

# Spatial analysis of *Salmonella* outbreaks in cattle in the United Kingdom

Pingping Zheng<sup>a</sup>, Willem D. Vink<sup>b\*</sup>, Peter J. Diggle<sup>c</sup>, Alasdair J. Cook<sup>d</sup> and Nigel P. French<sup>b</sup>

<sup>a</sup> Department of Mathematics and Statistics, Lancaster University, Lancaster, U.K.; <sup>b</sup> Institute of Veterinary, Animal and Biomedical Sciences, College of Sciences, Massey University, New Zealand; <sup>c</sup> Department of Medicine, Lancaster University, Lancaster, U.K.; <sup>d</sup> Centre for Epidemiology and Risk Analysis, Veterinary Laboratories Agency, Addlestone, Surrey, U.K.

\* Corresponding author: W.D.Vink@massey.ac.nz



## 1. Introduction

- Salmonella species are foodborne and direct-contact zoonoses which have a major impact on human and animal health.
- S. enterica* infections display substantial variation in their epidemiology; quantitative description of large- and small-scale spatial patterns of *Salmonella* serovars will help target surveillance and control activities towards high risk farms.

## 2. Data

- Statutory reports of outbreaks of two serotypes (*S. Dublin* and *S. Typhimurium* DT104, henceforth abbreviated as Dublin and DT104) in the U.K. cattle population, between 1 January 1994 and 31 December 1995, were used.
- Cases were defined as farms with a positive sample at any time ( $n=1083$  and  $1033$  for Dublin and DT104 respectively). Controls were defined as 77,019 cattle farms in the 1994 agricultural census. Spatial farm location data were obtained.
- The locations of 25 regional laboratories distributed over the U.K. were used to examine reporting bias.

## 3. Kernel smoothing: cases versus controls

- Kernel regression<sup>1</sup> was used to investigate spatial heterogeneity of cases (irrespective of serotype) versus controls.
- We investigated spatial patterns of incidence of cases using a nonparametric method<sup>2</sup>. An optimal bandwidth of 10,000 m was used in the kernel estimate; this was chosen by maximizing the cross-validated log-likelihood function:

$$L_c(h) = \sum_{i=1}^n \sum_{k=1}^m I(Y_i = k) \log\{\hat{p}_k^{(i)}(x_i)\} \quad (1)$$

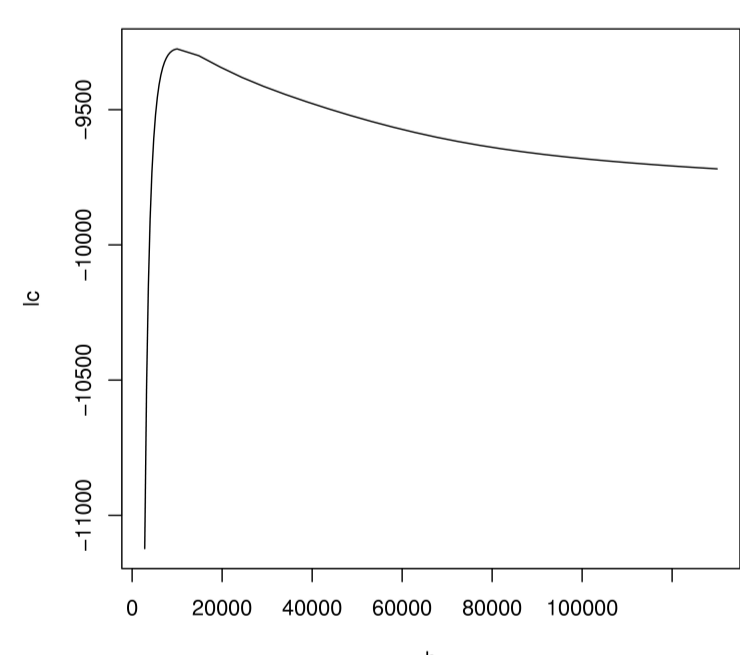


Figure 1. The cross-validated log likelihood function for risk estimation; horizontal bandwidth  $h$  in units of 1 metre.

- The resulting plot of the estimated risk of infection shows spatial heterogeneity, allowing for distribution of at-risk farms.

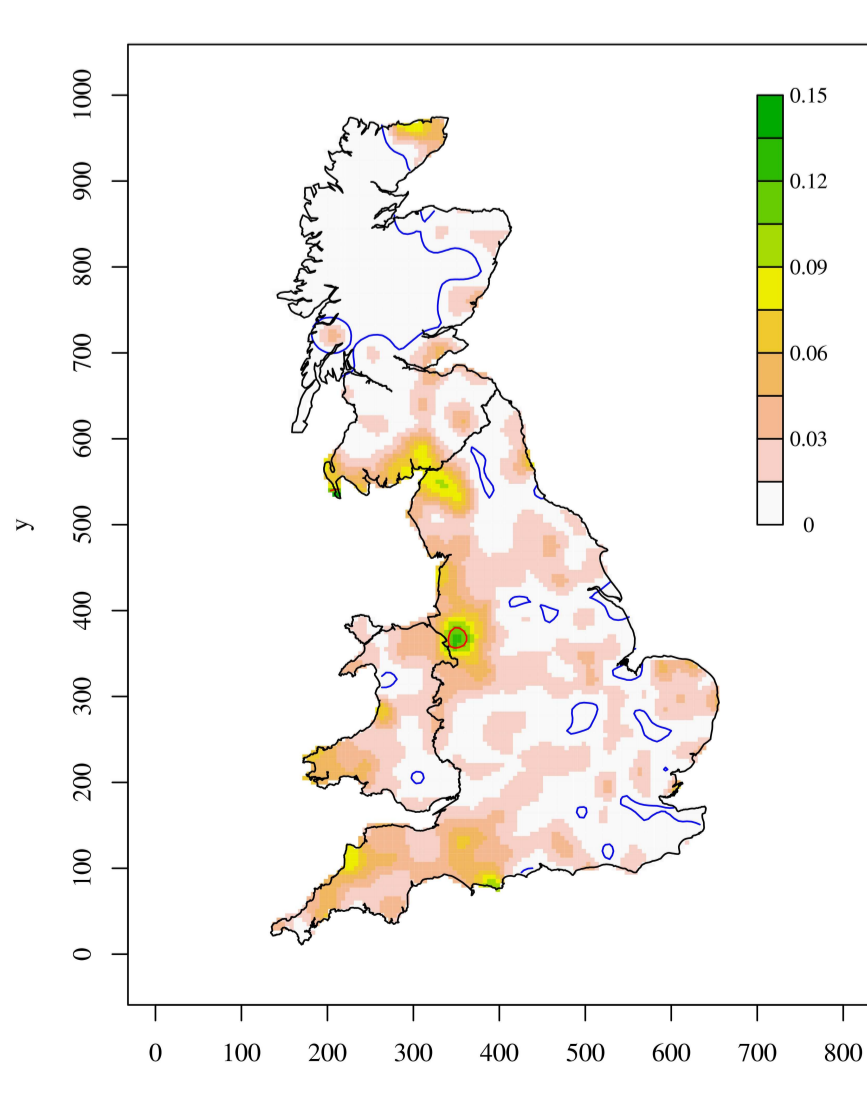


Figure 2. The estimated relative risk of *S. Dublin* and *S. Typhimurium* DT104, under the assumption of homogeneity of the risk. The contours show the  $z$  statistic, i.e. local areas where the risk probabilities are average + ( $2 * \text{s.e.}$ ) ( $z = 2$ , red) and average - ( $2 * \text{s.e.}$ ) ( $z = -2$ , blue).

## 4. Spatial segregation of serotypes

- A similar kernel smoothing approach was now applied to investigate spatial segregation of the serotypes Dublin and DT104.

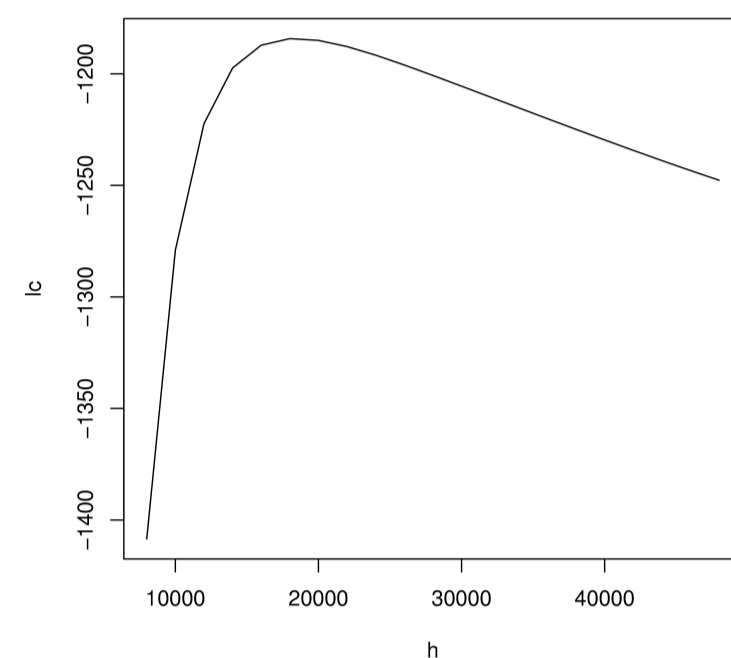


Figure 3. The cross-validated log likelihood function for segregation analysis of Dublin and DT104.

- Using the optimal bandwidth of 18,000m in the kernel regression, the estimated probabilities of the serotypes showed evidence of spatial segregation. The diagrams are complementary as the probabilities must add to 1 at each location.

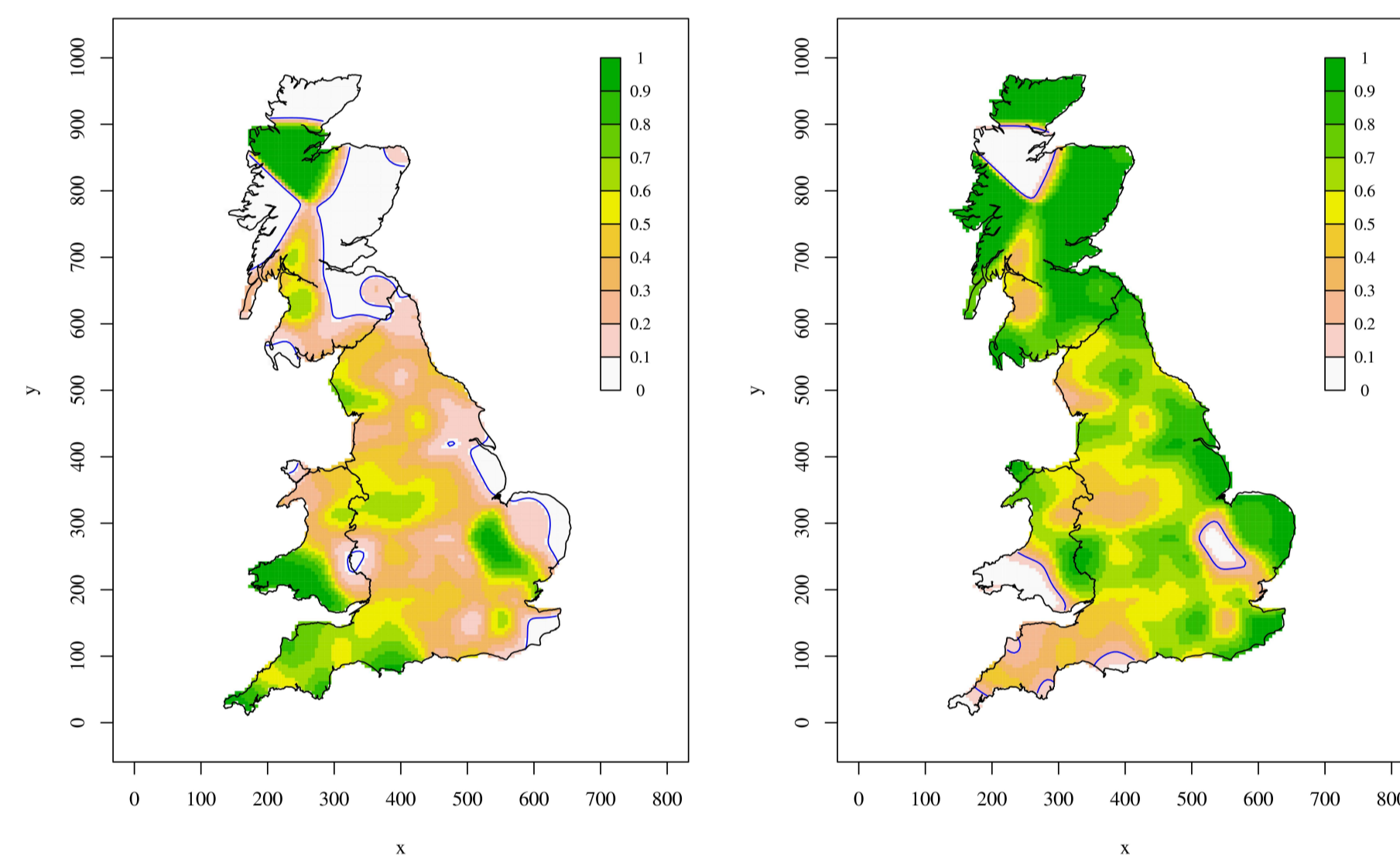


Figure 4. The estimated risk of Dublin (left) and DT104 (right) under the assumption of homogeneity of the risk of being a case of either serotype, with contours of  $z=-2$  (blue) and  $z=2$  (red).

## 5. Distance-related reporting bias: fitting Generalized Additive Models (GAMs)

- Exploratory analysis indicated an inverse relationship between the rates of reporting of farm-level outbreaks and the distances to the nearest regional laboratories.
- To formally investigate this relationship, a generalized additive model (GAM) was fitted.
- We assumed that the spatial location distributions of cases (irrespective of serotype) and controls are described by related inhomogeneous Poisson processes, with intensities satisfying

$$\lambda_1(x) = \lambda_0(x)\rho(x)f(d) \quad (2)$$

where  $\lambda_1(x)$  and  $\lambda_0(x)$  are the intensities of cases and controls at spatial location  $x$ ;  $\rho(x)$  is a spatial variation factor; and  $f(d)$  is a 'reporting factor' – a function of the distance covariate,  $d$ , between the farms and the nearest regional laboratory.

- The probability  $p(x)$  that an event is a case is then

$$p(x) = \frac{\rho(x)f(d)}{1 + \rho(x)f(d)} \quad (3)$$

- Subsequently, we defined  $s_1(x) = \log\{\rho(x)\}$  as a two-dimensional spatial smoothing term of  $x$ , and  $s_2(d) = \log\{f(d)\}$  as a one-dimensional smoothing term of  $d$ . Applying the logit function, substituting and rearranging,

$$\text{logit}\{p(x)\} = \log\{p(x)\}/(1 - \log\{p(x)\}) = s_1(x) + s_2(d) \quad (4)$$

## 5. Distance-related reporting bias (contd.)

- The plot of  $s_2(d)$  shows a large increase in the standard error beyond a distance of 50km.

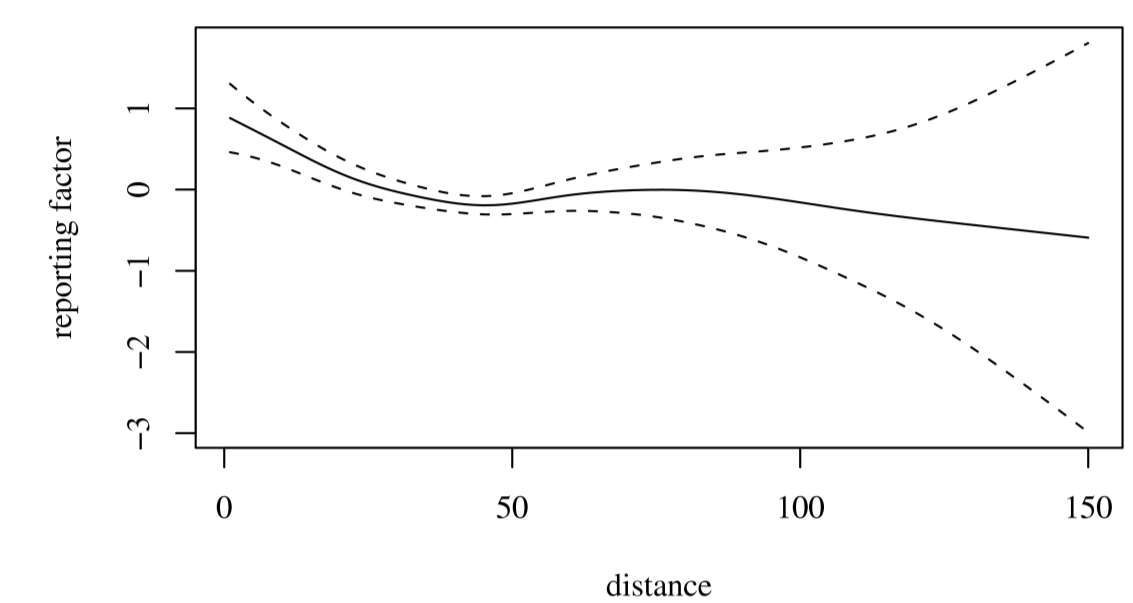


Figure 5. Predicted smoothing reporting factor  $s_2(d) = \log\{f(d)\}$ , with  $2 * \text{s.e.}$  in dashed lines (up to a constant).

- The corresponding  $\hat{f}(d)$  shows a large decrease in reporting rate to a distance of 50km; the subsequent rise may be an artefact due to uncertainty.

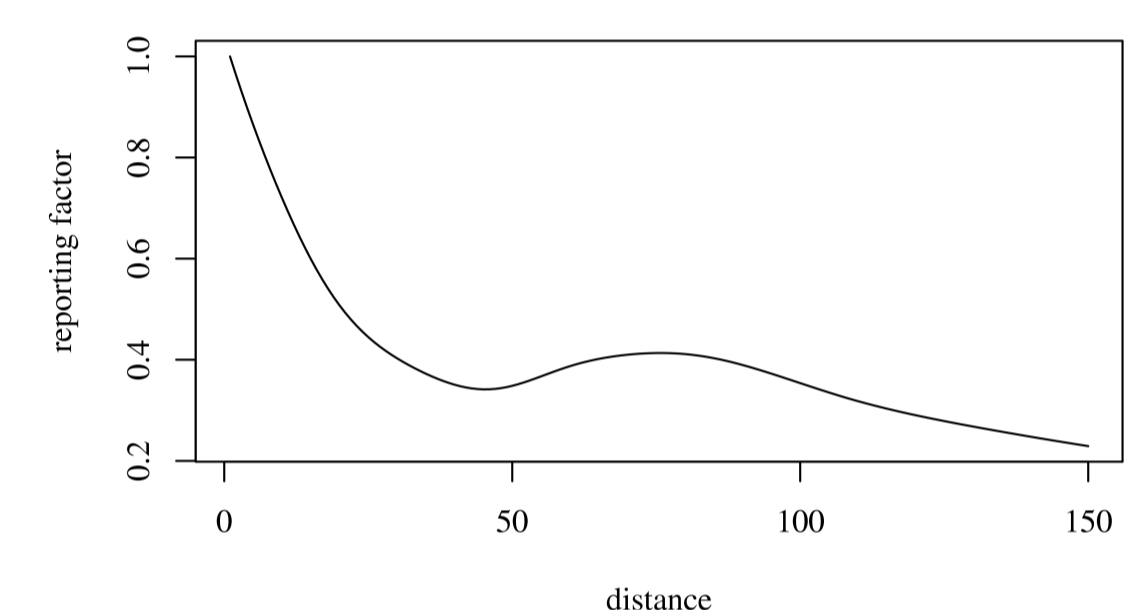


Figure 6. The corresponding reporting factor  $\hat{f}(d)$  rescaled to 1 at  $d = 0$ .

- The estimated spatial effect  $s_1(x)$  and corresponding case probabilities  $\hat{p}(x)$  show a similar pattern of spatial variation in risk as in Figure 2. However, the estimates in Figure 7 are adjusted for the distance-related reporting bias.

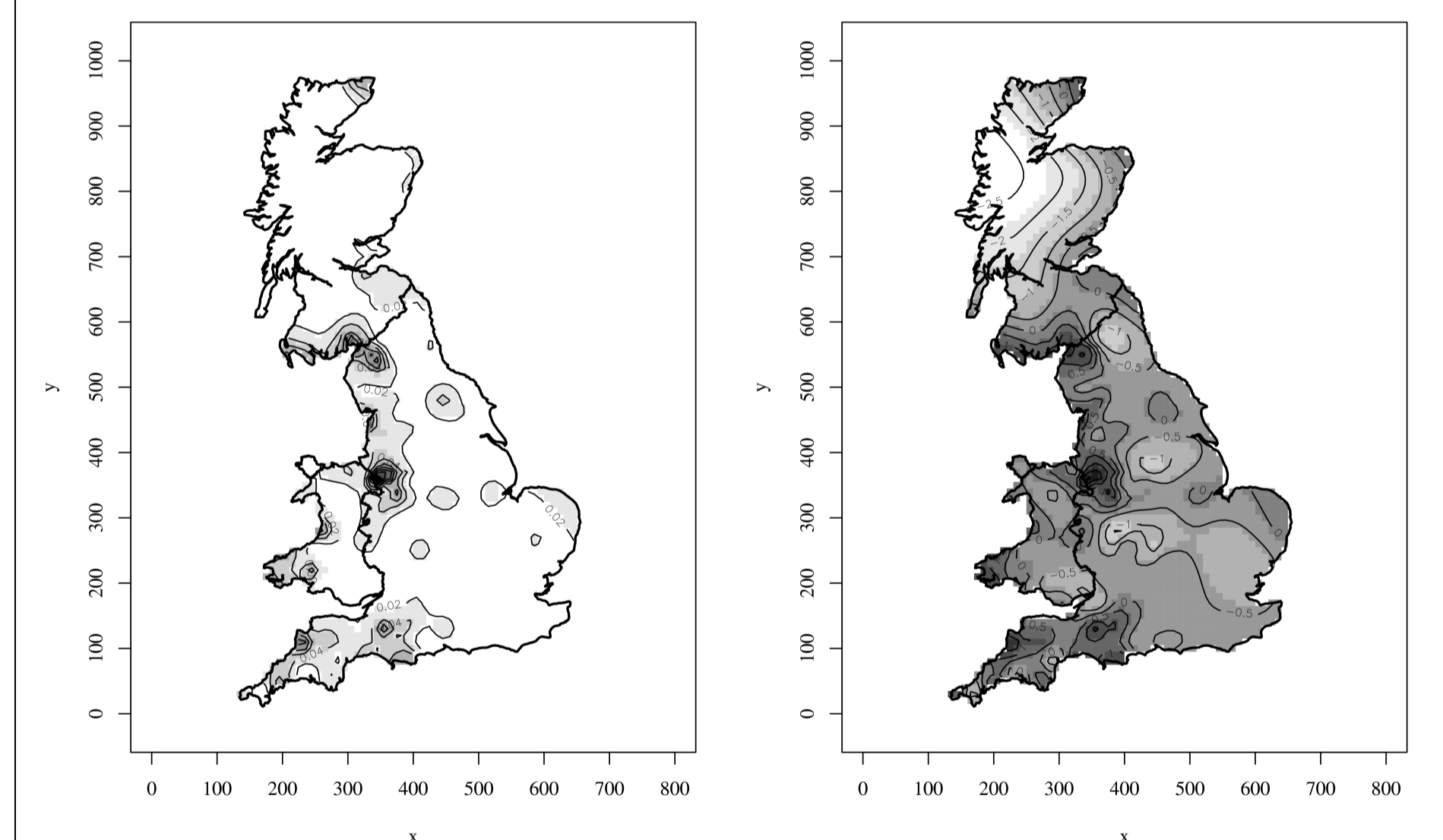


Figure 7. Predicted smoothing of relative case probabilities  $\hat{p}(x)$  (left) and spatial variation of  $s_1(x) = \log\{\rho(x)\}$  (right).

## 6. General discussion

- The kernel regression method and the GAM approach had comparable results, showing clustering of cases relative to spatial variation in intensity of control farms.
- Statistically significant spatial segregation patterns of Dublin and DT104 were identified, allowing for spatial distribution of at-risk farms.
- Evidence of reporting bias was found.
- This work provides evidence for heterogeneity in the distribution of *Salmonella* serotypes in the U.K. dairy cattle population.

This work was funded by Defra (Project OZ0318)

## References

- Kelsall, J. and Diggle, P. J. *Appl. Statist.* **47**(4), 559–573 (1998).
- Diggle, P. J., Zheng, P., and Durr, P. A. *Appl. Statist.* **54**(3), 645–658 (2005).