. State 1 is very closely aligned with the herd bearing but exhibits a large separation from the centre of the herd and other animals in general. The displacement for animals in this state is moderate. State 2 represents animals that move much less and while close

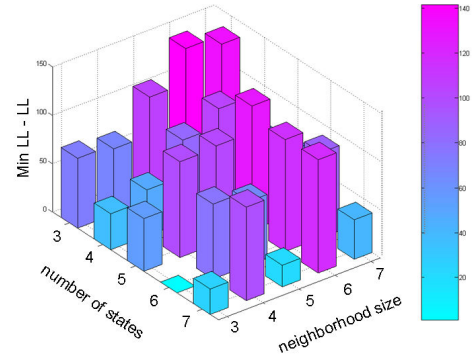


Fig. 10. Model order selection using maximum likelihood as a function of neighborhood size and state variable cardinality.

to other animals, remains distant from the centre of the herd and moves in a different direction. State 2 could be representative of a clique or sub-herd while state 1 could represent a general collective movement. State 3 is not aligned with the movement of the herd at all, is separated from other animals and moves the largest distance of all animals. This behavior state may be seen as a notional ‘exploratory’ behavior as postulated by [5]. State 4 is similar to state 3 but is closer aligned with the herd bearing. Similar to state 4 but for displacement, is state 5 – this may equate to foraging behavior or grazing again identified by [5]. Finally state 6 like state 1 represents the motion of a sub-herd albeit with more cohesion and a more sedate pace.

TABLE II  
MODEL PARAMETERS

| State | Herd Alignment | Displacement | Nearest Distance | Distance to Centroid |
|-------|----------------|--------------|------------------|----------------------|
| 1     | 27°            | 18m          | 15m              | 41m                  |
| 2     | 86°            | 6m           | 6m               | 15m                  |
| 3     | 119°           | 26m          | 12m              | 33m                  |
| 4     | 74°            | 9m           | 16m              | 43m                  |
| 5     | 76°            | 5m           | 9m               | 23m                  |
| 6     | 27°            | 5m           | 8m               | 20m                  |

While visual inspection of the animated herd traces demonstrates clear separations between animals at various times of the day, a more quantifiable metric was utilized to attempt to validate state explanations. By running a trained model over 24 hours worth of GPS traces, counts of fixes an animal was in a particular state were accumulated. To visualize the states over the entire observation sequence, count histograms are produced for each animal over another 24 hour period, this time from midnight on 8th August 2006; figure 14 shows how this represents the variety of behavior an individual animal exhibits over a given time, how this differs between animals and which states dominate. In figure 11, stacked histograms of these counts show the distribution of states for each animal over a single day.

#### B. Indicators of Unusual Events

Again, using the data from August 8th 2006, the model

observation likelihood was tracked over the 24 hour period. As figure 12 shows, the model exhibits a number of regions of low likelihood between 4am and 7am.

This represents a configuration that the model is unfamiliar with from the herd data it has already seen. To investigate the cause of this unlikely behavior, the animals positions were inspected manually revealing that there was a sudden large collective movement then dispersal of most animals in the

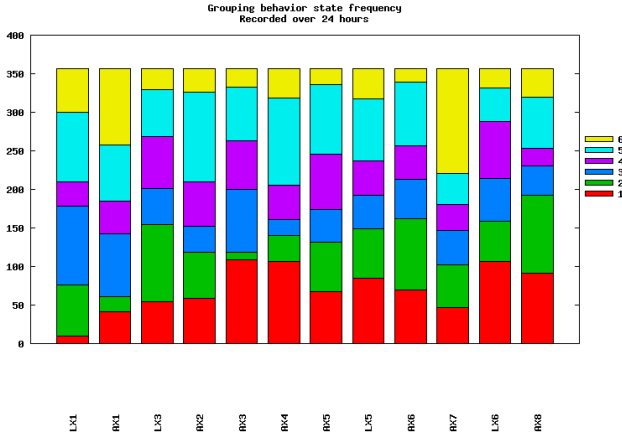


Fig. 11. Histogram of animal labeling over August 8<sup>th</sup> 2006.

herd. This was in a north easterly direction at around 4:45am. At around 6:50am most animals head south west, regrouping in the process led by animal LX5. This sequence of events may relate to the common observation that cattle outdoors tend to lie together in a group during darkness and then set off to graze or perhaps to a water source early in the morning when the sun

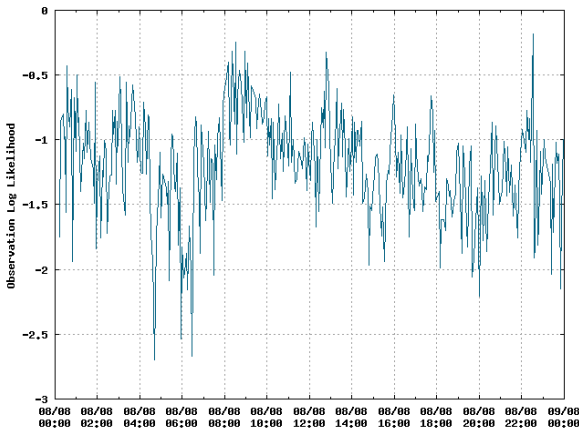


Fig. 12. Time series of model observation likelihood over August 8<sup>th</sup> 2006.

comes up. Such collective behavior may be an evolutionary adaptation to predation during darkness.

## IX. FUTURE DIRECTION

This paper has introduced a Hidden Markov Random Field based model of group behavior and has demonstrated through position data gathered from a herd of beef cows how such a model can: identify generalized behavior of individuals and in doing so identify generalized collections of individual

behaviors in groups which can then be used to track the behavior of a group with respect to its expected behavior as represented by the model through the use of the observation likelihood of the model.

Through this basic functionality, it has been shown how animal behavior characteristics can be identified over a set period of time. It also demonstrated how external influences that incur extraordinary behavior can be detected as herd configurations with low likelihood. The learned models can then go on to be applied to other herds regardless of their size without having to relearn their parameters.

Animal behaviors, movements and spatial distributions can be analyzed for either individual or herd based parameters such as distance traveled, time based grazing or rumination budgets, number and duration of social interactions with other animals and consequently aspects of social status may be inferred. This paper discusses some of the ways in which GPS datasets could be analyzed and potentially modeled using a variety of machine learning approaches that would help to quantify and hence introduce some objectivity into the determination of these parameters in real world situations. Use of these modeling approaches would assist animal behavioral scientists and commercial practitioners to gain new insights into the factors which determine key features of animal performance, health and welfare under defined management or experimental conditions

The work undertaken in this paper is intended to be a first step towards using a statistical pattern recognition approach to understanding animal grouping behaviors. In the time since the trials for the work reported in this paper were completed, sensor technology has become cheaper, more reliable and the platforms used to run them e.g. [20] have greater computing power and programmability. Monitoring capabilities can now be expanded more easily and allow additional sensors to be incorporated into telemetry. Accurate estimation of the bearing of an animal at rest or between two very close fixes is difficult using derived quantities. Directional tracking could be improved with the inclusion of magnetometers in the collar mounted telemetry rather than relying on deriving the bearing from consecutive fixes. Similarly, close proximities may be better measured with for example, proximity logging devices using UHF frequencies to alleviate potential errors from the fixes.

Hardware limitations aside, additional modeling improvements could lead to additional insight into both individual and animal grouping behavior, in particular, explicitly relating the site configurations to known herd behaviors. The model presented in this paper is a generalization of all possible behaviors seen during the training period. Segmenting this data using ground truth data such as physiological or farm staff observations of behaviors such as oestrus activity (mating behavior) into sets of distinct recognizable behaviors or interactions would allow individual models to be trained on the previously identified behaviors. The resulting ensemble could be used as a classifier for



automated recognition of future occurrences of these behaviors or events. Ultimately, higher level models can be anticipated based on the one proposed here that may be used to automatically detect key management stages requiring intervention during mating periods for example, the assessment of grazing area and herd interactions during the summer grazing of environmentally sensitive grasslands, models of disease transmission in animal herds or groups, and at an individual level, behaviorally sensitive illness events.

#### ACKNOWLEDGMENT

The authors wish to acknowledge the help and advice over the course of this research given by Dr Rob Baxter and Alisdair Tullo of Edinburgh Parallel Computing Centre (EPCC), Matthew Bell at Scottish Agricultural College (SAC) Easter Bush and Dr Hanna Miedema of the Royal (Dick) School of Veterinary Studies.

#### REFERENCES

- [1] D. A. Wieckert, "Social Behaviour in Farm Animals", *Journal of Animal Science* 6:32. 1971.
- [2] S. A. Rands, R. A. Pettifor, J. M. Rowcliffe and G. Cowlshaw, "State Dependent Foraging Rules for Social Animals in Selfish Herds", in *Proceedings of the Royal Society* 271. 2004.
- [3] M.F. Bouissou, Boissy, A., Le Neindre, P. and Veissier, I. "The social behavior of cattle" In: *Social Behaviour of Farm Animals* (ed. L. Keeling & H.W. Gonyou). CABI, Oxford, UK. 2001.
- [4] C. W. Reynolds, "Flocks, Herds, and Schools: A Distributed Behavioural Model", Published in *Computer Graphics*, 21(4), pp. 25-34. July 1987.
- [5] C. W. Reynolds, "Steering Behaviours for Autonomous Characters", in *Proceedings of Game Developers Conference 1999*, pp. 763-782. Miller Freeman Game Group 1999.
- [6] J. K. Hodgins and D. C. Brogan, "Robot Herds: Group Behaviours for Systems with Significant Dynamics", in *Proceedings of Artificial Life IV*. 1994.
- [7] M. Schwager, D. M. Anderson, Z. Butler and D. Rusa, "Robust Classification of Animal Tracking Data", *Computers and Electronics in Agriculture* 56. 2007.
- [8] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer Science + Business Media 2006.
- [9] *A Guide to Co-ordinate Systems in Great Britain*, version 1.6. Ordnance Survey, Crown Copyright 1996.
- [10] I. A. R. Hulbert and J. French, "The Accuracy of GPS for Wildlife Telemetry & Habitat Mapping", *Journal of Applied Ecology* 38, pp 869-878. British Ecological Society 2001.
- [11] G. J. McLachlan and D. Peel, *Finite Mixture Models*. Wiley-Interscience. 2000.
- [12] U. Grenander, *Elements of Pattern Theory*. John Hopkins University Press 1996.
- [13] D. Mumford, "Pattern Theory: The Mathematics of Perception", in *Proceedings of the International Conference on Mathematics*, pp. 401-422. 2002.
- [14] B. S. Everitt, *Introduction to Latent Class Models*. Chapman and Hill, London and New York. 1984.
- [15] K. P. Murphy, A Brief Introduction to Graphical Models and Bayesian Networks. Available: <http://www.cs.ubc.ca/~murphyk/Bayes/bnintro.html>
- [16] J. Zeng, J. and Z. Q. Liu, "Markov Random Field-Based Statistical Modelling for Handwritten Chinese Character Recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence* Volume 30 No 5. May 2008.
- [17] Z. Kato, T. C. Pong and J. C. M. Lee, "Colour Image Segmentation and Parameter Estimation in a Markovian Framework", *Pattern Recognition Letters*, 22(3-4) pp 309-321, March 2001.
- [18] A. Dempster, N. Laird and D. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm". *Journal of the Royal Statistical Society, Series B*. 1977.
- [19] J. Besag, "On the Statistical Analysis of Dirty Pictures", *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 48, No. 3, pp. 259-302. 1986.
- [20] A. Young, M. Ling and D. K. Arvind, "Orient-2: A Wireless Real-time Posture Tracking System Using Local Orientation Estimation", in *Proceedings of the 4th International Workshop on Embedded Networked Sensors (EmNets 2007)*, Cork, Ireland, 25-26 June 2007.



**Dr Bruce Stephen** currently holds the post of Research Fellow within the Institute for Energy and Environment. He received his BSc from Glasgow University and MSc and PhD degrees from the University of Strathclyde. His research interests include: Distributed Information Retrieval Systems, Condition Monitoring, Knowledge Based Systems, Latent Class Models and Unsupervised Machine Learning.



**Dr Cathy Dwyer** is a Reader in the Sustainable Livestock Systems Group at Scottish Agricultural College, Bush Estate. She received her BSc from University of Bristol and a PhD from the Royal Veterinary College, University of London. Her research interests include: mother offspring interactions, behavioural development, neonatal survival and the welfare of extensively managed farm animals.



**Dr Jimmy Hyslop** currently holds the post of Beef Specialist within SAC's Beef and Sheep Select Group developing services to support the beef sector and helping roll out the knowledge gained from SAC research and demonstration studies. He also has a major role in conducting research and development projects for Government and commercial clients with suckler beef systems typical of those commonly found across the UK.



**Matthew Bell** is currently a PhD Student within the Sustainable Livestock Systems Group at Scottish Agricultural College, Bush Estate. His research interests include investigating greenhouse gas emissions for livestock and animal monitoring.



**David Ross** currently holds the post of Research Engineer within the Sustainable Livestock Systems Group at Scottish Agricultural College, Bush Estate. He received his BSc from Edinburgh University and PDip from Strathclyde University. Research interests include animal monitoring for health, welfare and physiological condition and sensing systems for animal product quality.

**Dr Kae Hsiang Kwong** is a Research Fellow in the Centre for Intelligent Dynamic Communications at University of Strathclyde

**Dr Craig Michie** is Senior Lecturer within the Electronic and Electrical Engineering Department at the University of Strathclyde. He received his BSc and PhD from Glasgow University. His current research interests follow two independent themes – optical access technologies and in particular the application of semiconductor optical amplifiers (SOA) within access networks and wireless sensor systems for condition based monitoring focusing on animal monitoring.

**Prof. Ivan Andonovic** is the ITI Techmedia Professor of Broadband Networks and the Head of the Centre for Intelligent Dynamic Communications Systems at the Department of Electronic and Electrical Engineering, Strathclyde University. His research interests centre on the development of broadband networks, access and home networking and wireless sensor systems. He is Topical Editor for the 'IEEE Transactions on Communications' and was Technical Programme Co-Chair for the recent 'IEEE International Conference in Communications (ICC07)'.