LETTERS

Cattle movements and bovine tuberculosis in Great Britain

M. Gilbert[1], A. Mitchell[2], D. Bourn[3], J. Mawdsley[2], R. Clifton-Hadley[2] & W. Wint[3]

For 20 years, bovine tuberculosis (BTB) has been spreading in Great Britain (England, Wales and Scotland) and is now endemic in the southwest and parts of central England and in southwest Wales, and occurs sporadically elsewhere. Although its transmission pathways remain poorly understood, the disease's distribution was previously modelled statistically by using environmental variables and measures of their seasonality[1]. Movements of infected animals have long been considered a critical factor in the spread of livestock diseases, as reflected in strict import/export regulations, the extensive movement restrictions imposed during the 2001 foot-and-mouth disease outbreak[2,3], the tracing procedures after a new case of BTB has been confirmed and the Government's recently published strategic framework for the sustainable control on BTB[4]. Since January 2001 it has been mandatory for stock-keepers in Great Britain to notify the British Cattle Movement Service of all cattle births, movements and deaths[5]. Here we show that movements as recorded in the Cattle Tracing System data archive, and particularly those from areas where BTB is reported, consistently outperform environmental, topographic and other anthropogenic variables as the main predictor of disease occurrence. Simulation distribution models for 2002 and 2003, incorporating all predictor categories, are presented and used to project distributions for 2004 and 2005.

BTB, once almost eradicated from Great Britain, has been spreading since 1984, and the number of detected BTB cases continues to rise exponentially (Fig. 1). Because BTB transmission cycles remain poorly understood, the causes of this epidemic are keenly debated, with various possible explanations including transmission from wildlife reservoirs, inadequate control measures, agro-environmental factors and movement of infected animals[6–8]. Previous work has shown the value of environmental variables and measures of their seasonality[9], as encapsulated in Fourier-processed satellite imagery[10], in modelling the distributions of BTB in Great Britain[1]. The pattern of spread of BTB between 1984 and 2003 shows an expanding core area with outlying foci (Fig. 1). Such a pattern is commonly described in invasion ecology by stratified dispersal models that combine short-distance spread with long-distance dispersal events[11–13]. For infectious diseases of livestock, short-distance spread can be viewed as contagion to adjacent or nearby farms located within a few kilometres, by direct contact or borne by wind, insects, rodents or alternative hosts, and generally resulting in the local spatial clustering of cases. Long-distance jump-spread can be viewed as contagion occurring between locations separated by large areas of disease absence, and caused by movements of infectious individuals or infected material. The present study sought to evaluate the relative importance of cattle movement as a predictor of BTB distribution and invasion in Great Britain, and to use the established models for short-term predictions.

Two sets of analyses were undertaken. The first was designed to

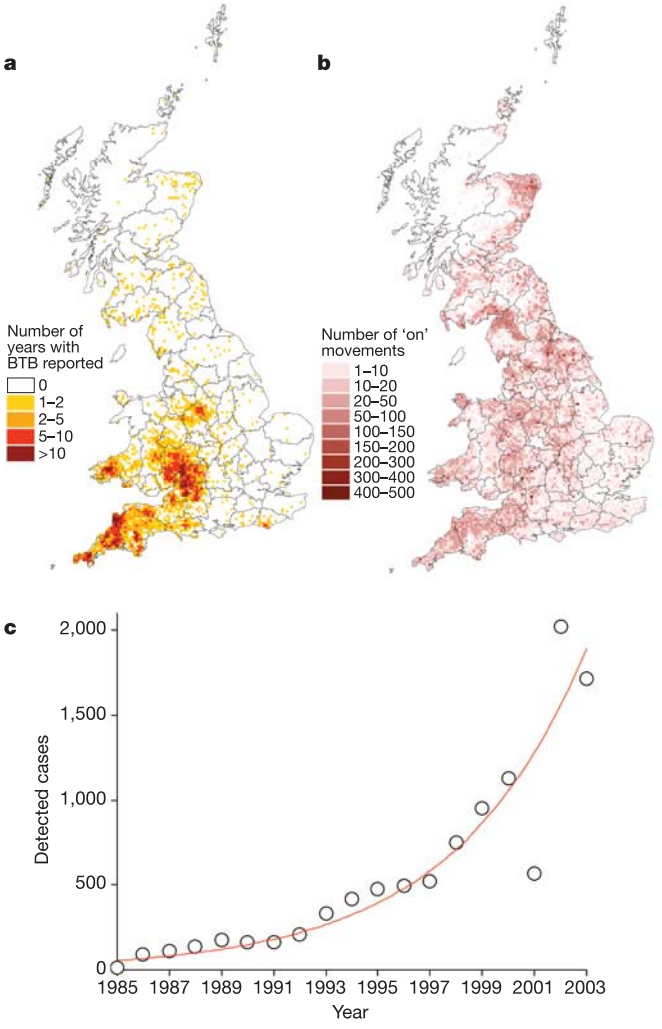

Figure 1 | Distribution and spread of BTB in 1985–2003 and distribution of annual cattle movements. **a**, Distribution of the number of years in which BTB has been reported. **b**, Distribution of cattle inward movements in 2002 by 5-km cells. **c**, Change in the number of BTB detected cases. Note that the number of cases recorded for 2001 was relatively low, probably because of the reduction in testing during the outbreak of foot and mouth disease. They seem to have been compensated for by a comparative increase in 2002. The equations of the curve are: $y = 8.70 \times 10^{-168} e^{0.1958x}$; $R^2 = 0.876$; $n = 19$; $P < 0.001$.

[1]Biological Control and Spatial Ecology CP160/12, Université Libre de Bruxelles, avenue F.D. Roosevelt 50, B-1050 Brussels, Belgium. [2]Veterinary Laboratories Agency, Weybridge, New Haw, Addlestone, Surrey KT15 3NB, UK. [3]Environmental Research Group Oxford Limited, PO Box 346, Oxford OX1 3QE, UK.

Table 1 | Summary statistics and top five predictors for models of BTB during 2002 and 2003

Parameter	Year			
	2002	2003	2002	2003
Movement from	2002	2003	2000	2001
Variables available to model	100	100	100	100
Variables included in model	18	21	20	20
−2 log-likelihood	2,671.8	3,637.5	3,000.3	3,657.3
χ^2	4,444.4	2,730.4	4,174.2	2,748.7
P	<0.001	<0.001	<0.001	<0.001
Correct presence (%)	90.04	87.07	91.28	85.51
Correct absence (%)	91.82	81.67	91.07	83.15
Overall correct (%)	90.94	84.39	91.17	84.33
Overall kappa	0.82	0.69	0.82	0.69
Variable 1	PMI (1,033.4)	PMI (990.4)	PMI (842.1)	PMI (970.0)
Variable 2	DBTD00 (48.4)	ATVM (74.0)	DBTD00 (61.5)	ATVM (73.7)
Variable 3	ATVARA (47.2)	ATMN (44.6)	NDVIPH1 (34.6)	ATMN (44.9)
Variable 4	NDVIMAX (29.1)	ATPH1 (30.5)	DBR (34.0)	NDVIMN (33.6)
Variable 5	ATRNG (16.9)	MIRMN (28.2)	NDVIMAX (29.1)	VPDPH2 (25.9)

PMI, proportion of movements from infected areas. DBTD00, distance to the nearest cell with BTB present in year 2000. ATVARA, air temperature (variance all). NDVIMAX, normalized deviation vegetation index (maximum). ATRNG, air temperature (range). ATVM, air temperature (variance of mean). ATMN, air temperature (mean). ATPH1, air temperature (phase 1). MIRMN, middle infrared reflectance (mean). NDVIPH1, normalized deviation vegetation index (phase 1). DBR distance to nearest recorded badger presence. NDVIMN, normalized deviation vegetation index (mean). VPDPH2, vapour pressure deficit (phase 2). Figures in parentheses are variable Wald statistics. The model with all predictors is provided in Supplementary Table 2.

extend the earlier modelling of known BTB presence using multiple logistic regressions, to compare cattle movement indices with other previously assessed environmental, demographic, agricultural and climatic parameters (detailed in Supplementary Information). Procedures were developed[14] to extract additional parameters from the 97-million-record Cattle Tracing System (CTS) data archive[5]. These cattle movement data for 2000–03 inclusive were incorporated into the suite of predictor variables available to model BTB distributions. With disease data for 2002 and 2003, the analyses showed unequivocally that movement parameters consistently outperform the other variables in predicting BTB distributions (as quantified by their Wald statistic compared with other predictors), and can produce model accuracies with kappa values of about 0.7 and above[15] (Table 1). Of the range of movement variables tested (total number of 'on' movements, number of movements from infected areas and the proportion of movements from infected areas) the proportion of movements from infected areas arriving at a location was found to be most closely associated with disease presence. The predictive power of this variable also exceeds that of distance to previous disease cases, used as an indicator of previous disease status in the vicinity. The analyses were repeated with movement variables from two years earlier (Table 1). These historical variables, which also provided accurate distribution models, were then replaced with concurrent values, and the models were rerun to generate short-term BTB predictions for 2004 and 2005. The modelled distributions and corresponding projections are shown in Fig. 2, together with insets of actual distributions. High kappa values indicate a good match between actual and modelled predictions. The model predictions indicate that, although below the usually accepted 50% threshold for presence, some areas in northeast Wales, Cumbria and the Scottish borders have a 20–30% risk level of BTB occurrence. These areas stand out more in the projections because the patches are larger and have somewhat higher projected risks, implying that, should current trends continue, the disease is more likely to be found there.

However, the projections are heavily dependent on disease distribution in the base year, and any abnormality is reflected in projections derived from it. This can only be avoided by developing process-based models—for which insufficient epidemiological detail is currently available—or by developing projection models based on more than one year's disease data.

Accordingly, a second set of analyses focused on assessing whether stochastic simulation methods of estimating disease spread could provide a dynamic disease spread model. This model, when applied to known starting BTB distributions, would replicate the observed disease spread and, given the use of predictors derived from a time series rather than a single year, could be projected into the medium-term future. Because the spatial and temporal patterns of movement were shown to be consistent from year to year, it was possible to assemble a model incorporating disease data from 1997 to 2003, with a set of predictor variables. To address the probably stratified nature of the disease's dispersal, multi-annual logistic regressions of BTB occurrence were produced for core and remote areas (core areas

Table 2 | Multi-annual multiple logistic regression of BTB occurrence in Great Britain from 1990 to 2003

Parameter	Core	Remote with movement data	Remote with transformed distance
Variables available to model	100	100	100
Variables included in model	2	6	6
−2 log-likelihood	10,612.3	7,196.0	6.255.1
χ^2	1,316.7	5,552.5	6,493.5
P	<0.001	<0.001	<0.001
Correct presence (%)	56.31	79.07	87.67
Correct absence (%)	78.03	86.39	84.51
Overall correct (%)	67.15	82.83	86.04
Overall kappa	0.34	0.66	0.72
Variable 1	NYBTB (689.1)	NYBTB (129.2)	NYBTB (80.7)
Variable 2	DNT (268.6)	DNT (313.5)	DNT (184.7)
Variable 3	-	PMI (451.8)	TDBTD (1,150.1)
Variable 4	-	NDVIMN (87.9)	NDVIMN (54.0)
Variable 5	-	CAD (145.9)	CAD (36.2)
Variable 6	-	PCU (135.0)	PCU (73.2)

NYBTB, number of years of past BTB infection in the 5-km cell. DNT, number of infected 1-km cells in the previous year in a doughnut window 5 km in radius. CAD, cattle density. PCU, percentage of cultivation and managed grassland. TDBTD, transformed distance to the nearest 1 km square with BTB reported in the previous year. Figures in parentheses are Wald statistics.

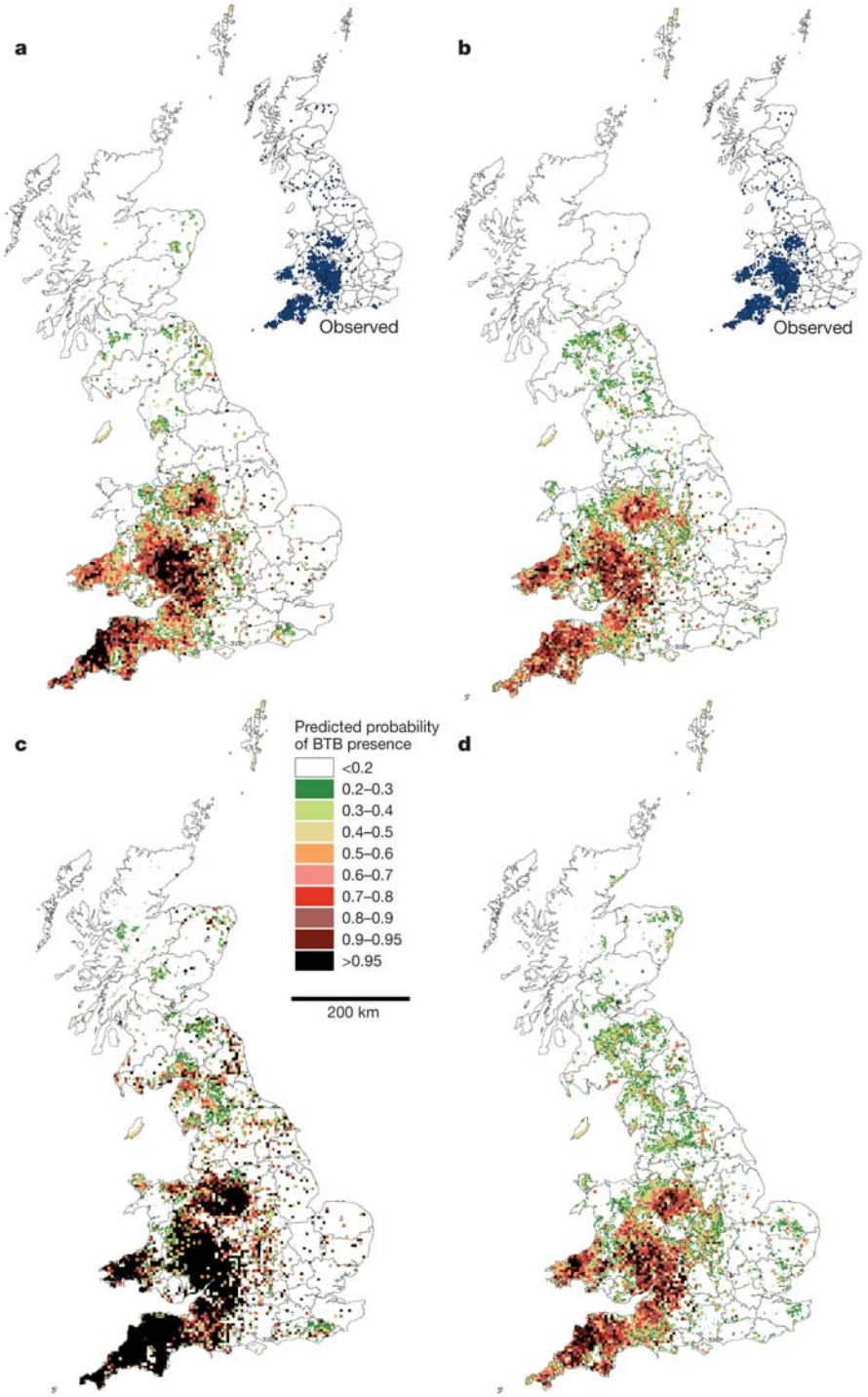

Figure 2 | **Modelled BTB distributions and projections.** **a**, 2002; **b**, 2003; **c**, 2004; **d**, 2005.

Table 3 | **Results of the simulation model**

Parameter estimate period	Overall correct (%)	Correct presence (%)	Correct absence (%)	New cases correct (%)	−2 log-likelihood	R^2	Threshold
PMI model							
1997–2002	84.6	83.3	84.7	81.6	−8,961	0.251	0.010
1997–2001	83.4	82.3	83.4	80.6	−9,408	0.214	0.003
1997–2000	83.9	84.2	83.9	82.6	−9,048	0.244	0.009
TDBTB model							
1997–2002	83.2	82.4	83.2	81.0	−9,165	0.234	0.010
1997–2001	83.5	81.7	83.5	80.4	−9,311	0.222	0.006
1997–2000	84.2	81.8	84.2	80.3	−9,110	0.239	0.010

All simulations used the observed distribution of BTB in 1997 as the starting distribution, and the distribution in 2003 as the predicted year. The threshold is the value that minimizes the difference between percentage of correct presence and percentage of correct absence.

being defined as those where BTB had been detected in two or more of the three previous years; see Table 2 and Supplementary Information).

BTB occurrence in core areas was associated with the previous year's BTB status and with disease status in the previous year in surrounding areas (Table 2). In remote areas, BTB occurrence was associated with the following: variables describing past BTB status; the status of the disease in the previous year in surrounding areas; the proportion of inward movements from infected areas; the density of cattle; the proportion of grassland; and the mean of the normalized deviation vegetation index. Local persistence, short-distance spread from adjacent locations, inward movements from infected areas, host density and two other variables relating to land use and vegetation seem to be significant factors associated with the establishment of new infections away from established foci. The predictions in remote areas are better than in core areas (as quantified by their higher kappa value; Table 2) probably because the latter represent regions of endemicity, where changes in disease status are more difficult to predict consistently by using environmental variables. The possible introduction of BTB by animal movements might also be less critical in such areas than in those where the disease has yet to become established.

The distribution modelling demonstrates the value of movement indicators in short-term projections, and the simulation analyses validate medium-term projections and suggest that the multi-annual models can reproduce the known spread of BTB effectively (Table 3, Fig. 3). Models using parameters derived from an analysis of BTB distribution in the periods 1997–2002 and 1997–2000 provided reasonably good predictions, whereas the model using the period 1997–2001 for determining the parameters tended to underestimate the distribution of BTB in 2003. This underestimation results from the smaller number of detected cases in 2001 (Fig. 1c), which reduces the probability of BTB presence in the modelled time-series. The model including the proportion of movement from infected areas (PMI) best reproduces the observed distribution in 2003, followed by the model in which the distance from nearest infected area is weighted by the frequency distribution of movement distances (TDBTB), which tends to underestimate the observed spread (Table 3, Fig. 3b, d). Although the TDBTB model could be used to generate predictions (Fig. 3e, f), a comparison of the outputs of the TDBTB and PMI models indicates that better predictions would be generated by a PMI model for which the proportion of inward movement from infected areas was estimated at each simulation run rather than extracted from a fixed distribution. This would require 'on-the-fly' extraction of inward movements from the CTS database, which, although computationally intensive, might justify further exploration. With a present maximum of only four predictors, other than variables concerning the past status of BTB, the predictive power of the simulations might also be improved by incorporating additional variables; the limitation is computational processing time.

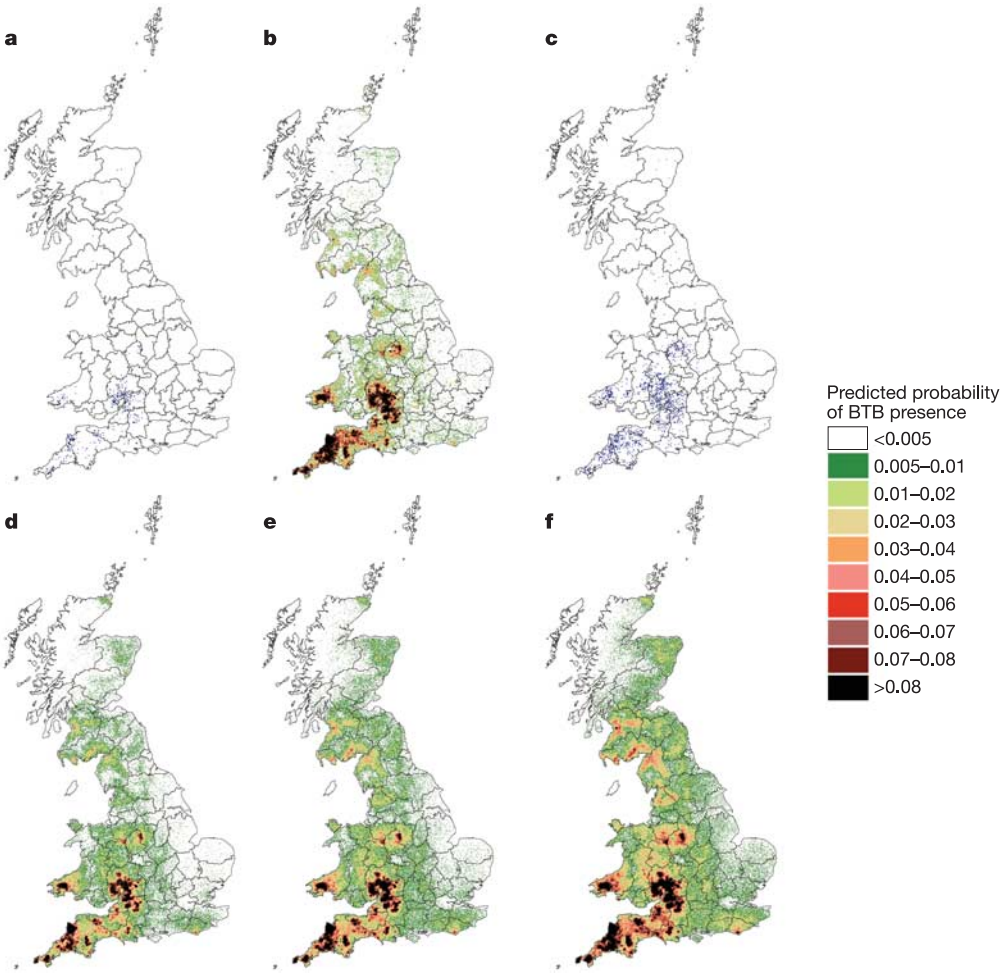

Figure 3 | Observed and predicted distributions of BTB. **a**, Starting distribution in 1997. **b**, Distribution predicted by the PMI model for 2003, using the 1997–2002 periods for estimates of parameters. **c**, Observed 2003 distribution. **d**–**f**, Predicted distribution for 2003 (**d**), 2004 (**e**) and 2005 (**f**) using the TDBTB model and the 1997–2002 periods for estimates of parameters. The average numbers of predicted cases by the TDBTB model were 1,750, 2,224 and 2,799 for 2003, 2004 and 2005, respectively.

The British Cattle Movement Service and the CTS were set up to ensure the identification and traceability of individual cattle during the recovery period after the BSE crisis; they were not intended to serve as a disease-control support system for fast-moving diseases, such as the outbreak of foot-and-mouth disease that occurred in 2001 (ref. 2). However, as the number of detected BTB cases continues to rise exponentially (Fig. 1c), the need to identify critical risk factors becomes ever more important. These results demonstrate that the movement of animals, especially from locations where BTB is present and particularly to locations outside endemic core areas, is one such critical factor.

These findings support the case for movement controls, especially from 'core' to 'remote' locations, as a disease control measure. The models developed here can be used to assess the potential impact of various movement control strategies and so provide underpinning for cost–benefit analyses of this approach to controlling BTB.

However, it should be pointed out that despite the demonstrated association of movement levels with the disease, there are several regions into which many animals are imported where the disease appears regularly but where it does not seem to persist. Potential explanations include the possibility that the critical movement categories (if any) do not occur in these areas, the fact that the imported animals may remain in them only briefly before being slaughtered or moved elsewhere, and that either suitable wildlife reservoirs do not exist in those areas or that conditions necessary for establishment in those species are not fulfilled.

This work has therefore established a clear requirement for further examination of cattle movements to define critical movement categories and to incorporate length of stay into the predictive models, both of which are readily feasible using the methods presented here. Combining cattle movement data, molecular typing data of *Mycobacterium bovis* isolates and knowledge of local wildlife BTB infection status would provide a sound basis for investigating why disease becomes more widespread or persistent in some areas and not others.

METHODS

Data. BTB data were derived from the UK Department for Environment, Food and Rural Affairs (DEFRA) VETNET database for the period 1984–2003 and included BTB monitoring data for more than 90,000 holdings in Great Britain. Two sources of uncertainty are acknowledged in the BTB monitoring data. First, imperfect sensitivity of BTB skin tests may result in a small fraction of false negatives. Positive skin tests in an animal are confirmed only if *M. bovis* typical lesions are identified, or if a culture positive for *M. bovis* is obtained. For a herd breakdown to be considered confirmed it must contain one or more confirmed animals. The proportion of confirmed herd breakdowns was 65.5%, 71.8%, 61.9% and 55.7% in the years 2000, 2001, 2002 and 2003, respectively. Second, BTB testing frequency is not homogeneously distributed in Great Britain: the a testing frequency is once every one to four years, which may result in uncertainty on the dating of detected cases in low-frequency testing areas.

Predictors included a broad range of anthropogenic, biological, demographic, climatic and topographic variables compiled and resampled at 1 km resolution (Supplementary Table 1). Variables representing disease persistence were added to the predictor suite, including the number of years of past BTB infection in a 5-km cell (a measure of temporal local persistence) and the number of infected cells in the previous year within a 5-km doughnut window around the sampling point (as a measure of short-distance spread).

Preliminary treatments of the raw CTS data included the geo-referencing of movement records (98% of locations associated with cattle movement were geo-referenced), pairing movement records[14] (all events except birth and death have two records: one for the location from which the animal had moved and the other for the location onto which the animal had moved; for any analysis of movement patterns, 'off' and 'on' records have to be 'paired'). Movement data from the period 2000–03 were mapped and added to the series of predictors obtained for each 1-km and 5-km cell. These variables included the total number of movements into a cell, the total number of movements from an infected area, and the proportion of movements that originated from infected areas.

Analysis. The association between BTB occurrence and the predictors was explored using a stepwise multiple logistic regression analysis of 2002 and 2003 BTB distribution data. Data values were extracted for data points corresponding to all BTB-positive locations and an equal number of BTB-negative locations systematically (that is, every *n*th negative value within a geographically sorted data set) covering the whole land mass. Models were assessed using the kappa index of agreement[15].

A simulation model was also developed that, when applied to a known starting distribution, would replicate the known spread of the disease and could be projected into the future. Simulation modelling has been used previously to identify landscape characteristics associated with the spread of diseases[16] or invading organisms[13] and to forecast their invasion pattern[17]. The simulation model was built in two main steps.

First, it was necessary to identify the predictors applying to the disease over several years. This involved building a multi-annual database of disease presence and absence, and defining the best predictors and their numeric multipliers. Cattle movement data were available only for the period 2000–03, but the spatial distribution of cattle movement as measured by the number of annual inward movements per 5-km pixel was relatively stable in this period (see Supplementary Information). It was therefore assumed that movement patterns were similar in the previous years, allowing a fixed movement index—the proportion of movement between 2000 and 2003 from infected areas—to be incorporated into the multi-annual predictor suite. In an attempt to differentiate the behaviour of the disease in its focal areas from that elsewhere, and as a first step towards pinpointing where new foci might become established, predictors were identified separately for 'core' and 'remote' areas, with the 'core' areas defined as those cells in which the disease had been found in at least two of the previous three years. Two separate multi-annual multiple logistic regression models were thus built for core and remote areas, respectively, using the observed distribution of BTB in the period 1990–2003 as a dependent variable. Variables relating to local persistence (number of years of past BTB infection in the 5-km cell), to short-distance spread (number of infected cells in the previous year in a doughnut-shaped window 5 km in radius), to animal movements (total movements in, total movements in from infected areas, proportion of movements from infected areas), and to cattle density were entered first; other variables were then entered into the model with a standard forward-entry stepwise procedure. We sought to restrict the model to variables with the highest predictive power, and only those presenting more than 1% of log-likelihood change after removal were retained. Finally, we ensured that all variables used were consistently associated with the presence or absence of BTB over time, in other words that were significant within each year when tested on annual models, by rejecting variables found not significant for more than one year in the period 1997–2003.

The second step involved using the selected variables to simulate the spread of BTB from a starting year. The parameter of each variable was established from multi-annual multiple logistic regression models detailed above, with the incorporation of the year as predictor to account for the trend in increased case numbers. The algorithm used to simulate the spread of BTB involved the following: first, the infection probability of each cell was estimated as a logistic function of the predictors identified through the multi-annual logistic regressions of disease data from 1997 to 2003; second, a layer of random numbers uniform over the interval [0–1] was generated and cells with a random number lower than their infection probability were set as BTB-present; and third, each cell's BTB status was updated and the algorithm reiterated to simulate the spread in the next year. Processing constraints and the current structure of the CTS database prevented movement data to be generated on the fly during the simulation process as a function of the previous year's simulated distribution; it was therefore necessary to identify a surrogate variable for animal movement from previously infected areas. The animal movement kernel, averaged over the years 2000–03, was modelled as a function of distance (see Supplementary Information). The function obtained was used at simulation time to transform geographical distance to the nearest square infected in the previous year into a surrogate index of potential inward movements. The algorithm started with the observed distribution of BTB in a first year and was iterated until the target year. This set of *n* hypothetical distributions of BTB in *n* consecutive years constitutes a single run, of which 500 were performed; these distributions were averaged over the 500 runs to derive the predicted probability of presence at 1 km resolution. This was compared to the observed distribution at the same scale by estimating the model -2 log-likelihood and McFadden's pseudo-R^2; the threshold used to compare predicted and observed presence and absence was determined so as to minimize the difference between the percentage of correct presence and the percentage of correct absence. The model was tested for its ability to predict the distribution in 2003, using parameters derived from the multi-annual logistic regression analysis of the periods 1997–2002, 1997–2001 and 1997–2000, so as to test the model's ability to predict the BTB distribution in the last year of the training set (2003), and then for 1–3 years ahead.

Received 20 January; accepted 18 March 2005.

1. Wint, G. R. *et al.* Mapping bovine tuberculosis in Great Britain using environmental data. *Trends Microbiol.* **10,** 441–444 (2002).
2. Gibbens, J. C. *et al.* Descriptive epidemiology of the 2001 foot and mouth disease epidemic in Great Britain: The first five months. *Vet. Rec.* **149,** 729–743 (2001).
3. Haydon, D. T. *et al.* The construction and analysis of epidemic trees with reference to the 2001 UK foot-and-mouth disease outbreak. *Proc. R. Soc. Lond. B* **270,** 121–127 (2003).
4. DEFRA. *Government Strategic Framework for the Sustainable Control of Bovine Truberculosis (bTB) in Great Britain* March (Department for Environment, Food and Rural Affairs, London, 2005).
5. NAO. *Identifying and Tracking Livestock in England* (Stationery Office, on behalf of National Audit Office, London, 2003).
6. Bourne, F. J. *et al. An Epidemiological Investigation into Bovine Tuberculosis—Towards a Sustainable Policy to Control TB in Cattle* (Department for Environment, Food and Rural Affairs, London, 2001).
7. Krebs, J. *et al. Bovine Tuberculosis in Cattle and Badgers* (Ministry of Agriculture, Fisheries and Food Publications, London, 1997).
8. Donnelly, C. A. *et al.* Impact of localized badger culling on tuberculosis incidence in British cattle. *Nature* **426,** 834–837 (2003).
9. Hay, S. I. & Lennon, J. J. Deriving meteorological variables across Africa for the study and control of vector-borne disease: a comparison of remote sensing and spatial interpolation of climate. *Trop. Med. Int. Health* **4,** 58–71 (1999).
10. Hay, S. I., Randolph, S. E. & Rogers, D. J. (eds) *Remote Sensing and Geographical Information Systems in Epidemiology (Advances in Parasitology* Vol. 47*)* (Academic, London, 2000).
11. Hengeveld, R. *Dynamics of Biological Invasions* (Chapman & Hall, London, 1989).
12. Shigesada, N. & Kawasaki, K. *Biological Invasions: Theory and Practices* (Oxford Univ. Press, Oxford, 1997).
13. Gilbert, M., Grégoire, J.-C., Freise, J. & Heitland, W. Long-distance dispersal and human population density allow the prediction of invasive patterns in the horse-chestnut leafminer *Cameraria ohridella*. *J. Anim. Ecol.* **73,** 459–468 (2004).
14. Mitchell, A. *et al.* Characteristics of cattle movements in Britain—an analysis of records from the Cattle Tracing Scheme. *J. Anim. Sci.* **80,** 265–273 (2005).
15. Congalton, R. G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* **37,** 35–46 (1991).
16. Smith, D. L., Lucey, B., Waller, L. A., Childs, J. E. & Real, L. A. Predicting the spatial dynamics of rabies epidemics on heterogeneous landscapes. *Proc. Natl. Acad. Sci. USA* **99,** 3668–3672 (2002).
17. Russell, C. A., Smith, D. L., Waller, L. A., Childs, J. E. & Real, L. A. A priori prediction of disease invasion dynamics in a novel environment. *Proc. R. Soc. Lond. B* **271,** 21–25 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Cox for comments on an early draft of the paper; D. Rogers and his TALA Research Group at Oxford for advice and for providing satellite imagery; and P. Roeder for help and encouragement. This work was funded by a DEFRA research project.

Author Contributions M.G. did the simulation modelling and spatial statistics. W.W. did the data management, analysis and regression modelling. A.M. and J.M. did the movement data acquisition, processing and analysis. D.B. and R.C.-H. did the interpretation, editing and project management. A.M., J.M. and R.C.-H. are Civil Servants and as such their work is subject to Crown Copyright.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to W.W. (william.wint@zoo.ox.ac.uk).