Programmatic detection of spatial behaviour in an agent-based model

William J. Chivers, William Gladstone and Ric D. Herbert

Abstract—The automated detection of aspects of spatial behaviour in an agent-based model is necessary for model testing and analysis. In this paper we compare four predictors of herding behaviour in a model of a grazing herbivore.

We find that a) the mean number of neighbours adjusted to account for population variation and b) the mean Hamming distance between rows of the two-dimensional environment can be used to detect herding. Visual inspection of the model behaviour revealed that herding occurs when the herbivore mobility reaches a threshold level. Using this threshold we identify a limits for these predictors to use in the program code.

These results apply only to one set of parameters and environment size; future research will involve a wider parameter space.

Index Terms—Agent-based model, Herding behaviour, Model testing, Spatial behaviour

I. Introduction

This paper describes the detection of herding behaviour in an agent-based model (ABM) of the movement of a herbivore in a two-dimensional space. The model was built to explore the mechanism of the predator-prey interaction between the herbivore and its "prey", the producer organism on which it feeds. The ecological findings and the model itself are described in [3, Ch. 5]. Herding behaviour is of interest to ecologists; examples of models built to investigate the mechanism of herding behaviour include those of [16], [4], [13], [9] and [14].

Of interest here is the programmatic detection of herding, necessary for model analysis. An important question in the research described in [3] concerned the model parameters which affected the development and persistence of herding behaviour. The investigation of herding sensitivity to each parameter involved thousands of executions of the model as each parameter was varied. Herding behaviour is visually evident in the two-dimensional display written into the model, but to implement the parameter analysis an automated method of detection of herding behaviour was needed.

The automated detection of herding in computer-based models raises a challenging problem of computer science, although object detection (is it a face?) is more readily implemented than object recognition (whose face is it?) [11], [10], [5]. For the computer code to be able to recognise an object or pattern such as a herd within the data structure representing the herd there must be some model of the search object in the code, and that model ideally should have high interclass variability but low intraclass variability [10]. Three broad categories of object detection systems are described in the literature:

1) Model-based systems, in which a model of the pattern is defined and the system attempts to match this model to parts of the image [17], [10], [5].
2) Image invariance methods, which define a set of image pattern characteristics or relationships (such as brightness levels) that uniquely define the search object [15], [10].
3) Example-based learning algorithms, in which an artificial neural network system learns the salient features of the search pattern from other images which are identified as positive or negative [11], [10].

The detection of herding described here falls into the first category above.

II. Model description

A. Introduction

The model description below follows the ODD protocol of [6].

B. Overview

1) Purpose: This model was built to explore the mechanism of herbivore-producer interaction in a two-dimensional space. The model was not built to represent any specific system, but is a parsimonious, generalised model of predator-prey interaction.

2) State variables and scales: The individuals in the ABM are a) herbivores and b) cells of a two-dimensional matrix. The members of the producer species are not represented as individuals, but rather as an attribute of individual cells. The biomass of the producer in each cell decreases as it is eaten and increases according to logistic growth, each cell having a growth rate (b_r) and max population (b_K) parameter[1]. Members of the herbivore species have a resource level attribute which represents the energy reserves carried by living organisms. Herbivores are born, consume the biomass of the cells, reproduce and may die of starvation. After eating, herbivores move towards cells with higher producer biomass.

The attributes of the herbivores and cells are listed in Table I. The resource units are not formally defined, nor is

William J. Chivers: Faculty of Science and Information Technology, The University of Newcastle, PO Box 127, Ourimbah, NSW 2259, Australia, phone: +61 2 4349 4473, fax: +61 2 4348 4145, email: william.chivers@newcastle.edu.au

William Gladstone: School of the Environment, University of Technology, Sydney, Ultimo, NSW, Australia, email: william.gladstone@uts.edu.au

Ric D. Herbert: Faculty of Science and Information Technology, The University of Newcastle, PO Box 127, Ourimbah, NSW 2259, Australia, email: rdh698@gmail.com

[1]The symbols are listed in Table I.

Name	Symbol	Default value	Units
Matrix			
Width	m_w	100	Cells
Height	m_h	100	Cells
Edge existence (values 0 or 1)	m_e	1	
Herbivore			
Initial population	p_n	100	Individuals
Initial resources	p_{ir}	1000	Resource units
Metabolic cost per ts	p_{mc}	100	Resource units
Trophic efficiency	p_{te}	20	
% of producer eaten on current cell per time step	p_{pe}	80	
Max move distance per time step	p_{md}	1	Cells
Move cost	p_{mvc}	10	Resource units
Move level of stochasticity	p_{ms}	2.0	
Reproductive cost	p_{rc}	1000	Resource units
Producer			
Growth rate of each cell	b_r	0.1	
Max population of each cell	b_K	1000	Producer units

TABLE I
DEFAULT INITIAL CONDITIONS AND PARAMETERS USED IN THIS CHAPTER.

the length of time represented by each simulation time step. The herbivores have no learning or social behaviour and have a minimal foraging strategy. Population sizes per time step and a graphical representation of herbivore positions and cell biomass levels are the primary outputs of the model.

3) Process overview and scheduling: During each discrete time step the cells and herbivores are processed individually. Processing involves the following:

1) *Growth of the producer organism:* The producer population on each cell advances towards the maximum carrying capacity of the cell (b_K) according to the growth rate of the cell (b_r) using the logistic equation.

2) *Herbivore movement:* Each herbivore moves to the cell with the highest effective producer population within the herbivore maximum move distance (p_{md}). This may mean staying on the current cell.

3) *Implementation of resource intake—herbivores:* The resource total of each herbivore increases according to the percentage of producer eaten on current cell parameter (p_{pe}), the number of herbivores on the cell and the herbivore trophic efficiency parameter (p_{te}). The cell producer population is reduced accordingly.

4) *Application of metabolic cost—herbivore:* The herbivore metabolic cost per time step (parameter p_{mc}) is subtracted from the resource level of each herbivore. Herbivores which then do not have sufficient resources to live one more time step die of starvation.

5) *Implementation of asexual reproduction—herbivores:* Herbivores with sufficient resources to live one more time step *and* reproduce will produce offspring asexually at a resource cost of p_{rc} per offspring. New individuals are placed on the same cell as the parent.

All variables are updated in real time, that is, during the processing of each individual predator or prey object within each time step.

C. Design concepts

Emergence

The system level outcomes which emerge from the interactions of the herbivores with each other and with individual cells are the population levels of the two species. These are the absolute herbivore population per time step and the mean producer population per cell per time step. In the case of the herbivore population, no system level phenomena such as population growth rate, carrying capacity or minimum population are imposed by the model.

Interaction

The interactions between individual herbivores and individual cells are:

- Herbivory: the herbivores consume a percentage of the producer population of the occupied cell. The producer population, an attribute of each cell, is stored in the computer memory as a real number with *double* precision as defined in [2].
- Occupation: Herbivores have x and y attributes to locate them in the two-dimensional matrix. Individual cells have a herbivore population attribute which stores the current number of herbivores occupying the cell.

The interactions between individual herbivores are:

- Asexual reproduction: herbivores reproduce asexually at a cost to their individual resource total.
- Co-occupancy: Two or more herbivores may occupy one cell, reducing the producer units available to the second herbivore to be processed in the time step.

Individual cells do not interact with each other.

Stochasticity

The model used pseudorandom numbers which were generated in the Java computer programming language using a linear congruential formula and a 48-bit seed [2] as described by [8, Section 3.2.1]. Unless otherwise noted, a uniform distribution was used as this best represented the selection of the default parameter values.

During initialisation the pseudorandom number generator is used to determine the producer population in each cell and the position in the two-dimensional matrix of each herbivore. Processing during each time step uses the pseudorandom number generator in the herbivore movement algorithm as described below: the cell to which a herbivore moves is not selected completely deterministically—there is a small element of chance involved.

Observation

The primary output of the model is the absolute herbivore population per time step and the mean producer population per cell per time step, as illustrated in Fig. 4a and e. More importantly for this discussion, a graphical display of the two-dimensional matrix of cells shows the position of each herbivore and the cell producer density each time step as illustrated in Figs. 1, 2 and 3.

The characteristics of matrix cells

The cells in the model are graphically represented in the two-dimensional display as squares. In the model algorithm, however, cells have zero dimensions: herbivores occupy a position in the Cartesian plane which is defined with integer x and y coordinates, but herbivores do not move within cells.

D. Details

1) Initialisation and input: At the start of each simulation the cells are created in a two-dimensional matrix of m_w width and m_h height. The matrix may or may not have an edge (m_e). Each cell is assigned a random producer population chosen from between zero and the maximum b_K value for the cell. The herbivores are then created and assigned to random cell addresses. The number of herbivores created is p_n, and each is given an initial resource level of p_{ir}.

2) Submodels: The parameter space in which the herbivore population consistently persists to 2000 time steps of the model is very large. The parameter set used here was not intended to represent any specific system, but rather to allow the herbivore maximum move distance per time step (p_{md}) parameter to be increased to over 30 cells without causing herbivore extinction.

Each cell and herbivore object is processed during each time step of the simulation: in an individual time step the entire cell population is processed before the entire herbivore population.

Processing the individuals involves:

Growth of the producer organism

Each cell producer population advances towards the maximum carrying capacity of the cell (b_K) according to the growth rate of the cell (b_r) and using the logistic growth model of classical ecology (see, for example, [1, pages 224-226]. The logistic growth model used is expressed here as the stepping function used in the computer code:

$$N_{t+1} = N_t + N_t b_r (1 - \frac{N_t}{b_K}) \quad (1)$$

Where N_t is the biomass at time t. Note that the values of b_r and b_K are uniform across the matrix in the simulations reported here.

Herbivore movement

Each herbivore compares its current cell producer population with the neighbouring cells and moves to the cell with the highest effective producer population. This may mean staying on the current cell. The number of neighbouring cells considered depends on the herbivore maximum move distance per time step (p_{md}). Herbivores are unaware of cells beyond this distance.

The effective producer population of each candidate cell is calculated using the actual producer population of the target cell and related factors: the number of other herbivores already on the cell (resources are shared equally) and the cost of moving to the cell ($p_{mvc}\times$ the Cartesian distance).

After calculating the effective producer populations on all target cells, the algorithm uses the p_{ms} parameter to randomise $p_{ms}\%$ of the calculated effective populations. Using a level of stochasticity in this way has precedents in the modeling literature, for example in the fish schooling model of [16].

(a)   (b)
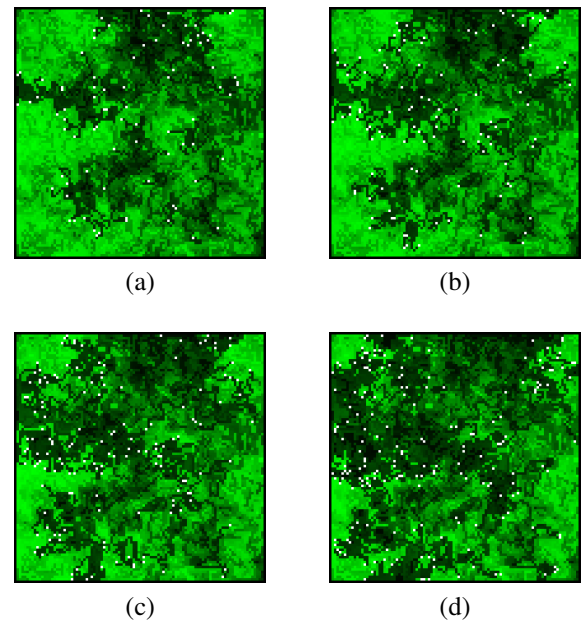
(c)   (d)

Fig. 1. Screenshots with ten time steps between images (a) to (d) illustrating an increasing herbivore population consuming the large areas of high producer density on the lower left of the matrix. The herbivores are represented by the white cells, the producers by the green cells. The higher the producer population, the brighter the green of the cell.

Implementation of resource intake—herbivores

The number of producer units consumed by a herbivore is calculated using p_{pe} and the number of herbivores on the cell. The cell producer population is reduced accordingly. The herbivore resource level is increased taking into account p_{te}).

Application of metabolic cost—herbivores

The cost for each herbivore of living one time step is represented by p_{mc}. This is subtracted from the resource level carried by each herbivore each time step. Herbivores which then do not have sufficient resources to live one more time step (resource level $< p_{mc}$) die of starvation and are removed from the simulation.

Implementation of asexual reproduction—herbivores

Any herbivores with sufficient resources to live one more time step *and* reproduce (resource level $\geq p_{mc} + p_{rc}$) will then produce offspring asexually at a resource cost of p_{rc} per offspring. The new individuals, with the default initial resource level are placed on the same cell as the parent. This model was built with two related parameters, p_{ir} and p_{rc}. In all simulations reported here, $p_{ir} = p_{rc}$.

E. Graphical output

The graphical display includes

- Producer population density: the brightness of each cell indicates the cell producer population—the brighter the cell, the closer the population is to the maximum b_K value of the cell.
- Herbivore positions: the white cells indicate the presence of a herbivore.

Fig. 1 consists of four screenshots of the matrix in sequence. The dark areas have low producer population and are areas

(a)   (b)

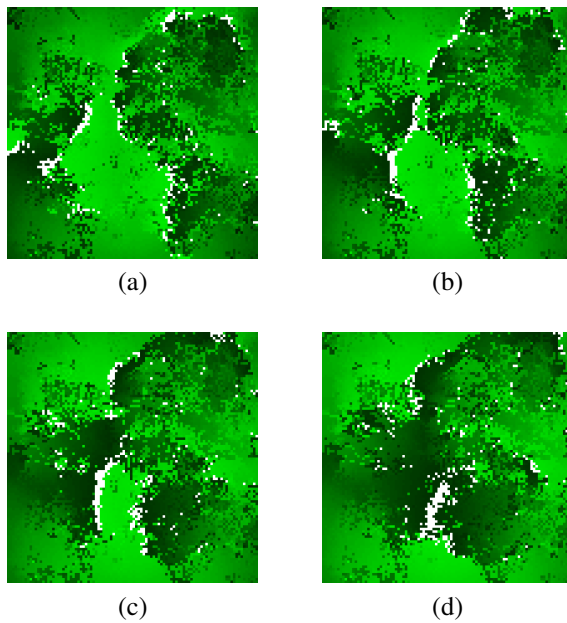(c)   (d)

Fig. 2. Time steps 591, 594, 597 and 600 of a single execution of the model where $p_{md} = 10$ cells. Herding behaviour is evident as the herbivores consume the resources in the middle and lower middle of the images.
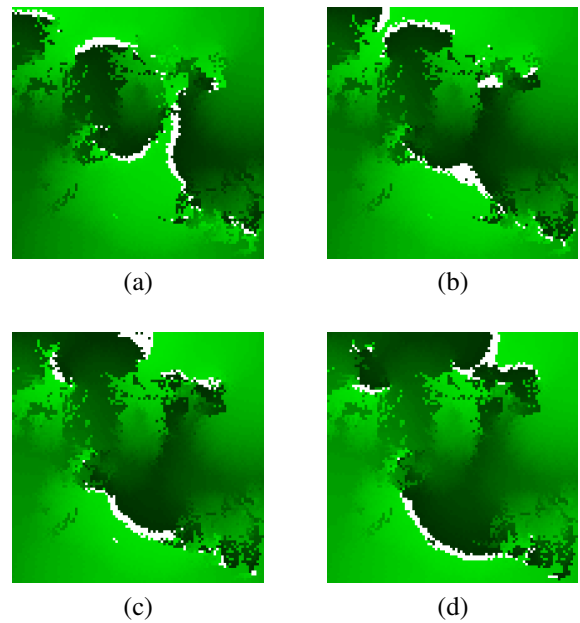

(a)   (b)

(c)   (d)

Fig. 3. Time steps 2133, 2136, 2139 and 2142 of a single execution of the model where $p_{md} = 20$ cells. Herding behaviour is evident as the herbivores consume resources in the lower left and upper right of the images.

from which the herbivores have recently moved. The brighter areas have had longer to recover since herbivore grazing, and are the areas towards which the herbivores are moving.

III. DESCRIPTION OF HERDING BEHAVIOUR

Chivers[3] found that an increase in herding behaviour is positively correlated with the herbivore maximum move distance per time step (p_{md}). To illustrate this, Figs. 2 and 3 are taken from two simulations with $p_{md} = 10$ cells and $p_{md} = 20$ cells respectively. These screenshots were taken after many hundreds of time steps and illustrate well-developed herding behaviour. The herding is more pronounced in Fig. 3, when $p_{md} = 20$ cells, than in Fig. 2, when $p_{md} = 10$ cells.

Herding takes tens, if not hundreds, of time steps to appear. Fig. 1 illustrates a typical execution of the model early in its lifespan.

A. Qualitative definition

Chivers[3] defines herding as behaviour with the following characteristics:

- Herding takes many time steps to appear in any given simulation.
- Once herding behaviour has commenced in any given simulation, it is persistent.
- Herding behaviour involves the herbivores moving into areas of high producer density in herds which can be many herbivores deep and which do not extend along the entire front of the areas of high producer density.

IV. DETECTION OF HERDING BEHAVIOUR

A. Visual detection

Herding behaviour was initially detected by watching the graphical display of the two-dimensional matrix, which is part

of the graphical user interface of the computer model, using the above definition. The graphical display of the matrix is built from a two-dimensional array data structure maintained in the computer memory during each simulation. Observation of the graphical display led to the selection of $p_{md} = 4$ cells as the minimum value for the development of herding behaviour.

B. Programmatic detection

The following potential methods of detection of herding are discussed here:

1) Mean number of neighbours for each herbivore.
2) Mean Hamming[2] distance between adjacent rows[3].

These methods are based on the assumptions that the mean number of neighbours of each herbivore will rise as herding becomes more pronounced and that that the average Hamming distance between adjacent rows and columns will fall as herding behaviour becomes more pronounced.

Figs. 4 and 5 illustrate these methods. Fig. 4 illustrates two separate executions of the model with $p_{md} = 10$ cells: the mean populations, mean number of neighbours and mean Hamming distances between rows and columns per time step are graphed. Fig. 5 graphs the same output from two executions of the model where $p_{md} = 20$ cells. These figures indicate that the mean neighbour and Hamming distance measures *may* be positively and negatively correlated with

---

[2]The Hamming distance or Hamming norm between two strings (usually of binary digits) of equal length is the number of positions in which the two strings differ [7]. For example, the two binary strings 11001100 and 11000000 have a Hamming distance of 2, as they differ in two positions (positions five and six counting from the left). To use the Hamming distance here as a measure of herding, cells containing at least one herbivore are given a value of 1 and cells without a herbivore are given a value of 0.

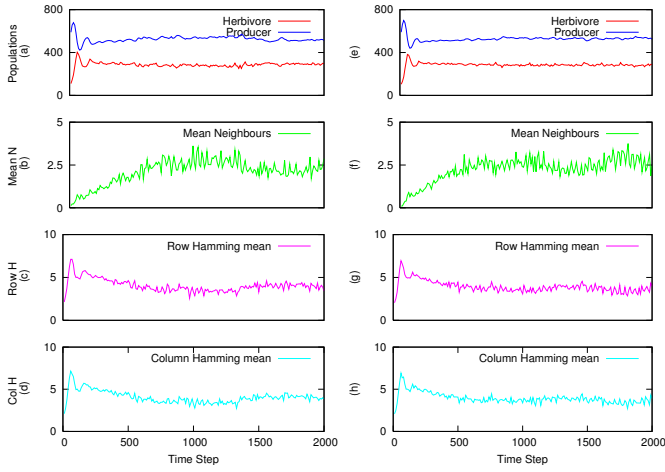[3]The columns display the same patterns (Spearman's rho values typically over 0.9).

Fig. 4. ($p_{md} = 10$ cells) Population levels of herbivores and producers over 2000 time steps of two independent executions of the model. The caption in Fig. 5 describes each individual graph.
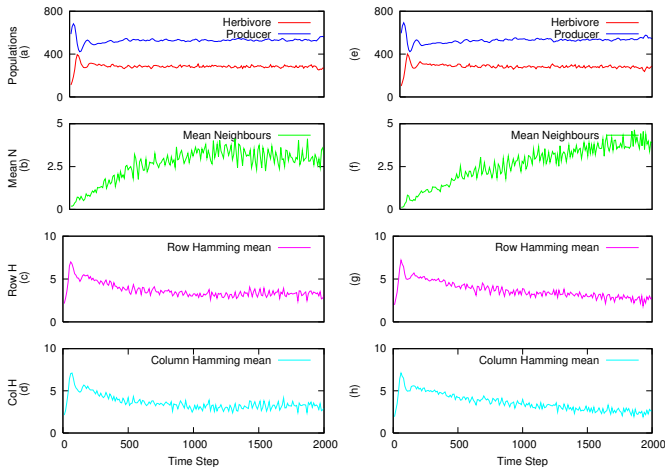

Fig. 5. ($p_{md} = 20$ cells) Population levels of herbivores and producers over 2000 time steps of two independent executions of the model. Graphs (a) to (d) illustrate the first execution, graphs (e) to (h) the second. In graphs (a) and (e) the red line represents the absolute herbivore population per time step; the blue line represents the mean producer population or biomass per cell per time step. Graphs (b) and (f) illustrate the mean number of neighbours for each herbivore per time step, and graphs (c), (g), (d) and (h) illustrate the mean row and column Hamming distances per time step.

p_{md} respectively (herding behaviour, observed in the graphical display, developed as these simulations progressed but is *not* evident in the population levels, graphs (a) and (e) in Figs. 4 and 5).

1) Adjusted measures: The validity of the Hamming distance as a measure of herding behaviour was threatened by the changing absolute number of herbivores per time step in the fixed number of cells of the two-dimensional matrix. Very high or very low herbivore populations may result in low mean Hamming distances with or without the presence of herding behaviour. Similarly, very high populations of herbivores may result in high mean neighbour scores with or without the presence of herding behaviour.

To address these threats, an "adjusted" Hamming distance

Predictor	R^2	Adjusted R^2
Hamming row adjusted	0.878	0.878
Hamming row mean	0.8803	0.8803
Mean neighbours	0.8923	0.8923
Mean neighbours adjusted	0.8993	0.8993

TABLE II
BIVARIATE REGRESSION ANALYSES OF FOUR PREDICTORS SORTED BY ADJUSTED R^2, $1 \leq p_{md} \leq 10$.

Predictor	R^2	Adjusted R^2
Hamming row adjusted	0.5382	0.538
Mean neighbours	0.5637	0.5635
Mean neighbours adjusted	0.6232	0.6231
Hamming row mean	0.6756	0.6755

TABLE III
BIVARIATE REGRESSION ANALYSES OF FOUR PREDICTORS SORTED BY ADJUSTED R^2, $3 \leq p_{md} \leq 5$.

was calculated, the quotient of the original measures and the mean number of herbivores, and an "adjusted" mean number of neighbours was calculated, the quotient of the original measures and the mean number of herbivores. These adjusted measures were analysed in the same way as the raw scores.

V. METHODOLOGY

To test the proposed methods of detecting herding, the model was executed ten times at each parameter level; $1 \leq p_{md} \leq 10$. The data collected are listed below. These data were collected from time steps 1010–2000 (each tenth time step) of each simulation as herding behaviour takes up to 1000 time steps to develop in the simulations observed with this parameter set.

Data collected:

- time step,
- move distance,
- the mean absolute herbivore population per time step,
- the mean producer population or biomass per cell per time step,
- the mean Hamming distance between adjacent rows per time step and
- the mean number of neighbours per herbivore per time step.

A. Data analysis

The R statistical language[12] was used to perform bivariate regression analyses to determine the strength of each of the following predictors of herding behaviour:

- mean Hamming distance between adjacent rows and columns,
- adjusted mean Hamming distance between adjacent rows and columns,
- mean number of neighbours for each herbivore and
- adjusted mean number of neighbours for each herbivore.

VI. RESULTS

Bivariate regression analyses revealed the mean neighbours adjusted figure as the strongest individual predictor ($R^2 =$
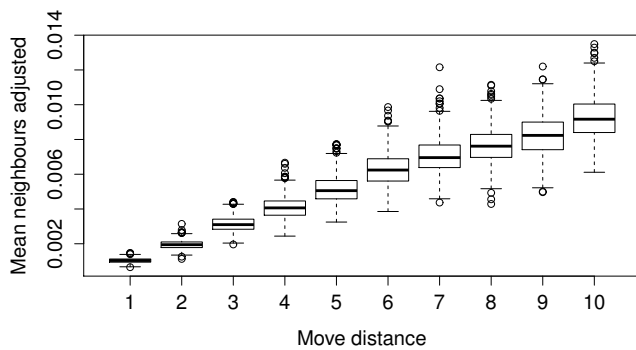
Fig. 6.　Mean neighbours adjusted for $1 \leq p_{md} \leq 10$. The thick line is the median; the bottom of the box is quartile 1, the top quartile 3; the whiskers show the range excluding outliers, values farther than 1.5IQR away from the box. The circles identify outliers.

Fig. 7.　Hamming row means for $3 \leq p_{md} \leq 5$. The thick line is the median; the bottom of the box is quartile 1, the top quartile 3; the whiskers show the range excluding outliers, values farther than 1.5IQR away from the box. The circles identify outliers.
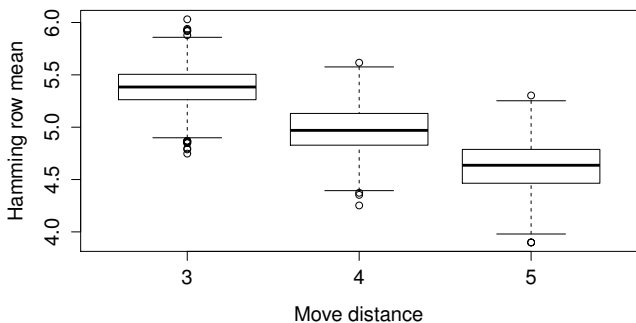
0.8993, adjusted $R^2 = 0.8993$) for herding over the range $1 \leq p_{md} \leq 10$, this predictor explaining 90% of herding behaviour. The R^2 and adjusted R^2 scores for each potential predictor are listed in Table II and median, Q1 and Q3, range and outliers are illustrated in Fig. 6.

The threshold for herding behaviour, determined by visual inspection of the model 2D display, is $p_{md} = 4$. Bivariate regression analyses of potential predictors over the range $3 \leq p_{md} \leq 5$ found that the Hamming row mean was the strongest individual predictor ($R^2 = 0.6756$, adjusted $R^2 = 0.6755$) for herding, explaining 68% of herding behaviour. The R^2 and adjusted R^2 scores for each potential predictor in this parameter range are listed in Table III and median, Q1 and Q3, range and outliers are illustrated in Fig. 7.

VII. Discussion

In the $1 \leq p_{md} \leq 10$ range all four predictors appear to provide a reasonable indication of herding behaviour. In the $1 \leq p_{md} \leq 10$ range there is less separation between predictors than in the $3 \leq p_{md} \leq 5$ range, however the latter

is more important because the visual threshold for herding behaviour was at $p_{md} = 4$.

In all simulations in which herding behaviour was observed, the adjusted mean neighbour measure was ≥ 0.00244 and the mean Hamming distance was ≤ 5.6162, these two predictors having the highest adjusted R^2 values for $1 \leq p_{md} \leq 10$ and $3 \leq p_{md} \leq 5$ respectively. We propose, therefore, that with this set of parameters and environment size these two predictors could be used to detect herding. The model can be set up to execute in batch mode without the graphical display as part of a parameter sensitivity analysis. The program code then could record herding behaviour based on these predictor levels.

The potential threat to the Hamming row measure was not realised, the raw predictor more accurate than the adjusted figure for both p_{md} ranges. The mean neighbours adjusted predictor, however, was in both cases a more accurate measure than the corresponding raw score.

Future research will involve a range of parameter sets and environment sizes. Important questions include the relative accuracy of the four measures and a comparison between the raw and adjusted scores across a wide parameter space.

References

[1]　F. R. Adler. *Modelling the Dynamics of Life: Calculus and Probability for Life Scientists*. Brooks/Cole, New York, 1998.

[2]　K. Arnold, J. Gosling, and D. Holmes. *The Java Programming Language*. Addison-Wesley, Reading, Massachusetts, 3rd edition, 2000.

[3]　W. J. Chivers. *Generalised, parsimonious, individual-based computer models of ecological systems*. PhD thesis, University of Newcastle, NSW, Australia, 2009. Thesis available at http://hdl.handle.net/1959.13/39553.

[4]　B. Dumont and D. R. C. Hill. Spatially explicit models of group foraging by herbivores: what can agent-based models offer? *Animal Research*, 53:419–428, 2004.

[5]　C. Eckes, J. Triesch, and C. von der Malsburg. Analysis of cluttered scenes using an elastic matching approach for stereo images. *Neural Computation*, 18:1441–1471, 2006.

[6]　V. Grimm, U. Berger, D. L. DeAngelis, J. G. Polhill, J. Giske, and S. F. Railsback. The ODD protocol: A review and first update. *Ecological Modelling*, 221:2760–2768, 2010.

[7]　R. W. Hamming. Error detecting and error correcting codes. *Bell System Technical Journal*, 26(2):147–160, 1950.

[8]　D. Knuth. *The Art of Computer Programming, Volume 2: Seminumerical Algorithms*. Addison-Wesley, Reading, Massachusetts, 3rd edition, 1997.

[9]　S. H. Lee, H. K. Pak, and T. S. Chon. Dynamics of prey-flock escaping behavior in response to predator's attack. *Journal of Theoretical Biology*, 240(2):250–259, 2006.

[10]　A. Mohan, C. Papageorgiou, and T. Poggio. Example-based object detection in images by components. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(4):349–361, 2001.

[11]　C. Papageorgiou and T. Poggio. A trainable system for object detection. *International Journal of Computer Vision*, 38(1):15–33, 2000.

[12]　R Core Team. R: A language and environment for statistical computing, 2013. Last accessed 20 June 2013.

[13]　S. A. Rands, R. A. Pettifor, J. M. Rowcliffe, and G. Cowlishaw. State-dependent foraging rules for social animals in selfish herds. *Proceedings of the Royal Society of London, B*, 271:2613–2620, 2004.

[14]　S. A. Rands, R. A. Pettifor, J. M. Rowcliffe, and G. Cowlishaw. Social foraging and dominance relationships: the effects of socially mediated interference. *Journal of Behavioral Ecology and Sociobiology*, 60(4):572–581, 2006.

[15]　P. Sinha. Object recognition via image invariants: a case study. *Investigative Ophthalmology and Visual Science*, 35:1735–1740, 1994.

[16]　R. Vabø and L. Nøttestad. An individual based model of fish school reactions: predicting antipredator behaviour as observed in nature. *Fisheries Oceanography*, 6(3):155–171, 1997.

[17]　A. L. Yuille. Deformable templates for face recognition. *Journal of Cognitive Neuroscience*, 3(1):59–70, 1991.