Supporting Information

Appendix 1. Supplementary information regarding TB in cattle

Table S.1. Frequency of herd testing in the National Program for the Bovine Tuberculosis Eradication (Anon, 2013, 2021). Herd prevalence is defined as the percentage of herds in the Autonomous Community with at least one animal testing positive, while animal prevalence is the percentage of animals testing positive in the given herd.

Herd prevalence in the Autonomous Community	Animal prevalence in herd	Frequency of sampling
0 %	0 %	Once every two years
< 1 %	0 % during at least 1 year	Once every year
	Otherwise	Twice every year
> 1 %	0 % during at least 1 year	Once every year
	Otherwise	Three times every year

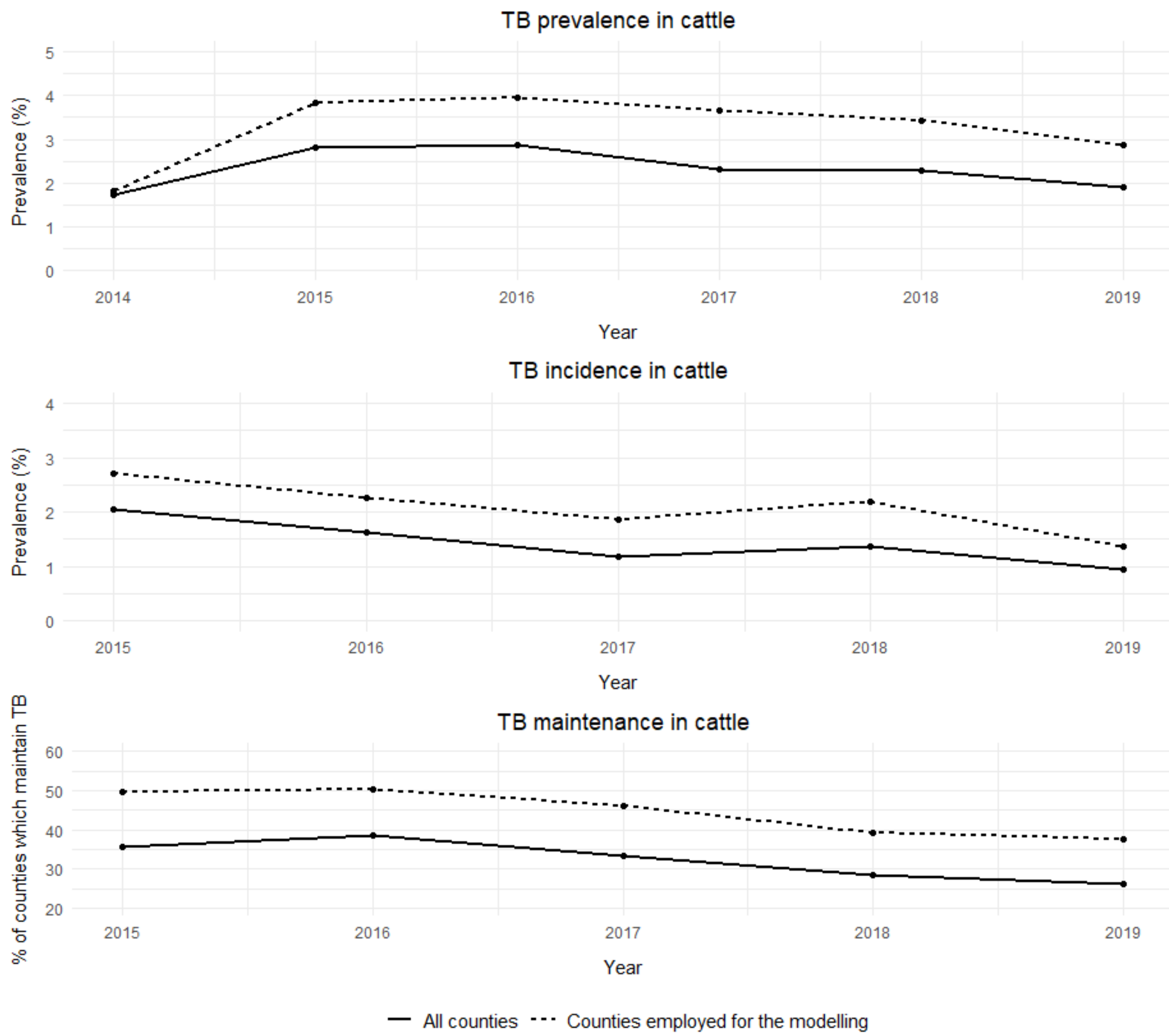

Figure S.1. Temporal pattern of the cattle prevalence (upper graph), incidence (middle graph), and maintenance (lower graph) in all the Spanish counties (continuous line) and the counties employed for the modelling (dashed line).

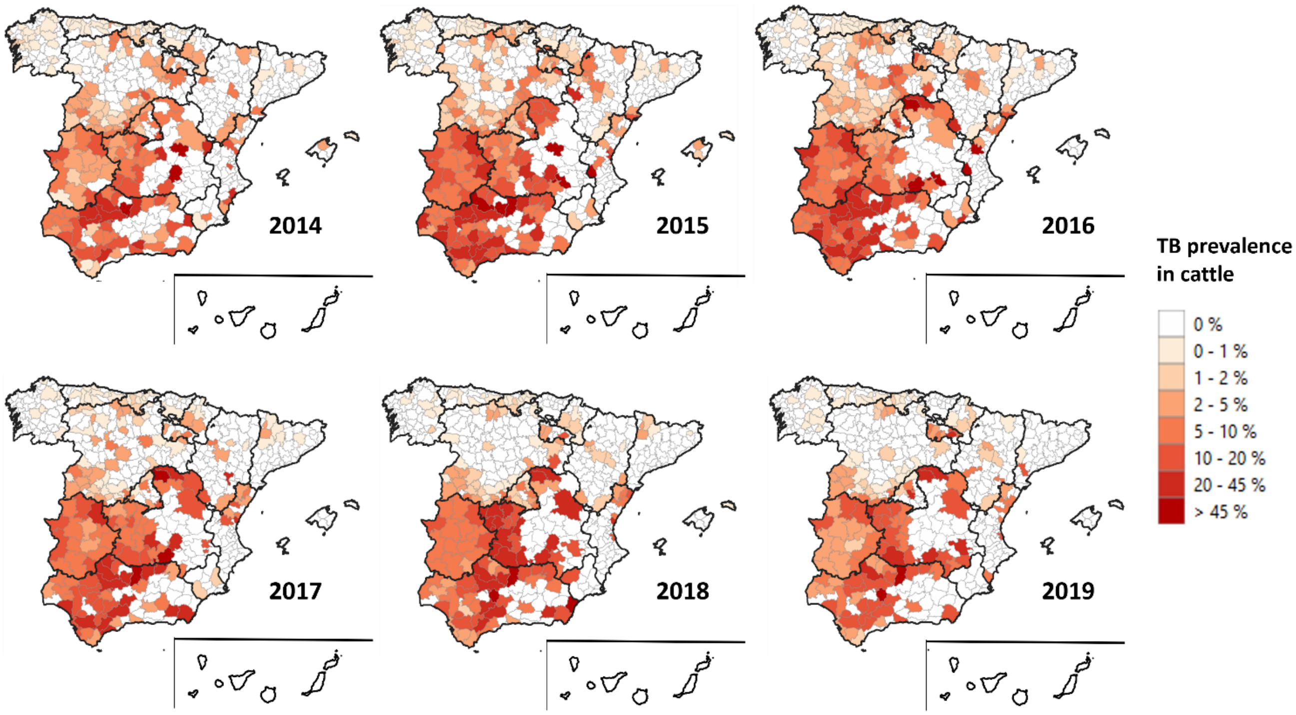

Figure S.2. Spatial distribution of prevalence (% positive herds) of animal tuberculosis in cattle herds during the studied period in Spain. Canary Islands are represented in the box under the map.

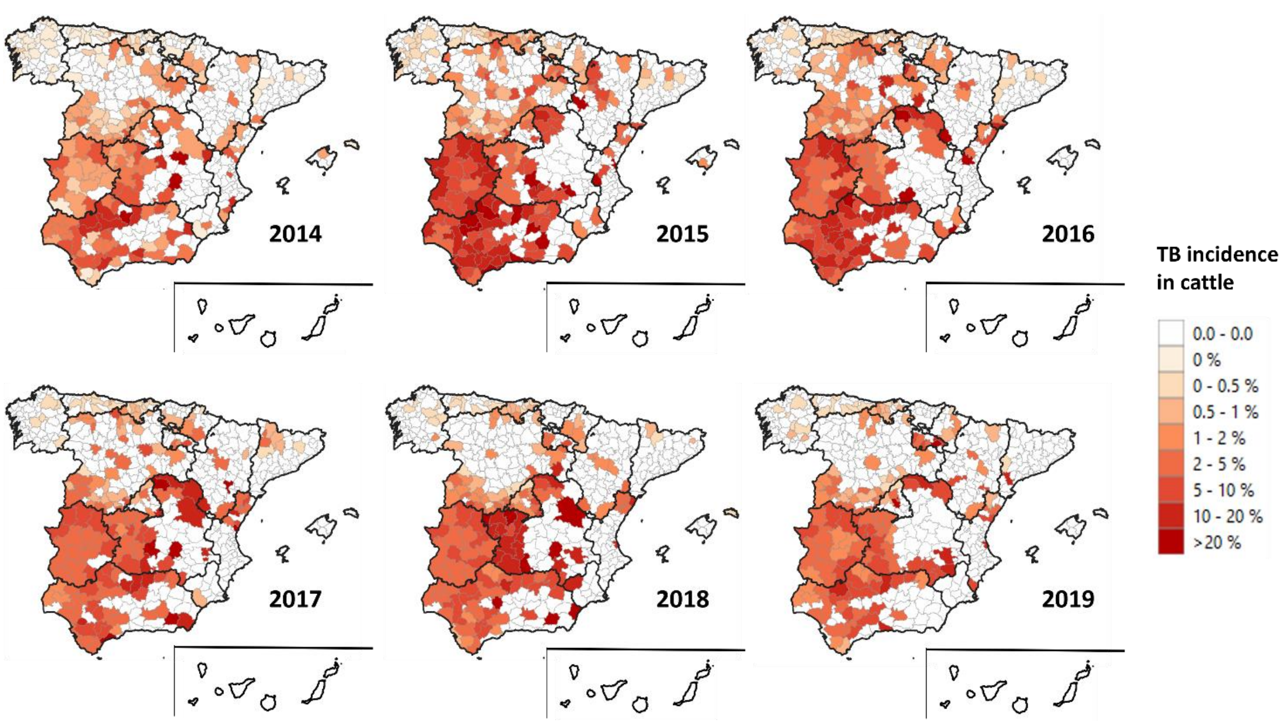

Figure S.3. Spatial distribution of incidence (% new positive herds) of animal tuberculosis in cattle herds during the studied period in Spain. Canary Islands are represented in the box under the map.

Appendix 2. Structured spatial effect definition

The structured spatial effect was defined both by Intrinsic Conditional Autoregressive Models (ICAR; Besag, 1974) and by Stochastic Partial Differential Equations (SPDE) using Delaunay triangulation (Lindgren et al., 2011). The first one creates a neighborhood matrix based in adjacency, while the second one creates a triangular mesh continuous in the space, being able to assign coefficients to any place inside the mesh based on the whole mesh structure (see Figure S.4; Blangiardo and Cameletti, 2015). The SPDE was also tested in three different approaches: (1) Simple Delaunay Triangulation (SDT), with a vertex in each county centroid and additional vertices to complete the mesh limits; (2) Constrained Refined Delaunay Triangulation (CRDT), in which additional vertices are incorporated to obtain triangles as regular as possible in shape and size (Huang et al., 2017; Krainski et al., 2018); and (3) CRDT grouped by year (becoming in a spatiotemporal effect; Blangiardo and Cameletti, 2015). The R code for the different approaches of the structured spatial effect is shown in Appendix 5.1.

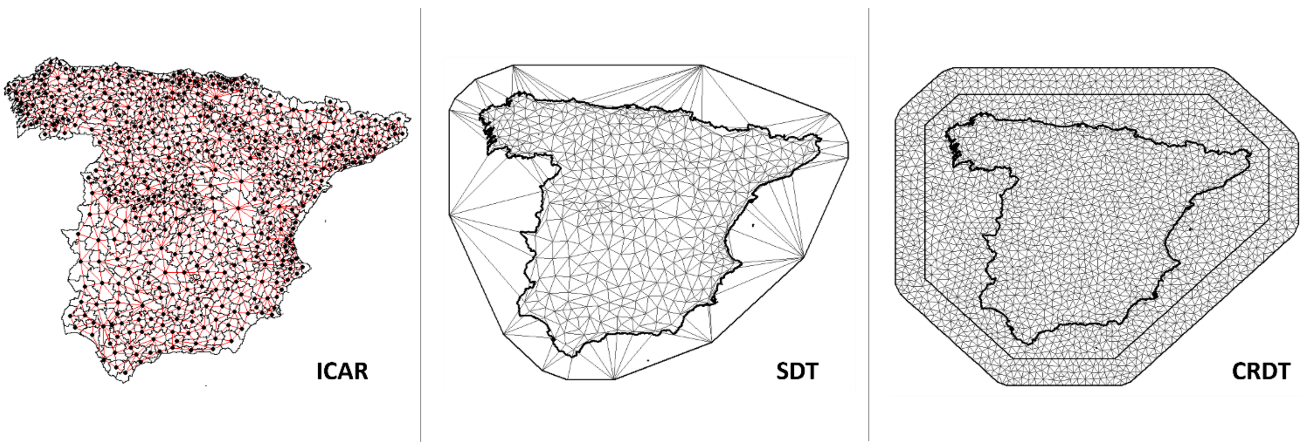

Figure S.4. Structured spatial effect approaches: Intrinsic Conditional Autoregressive Models (ICAR; Besag, 1974) creates a neighborhood matrix in which each county centroid gets a coefficient based on neighbor counties, considering two counties as neighbors if their borders touch at least in one point. Simple Delaunay Triangulation (SDT) and Constrained Refined Delaunay Triangulation (CRDT) are Stochastic Partial Differential Equations (SPDE) approaches for implementing spatial effect. In this SPDE approaches a triangular mesh is created, forming an element which is continuous in the space. This way, coefficients are estimated based in the whole mesh structure, and can be calculated for any place inside the mesh, being more conditioned by values in closer vertices and less conditioned by values in farer vertices (Lindgren et al., 2011). In the SDT each county centroid is a vertex and additional vertices are implemented to complete the mesh boundaries. In the CRDT vertices are added to the SDT mesh in order to obtain triangles as regular as possible in shape and size and mesh limited are extended for avoiding boundary effects, following recommendations described in Huang et al. (2017). CRDT groped by year follows the same spatial structure of CRDT, but the coefficients are calculated for each year instead of for the whole period, becoming in a spatio-temporal effect (Blangiardo and Cameletti, 2015).

Appendix 3. Predictor selection procedure

The step-forward procedure for predictors selection worked as follows:

1. We created a correlation chart for analyzing the collinearity with the chart.Correlation function from the PerformanceAnalytics package (Carl et al., 2010).

2. We adapted the INLAstep function from the INLAutils package (Redding et al., 2017) to work in binomial distribution with our dataset and using Deviance Information Criterion (DIC; Spiegelhalter et al., 2002) as selection criterion (see R code at the Appendix 5.2). Then we executed the function in our dataset to test the possible predictors. When on a given step a model with a DIC value at least five units lower than the previous one was not obtained, the step-forward was stopped and the model with the lowest DIC was selected (Statisticat, 2013; Donovan and Mickey, 2019).

Table S.2. Deviation Information Criterion (DIC) and Conditional Predictive Ordinate (CPO) obtained during the step-forward procedures for the fixed effects selection.

Prevalence model			Incidence model		
Variable	DIC	$-\sum_{i=1}^{n}\log(CPO_i)$	Variable	DIC	$-\sum_{i=1}^{n}\log(CPO_i)$
LR movements [M1]	10781.34	5418.50	LR movements [M1]	6759.13	3395.86
M1 + Red deer ab. [M2]	8696.07	4404.89	M1 + Red deer ab. [M2]	5577.34	2808.84
M2 + Bullfight herds [M3]	7906.65	4034.95	M2 + Bullfight herds [M3]	5281.60	2669.57
M3 + TB in wild boar [M4]	7703.00	3979.26	M3 + TB in wild boar [M4]	5108.38	2614.81
M4 + Wild boar ab. [M5]	7565.35	3918.45	M4 + Wild boar ab. [M5]	5014.60	2575.66
M5 + EHR movements	7498.40	3921.82	M5 + EHR movements	4962.42	2581.27
Maintenance model			**Persistence model**		
Variable	DIC	$-\sum_{i=1}^{n}\log(CPO_i)$	Variable	DIC	$-\sum_{i=1}^{n}\log(CPO_i)$
HR movements [M1]	634.27	2442.99	HR movements [M1]	843.74	424.72
M1 + Beef herds [M2]	619.84	2435.78	M1 + Bullfight herds [M2]	826.31	416.67
M2 + TB in wild boar [M3]	611.48	2431.58	M2 + Beef herds [M3]	803.21	405.35
M3 + Bullfight herds	602.01	2427.09	M3 + TB wild boar [M4]	779.75	393.88
			M4 + Red deer ab.	769.88	389.45

3. For the set of selected predictors, we checked the collinearity in the correlation chart. If two or more predictors were highly correlated (correlation coefficient > |0.6|) the one which was selected earlier in the step-forward process remained, and the rest were excluded. Then, we executed the function again, now without the excluded predictors. We repeated this process until no correlated predictors were included by the step-forward procedure. For prevalence and incidence models, we repeated this process

once. For maintenance and persistence models, no repetitions were needed since no highly correlated predictors were included by the first execution.

4. Once the fixed effects were selected, we carried out the step-forward procedure for the latent effect selection. In this case, given the four possibilities for the definition of the structured spatial effect (ICAR, SDT, CRDT, and CRDT by year), we repeated the step-forward procedure four times, one for each one of them. Hence, the first step of the latent effects selection procedure for prevalence, incidence, and maintenance models, for each one of the structured spatial effects possibilities, was:

   i. Fixed effects + Structured spatial (ICAR/SDT/CRDT/CRDT by Year)
   ii. Fixed effects + Temporal (Year, defined by ARIMA order 1)
   iii. Fixed effects + Unstructured spatial (County, defined by iid)
   iv. Fixed effects + Space-time interaction (County·Year, defined byARIMA order 1)

   For persistence, since the response variable has no temporal variation, the temporal latent effects are not considered. Therefore, the possibilities were:

   i. Fixed effects + Structured spatial effect (ICAR/SDT/CRDT)
   ii. Fixed effects + Unstructured spatial (County, defined by iid)

   As in the fixed factors selection, when on a given step a model with a DIC value at least five units lower than the previous one was not obtained, the step-forward was stopped and the model with the lowest DIC was selected (Statisticat, 2013; Donovan and Mickey, 2019). When the process was finished, four final models were obtained for each disease parameter (three for persistence), one for each spatial structured effect definition methods. The model with the lowest DIC was selected for each one of the parameters. The DIC and Conditional Predictive Ordinate (CPO; Pettit, 1990) of each one of the final models for each parameter and structured spatial effect approaches are shown in Table S.3.

Table S.3. Deviation Information Criterion (DIC) and Conditional Predictive Ordinate (CPO) values for all the models by the different approximations for implementing the structured spatial effect. Abbreviations: Fixed: Fixed effects selected in the step-forward procedure; Int: Space-time interaction (County·Year, defined byARIMA order 1); Temp: Temporal autocorrelation (Year, defined by ARIMA order 1).

Prevalence model			Incidence model		
Spatial autocorrelation implementation method	DIC	$-\sum_{i=1}^{n}\log(CPO_i)$	Spatial autocorrelation implementation method	DIC	$-\sum_{i=1}^{n}\log(CPO_i)$
Fixed + Int + SDT + Temp	2728.68	7717.61	Fixed + Int + SDT	2462.59	7782.04
Fixed + Int + ICAR + Temp	2732.72	8139.04	Fixed + Int + ICAR	2464.89	8240.62
Fixed + Int + CRDT + Temp	2794.47	9017.62	Fixed + Int + CRDT	2535.27	8622.86
Fixed + Int + CRDT by year	2856.08	9501.07	Fixed + Int + CRDT by year	2554.87	8145.05
Maintenance model			**Persistence model**		
Spatial autocorrelation implementation method	DIC	$-\sum_{i=1}^{n}\log(CPO_i)$	Spatial autocorrelation implementation method	DIC	$-\sum_{i=1}^{n}\log(CPO_i)$
Fixed + Int + SDT + Temp	469.29	2390.19	Fixed + SDT	586.26	1713.87
Fixed + Int + ICAR + Temp	469.44	2422.71	Fixed + ICAR	591.16	1430.17
Fixed + Int + CRDT + Temp	483.36	6592.32	Fixed + CRDT	769.88	389.45
Fixed + Int + CRDT by year	627.68	8720.88			

Appendix 4. TB prevalence in wild boar temporal trend

The temporal trend of the TB prevalence in wild boar was assessed by statistical modelling assuming a binomial distribution in the response variable and considering county as spatial latent effect to reduce bias as consequence of the sampling design. Two models were performed: (1) year as categorical predictor, oriented to see the variations between years, and (2) year as continuous predictor, to identify general decreasing or increasing trend patterns through the entire studied period.

Table S.4. Models for analyze (1) annual changes in the TB prevalence in wild boar and (2) trend of TB prevalence in wild boar for the whole study period.

Interannual trend model				
Variable	Coefficient	SD	Credible 2.5 %	Credible 97.5%
Year 2015	1.860	0.089	1.688	2.037
Year 2016	1.608	0.089	1.436	1.785
Year 2017	1.319	0.100	1.123	1.517
Year 2018	1.266	0.087	1.097	1.440
Year 2019	1.660	0.094	1.478	1.846
Spatial effect (σ^2)	2.486	0.260	2.024	3.045
Study period trend model				
Variable	Coefficient	SD	Credible 2.5 %	Credible 97.5%
Year	0.024	0.010	0.004	0.045
Spatial effect (σ^2)	2.590	0.268	2.116	3.168

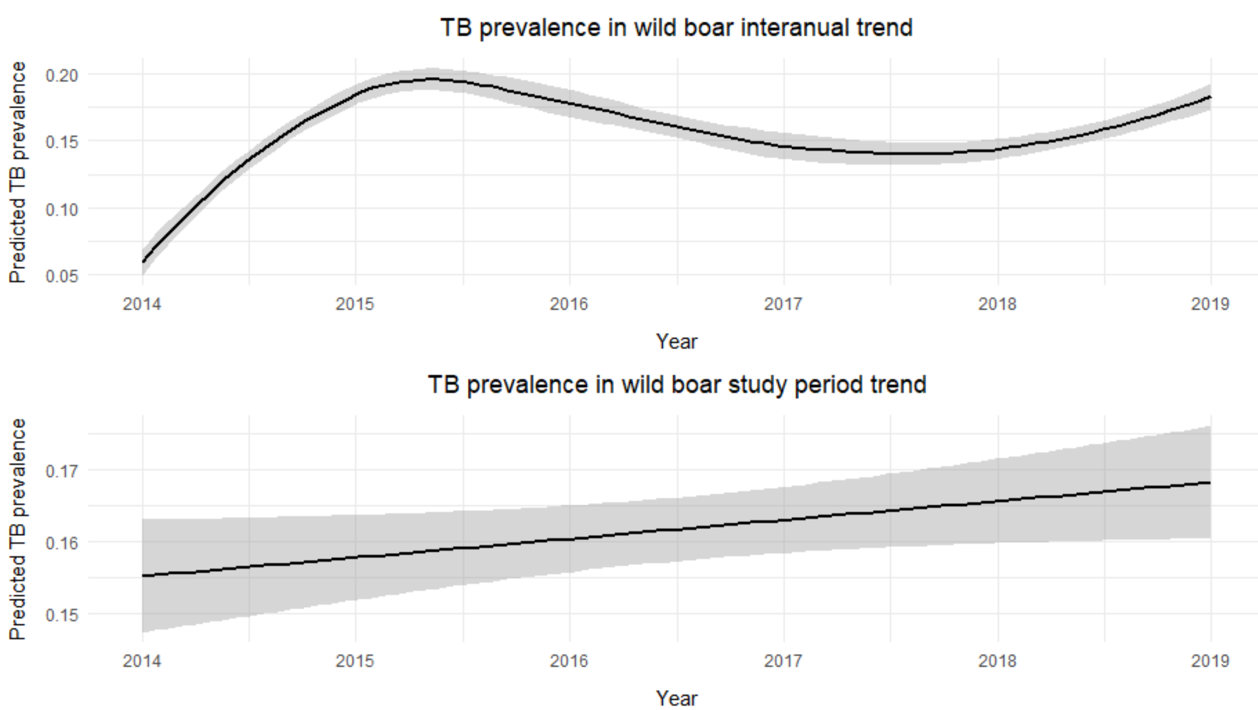

Figure S.5. Interannual trend of TB prevalence in wild boar (upper graph) and general trend of TB prevalence in wild boar for the whole study period (lower graph) based in the models detailed in table S.4. Shadow areas represent 95% confidence intervals.

Appendix 5. R code

R code is available at: https://github.com/cherraiz/Spatio-temporal-modelling-TB.

Appendix 6. Supporting information bibliography

Anon, 2013. Programa Nacional de Erradicación de Tuberculosis Bovina 2013. In: Dirección General de la Sanidad de la Producción Agraria (Ed.) Ministerio de Agricultura, Alimentación y Medio Ambiente, Madrid.

Anon, 2021. Programa Nacional de Erradicación de Tuberculosis Bovina 2021. In: Dirección General de la Sanidad de la Producción Agraria (Ed.) Ministerio de Agricultura, Pesca y Alimentación, Madrid.

Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society: Series B (Methodological) 36, 192-225.

Blangiardo, M., Cameletti, M., 2015. Spatial and spatio-temporal Bayesian models with R-INLA. John Wiley & Sons.

Carl, P., Peterson, B.G., Peterson, M.B.G., 2010. Package 'PerformanceAnalytics'. Retrieved March 29, 2011.

Huang, J., Malone, B.P., Minasny, B., McBratney, A.B., Triantafilis, J., 2017. Evaluating a Bayesian modelling approach (INLA-SPDE) for environmental mapping. Sci Total Environ 609, 621-632. https://doi.org/10.1016/j.scitotenv.2017.07.201.

Krainski, E., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F., Rue, H., 2018. Advanced spatial modeling with stochastic partial differential equations using R and INLA. Chapman and Hall/CRC.

Lindgren, F., Rue, H., Lindström, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73, 423-498. https://doi.org/https://doi.org/10.1111/j.1467-9868.2011.00777.x.

Redding, D.W., Lucas, T.C.D., Blackburn, T.M., Jones, K.E., 2017. Evaluating Bayesian spatial methods for modelling species distributions with clumped and restricted occurrence data. PLOS ONE 12, e0187602. https://doi.org/10.1371/journal.pone.0187602.