

Modelling the spatial distribution of *Fasciola hepatica* in dairy cattle in Europe

Els Ducheyne¹, Johannes Charlier², Jozef Vercruyse², Laura Rinaldi³, Annibale Biggeri⁴, Janina Demeler⁵, Christina Brandt⁵, Theo de Waal⁶, Nikolaos Selemetas⁶, Johan Höglund⁷, Jaroslaw Kaba⁸, Slawomir J. Kowalczyk⁸, Guy Hendrickx¹

¹Avia-GIS, Zoersel, Belgium; ²Department of Virology, Parasitology and Immunology, Faculty of Veterinary Medicine, Ghent University, Ghent, Belgium; ³Department of Veterinary Medicine and Animal Productions, University of Naples Federico II, Naples, Italy; ⁴Department of Statistics, Informatics and Applications, University of Florence, Florence, Italy; ⁵Institute for Parasitology and Tropical Veterinary Medicine, Freie Universität Berlin, Berlin, Germany; ⁶UCD School of Veterinary Medicine, University College Dublin, Dublin, Ireland; ⁷Department of Biomedical Sciences and Veterinary Public Health, Section for Parasitology, Swedish University of Agricultural Sciences, Uppsala, Sweden; ⁸Department of Large Animal Diseases, Faculty of Veterinary Medicine, Warsaw University of Life Sciences, Warsaw, Poland

Abstract. A harmonized sampling approach in combination with spatial modelling is required to update current knowledge of fasciolosis in dairy cattle in Europe. Within the scope of the EU project GLOWORM, samples from 3,359 randomly selected farms in 849 municipalities in Belgium, Germany, Ireland, Poland and Sweden were collected and their infection status assessed using an indirect bulk tank milk (BTM) enzyme-linked immunosorbent assay (ELISA). Dairy farms were considered exposed when the optical density ratio (ODR) exceeded the 0.3 cut-off. Two ensemble-modelling techniques, Random Forests (RF) and Boosted Regression Trees (BRT), were used to obtain the spatial distribution of the probability of exposure to *Fasciola hepatica* using remotely sensed environmental variables (1-km spatial resolution) and interpolated values from meteorological stations as predictors. The median ODRs amounted to 0.31, 0.12, 0.54, 0.25 and 0.44 for Belgium, Germany, Ireland, Poland and southern Sweden, respectively. Using the 0.3 threshold, 571 municipalities were categorized as positive and 429 as negative. RF was seen as capable of predicting the spatial distribution of exposure with an area under the receiver operation characteristic (ROC) curve (AUC) of 0.83 (0.96 for BRT). Both models identified rainfall and temperature as the most important factors for probability of exposure. Areas of high and low exposure were identified by both models, with BRT better at discriminating between low-probability and high-probability exposure; this model may therefore be more useful in practise. Given a harmonized sampling strategy, it should be possible to generate robust spatial models for fasciolosis in dairy cattle in Europe to be used as input for temporal models and for the detection of deviations in baseline probability. Further research is required for model output in areas outside the eco-climatic range investigated.

Keywords: fasciolosis, dairy cattle, animal health, spatial modelling, Europe.

Introduction

Fasciola hepatica is a trematode parasite with a high economic impact on livestock production worldwide (Mas-Coma et al., 2005). For example, in northern Belgium (Flanders), the yearly cost of infections with this parasite in dairy cattle has been estimated at € 8.2 million or € 30 per adult dairy cow present in the population (Charlier et al., 2007, 2009). Similarly, Schweizer et al. (2005) estimated, using Monte Carlo

simulation, that the median financial loss due to bovine fasciolosis in Switzerland amounts to approximately € 52 million representing a median loss of € 299 per infected animal. Most of these losses arise from reduced milk yield and reduced fertility, with smaller losses due to reduced meat production and liver damage.

In order to identify high-risk zones, several spatial modelling approaches have been suggested, either directly by modelling of the presence/absence of the disease on farms in Europe (Rapsch et al., 2008; McCann et al., 2010; Bennema et al., 2011; Fox et al., 2011; Kantzoura et al., 2011) and worldwide (Fuentes 2004, 2006; Asrat et al., 2007; Valencia-Lopez et al., 2012), or indirectly through species distribution modelling of the intermediate host *Galba truncatula* (De Roeck et al., 2014). These models usually focus on a

Corresponding author:
Guy Hendrickx
Avia-GIS
Risschotlei 33, 2980 Zoersel, Belgium
Tel/Fax +32 3458 2979
E-mail: ghendrickx@avia-gis.com

region within a single country and have often a low spatial resolution. McCann et al. (2010), for example, described the prevalence of fasciolosis in England and Wales at the postcode level with each polygon having a mean surface of 20,000 km². On the other end of the spectrum, some examples can be found describing the potential habitat sites on a single farm with a very high spatial resolution (Charlier et al., 2011, 2014; De Roeck et al., 2014).

Few studies focus on the development of a pan-European model, probably due to the lack of comparable epidemiological surveys of the disease between different countries and study sites. Thanks to the EU-funded project GLOWORM, a harmonised sampling strategy and diagnostic method enabled the creation of a baseline dataset that can be used to model the probability of exposure to liver fluke at the European level. The aim of this paper was to create a presence/absence modelling approach describing the observed patterns of fasciolosis throughout Europe at a medium spatial resolution. This would identify the most important eco-climatic factors determining exposure and allow the exploration of the differences in impact of environmental and meteorological factors between biomes, e.g. of the Mediterranean, temperate or northern variety.

Materials and methods

Five countries (Belgium, Germany, Ireland, Poland, Sweden) were included in a cross-sectional survey. Within each study region, a two-stage sampling strategy was followed. In the first stage, farms from 849 municipalities located in the countries mentioned above were selected and stratified according to three dairy cattle density categories. Within the municipalities, a random sample of dairy farms was taken. In total, bulk milk tank (BMT) samples were collected from 3,359 randomly selected farms (Table 1 and 2, Fig. 1). All samples were tested using the SVANOVIR® *F. hepatica*-Ab ELISA (Boehringer Ingelheim Svanova, Uppsala, Sweden). This test quantifies antibodies to excretory-secretory products of *F. hepatica* and is based on the protocol as described by Charlier et al. (2007). For each of the study regions, a separate histogram was calculated to compare the optical density ratios (ODRs) obtained. In the next step, the maximum ODR value per municipality was retained, as we assumed that this represented the farms with the lowest management impact with respect to fasciolosis. Additionally, data from a cross-sectional survey conducted by Kuerpick et al. (2013) were digitised on-screen and added to the database to improve the data collected in Germany.

Table 1. Administrative units included in GLOWORM sampling strategy.

| Study area | Surface (km ²) | LAU2 (N) | LAU2 (km ²) |
|--------------|----------------------------|----------|-------------------------|
| Belgium | | | |
| Flanders | 13,521 | 308 | 44 |
| Germany | | | |
| Lower Saxony | 47,624 | 1,022 | 45 |
| Mecklenburg | 23,174 | 818 | 28 |
| Ireland | | | |
| Ireland | 70,273 | 3,431 | 20 |
| Poland | | | |
| Podlaskie | 20,187 | 118 | 171 |
| Łódzkie | 18,219 | 177 | 103 |
| Sweden | | | |
| South Sweden | 83,186 | 123 | 604 |

LAU: Local Area Unit.

Table 2. Number of samples and municipalities in every country.

| | Samples | Municipalities |
|---------|---------|----------------|
| Belgium | 1,288 | 281 |
| Germany | 768 | 315 + 849* |
| Ireland | 189 | 54 |
| Poland | 773 | 75 |
| Sweden | 339 | 124 |

*Extra data digitised from Kuerpick et al. (2013).

MODerate-resolution Imaging Spectroradiometer (MODIS) imagery (<http://modis.gsfc.nasa.gov/>) was used to derive daytime and night-time Land Surface Temperature (LST_{DAY} and LST_{NIGHT}, respectively) and two vegetation indices, the normalised difference vegetation index (NDVI) and the enhanced vegetation index (EVI). The temporal Fourier transformed data were made available by Prof D. Rogers, Oxford, UK through the EDENext network (FP7) (www.edenext-data.com). For the MODIS data, the three harmonics, corresponding to the annual, bi-annual and tri-annual seasonal cycles, were extracted and saved. The characteristics of these final Fourier fits were saved as a series of output layers, more specifically the harmonic amplitudes (A1-A3) and phases (P1-P3), and the mean (Mean), minimum (Min) and maximum (Max) over the entire time series. Additionally, Bioclim variables from Hijmans et al. (2005) were added to the set of predictors. The bioclim variables describe seasonal temperature and precipitation patterns for the entire

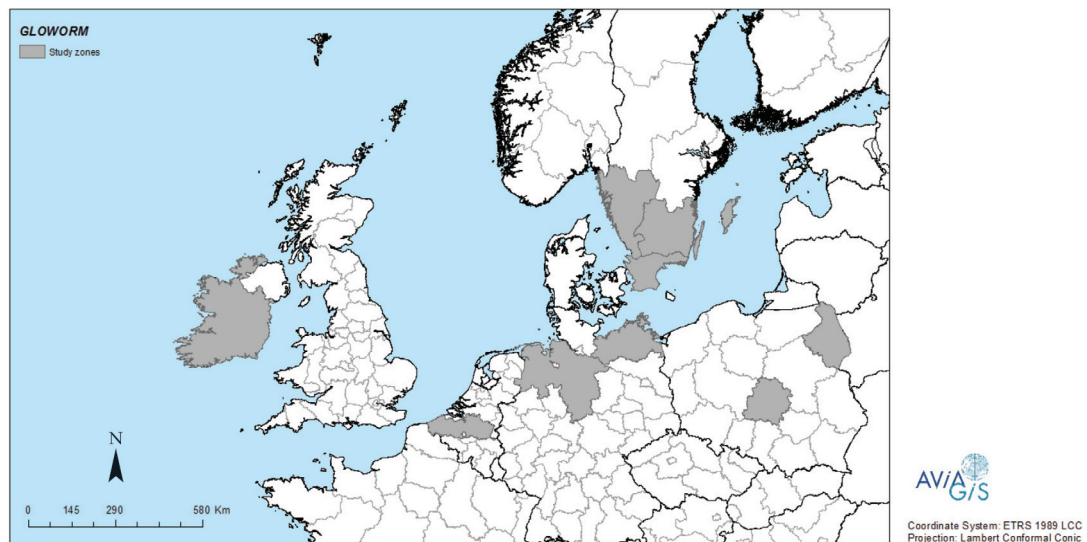


Fig. 1. GLOWORM study zones.

study area. In total, the predictor set comprised 49 variables.

Prior to the modelling stage, the centroid of each of the municipalities was calculated. For each of these centroids, the predictor values were extracted using the EXTRACT function in VECMAP, version 0.9 (Avia-GIS, 2014). An intermediate step involved clustering the data points according to the different eco-climatic patterns allowing recognition of any patterns in the observed presence and absence data. This was especially relevant because we modelled over a large spatial area, where the presence/absence discrimination could be affected by different limiting factors in different geographical zones (e.g. dry limit in the South versus cold limit in the North). Two modelling techniques derived from machine-learning were used to model the probability of exposure to *F. hepatica*. In contrast to statistical techniques, which focus on what model to postulate, how the response is distributed and if observations are independent, the machine learning-techniques assume that the data-generating process is complex and unknown and try to learn the response by observing the inputs and responses and finding dominant patterns (Elith et al., 2008).

Random Forests (RF) is an empirical modelling technique using field observations to establish the relationship between vector occurrence and prevailing environmental conditions. This technique generates various classifications, i.e. in case of a categorical response, a set of classification trees or, alternatively, in case of a continuous response variable, many regression trees (Breiman et al., 1984). A random classification forest approach is a robust ensemble learning-technique, which can be applied to model probability

maps, expressing the probability of occurrence. Each classification tree will provide a predicted model output and the final probability is measured as the proportion of classification trees predicting a presence value for the pixel over the total number of generated classification trees. A major advantage of RF over traditional statistically techniques is that input data do not have to adhere to statistical constraints (e.g. homogeneity of variance, uncorrelated predictor dataset). The technique consistently outperforms traditional modelling techniques such as logistic regression (Cutler et al., 2007; Peters et al., 2011) and has successfully been applied for ecological research projects, e.g. to assess if temperature and precipitation affect the minimum infection rate of *Culex* species for West Nile Virus in Illinois (Ruiz et al., 2010), to model the current spatial distribution of *Aedes albopictus* in Europe using a wide set of predictor variables (ECDC, 2009), and to estimate the probability of presence for *Culex pipiens* and *Aedes vexans* in Belgium (Versteirt et al., 2011).

Boosted Regression Trees (BRT) has been developed by Schapire (2003) and proved successful in modelling extreme values, which can be difficult for RF. Still, this technique has rarely been applied to ecological applications (Elith et al., 2008).

The probability of exposure to *F. hepatica* was first modelled using RF in VECMAP (version 0.9). The number of trees was set to 500, the points sampled for each tree to 49 and the number of variables used at each node to 7. The BRT required a preliminary step to optimise the learning rate, bag fraction and number of trees using a cross-validation accuracy test with 10 folds. Learning-rate and bag fraction were varied to

determine the best configuration for the given data set. The learning-rate was evaluated at 0.001, 0.01 and 0.005 and the bag fraction at values of 0.5, 0.6, 0.7 and 0.75. For each combination the optimal number of trees was determined. These results indicated that a configuration using a learning-rate of 0.005, a bag fraction of 0.5 and the number of trees = 2,250 should yield the highest area under the receiver operating characteristic (ROC) curve (AUC) (Fielding and Bell, 2007), i.e. 0.832 ± 0.0012 .

All model outputs were only generated in areas within the range of the observed eco-climatic dataset to avoid extrapolation. Any results beyond the limits of observation were masked out.

The predictor variable importance was assessed through the measurement of the decline in performance if the model is run without the variable. The performance decline was expressed as the mean decrease of the GINI index (Breiman, 2001). This index is unitless and can only be interpreted as a relative indication of variable importance. It does not allow comparing the importance of a given variable between two models, but rather indicates the relative importance of different variables within one model.

The model accuracy for both techniques was assessed using the AUC, which can be roughly interpreted as the probability that a model will correctly distinguish true presence and true absence from false ones (Fielding and Bell, 2007). Additionally, the model accuracy was assessed using sensitivity and specificity. All indices were calculated using the presence/absence package (1.1.9) (Freeman and Moisen, 2008) in R 3.0.2 (R Core Team, 2014).

Results

Observed field data

The histograms for the different countries are shown in Fig. 2. For all countries, the distribution of the ODR values are skewed to the right. The median ODR amounts to 0.31 in Belgium, 0.12 in Germany, 0.54 in Ireland, 0.25 in Poland and 0.44 in southern Sweden. Given the cut-off of 0.3 for distinguishing between positive and negative samples, it should be assumed that the a total number of 571 villages must be categorised as positive and 427 as negative (Fig. 3).

RF-predicted probability of exposure

Fig. 4a shows the spatial distribution of the probability of being exposed to *F. hepatica* according to

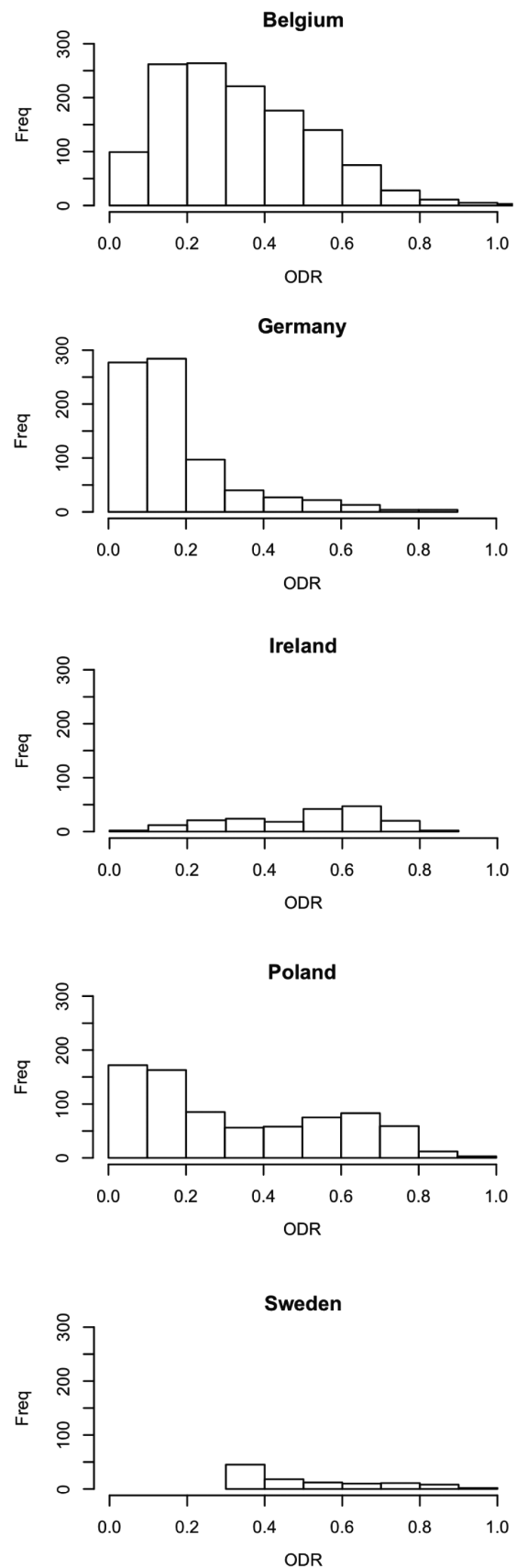


Fig. 2. Histogram of ODR values by country.

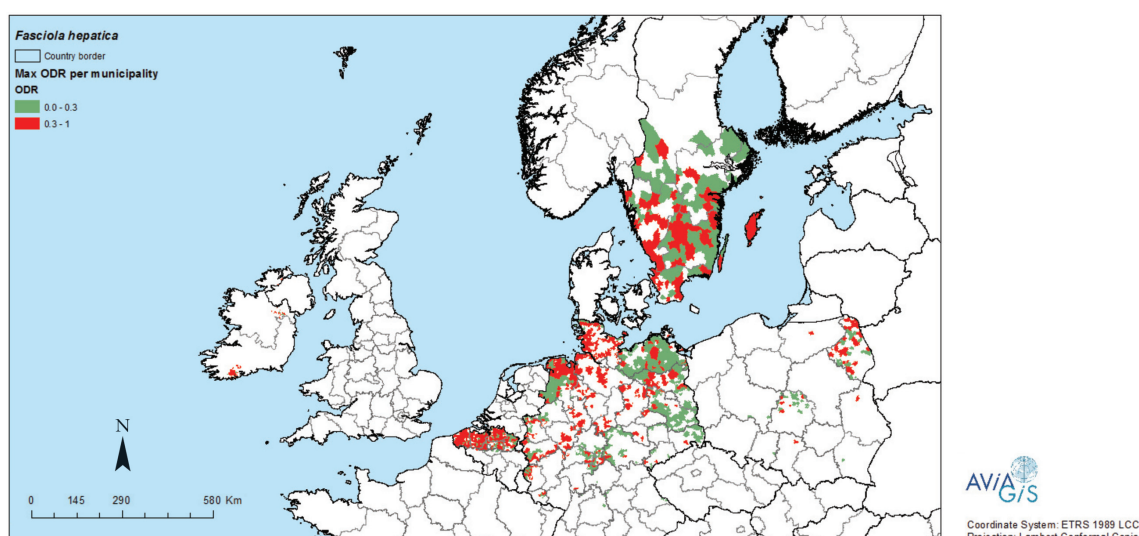


Fig. 3. Results of *F. hepatica* antibody detection in bulk tank milk (BTM) from dairy cattle in sampled municipalities in the GLOWORM study area.

RF. High probability of exposure is predicted in Ireland, the west-coast of United Kingdom, northern France, northern Belgium, The Netherlands, north-west Germany, Denmark, southern Sweden and Norway.

The five most important variables in the case of RF are the mean temperature of the driest quarter, the temperature seasonality (from Bioclim), the mean temperature of the coldest quarter, the precipitation of the driest quarter and the annual precipitation. The model accuracy is shown in Fig. 4b. The histogram plot on the top left indicates the frequency of observed presence and absence sites per predicted probability range. In a perfect model all observed presence sites would be predicted with a probability higher than 0.5, whilst the negatives would be attributed with a probability inferior to 0.5. The histogram shows that there is some confusion between the observed presence sites, which are predicted with a lower probability. The same is true for some absence sites that are modelled as presence. The top right graph displays the number of observed sites proportioned per bin. The ROC plot indicates that the RF is capable of predicting the probability well (AUC = 0.83). Finally, the lower right plot visualises the sensitivity, specificity and Kappa index of agreement in function of the cut-off threshold to classify the probability into actual presence and absence sites. The optimal cut-off, where sensitivity equals specificity is at 0.64. At this level the sensitivity and specificity, amount to 0.74 whilst the Kappa is 0.47. The maximum Kappa index of agreement (0.50) is obtained when using a cut-off of 0.57.

BRT-predicted probability of exposure

Fig. 5a displays the probability of being exposed modelled using BRT. The output map features a similar pattern as for RF. The same areas are indicated as either high and low probability. The biggest difference between the two approaches is the contrast between the zones of predicted presence and absence. Whereas the RF smoothens the output, a distinct contrast is shown by the BRT.

In the case of BRT, the five most important factors are temperature seasonality, the second amplitude of the middle infrared band, the precipitation of the coldest quarter, the second phase of the daytime temperature and the minimum temperature of the coldest quarter. When looking at the accuracy plots (Fig. 5b), the top left plot indicates that presence and absence sites are well discriminated. Indeed, few of the presence sites are assigned a low probability; conversely the absence sites are rarely assigned a higher value. The AUC value amounts to 0.96. The sensitivity and specificity are equal (0.83) when using a cut-off of 0.63, at which point the Kappa (0.77) index of agreement is 0.75. The maximum Kappa can be reached when using a cut-off of 0.49.

Discussion

Within GLOWORM, a unique set of ground-measured data was collected via a harmonised sampling strategy. The sampling strategy did not focus directly on areas with a high prevalence of *F. hepatica*, but used instead a two-stage sampling approach, where

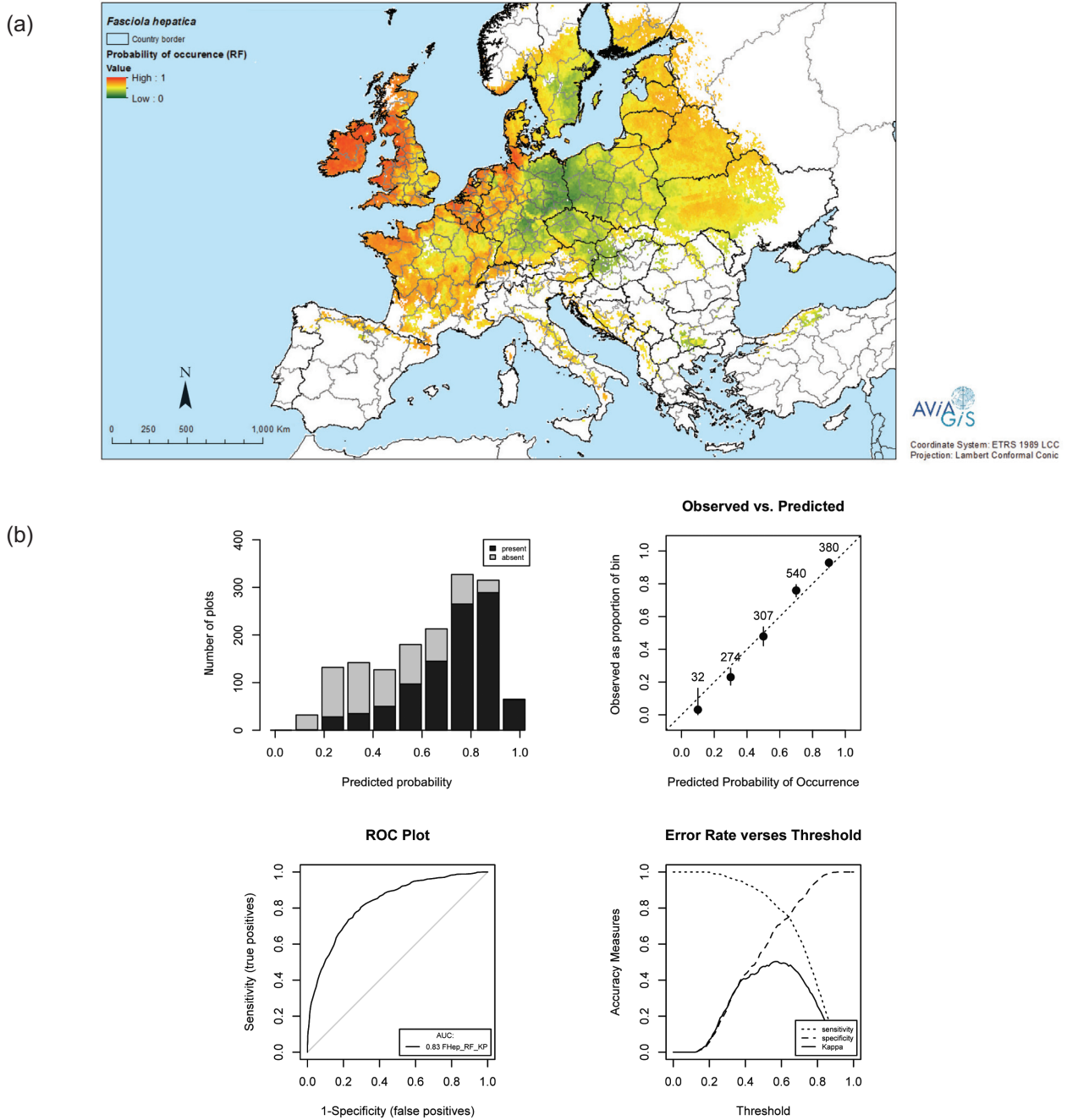


Fig. 4a-b. Probability of being exposed to *Fasciola hepatica* modelled using RF (a) and the spatial distribution of the probability of being exposed to *F. hepatica* according to the RF approach (b).

the municipalities were first randomly chosen followed by selection of farms within these municipalities. While this approach theoretically avoids bias, some bias may in reality have been introduced due to the way that the farms were selected. Due to the diagnostic testing, all samples originated from dairy farms linked to dairy processing companies. Therefore the size of the dairy herds and the management of the dairy farms may differ from farms not linked to dairy processors.

Comparing the ODR values with previous studies such as by Charlier et al. (2007) or Kuerpick et al. (2013) is difficult because they show the outcome from an in-house ELISA test. All results shown here were obtained via the commercially available diagnostic test. In general, the observation of a skewed distribution of ODR is expected because the ODR distribution includes farms being both positive and negative with respect to *F. hepatica* due to the random sampling protocol.

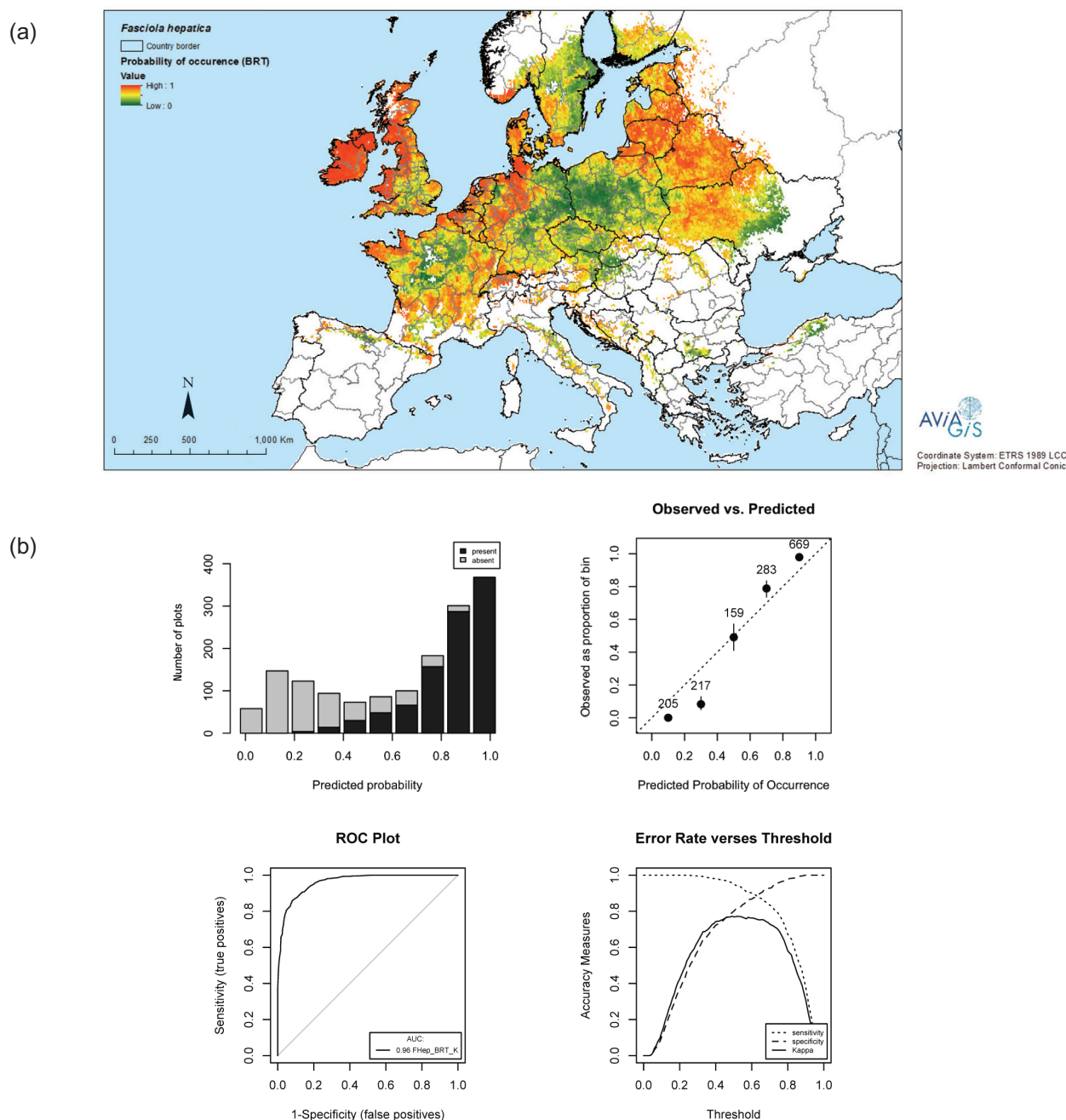


Fig. 5a-b. Probability of being exposed to *Fasciola hepatica* modelling using BRT (a) and the spatial distribution of the probability of being exposed to *F. hepatica* according to the BRT approach (b).

The modelling techniques used here, RF and BRT, proved to be robust while still allowing a biological interpretation. The model yields the probability of exposure as measured by BTM serology. Thus, the model output only indirectly conveys a message about current infection and even more indirectly about the presence/absence of the vector host, *G. truncatula*. Further, it neither accounts for population dynamics of the snail hosts nor the number of generations of the intermediate host. The geographical differences in the

development of the parasite within the snail host are captured via proxies such as rainfall and temperature. Mas-Coma et al. (2009) present a list of useful environmental factors that can be measured from Earth-observing space platforms including variables with respect to temperature and soil characteristics, such as surface water, standing water, soil moisture, vegetation and topography. The set of variables used as input for RF and the BRT is similar. For both RF and the BRT outputs, the most important variables were

found to be related to rainfall and temperature. Different predictors were selected for the two models, but this is probably due to the high correlation between, for example, the various rainfall parameters. Indeed, the absolute value of e.g. precipitation of driest quarter (RF) *versus* the precipitation of the coldest quarter (BRT) should not be overestimated. Because both techniques use bootstrapping, one or the other variable should be selected. The final message is therefore that precipitation and temperature are the most important variables, whereas vegetation is not considered important. This is in accordance with previous studies where an index, such as the one based on temperature and rainfall introduced by Ollerenshaw and Rowlands (1959), is either used to predict fasciolosis directly or employed as a predictor in a statistical model (McCann et al., 2010). The Ollerenshaw index has even been used to assess the impact of climate change on the prevalence of fasciolosis (Fox et al., 2011). Soil type was included as a factor by Bennema et al. (2011), and this helped to predict the spatial distribution if no management data were included. European soil maps are available from the Joint Research Centre (http://eusoiils.jrc.ec.europa.eu/projects/soil_atlas/index.html) at 1-km² resolution, yet they are raster versions of a soil map of the scale 1:1,000,000, which is a simplified representation of the diversity and variability of the soil coverage. Because of this, they were not included in the model.

Both the RF and BRT models are capable of distinguishing areas of high *F. hepatica* exposure from those of low exposure. Countries or regions such as Ireland and the west coast of UK (McCann et al., 2010, Fox et al., 2011) are well known with regard to fasciolosis. Outside the countries sampled, our results are concurrent with previous studies. The Netherlands is predicted to have a high probability overall. This is corroborated by a cross-sectional study carried out in 2012 on 2,500 farms in the Netherlands, where 30.2% of farms from all over the country harboured seropositive animals (Mons, 2012), which was conducted by GD Animal Health (<http://www.gdanimalhealth.com>), a leading Dutch organization in active animal health and animal production. In Switzerland for example, Rapsch et al. (2008) found that fasciolosis remains limited to the northern parts of the country, which they visualise in the Swiss Liver Fluke forecast (http://www.carto.net/rapsch/Schneckenkarte/index_de.svg). Our model predicts moderate to high probability of exposure in Denmark. Ersbøll et al. (2006) investigated the geographic distribution and risk factors for dairy and beef cattle. While dairy cattle overall has a

low prevalence (median prevalence of 0.7%, range 0-14%), beef cattle has a much higher prevalence (median prevalence of 2.6%, range 0-27%).

The models showed here only present results in areas that are similar to the sampled regions. In some countries meteorological and/or climatological conditions are too different and the probability of exposure cannot therefore be predicted. In Italy, large parts fall outside the eco-climatic range from the samples, and are therefore represented as “no data”. In the few areas that are predicted, results indicate that the model is correct, e.g. in Campania in southern Italy, *F. hepatica* appears in a small cluster with both positive bovine and ovine farms corroborating previous results (Cringoli et al., 2002).

In addition to the continental map outputs, there is a need for country-specific models. Indeed, in the case of Ireland, where the risk of exposure to fasciolosis is very high as compared to other European countries, there is still variation between regions within the country itself with the west coast subject to higher risk (Selemetas et al., 2015). Similarly, in the case of Poland, a moderate risk of fasciolosis follows from the model, but the distinction between the high-risk zones and low-probability zones is much more evident within the country.

Conclusions

The use of a harmonised sampling approach in five countries allowed the creation of two pan-European ensemble models, RF and BRT, that delineate the probability of exposure to *F. hepatica* with high accuracy. The two machine-learning models are capable of distinguishing areas with high probability of exposure from those with intermediate and low probability. Still the BRT model is better at discriminating low probability and high probability areas and may therefore be of better practical use. Rainfall and temperature are the most important drivers for the probability of exposure. During the model generation, areas outside the eco-climatic range of the samples were masked out, requiring further field study to increase the spatial extent of the model output. This baseline model can now be used as input to temporal models, to detect deviations in the baseline probability and to assess the economic cost of fasciolosis in Europe.

Acknowledgements

This study was supported by the 7th Framework Programme of the EU (GLOWORM, Project FP7-KBBE-2012-288975). All

partners from the GLOWORM project are acknowledged for their field work and enthusiasm during the project.

References

- Asrat M, Peden D, Jobre Y, Abebe G, Gideyelew T, 2007. Evaluation of the spatial distribution of *Fasciola hepatica* and *Fasciola gigantica* using geographical information system (GIS) in the Nile River Basin, Ethiopia. *Ethiopian Vet J* 11, 41-57.
- Avia-GIS, 2014. VECMAP™: a one-stop-shop for vector mapping, Zoersel, Belgium. Available at: <http://www.avia-gis.com/vecmap> (accessed on June 2014).
- Bennema SC, Ducheyne E, Vercruyse J, Claerebout E, Hendrickx G, Charlier J, 2011. Relative importance of management, meteorological and environmental factors in the spatial distribution of *Fasciola hepatica* in dairy cattle in a temperate climate zone. *Int J Parasitol* 41, 225-233.
- Breiman L, 2001. Random Forests. *Mach Learn* 45, 5-32.
- Breiman L, Friedman JH, Olshen RA, Stone CG, 1984. Classification and Regression Trees. Wadsworth International Group, Belmont, 368 pp.
- Charlier J, Bennema SC, Caron Y, Counotte M, Ducheyne E, Hendrickx G, Vercruyse J, 2011. Towards assessing fine-scale indicators for the spatial transmission risk of *Fasciola hepatica* in cattle. *Geospat Health* 5, 239-245.
- Charlier J, De Cat A, Forbes A, Vercruyse J, 2009. Measurement of antibodies to gastrointestinal nematodes and liver fluke in meat juice of beef cattle and associations with carcass parameters. *Vet Parasitol* 166, 235-240.
- Charlier J, Duchateau L, Claerebout E, Williams D, Vercruyse J, 2007. Associations between anti-*Fasciola hepatica* antibody levels in BTM samples and production parameters in dairy herds. *Prev Vet Med* 78, 57-66.
- Charlier J, Soenen K, De Roeck E, Hantson W, Ducheyne E, Van Coillie F, De Wulf R, Hendrickx G, Vercruyse J, 2014. Longitudinal study on the temporal and micro-spatial distribution of *Galba truncatula* in four farms in Belgium as a base for small-scale risk mapping of *Fasciola hepatica*. *Parasit Vectors* 7, 528.
- Cringoli G, Rinaldi L, Veneziano V, Capelli G, Malone JB, 2002. A cross-sectional coprological survey of liver flukes in cattle and sheep from an area of the southern Italian Apennines. *Vet Parasitol* 108, 137-143.
- Cutler DR, Edwards Jr TC, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ, 2007. Random forests for classification in ecology. *Ecology* 88, 2783-2792.
- De Roeck E, Van Coillie F, De Wulf R, Soenen K, Charlier J, Vercruyse J, Hantson W, Ducheyne E, Hendrickx G, 2014. Fine-scale mapping of vector habitats using very high resolution satellite imagery: a liver fluke case-study. *Geospat Health* 8, S671-S683.
- ECDC, 2009. Development of *Aedes albopictus* risk maps. European Center for Disease Prevention and Control, 45 pp.
- Elith J, Leathwick JR, Hastie T, 2008. A working guide to boosted regression trees. *J An Ecol* 77, 802-813.
- Ersbøll AK, Kähler J, Pederson ND, Thamsborg SM, Larsen M, 2006. Modelling spatial risk factors for occurrence of *Fasciola hepatica* in Danish cattle. Proceedings of the 11th International Symposium on Veterinary Epidemiology and Economics 2006, Cairns, Australia, 920 pp.
- Fielding AH, Bell JF, 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ Conserv* 24, 38-49.
- Fox NJ, White PC, McClean CJ, Marion G, Evans A, Hutchings MR, 2011. Predicting impacts of climate change on *Fasciola hepatica* risk. *PLoS One* 6, e16126.
- Freeman EA, Moisen G, 2008. PresenceAbsence: An R package for presence-absence analysis. *J Stat Softw* 23, 1-31.
- Fuentes MV, 2004. Proposal of a geographic information system for modeling zoonotic fasciolosis transmission in the Andes. *Parasitología latinoamericana* 59, 51-55.
- Fuentes MV, 2006. Remote sensing and climate data as a key for understanding fasciolosis transmission in the Andes: review and update of an ongoing interdisciplinary project. *Geospat Health* 1, 59-70.
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A, 2005. Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* 25, 1965-1978.
- Kantzoura V, Kouam MK, Feidas H, Teofanova D, Theodoropoulos G, 2011. Geographic distribution modelling for ruminant liver flukes (*Fasciola hepatica*) in south-eastern Europe. *Int J Parasitol* 41, 747-753.
- Kuerpick B, Conraths FJ, Staubach C, Froehlich A, Schnieder T, Strube C, 2013. Seroprevalence and GIS-supported risk factor analysis of *Fasciola hepatica* infections in dairy herds in Germany. *Parasitology* 140, 1051-1060.
- Mas-Coma S, Bargues MD, Valero MA, 2005. Fascioliasis and other plant-borne trematode zoonoses. *Int J Parasitol* 35, 1255-1278.
- Mas-Coma S, Valero MA, Bargues MD, 2009. Climate change effects on trematodiasis, with emphasis on zoonotic fascioliasis and schistosomiasis. *Vet Parasitol* 163, 264-280.
- McCann CM, Baylis M, Williams DJ, 2010. Seroprevalence and spatial distribution of *Fasciola hepatica*-infected dairy herds in England and Wales. *Vet Rec* 166, 612-617.
- Mons G, 2012. Leverbot rukt op, *Melkveemagazine* 1, 12-13.
- Ollerenshaw CB, Rowlands WT, 1959. A method of forecasting the incidence of fascioliasis in Anglesey. *Vet Rec* 71, 591-598.
- Peters J, De Baets B, Van doninck J, Calvete C, Lucientes J, De Clercq EM, Ducheyne E, Verhoest NEC, 2011. Absence reduction in entomological surveillance data to improve niche-based distribution models for *Culicoides imicola*. *Prev Vet Med* 100, 15-28.
- R Core Team, 2014. R: A language and environment for statis-

- tical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: <http://www.R-project.org> (accessed on September 2014).
- Rapsch C, Dahinden T, Heinzmann D, Torgerson PR, Braun U, Deplazes P, Hurni L, Hansruedi B, Knubben-Schweizer G, 2008. An interactive map to assess the potential spread of *Lymnaea truncatula* and the free-living stages of *Fasciola hepatica* in Switzerland. *Vet Parasitol* 154, 242-249.
- Ruiz MO, Chaves LF, Hamer GL, Sun T, Brown WM, Walker ED, Haramis L, Goldberg TL, Kitron UD, 2010. Local impact of temperature and precipitation on West Nile virus infection in *Culex* species mosquitoes in northeast Illinois, USA. *Parasit Vectors* 3, 19.
- Schweizer G, Braun U, Deplazes P, Torgerson PR, 2005. Estimating the financial losses due to bovine fasciolosis in Switzerland. *Vet Rec* 157, 188-193.
- Schapiro R, 2003. The boosting approach to machine learning - an overview. In: MSRI Workshop on Nonlinear Estimation and Classification, 2002. Denison DD, Hansen MH, Holmes C, Mallick B, Yu B (eds). Springer, New York, 149-171 pp.
- Selemetas N, Ducheyne E, Phelan P, O'Kiely P, Hendrickx G, de Waal T, 2015. Spatial analysis and risk mapping of *Fasciola hepatica* infection in dairy herds in Ireland. *Geospat Health* 9, 281-291.
- Valencia-López N, Malone JB, Carmona CG, Velásquez LE, 2012. Climate-based risk models for *Fasciola hepatica* in Colombia. *Geospat Health* 6, S75-S85.
- Versteirt V, Boyer S, Damiens D, De Clercq EM, Dekoninck W, Ducheyne E, Grootaert P, Garros C, Hance T, Hendrickx G et al., 2013. Nationwide inventory of mosquito biodiversity (Diptera: Culicidae) in Belgium, Europe. *Bull Entomol Res* 103, 193-203.