

Where the fluke meets the snail: climate, ensemble models, and fasciolosis risk in Ukraine

Volodymyr Tytar, DSc

I. I. Schmalhausen Institute of Zoology, National Academy of Sciences of Ukraine, 15 Bohdan Khmelnytsky Street, Kyiv, 01054 Ukraine; email: vtytar@gmail.com; orcid:0000-0002-0864-2548

Abstract

Fasciolosis, caused by the liver fluke *Fasciola hepatica*, is an economically important zoonotic disease that is endemic in Ukraine, yet the environmental drivers of its distribution remain insufficiently understood and no national-scale risk model has been produced. We compiled occurrence records for *F. hepatica* and its intermediate snail host *Galba truncatula* and used an ensemble species distribution modelling framework to map their potential distributions across Ukraine. Climatic predictors were drawn from the *CliMond* database and screened for collinearity before modelling. Several machine learning algorithms were tuned, evaluated, and combined into weighted ensembles, and the resulting suitability surfaces were compared spatially to identify areas of agreement and divergence. SHAP analysis (SHapley Additive exPlanations) was then applied to both ensembles to quantify and compare the climatic drivers of the parasite and its host. The *F. hepatica* model performed well, whereas the *G. truncatula* model was considerably weaker, indicating that climate alone does not fully explain the snail's distribution at the considered resolution. Despite this, the two suitability surfaces were strongly and positively correlated, pointing to a significant spatial concordance between parasite and snail host suitability. Temperature seasonality emerged as the dominant climatic driver for both species, and permutation tests showed that its net directional effect was similar for the two, while its overall magnitude of influence was significantly stronger for the parasite. Both species showed unimodal responses to temperature seasonality, with intermediate values most favourable; the parasite exhibited a stronger but narrower optimum, whereas the snail tolerated a broader range before the effect turned negative. Overlay analysis identified joint high-risk zones in western and northern Ukraine, parasite-only zones in the north, and snail-only zones in the south that represent potential introduction areas. No region was entirely free of climatic suitability for at least one of the two species. These results indicate that fasciolosis risk in Ukraine is concentrated in the west and north, and that the parasite's distribution is not simply a subset of its host's. Combining ensemble SDMs with explainable AI provides a transferable approach for understanding the environmental drivers of host-dependent zoonoses and for targeting surveillance and control.

Keywords: *Fasciola hepatica*; *Galba truncatula*; species distribution modelling; ensemble models; SHAP; Ukraine; fasciolosis; risk mapping; temperature seasonality; intermediate host

1. Introduction

1.1 Background

Fasciola hepatica (the common liver fluke) is a trematode parasite with a global distribution, infecting a wide range of definitive hosts including cattle, sheep, and other ruminants, as well as humans (Mas-Coma, 2005; Mas-Coma et al., 2022; Lan et al., 2024). The parasite's life cycle requires an intermediate

snail host, typically *Galba truncatula* (formerly *Lymnaea truncatula*), and is strongly constrained by environmental conditions — particularly temperature and moisture — that govern both free-living larval stages and snail population dynamics (Ollerenshaw & Rowlands, 1959; Malone et al., 1987).

In Ukraine, fasciolosis is endemic in many regions, with prevalence varying geographically and seasonally. The disease causes significant economic losses through reduced milk yield, weight gain, and fertility, as well as liver condemnation at slaughter (Honcharov et al., 2024; Lytvynenko et al., 2024). Recent studies have documented an average infection rate of approximately 1.5% in cattle across Ukraine from 2021 to 2023, with the highest rates in the northwestern regions (Honcharov et al., 2024; Lytvynenko et al., 2024). Despite this, the environmental drivers of *F. hepatica* distribution in Ukraine remain insufficiently understood, and no spatially explicit risk model has been developed at the national scale.

1.2 The Role of the Intermediate Host

The distribution and abundance of *F. hepatica* are inextricably linked to the presence and population dynamics of its intermediate snail host, *G. truncatula*. This amphibious snail inhabits small, shallow water bodies, wet meadows, ditches, and the edges of streams and ponds — microhabitats that are highly sensitive to temperature and moisture regimes (Mas-Coma et al., 2009a; Mas-Coma et al., 2009b). The snail's life cycle, reproductive rate, and survival are strongly constrained by environmental conditions, particularly:

- **Temperature:** *G. truncatula* requires temperatures above approximately 10°C for development and reproduction, with optimal temperatures around 15–20°C. Prolonged exposure to temperatures below 5°C or above 25°C reduces survival and reproductive output (Kendall & McCullough, 1951; Poulin, 2006).
- **Moisture:** The snail requires persistent moisture, typically in the form of small water bodies, saturated soil, or high humidity. Drought conditions can cause aestivation or mortality, while excessive flooding can wash away snail populations (Ollerenshaw & Rowlands, 1959; Malone et al., 1987).
- **Vegetation and soil:** The snail is often associated with grassland, wetland, and riparian habitats with dense vegetation and poorly drained soils (Qin et al., 2016).

Because the parasite's free-living stages (miracidia and cercariae) also depend on moisture for dispersal and survival, the environmental drivers of the parasite and its snail host are expected to overlap substantially. However, the snail host may have narrower tolerances or additional microhabitat requirements that are not captured by the parasite's occurrence data alone. Comparative SDM and SHAP analyses of the two species can therefore reveal whether the same climatic variables drive their distributions, and whether their relative importance differs.

1.3 Species Distribution Modelling and *F. hepatica*

Species distribution models (SDMs) have been widely used to map the potential distribution of *F. hepatica* and its intermediate hosts (Ducheyne et al., 2015; Valencia-López et al., 2012; Qin et al., 2016; Selemetas et al., 2015; Assefa et al., 2026). These models typically relate occurrence records to environmental predictors — most commonly temperature and precipitation variables — to estimate habitat suitability across space and time. In Europe, Ducheyne et al. (2015) used ensemble modelling (Random Forests and Boosted Regression Trees) to map the probability of exposure to *F. hepatica* in dairy cattle across five countries, identifying rainfall and temperature as the most important drivers. Similarly, Valencia-López et al. (2012) developed climate-based risk models for *F. hepatica* in Colombia using growing degree day–water budget (GDD-WB) indices and MaxEnt ecological niche modelling. Qin et al. (2016) applied MaxEnt to map *F. hepatica* infestation in sheep in Qinghai province, China, and found that temperature, precipitation, solar radiation, elevation, and slope were significant predictors. Assefa et al., 2026 highlight the usefulness of ecological niche modelling in

guiding targeted surveillance and control strategies in endemic, resource-constrained environments such as northwest Ethiopia.

Comparatively fewer studies have explicitly modelled the distribution of *G. truncatula* itself. De Roeck et al. (2014) used very high-resolution satellite imagery to map fine-scale habitat suitability for *G. truncatula* in Belgium, while Rapsch et al. (2008) developed an interactive map to assess the potential spread of *Lymnaea truncatula* and the free-living stages of *F. hepatica* in Switzerland. These studies highlight the importance of moisture-related variables and local habitat features for the snail host. A comparative, national-scale analysis of both the parasite and its snail host using consistent environmental predictors and modelling methods would therefore fill an important gap.

1.4 The Case for Ensemble Modelling and SHAP Interpretability

While individual SDM algorithms (e.g., Maxent, Random Forest, Boosted Regression Trees) have been successfully applied to *F. hepatica*, ensemble modelling — which combines predictions from multiple algorithms — often yields more robust and transferable predictions (Araújo & New, 2007; Hao et al., 2019). The 'flexsdm' R package (Velazco et al., 2022) provides a comprehensive framework for building ensemble SDMs, including tools for data partitioning, model tuning, evaluation, and ensemble construction.

However, even the best-performing ensemble model remains a "black box" unless its predictions can be interpreted in terms of the underlying environmental drivers. SHAP (SHapley Additive exPlanations) analysis, borrowed from cooperative game theory, offers a robust method to quantify the marginal contribution of each predictor to model output (Lundberg et al., 2018). Unlike traditional variable importance metrics, SHAP values account for interactions among variables and provide both global importance rankings and local explanations for individual predictions. Recent applications in ecology have demonstrated the value of SHAP for understanding SDMs under collinearity (Nikraftar et al., 2025).

By applying SHAP analysis to both the parasite and the snail host models, we can directly compare which climatic variables are most important for each species, and whether the direction and magnitude of their effects differ. This comparative approach is novel in fasciolosis research and has the potential to reveal whether the snail host is the primary constraint on parasite distribution or whether the parasite's free-living stages impose additional climatic limitations.

1.5 Objectives and Hypotheses

The objectives of this study were:

1. To compile and curate datasets of *F. hepatica* and *G. truncatula* occurrence records in Ukraine from the GBIF database, published and other available web sources.
2. To select a set of climatically meaningful, non-collinear environmental predictors from the *CliMond* database using VIF analysis, shared across both species.
3. To build and evaluate multiple SDMs for both *F. hepatica* and *G. truncatula* using the 'flexsdm' R package, testing a range of machine learning algorithms.
4. To construct ensemble models for both species based on weighted averages of individual models, using AUC, TSS, and the continuous Boyce index as evaluation metrics.
5. To assess the spatial correlation between the parasite and host suitability maps using SAGA GIS.
6. To apply SHAP analysis to both ensemble models to quantify and compare the importance and directionality of climatic drivers for the parasite and its intermediate host.
7. To produce national-scale risk maps for *F. hepatica* in Ukraine and discuss implications for surveillance and control.

We tested two underlying hypotheses based on the the following considerations:

- **H1:** The spatial distributions of the two species would be significantly correlated. This is essentially a spatial corollary of the obligate host relationship. *F. hepatica* cannot complete its life cycle without *G. truncatula*. Sustained transmission therefore requires overlap between suitable parasite and snail environments. At a broad scale, one would expect a positive correlation between their suitability surfaces, especially in the wetter west and northwest of Ukraine.
- **H2:** Temperature seasonality, mean annual temperature, and annual precipitation would emerge as the most important drivers for both *F. hepatica* and *G. truncatula*, consistent with their thermal and moisture requirements. Both the snail and the parasite's free-living stages — miracidia and cercariae — are ectothermic and moisture-sensitive. This creates a strong expectation of shared climatic drivers.

2. Materials and Methods

2.1 Study Area

Ukraine is situated in eastern Europe, spanning approximately 44°N to 52°N latitude and 22°E to 40°E longitude, with a total area of about 603,500 km². The country has a predominantly temperate continental climate, but this masks a pronounced west–east and north–south gradient in both temperature and precipitation (Marynych et al., 1985; Zastavniy, 1994; Marynych & Shyshchenko, 2005).

Mean annual temperature ranges from approximately 7–8°C in the northeast to 10–11°C in the south and southwest, with the Crimean Peninsula and the Black Sea coast being the warmest regions. The Carpathian Mountains in the west form a distinct orographic zone, with cooler temperatures and substantially higher precipitation than the surrounding lowlands. Annual precipitation declines from more than 1,000 mm in the Carpathians to roughly 400–500 mm in the southern and southeastern steppe regions. Snow cover is common in winter in the north and northeast, and the duration of snow cover and soil frost is an important determinant of overwintering survival for both the snail host and the parasite's free-living stages.

The country can be broadly divided into three physiographic–climatic zones that are relevant to fasciolosis epidemiology: (i) the forest and forest-steppe zones of the north, northwest, and centre, characterised by cooler, wetter conditions, extensive grassland, and poorly drained soils; (ii) the steppe zone of the south and east, characterised by warmer, drier conditions and limited surface water; and (iii) the Carpathian and Transcarpathian region in the west, with a montane climate, high precipitation, and extensive pasture-based livestock systems.

These climatic gradients are directly relevant to the present study. *G. truncatula* requires persistent moisture and moderate temperatures, and its distribution is therefore expected to be concentrated in the cooler, wetter western and northwestern regions and in localised wet microhabitats (ditches, riparian zones, wet meadows) elsewhere. *F. hepatica* transmission is likewise favoured by these conditions, but the parasite's free-living stages may impose additional thermal and moisture constraints. Ukraine's large agricultural sector — particularly cattle and sheep rearing, which are concentrated in the west, northwest, and centre — provides the definitive-host population necessary for transmission. Previous surveillance has documented an average cattle infection rate of approximately 1.5% for 2021–2023, with the highest rates in the northwestern regions (Lytvynenko et al., 2024), which is consistent with the climatic gradient described above.

2.2 Occurrence Data

2.2.1 *Fasciola hepatica*

Occurrence records of *F. hepatica* in domestic animals (cattle, sheep, goats) in Ukraine were downloaded from the Global Biodiversity Information Facility (GBIF) database (Skorobogatov, Vasyliuk, 2023; occurrence dataset <https://doi.org/10.15468/k5djvr> accessed via GBIF.org on 2024-03-

14). Records were filtered to remove duplicates, records with uncertain taxonomic identification, and records with missing or imprecise coordinates. Spatial thinning was applied to reduce sampling bias and spatial autocorrelation, using a minimum distance threshold (e.g., 10 km) to ensure that no grid cell contained more than one occurrence record (Boria et al., 2014; Radosavljevic & Anderson, 2014). Eventually, 158 georeferenced records were retained for analysis.

2.2.2 *Galba truncatula*

Occurrence records of *G. truncatula* in Ukraine were also downloaded from GBIF (GBIF.org [19 September 2026] GBIF Occurrence Download <https://doi.org/10.15468/dl.m5b2er>). Records were filtered and thinned using the same protocol as for the parasite. Where the number of records was insufficient for robust modelling (e.g., < 25), we supplemented with records from adjacent countries (e.g., Poland, Slovakia, Romania). Together 50 georeferenced records were used.

Longitude and latitude coordinates (WGS84 datum) of the sample records were stored in an Excel database and converted into CSV format for the establishment of the SDM models

2.3 Environmental Predictors

Climate data were obtained from the *CliMond* database (Kriticos et al., 2012), which provides 35 globally gridded, high-resolution (10' or ~18.5 km) climate layers for both current and future conditions.

To avoid multicollinearity, variance inflation factors (VIF) were calculated for all candidate variables, and variables with VIF>10 were removed iteratively (Dormann et al., 2013). The final set of predictors was used for both species to ensure comparability (Table. 1).

Table 1. VIF-selected variables from the CliMond database

Abbreviation	Variable
Bio02	Mean Diurnal Range
Bio04	Temperature Seasonality
Bio09	Temperature of Driest Quarter
Bio14	Precipitation of Driest month
Bio15	Precipitation Seasonality
Bio23	Radiation Seasonality
Bio24	Radiation of Wettest Quarter
Bio26	Radiation of Warmest Quarter
Bio29	Highest Weekly Moisture Index

2.4 Modelling Framework

2.4.1 Data Partitioning

Occurrence data for each species were partitioned into training (75%) and testing (25%) sets using spatial block partitioning to reduce spatial autocorrelation (Roberts et al., 2017). Spatial block partitioning divides the study area into a set of spatially independent blocks and assigns whole blocks to either the training or the testing set, so that training and testing records are not spatially proximate. This approach produces a more conservative and realistic estimate of model transferability than random

partitioning, which can inflate performance metrics when occurrence records are spatially autocorrelated (Roberts et al., 2017).

We used the *part_sblock()* function in the 'flexsdm' R package to implement spatial block partitioning (Velazco et al., 2022). This function requires the user to specify a range of candidate block sizes (in the units of the projection), and it evaluates each candidate by fitting a semi-variogram to the occurrence data and selecting the block size that best reduces spatial dependence between the training and testing folds. We tested block sizes from ranging from 50 to 500 km and selected the size that minimised the spatial autocorrelation of model residuals, following the recommendation of Velazco et al. (2022).

2.4.2 Pseudo-absence and Background Generation

Because presence-only occurrence data do not provide information on where a species is absent, we generated pseudo-absence (or background) points to characterise the environmental conditions available to each species. The choice of pseudo-absence strategy strongly influences model calibration and evaluation, and we therefore adopted a spatially constrained approach consistent with the ecology of both species.

For each species, we generated 10,000 pseudo-absence points using the *sample_pseudoabs()* function in 'flexsdm' (Velazco et al., 2022). We applied three constraints:

1. **Geographic extent:** Pseudo-absences were drawn from within the same calibration extent as the occurrence records, defined as a buffer of 200 km around the occurrence records, to avoid extrapolating into environments that the species could not plausibly reach.
2. **Spatial exclusion:** Pseudo-absences were excluded from a buffer of 20 km around any occurrence record, to reduce the risk of labelling suitable habitat as absent.
3. **Environmental stratification:** Where possible, pseudo-absences were sampled to span the range of environmental conditions within the calibration extent, so that the model could discriminate between suitable and unsuitable environments rather than simply learning the environmental mean.

We generated a single set of pseudo-absences for each species and used the same set across all algorithms, so that differences in performance reflect algorithm behaviour rather than sampling variation. Model evaluation was based on the spatial block test set, which contained only real occurrence records and therefore provided an independent assessment of discrimination ability.

2.4.3 Algorithms

A range of machine learning algorithms was tested for both species, including:

- Maximum Entropy (Maxent; Phillips et al., 2006)
- Random Forest (RF; Breiman, 2001)
- Boosted Regression Trees (BRT; Elith et al., 2008)
- Generalized Boosted Models (GBM)
- Generalized Linear Models (GLM)
- Generalized Additive Models (GAM)

Model tuning was performed using the 'flexsdm' framework, with regularization multipliers, feature classes, and hyperparameters optimized via cross-validation.

2.4.4 Model Evaluation

Model performance was assessed using:

- Area Under the Receiver Operating Characteristic Curve (AUC; Fielding & Bell, 1997)
- True Skill Statistic (TSS; Allouche et al., 2006)
- Continuous Boyce Index (CBI; Boyce et al., 2002; Hirzel et al., 2006)

Models with AUC > 0.7, TSS > 0.6, and CBI > 0.8 were considered "good" to "excellent."

2.4.5 Ensemble Construction

Ensemble models were constructed for each species as weighted averages of individual models, with weights proportional to their evaluation metrics (e.g., TSS or Boyce index). The ensembles were built using the 'flexsdm' R package, following the approach of Hao et al. (2019).

2.5 Spatial Correlation Analysis in SAGA GIS

The predicted habitat suitability maps for *F. hepatica* and *G. truncatula* were exported as GeoTIFF files and imported into SAGA GIS (Conrad et al., 2015). Before analysis, all raster layers were checked for identical extent, cell size, projection, and grid alignment; cells with missing data in either raster were excluded from pairwise calculations. Spatial analyses were conducted in three stages: (i) global spatial correlation, (ii) continuous difference mapping, and (iii) binary overlay and risk-zone classification. SAGA GIS tools used included 'Grid Calculator', 'Grid Difference', 'Reclassify Grid Values', and the correlation/scatterplot modules.

2.5.1 Global Spatial Correlation

Spatial correlation between the two continuous suitability surfaces was assessed using two complementary coefficients:

- Pearson correlation coefficient (r): To quantify the linear relationship between *F. hepatica* and *G. truncatula* suitability values at corresponding grid cells. Pearson's r is sensitive to non-normal distributions and outliers, so it was interpreted alongside the non-parametric coefficient below.
- Spearman rank correlation (ρ): To assess the monotonic relationship between the two surfaces without assuming normality. This is more robust when suitability values are bounded, skewed, or contain many zeros.

Correlation analyses were performed cell-by-cell across the full study area. Because the number of grid cells is large, p -values were interpreted alongside effect sizes; a statistically significant correlation was not considered biologically meaningful unless the coefficient indicated a moderate or strong relationship. A positive correlation indicated spatial concordance between parasite and snail host suitability, whereas a negative or weak correlation indicated spatial mismatch. These global statistics provided a first indication of whether the two species were broadly limited by the same environmental gradients.

However, we additionally applied the toroidal-shift null model (or torus-translation test), a non-parametric method used to evaluate whether the observed raster-raster correlation is genuinely meaningful or just an artifact of space (Segurado et al., 2006; Assunção et al., 2025). $p < 0.05$ means the observed correlation is higher than expected given autocorrelation, providing evidence for a genuine association (niche or biotic).

2.5.2 Difference Maps

To visualise where the two suitability surfaces diverged, we computed a continuous difference map:

$$D = S_F - S_G$$

where S_F is the ensemble suitability for *F. hepatica* and S_G is the ensemble suitability for *G. truncatula*. Positive values of D indicated cells where the parasite was more suitable than the snail host; negative values indicated cells where the snail host was more suitable; values near zero indicated similar suitability for both species. We also computed the absolute difference $|D|$ to identify the magnitude of divergence irrespective of direction. These difference maps helped identify regions where the parasite's potential distribution was not simply a subset of the snail host's distribution.

2.5.3 Overlay Analysis and Risk Zone Classification

Overlay analysis was used to identify areas of joint high suitability, parasite-only suitability, snail-only suitability, and joint low suitability. We distinguished between **continuous joint suitability** and **binary risk zones**, because the two approaches answer different questions and require different interpretations.

Continuous Joint Suitability

Using the continuous suitability rasters (scaled 0–1), we computed a joint suitability index:

$$C = S_F \times S_G$$

where S_F and S_G are the continuous ensemble suitability values for *F. hepatica* and *G. truncatula*, respectively. This product acts as a limiting-factor surface: it is high only where both species are highly suitable, and near zero if either species has low suitability. The highest values of C represent the strongest joint climatic suitability for the parasite and its intermediate host.

Binary Risk Zones

To produce discrete, interpretable risk zones, we first reclassified each continuous suitability map into a binary map using a threshold T :

- 1 = suitable (suitability $\geq T$)
- 0 = unsuitable (suitability $< T$)

The threshold T was selected using the TSS-optimised threshold to generate binary suitability maps (Guisan & Zimmermann, 2000; Farquhar et al., 2026) derived from the ensemble model evaluation. The same thresholding approach was applied to both species to ensure comparability. Let $g1$ be the binary suitability map for *F. hepatica* and $g2$ the binary suitability map for *G. truncatula*. Using the 'Grid Calculator' in SAGA GIS, we computed the following layers representing corresponding risk zones (Table 2):

Table 2. Risk zones for *F. hepatica* and their interpretation

Zone	Formula	Interpretation
Both suitable (joint high-risk)	$g1 \times g2$	Highest-risk zones where both parasite and snail host are predicted suitable
Parasite-only	$g1 \times (1 - g2)$	Areas suitable for <i>F. hepatica</i> but not for <i>G. truncatula</i>
Snail-only	$(1 - g1) \times g2$	Areas suitable for <i>G. truncatula</i> but not for <i>F. hepatica</i>
Neither suitable	$(1 - g1) \times (1 - g2)$	Lowest-risk zones where neither species is predicted suitable

Interpretation of Risk Zones

The resulting zones can be interpreted as follows:

- **Both suitable (joint high-risk):** These are the **highest-risk areas** for fasciolosis transmission. Because *F. hepatica* cannot complete its life cycle without a suitable intermediate snail host, areas where both species are predicted suitable represent the core transmission zones where sustained endemic transmission is most likely. These zones should be prioritised for surveillance and control.
- **Parasite-only:** These are **lower-risk areas** but remain a **surveillance priority**. The parasite is predicted to be climatically suitable, but the snail host is not. This may indicate either true absence of the snail host, precluding local transmission, or incomplete characterisation of snail habitat

requirements (e.g., small water bodies, wet pastures, or irrigation ditches not captured by coarse-resolution climate data). Targeted snail surveys in these areas could clarify transmission potential, so they are worth checking.

- **Snail-only:** These are **potential introduction zones**. The snail host is predicted suitable, but the parasite is not currently predicted. If infected livestock were introduced into these areas, transmission could potentially establish. These zones are relevant for risk assessment of livestock movement and trade.
- **Neither suitable:** These are the **lowest-risk areas** where neither species is predicted suitable. If the analysis is based on binary maps, these cells represent true “neither suitable” zones. If based on continuous maps, they represent relative low suitability rather than absolute absence.

2.5.4 Continuous Joint Unsuitability Index

In addition to the binary risk zones, we computed a continuous joint unsuitability index:

$$U=(1-S_F)\times(1-S_G)$$

where S_F and S_G are the continuous suitability values. Low values of U indicate high joint suitability (i.e., high risk), while high values indicate low joint suitability (i.e., low risk). Importantly, when derived from continuous rasters, this index does **not** represent “absence” or “zero risk” unless the input maps have been binarised first. A maximum value below 1.0 indicates that no grid cell is completely unsuitable for both species simultaneously.

2.5.5 Threshold Sensitivity

The choice of threshold T affects the extent of the exclusive zones (parasite-only, snail-only, neither). We therefore report results using the primary threshold (TSS-optimised threshold) and note that alternative thresholds (for instance, the 10th percentile training presence) may shift the boundaries between risk categories. In the Discussion, we consider the implications of threshold choice for risk mapping and surveillance prioritisation.

2.5.6 Important Caveats

The risk zones represent **climatic suitability** based on the environmental predictors included in our models. They do not account for:

- **Livestock density and management:** Actual transmission risk depends on the presence and density of definitive hosts (cattle, sheep), grazing practices, and veterinary control measures.
- **Local habitat features:** *Galba truncatula* occupies microhabitats (small water bodies, ditches, wet meadows) that may not be fully captured by coarse-resolution climate data.
- **Snail model uncertainty:** Parasite-only zones may represent areas where the snail model underestimates true suitability due to missing microhabitat variables or incomplete occurrence data.
- **Parasite introduction potential:** Snail-only zones may become transmission zones if infected livestock are introduced.

These caveats are discussed further in Section 4 (Discussion).

2.6 SHAP Analysis

SHAP analysis was applied to the ensemble models for both *F. hepatica* and *G. truncatula* to quantify the contribution of each environmental predictor to model predictions. We used the 'shapviz' R package (Mayer, 2026) and the 'shap-values' R package (Casas, 2024) to compute:

- **Global variable importance:** Mean absolute SHAP values, ranking variables by overall influence for each species.

- **Response curves:** Partial dependence-like plots showing how predicted suitability varied with each predictor, based on SHAP values, for each species.
- **Comparative analysis:** Direct comparison of SHAP importance rankings and response curves between *F. hepatica* and *G. truncatula* to identify shared and species-specific drivers.

3. Results

3.1 Model Performance

The ensemble suitability maps for *F. hepatica* and *G. truncatula* are presented in figures 1 & 2. For *F. hepatica* the mean Area Under the Receiver Operating Characteristic Curve (AUC) reached $0.83 \pm 0.12SD$, indicating good discriminatory capacity in distinguishing suitable from unsuitable habitats. The True Skill Statistic (TSS) yielded a value of $0.62 \pm 0.06SD$, indicative of good model performance. Furthermore, the continuous Boyce index (CBI) registered $0.92 \pm 0.03SD$, confirming that the predicted suitability values were consistently and positively correlated with the distribution of presence records in the evaluation dataset. Collectively, these metrics attest to the model's strong predictive reliability and its suitability for current-climate projection.

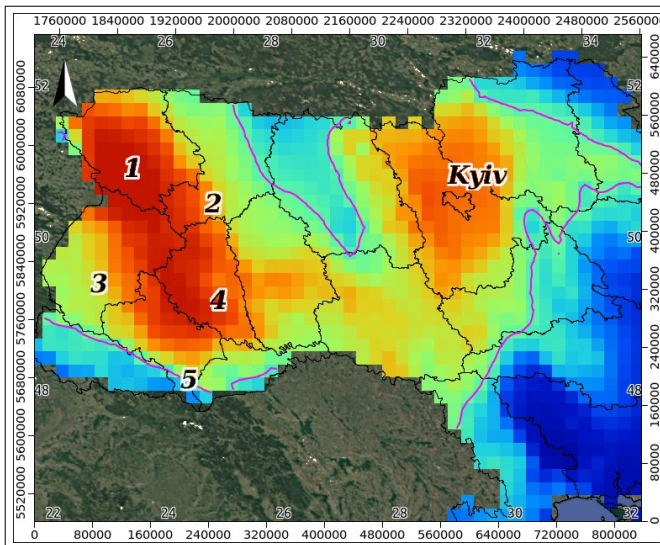


Fig. 1. Ensemble suitability map for *Fasciola hepatica*; colours show potential habitat suitability ranging from high (red) to low (blue); the fuchsia-coloured contour line represents the maximum training sensitivity plus specificity threshold (maxTSS=0.46); 1 - Volynska, 2 - Rivnenska, 3 - Lvivska, 4 - Ternopil'ska, 5 - Ivano-Frankiv'ska oblasts; coordinate reference system: Mollweide projection.

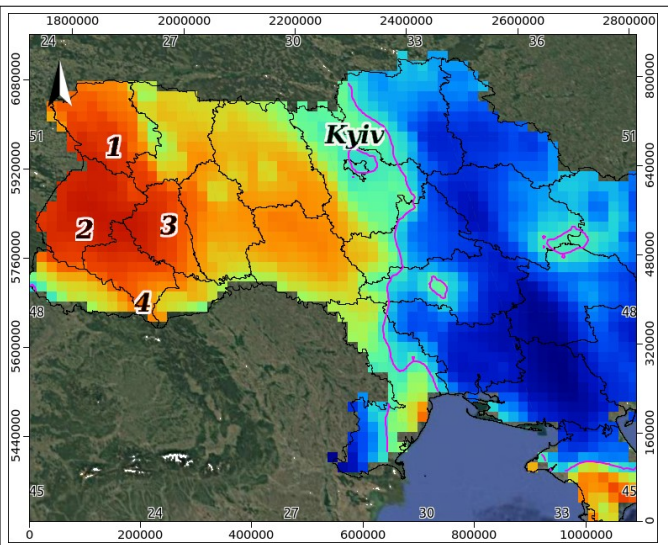


Fig. 2. Ensemble suitability map for *Galba truncatula*; colours show potential habitat suitability ranging from high (red) to low (blue); the fuchsia-coloured contour line represents the maximum training sensitivity plus specificity threshold (maxTSS=0.50); 1 - Volynska, 2 - Lvivska, 3 - Ternopil'ska, 4 - Ivano-Frankiv'ska oblasts; coordinate reference system: Mollweide projection.

For *G. truncatula* performance indices showed in general weaker predictive reliability: AUC= $0.60 \pm 0.01SD$, TSS= $0.35 \pm 0.07SD$, and only CBI reaching an acceptable level of $0.84 \pm 0.16SD$.

3.2 Spatial Distribution and Correlation

Despite these discrepancies, both species showed high predicted suitability in the western and northwestern regions, the Carpathian foothills, and parts of the forest-steppe zone, while the lowest suitability was predicted in the southern steppe and southeastern regions.

3.2.1 Global Spatial Correlation

Spatial correlation analysis revealed a strong positive correlation between the two suitability surfaces:

- Pearson correlation coefficient: $r = 0.792$, $p < 0.05$
- Spearman rank correlation: $\rho = 0.799$, $p < 0.05$

These coefficients indicate a statistically significant and strong spatial concordance between parasite and snail host suitability. The correlation is higher than expected under a hypothesis of complete spatial independence, suggesting that the two species share some, but perhaps not all, climatic drivers.

In terms of the toroidal-shift null model: $r_{obs} = 0.721$, null = -0.015 ± 0.235 , $p = 0.001$, confirming the validity of the strength of the raster-raster correlation.

3.2.2 Difference Maps

The continuous difference map (D) is shown in Figure 3. Positive values (parasite suitability > snail suitability) were most extensive in northern regions, particularly Kyivska, Chernihivska and Sumska oblasts. Negative values (snail suitability > parasite suitability) were concentrated in predominantly in the south and southeast. The absolute difference map (Figure 4) highlighted regions of greatest divergence, particularly Transcarpathia, Odeska oblasts and Crimea.

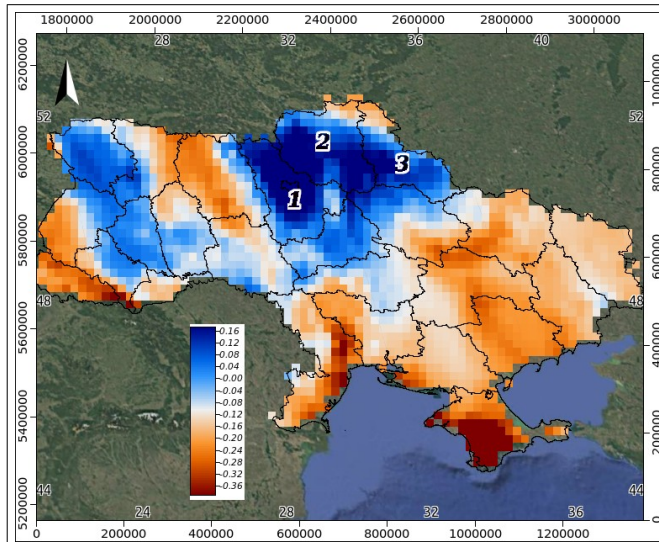


Fig. 3. The continuous difference map (D). 1 - Kyivska, 2 - Chernihivska, 3 - Sumska oblasts. Strong positive values (parasite suitability > snail suitability) are coloured in blue.

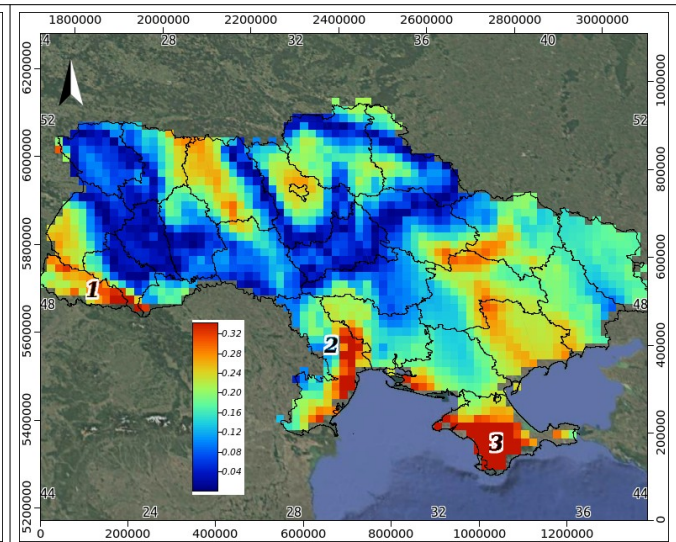


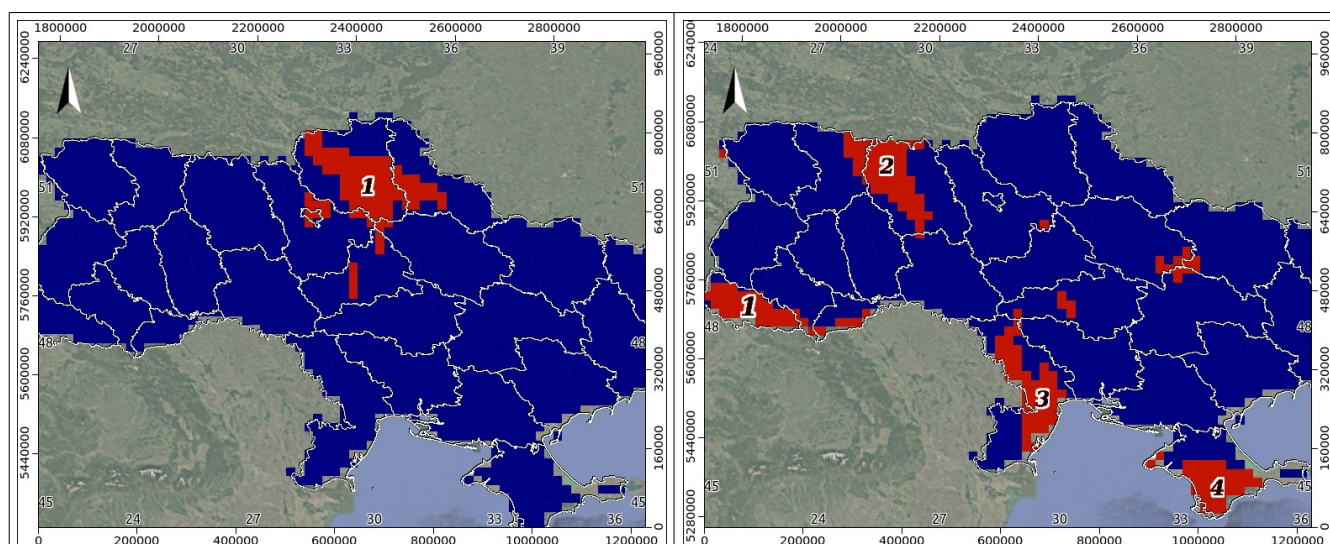
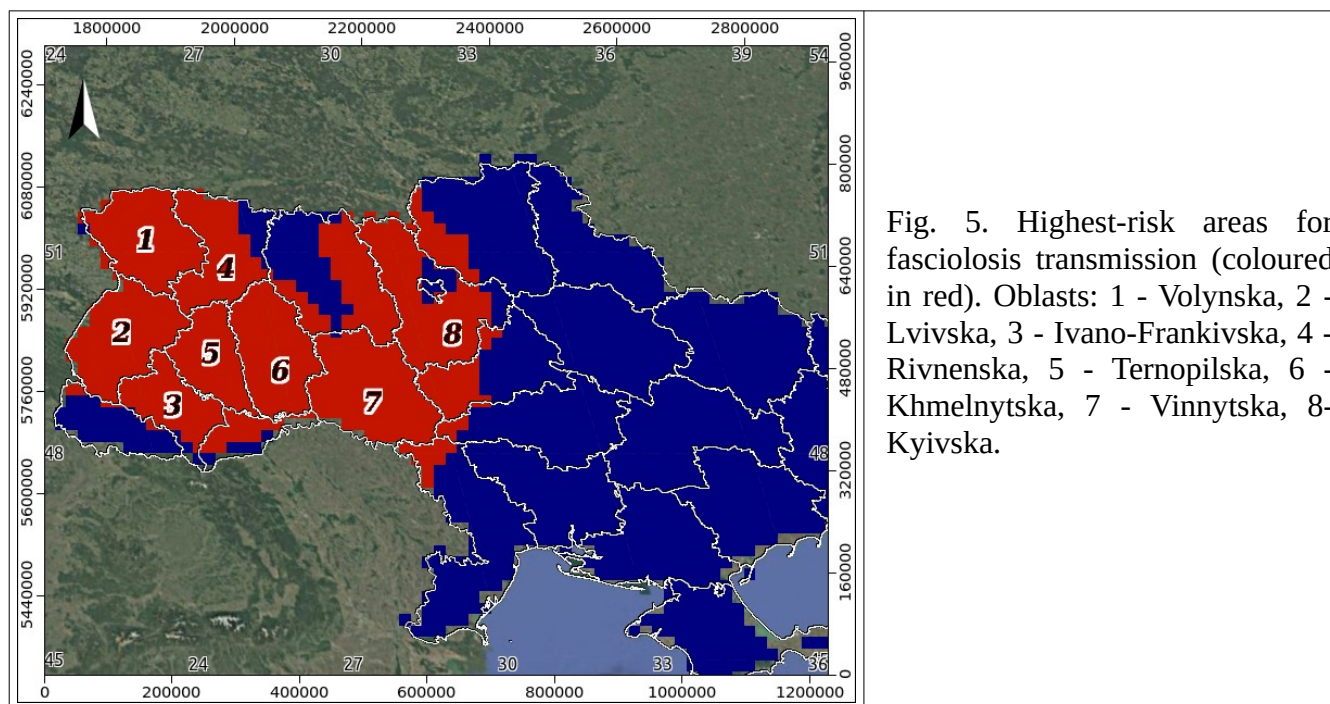
Fig. 4. The absolute difference map ($|D|$); highlights regions of greatest divergence between parasite suitability and snail suitability (coloured in red). 1 - Transcarpathia, 2 - Odeska oblasts, 3 - Crimea.

3.2.3 Risk Zone Classification

Binary overlay analysis using the max-TSS threshold of 0.46 produced four discrete risk zones.

1. **Both suitable (joint high-risk):** areas in Ukraine classified as suitable for both species (Figure 5). These zones are concentrated in the west and north of Ukraine. This represents the core transmission zone where the full *F. hepatica* life cycle is climatically supported.
2. **Parasite-only:** areas suitable for *F. hepatica* but not for *G. truncatula* (Figure 6). These areas are mainly located in the north of the country, especially in Chernihivska Oblast. They represent lower-risk areas but remain surveillance priorities because the snail model may underestimate suitability in localised wet microhabitats.

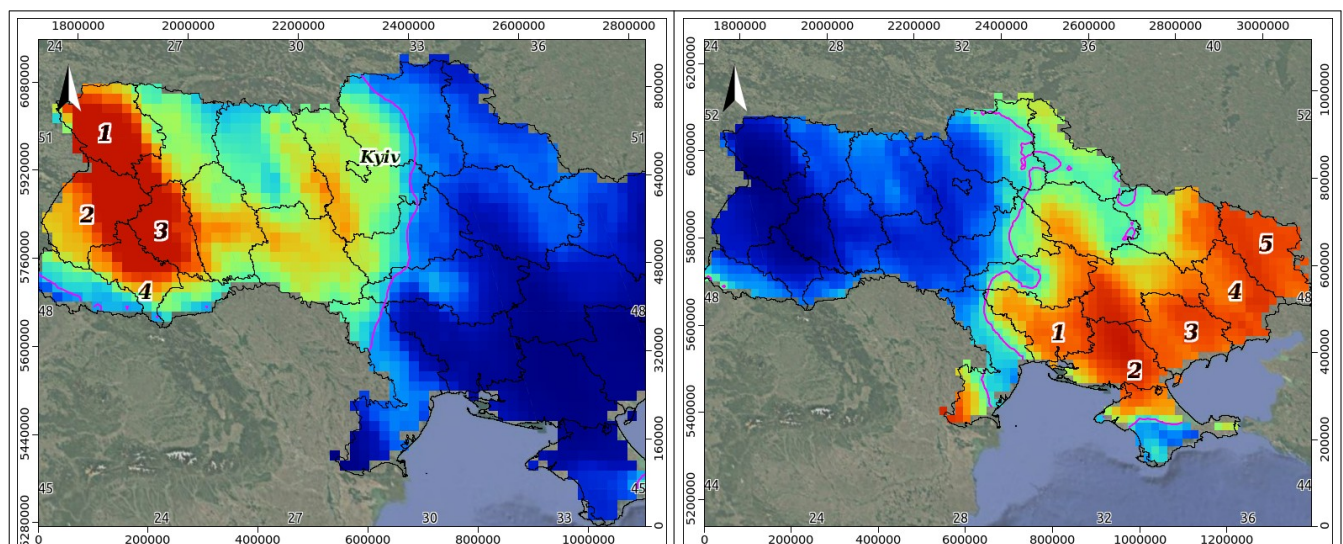
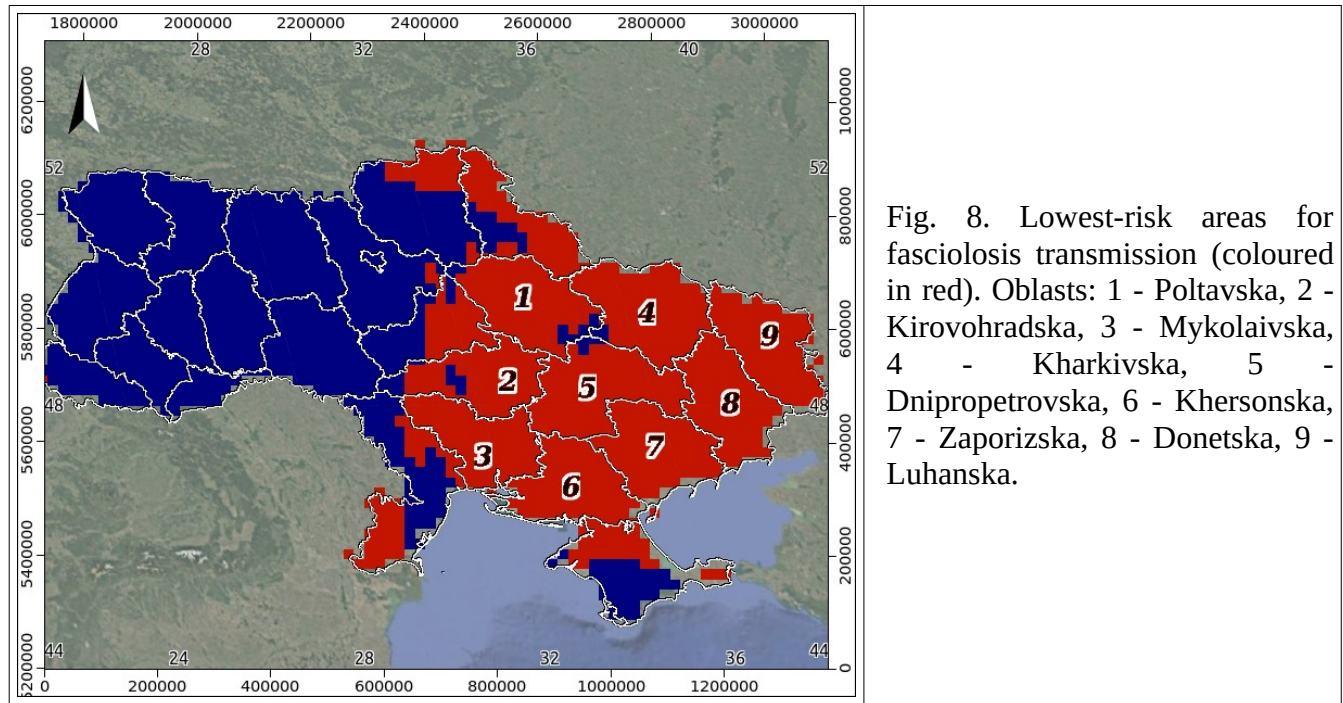
3. **Snail-only:** areas suitable for *G. truncatula* but not for *F. hepatica* (Figure 7). These zones were concentrated in the south and in Zhytomyrska Oblast in the north and represent potential introduction zones should infected livestock be moved into them.
4. **Neither suitable:** areas classified as unsuitable for both species (Figure 8). These areas were found primarily in the east of Ukraine. However, because the continuous joint unsuitability index (see below) never reached 1.0, even these areas retained some degree of climatic suitability for at least one species.



3.2.4 Continuous Joint Suitability and Unsuitability

The continuous joint suitability index (C) ranged from a minimum of 0.02 to a maximum of 0.72, averaged 0.22, and reached highest values in the west of the country (Figure 9).

The continuous joint unsuitability index (U) (Figure 10) ranged from a minimum of 0.02 to a maximum of 0.71, indicating that no grid cell was completely unsuitable for both species simultaneously. This suggests that at least one of the two species retains some climatic suitability throughout Ukraine.



3.2.5 Summary of Risk Zones and Implications

The risk zone classification identified four major clusters of joint high suitability, primarily in the west and north of the country. These areas should be prioritised for veterinary surveillance and snail habitat monitoring. Parasite-only zones, covering primarily northern oblasts, warrant targeted snail surveys to determine whether the snail host is truly absent or merely undetected. Snail-only zones represent vulnerable areas where livestock movement could introduce the parasite. The neither-suitable zone had the lowest climatic risk but was not completely risk-free.

These results are discussed in further detail in Section 4, where we consider their implications for fasciolosis surveillance and control in Ukraine.

3.3 SHAP analysis of ensemble models (Parasite vs. Snail Host)

3.3.1 Variable Importance

Based on the SHAP analysis, the following patterns emerged:

- **Shared drivers:** Temperature seasonality (Bio04), Radiation of wettest quarter (Bio24), Mean Diurnal Range (Bio02). Especially Bio04 notably ranked highly for both the parasite and intermediate host (Figures 11 & 12).

Permutation tests ($N=9,999$) comparing SHAP contributions for Bio04 showed no significant difference in mean signed SHAP values between *F. hepatica* and *G. truncatula* ($p = 0.846$), indicating similar net directional effects. However, mean absolute SHAP values differed significantly ($p = 0.006$), with a larger average contribution in *F. hepatica* (0.111) than in *G. truncatula* (0.087). Thus, temperature seasonality had a stronger overall influence on the parasite model, while its average signed effect was similar between species. The same relates to the other top contributing to the ensemble model variables shared by both species, namely Bio24 and Bio02.

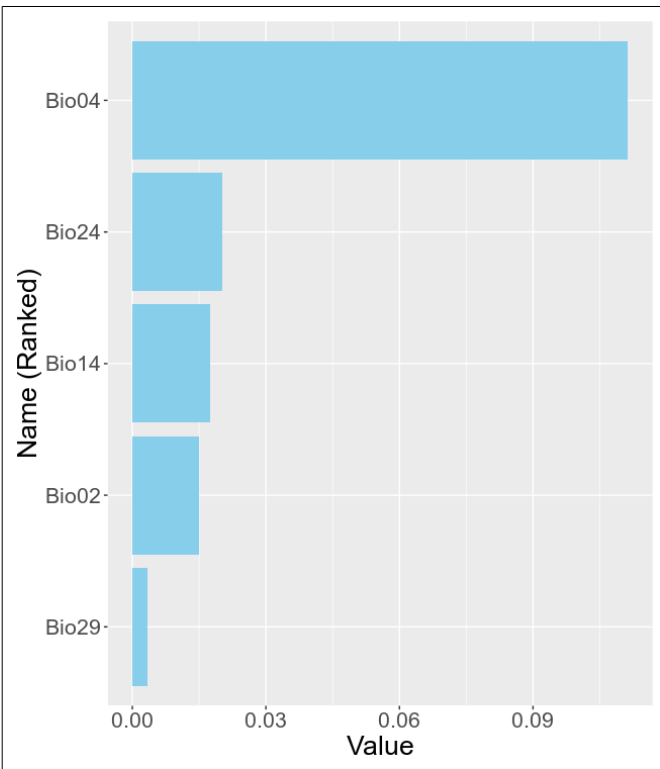


Fig. 11. Summary bar-plots of the average absolute SHAP values of the five top contributing to the ensemble

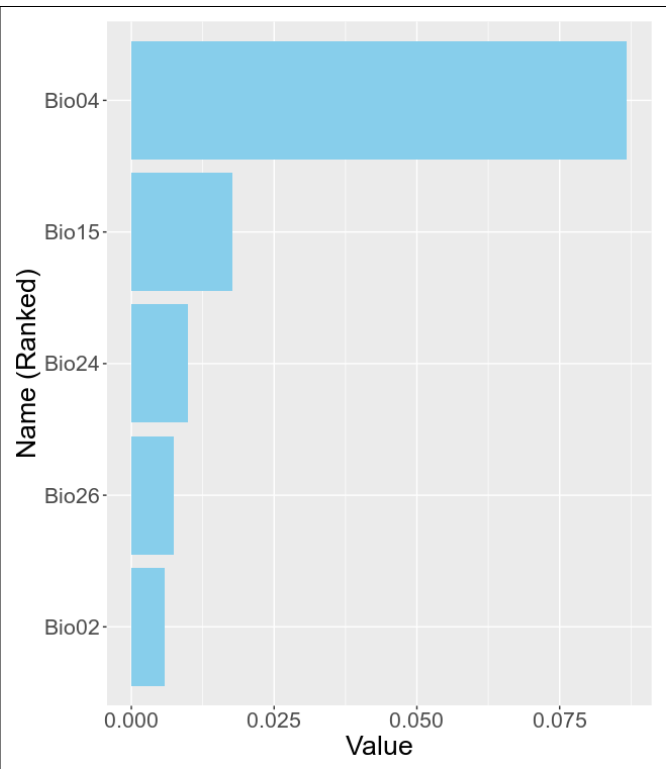


Fig. 12. Summary bar-plots of the average absolute SHAP values of the five top contributing to the ensemble model

model variables. The y-axis represents variables used in the study: Temperature seasonality (Bio04), Radiation of wettest quarter (Bio24), Precipitation of Driest month (Bio14), Mean Diurnal Range (Bio02), Highest weekly moisture index (Bio29). The x-axis represents the corresponding SHAP value.	variables. They-axis represents variables used in the study: Temperature seasonality (Bio04), Precipitation Seasonality (Bio15), Radiation of wettest quarter (Bio24), Radiation of warmest quarter (Bio26), Mean Diurnal Range (Bio02). The x-axis represents the corresponding SHAP value.
--	--

3.3.2 SHAP dependence plots for temperature seasonality

The SHAP dependence plots for temperature seasonality (Bio04) revealed non-linear, unimodal relationships for both species (Figs. 13 and 14). Because Bio04 overwhelmingly dominated the SHAP importance rankings for both species (Figs. 11 and 12), we focus the dependence analysis on this variable alone. For *F. hepatica*, predicted suitability increased with Bio04 up to an intermediate optimum (scaled Bio04 ≈ 0.0283), where the mean SHAP value peaked at approximately +0.152. Above this optimum, the SHAP values declined steeply, crossing zero at scaled Bio04 ≈ 0.0299 and reaching a minimum of approximately -0.200 at the highest Bio04 values.

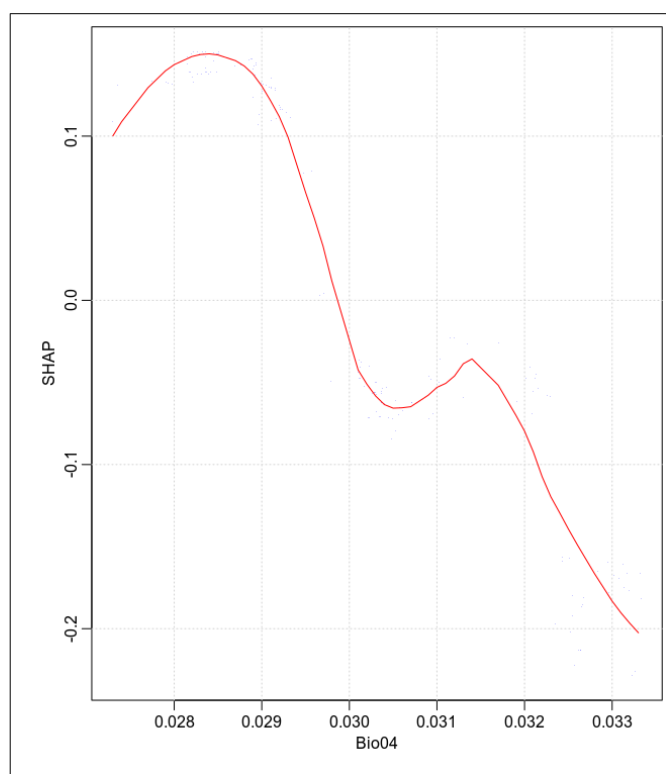


Fig. 13. The SHAP dependence plot for temperature seasonality (Bio04) in the ensemble model for *Fasciola hepatica*. The x-axis shows scaled Bio04 values; the y-axis shows SHAP values (contribution to predicted suitability). Positive SHAP values indicate an increase in predicted suitability, negative values a decrease. The curve peaks at intermediate Bio04 values and declines sharply at higher seasonality.

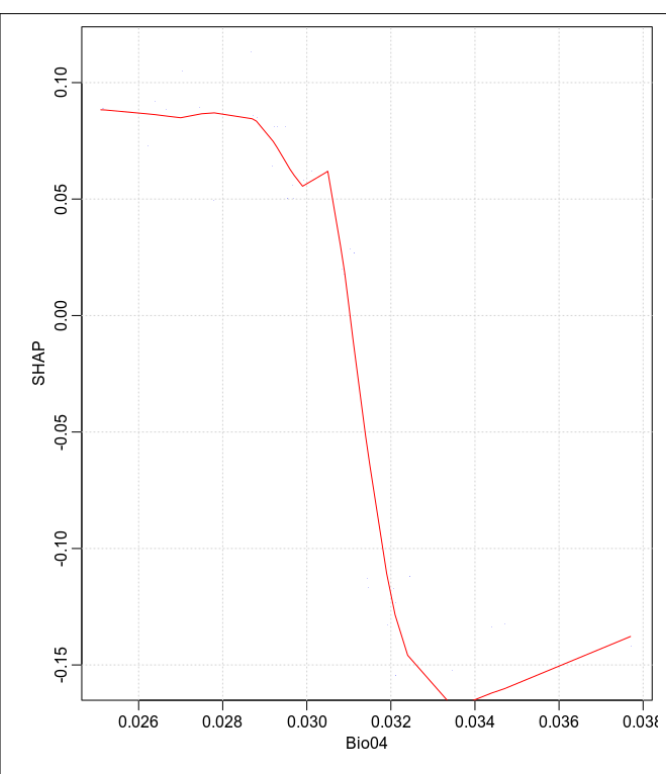


Fig. 14. The SHAP dependence plot for temperature seasonality (Bio04) in the ensemble model for *Galba truncatula*. The x-axis shows scaled Bio04 values; the y-axis shows SHAP values. The positive effect at low seasonality is weaker than for *F. hepatica*.

For *G. truncatula*, the response was also unimodal but differed in magnitude and shape. The positive effect of low Bio04 was weaker, with a maximum SHAP value of only $\approx +0.085$ at scaled Bio04 $\approx$

0.0252–0.0275. The curve crossed zero at scaled Bio04 $\approx$ 0.0308 and declined to a minimum of approximately -0.233 at scaled Bio04 $\approx$ 0.0333. Thus, while both species responded negatively to high temperature seasonality, the snail host showed a stronger negative response at the upper end of the gradient, whereas the parasite showed a stronger positive response at the lower end.

These patterns indicate that temperature seasonality acts as a limiting factor for both *F. hepatica* and *G. truncatula*, with intermediate values being most favourable. The comparatively stronger positive SHAP values for *F. hepatica* at low seasonality suggest that stable temperatures may benefit the parasite's free-living stages more than the snail host, while the more negative SHAP values for *G. truncatula* at high seasonality suggest that the snail is more sensitive to seasonal temperature extremes. However, these interpretations should be treated cautiously because the *G. truncatula* ensemble had weaker discrimination performance (AUC $\approx$ 0.60), and SHAP values describe model behaviour rather than direct ecological causation.

Overall, it can be asserted that temperature seasonality has a stronger but more narrowly optimal influence on the fluke, whereas the snail host shows a weaker but broader tolerance before the effect becomes negative.

4. Discussion

4.1 Model Performance and the Asymmetry Between Parasite and Host Models

The ensemble model for *F. hepatica* performed well across all three evaluation metrics (AUC = 0.83 ± 0.12 SD; TSS = 0.62 ± 0.06 SD; CBI = 0.92 ± 0.03 SD), indicating that climatic variables alone capture a substantial proportion of the parasite's realised niche in Ukraine. This is consistent with the strong environmental constraints imposed on the parasite's free-living stages and its intermediate host (Ollerenshaw & Rowlands, 1959; Malone et al., 1987; Ducheyne et al., 2015). The *G. truncatula* ensemble, by contrast, showed much weaker discrimination (AUC = 0.60 ± 0.01 SD; TSS = 0.35 ± 0.07 SD), although the continuous Boyce index remained acceptable (CBI = 0.84 ± 0.16 SD). This asymmetry has at least three plausible explanations.

First, the snail dataset was small (50 records) and spatially biased towards the west of Ukraine, supplemented by records from neighbouring countries. Small sample sizes reduce the ability of machine learning algorithms to characterise the full environmental niche and inflate the influence of individual records. Second, *G. truncatula* occupies microhabitats — small water bodies, ditches, wet meadows, and riparian zones — that are not resolved by the ~ 18.5 km *CliMond* grid (De Roeck et al., 2014). Coarse-resolution climate layers may therefore miss the local moisture conditions that ultimately determine snail presence. Third, the snail's distribution may be more strongly influenced by non-climatic factors, such as soil type, vegetation structure, grazing management, and water chemistry, which were not included as predictors. The weaker model performance should therefore be interpreted as a signal that climate alone is insufficient to fully predict *G. truncatula* distribution at this scale, rather than as evidence that climate is unimportant.

4.2 Spatial Concordance and the Obligate Host Relationship (H1)

H1 predicted that the spatial distributions of the two species would be significantly correlated because *F. hepatica* cannot complete its life cycle without *G. truncatula*. This hypothesis was supported: the two suitability surfaces were strongly and significantly correlated (Pearson $r = 0.792$; Spearman $\rho = 0.799$; $p < 0.05$; $r_{\text{obs}} = 0.721$, null = -0.015 ± 0.235 , $p = 0.001$). The strength of this correlation is remarkable given the weak performance of the snail model and the coarse resolution of the climate data, and it suggests that the broad-scale climatic gradients that favour the parasite also favour its intermediate host. Both species showed highest suitability in the western and northwestern regions, the Carpathian foothills, and parts of the forest-steppe zone, and lowest suitability in the southern steppe

and southeastern regions. This pattern is consistent with the known epidemiology of fasciolosis in Ukraine, where the highest cattle infection rates have been recorded in the northwestern oblasts (Lytvynenko et al., 2024).

However, the correlation was not perfect, and the difference maps revealed systematic divergence. Positive values of D (parasite suitability > snail suitability) were most extensive in northern Ukraine, particularly in Kyivska, Chernihivska, and Sumska oblasts, whereas negative values (snail suitability > parasite suitability) were concentrated in the south and southeast. The absolute difference map highlighted Transcarpathia, Odeska oblast, and Crimea as regions of greatest divergence. These patterns indicate that the parasite's potential distribution is not simply a subset of the snail's distribution; rather, the two species respond to overlapping but non-identical sets of climatic constraints.

The existence of parasite-only zones (Figure 6), particularly in Chernihivska Oblast, is epidemiologically important. These are areas where the parasite is predicted to be climatically suitable but the snail host is not. Two interpretations are possible. The first is that the snail is genuinely absent, in which case local transmission cannot occur despite the presence of suitable conditions for the parasite's free-living stages. The second is that the snail is present but undetected, either because of insufficient sampling or because the coarse-resolution climate data fail to capture suitable microhabitats. Targeted malacological surveys in these regions would be required to distinguish between these possibilities. Until then, parasite-only zones should be treated as areas of uncertain but non-negligible risk.

Conversely, snail-only zones (Figure 7), concentrated in the south and in Zhytomyrska Oblast, represent potential introduction zones. The snail host is predicted to be suitable, but the parasite is not currently predicted. If infected livestock were moved into these areas, the parasite could potentially establish, provided that the free-living stages can survive the local climate. These zones are therefore relevant for risk assessment of livestock trade and movement, and they merit surveillance attention even though current climatic suitability for the parasite is low.

The continuous joint suitability index (C) reached a maximum of 0.72 and averaged 0.22, with the highest values in the west of the country (Figure 9). The continuous joint unsuitability index (U) reached a maximum of 0.71, indicating that no grid cell was completely unsuitable for both species simultaneously. This suggests that at least one of the two species retains some climatic suitability throughout Ukraine, and that the concept of "zero risk" is not supported by the model. Even in the east, where the binary classification identified "neither suitable" zones, the continuous surfaces indicate low but non-zero suitability for at least one species. This has implications for surveillance: risk is a matter of degree, not of presence or absence, and control efforts should be proportional to the level of joint suitability rather than focused exclusively on high-risk zones.

4.3 Climatic Drivers and the Role of Temperature Seasonality (H2)

H2 predicted that temperature seasonality, mean annual temperature, and annual precipitation would emerge as the most important drivers for both species. This hypothesis was partially supported. Temperature seasonality (Bio04) was overwhelmingly the most important predictor for both *F. hepatica* and *G. truncatula*, consistent with the thermal requirements of the parasite's free-living stages and the snail host (Kendall & McCullough, 1951; Poulin, 2006; Mas-Coma et al., 2009). However, mean annual temperature and annual precipitation were not retained in the final VIF-selected predictor set, and the secondary drivers differed between species: *F. hepatica* showed stronger responses to precipitation of the driest month (Bio14) and highest weekly moisture index (Bio29), whereas *G. truncatula* showed stronger responses to precipitation seasonality (Bio15) and radiation of the warmest quarter (Bio26).

The dominance of Bio04 in both models is ecologically plausible. Temperature seasonality reflects the magnitude of seasonal temperature variation, which influences the duration of the transmission window, the survival of overwintering stages, and the rate of snail development. For *F. hepatica*, the free-living miracidia and cercariae are highly sensitive to temperature extremes, and the parasite's development within the snail host is temperature-dependent (Ollerenshaw & Rowlands, 1959). For *G. truncatula*, seasonal temperature variation affects reproductive output, growth rates, and aestivation behaviour (Kendall & McCullough, 1951). The shared importance of Bio04 therefore reflects a common thermal constraint on both the parasite and its intermediate host.

The permutation tests revealed an important nuance. Mean signed SHAP values for Bio04 did not differ significantly between species ($p = 0.846$), indicating that the net directional effect of temperature seasonality is similar for both. However, mean absolute SHAP values differed significantly ($p = 0.006$), with a larger average contribution in *F. hepatica* (0.111) than in *G. truncatula* (0.087). This means that temperature seasonality acts in the same direction for both species — low seasonality is favourable, high seasonality is unfavourable — but its overall influence is stronger in the parasite model. The same pattern was observed for the other shared predictors, Bio24 and Bio02.

The SHAP dependence plots for Bio04 (Figures 13 and 14) clarified the shape of these responses. Both species showed unimodal relationships, with intermediate values of temperature seasonality being most favourable. For *F. hepatica*, the positive effect was stronger but more narrowly optimal: SHAP values peaked at approximately +0.152 at scaled Bio04 ≈ 0.0283 , crossed zero at ≈ 0.0299 , and declined to ≈ -0.200 at the highest values. For *G. truncatula*, the positive effect was weaker but broader: SHAP values remained below +0.085 across a wider range (scaled Bio04 ≈ 0.0252 – 0.0308) before declining to ≈ -0.233 at the highest values. These patterns suggest that the parasite benefits more from stable temperatures but is more severely penalised when seasonality increases, whereas the snail tolerates a wider range of seasonal variation but is more strongly affected at the extreme end of the gradient.

This asymmetry has implications for understanding the relative roles of the snail host and the parasite's free-living stages in limiting transmission. If the snail were the primary constraint on parasite distribution, one would expect the snail model to show stronger and more consistent responses to climatic variables than the parasite model. The opposite was observed: the parasite model performed better and showed stronger SHAP responses to the dominant climatic driver. This suggests that the parasite's free-living stages impose additional climatic constraints beyond those imposed by the snail host, and that the two species do not simply track the same environmental gradient.

4.4 Species-Specific Drivers

The divergence in secondary drivers between the two species is ecologically informative. *Fasciola hepatica* showed stronger relative importance of precipitation of the driest month (Bio14) and highest weekly moisture index (Bio29), both of which are direct measures of moisture availability. This is consistent with the parasite's dependence on moisture for the survival and dispersal of miracidia and cercariae, and with the well-documented role of rainfall in driving fasciolosis outbreaks (Ollerenshaw & Rowlands, 1959; Malone et al., 1987). The parasite model appears to capture the moisture constraints on the free-living stages more effectively than the snail model captures the moisture constraints on the snail itself.

Galba truncatula, by contrast, showed stronger relative importance of precipitation seasonality (Bio15) and radiation of the warmest quarter (Bio26). Precipitation seasonality reflects the variability of rainfall throughout the year, and high seasonality may create dry periods that limit snail survival, particularly in shallow water bodies and wet meadows that are prone to desiccation. Radiation of the warmest quarter reflects summer heat load, which may influence water temperature, evaporation rates, and snail aestivation. These results are consistent with the snail's known sensitivity to desiccation and its preference for persistently wet, shaded microhabitats (Mas-Coma et al., 2009; De Roeck et al., 2014).

The fact that the snail model emphasises different moisture-related variables than the parasite model may reflect the different spatial and temporal scales at which moisture affects the two organisms. The parasite's free-living stages are short-lived and highly dependent on immediate moisture conditions, whereas the snail is a longer-lived organism whose distribution is influenced by the persistence of suitable microhabitats over time. Coarse-resolution climate data may capture the former more effectively than the latter, which may partly explain the weaker performance of the snail model.

4.5 Implications for Surveillance and Control

The risk maps produced in this study have several practical implications for fasciolosis surveillance and control in Ukraine.

Joint high-risk zones (Figure 5), concentrated in the west and north of the country, represent the core transmission zones where the full life cycle is climatically supported. These areas should be prioritised for veterinary surveillance, including regular faecal egg counts, liver inspection at slaughter, and snail habitat monitoring. The oblasts identified — Volynska, Lvivska, Ivano-Frankivska, Rivnenska, Ternopilska, Khmelnytska, Vinnytska, and Kyivska — are consistent with the known distribution of fasciolosis in Ukraine (Lytvynenko et al., 2024) and with the climatic gradients described in Section 2.1. Control measures in these areas should include strategic anthelmintic treatment, grazing management to avoid wet pastures during high-risk periods, and drainage or fencing of known snail habitats.

Parasite-only zones (Figure 6), particularly in Chernihivska Oblast, warrant targeted snail surveys to determine whether *G. truncatula* is truly absent or merely undetected. If the snail is present, these areas may represent unrecognised transmission zones; if it is absent, they represent areas where transmission cannot currently occur despite suitable conditions for the parasite. In either case, they are worth monitoring, particularly if livestock movement or climate change could alter the current situation.

Snail-only zones (Figure 7), concentrated in the south and in Zhytomyrska Oblast, represent potential introduction zones. These areas are currently unsuitable for the parasite but suitable for the snail host. If infected livestock were introduced, transmission could potentially establish, particularly if the parasite's free-living stages can survive the local climate. These zones are relevant for risk assessment of livestock trade and movement, and they should be included in surveillance programmes even though current risk is low.

Neither suitable zones (Figure 8), concentrated in the east, have the lowest climatic risk but are not completely risk-free. The continuous joint unsuitability index never reached 1.0, indicating that at least one species retains some climatic suitability throughout the study area. This means that even in low-risk zones, localised microhabitats — irrigation ditches, small ponds, wet depressions — could support transmission if the snail host is present. Surveillance in these areas should focus on identifying such microhabitats rather than on broad-scale climatic suitability.

4.6 Limitations

Several limitations should be acknowledged.

First, the *G. truncatula* model performed poorly ($AUC \approx 0.60$), which limits the reliability of its predictions and of the comparative SHAP analysis. The snail dataset was relatively small (50 records) and spatially biased, and the coarse-resolution climate data may not capture the microhabitat conditions that determine snail presence. Future work should prioritise expanded snail sampling, particularly in undersampled regions, and the integration of finer-resolution environmental data, such as soil moisture, land cover, and topography.

Second, the study relied on occurrence records from GBIF, which are subject to sampling bias. Spatial thinning and spatial block partitioning were used to reduce this bias, but they cannot eliminate it

entirely. The parasite records, in particular, are derived from domestic animal surveillance and may reflect veterinary sampling effort rather than true occurrence.

Third, the climate data from *CliMond* are at ~18.5 km resolution, which is coarse relative to the microhabitat scale at which *G. truncatula* operates. This mismatch may partly explain the weaker performance of the snail model and the divergence between the two suitability surfaces.

Fourth, the models used only climatic predictors and did not include livestock density, grazing management, soil type, vegetation, or water chemistry. These factors are known to influence fasciolosis transmission and snail distribution, and their omission may lead to under- or over-estimation of risk in some areas.

Fifth, the SHAP analysis describes the behaviour of the fitted models, not direct ecological causation. SHAP values are influenced by the choice of algorithms, hyperparameters, and pseudo-absence sampling, and they should be interpreted as model-derived contributions rather than as causal effects.

Sixth, the risk zone classification depends on the choice of threshold. The TSS-optimised threshold was used for consistency, but alternative thresholds would shift the boundaries between risk categories. The continuous indices (C and U) provide a threshold-independent complement to the binary classification, but they are not directly comparable to it.

Finally, the study did not account for future climate change. The models were calibrated on current climate conditions, and the risk maps should be interpreted as representing current climatic suitability. Future work should project these models onto climate change scenarios to assess how the distribution of *F. hepatica* and *G. truncatula* may shift in the coming decades.

4.7 Future Directions

Several avenues for future research emerge from this study. First, expanded snail sampling across Ukraine, particularly in the south and east, would improve the snail model and allow more robust comparative analyses. Second, the integration of finer-resolution environmental data — soil moisture, land cover, topography, and hydrology — would likely better capture the microhabitat conditions that determine *G. truncatula* distribution. Third, the inclusion of livestock density and management variables would allow the models to distinguish between climatic suitability and actual transmission risk. Fourth, the application of the ensemble framework to future climate scenarios would provide insights into how fasciolosis risk may change in Ukraine over the coming decades. Fifth, the comparative SHAP approach developed here could be extended to other host–parasite systems where the intermediate host is thought to constrain parasite distribution.

5. Conclusions

This study developed the first national-scale, spatially explicit ensemble species distribution models for *F. hepatica* and its intermediate snail host *G. truncatula* in Ukraine, and applied SHAP analysis to compare their climatic drivers. The main conclusions are as follows.

The *F. hepatica* ensemble model performed well (AUC = 0.83; TSS = 0.62; CBI = 0.92), demonstrating that climatic variables capture a substantial proportion of the parasite's realised niche in Ukraine. The *G. truncatula* model performed more weakly (AUC = 0.60; TSS = 0.35; CBI = 0.84), indicating that climate alone is insufficient to predict the snail's distribution at the resolution used here.

The two suitability surfaces were strongly and significantly correlated (Pearson $r = 0.792$; Spearman $\rho = 0.799$; $r_{\text{obs}} = 0.721$, null = -0.015 ± 0.235 , $p = 0.001$), supporting the hypothesis that the parasite and its obligate intermediate host share broad-scale climatic drivers. However, the correlation was not perfect, and systematic divergence was observed in northern, southern, and southeastern regions.

Temperature seasonality (Bio04) was the dominant climatic driver for both species, consistent with the thermal requirements of the parasite's free-living stages and the snail host. Its net directional effect was similar between species ($p = 0.846$), but its overall magnitude of influence was significantly

stronger for *F. hepatica* ($p = 0.006$), indicating that temperature seasonality is a stronger constraint on the parasite model than on the snail model.

The two species showed divergent secondary drivers: *F. hepatica* responded more strongly to moisture-limitation variables (Bio14, Bio29), whereas *G. truncatula* responded more strongly to precipitation seasonality (Bio15) and summer radiation (Bio26). These differences suggest that the parasite and its host are not limited by identical climatic factors and that the parasite's free-living stages impose additional environmental constraints.

The SHAP dependence plots revealed unimodal responses to Bio04 for both species, with intermediate values being most favourable. The parasite showed a stronger but narrower optimum, whereas the snail showed a weaker but broader tolerance before the effect became negative. This asymmetry suggests that stable temperatures benefit the parasite more than the snail, while the snail is more sensitive to extreme seasonality.

Four risk zones were identified: joint high-risk areas in the west and north, parasite-only areas in the north, snail-only areas in the south and in Zhytomyrska Oblast, and neither-suitable areas in the east. The continuous joint suitability index reached a maximum of 0.72 and averaged 0.22, and the continuous joint unsuitability index never reached 1.0, indicating that at least one species retains some climatic suitability throughout Ukraine.

The combination of ensemble SDMs with SHAP analysis provides a transferable framework for understanding the environmental drivers of zoonotic parasites and their intermediate hosts. This approach can be applied to other host–parasite systems and can support risk-based surveillance and control programmes.

In summary, this study demonstrates that fasciolosis risk in Ukraine is concentrated in the western and northern regions, where both the parasite and its snail host are climatically suitable. The parasite's distribution is not simply a subset of the snail's distribution, and temperature seasonality emerges as the dominant shared climatic constraint. These findings provide a basis for targeted surveillance and control efforts and highlight the value of explainable AI for understanding the ecology of vector-borne and host-dependent zoonoses.

Funding: This research received no external funding.

Data Availability Statement: The data supporting this study are available within the article.

Declaration: The author acknowledges the use of artificial intelligence (AI)–assisted tools in the preparation of this manuscript. Specifically, AI-based language models were used exclusively for language editing and stylistic improvement, including grammar correction, sentence restructuring, and enhancement of readability. This assistance was necessary because English is not the author's native language. AI tools were not used for: (i) the conception or design of the study; (ii) the collection, processing, or analysis of data; (iii) the generation or interpretation of results; (iv) the drawing of scientific conclusions; or (v) the creation of figures, tables, or references. All scientific content, analyses, interpretations, and conclusions presented in this manuscript are the sole work and responsibility of the author. The author has reviewed and verified all AI-assisted text and takes full responsibility for the accuracy, integrity, and originality of the final manuscript.

References

- Allouche, O., Tsoar, A., Kadmon, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, 43: 1223–1232. <https://doi.org/10.1111/j.1365-2664.2006.01214.x>
- Araújo, M.B., New M. (2007). Ensemble forecasting of species distributions. *Trends Ecol Evol.*, Jan; 22(1):42–7. doi: 10.1016/j.tree.2006.09.010
- Assefa, A.Y., Anteneh, S.T., Maru, M. et al. (2026). Prevalence and spatial distribution modelling of *Fasciola* species infection in sheep in northwest Ethiopia. *Vet J.*, Oct; 319:106828. doi: 10.1016/j.tvjl.2026.106828

- Assunção, R., Butler, K., Krause, E., et al. (2025). Spatially thinned bootstrap and random toroidal shift methods for correct estimation of P-values of maps similarity. *International Journal of Geographical Information Science*, 39(7), 1597–1621. <https://doi.org/10.1080/13658816.2025.2461603>
- Boria, R.A., Olson, L.E., Goodman, S.M., et al. (2014). Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecological Modelling*, 275, 73–77. <https://doi.org/10.1016/j.ecolmodel.2013.12.012>
- Boyce, M.S., Vernier, P.R., Nielsen, S.E., et al. (2002). Evaluating resource selection functions. *Ecol. Model.* 157,281–300. DOI: 10.1016/S0304-3800(02)00200-4
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Casas Pablo. (2024). A gentle introduction to SHAP values in R. Data Science Heroes blog, 2019-03-18, <https://blog.datascienceheroes.com/author/pabloc/>
- Conrad, O., Bechtel, B., Bock, M., et al. (2015). System for automated geoscientific analyses (SAGA) v. 2.1. 4. *Geoscientific model development*, 8(7), 1991–2007. doi:10.5194/gmd-8-1991-2015
- De Roeck, E., Van Coillie, F., De Wulf, R., et al. (2014). Fine-scale mapping of vector habitats using very high resolution satellite imagery: A liver fluke case-study. *Geospatial Health* 8, S671–S683. <https://doi.org/10.4081/GH.2014.296>
- Dormann, C. F., Elith, J., Bacher, S., et al. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
- Ducheyne, E., Charlier, J., Vercruysse, J., et al. (2015). Modelling the spatial distribution of *Fasciola hepatica* in dairy cattle in Europe. *Geospat Health*, Mar 26; 9(2):261–70. doi: 10.4081/gh.2015.348
- Elith, J., Leathwick, J.R., Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77: 802–813. <https://doi.org/10.1111/j.1365-2656.2008.01390.x>
- Farquhar, J.E., Chapple, D.G., Parkin, T. (2026). Beyond the mountains: incorporating newly documented low-elevation records reshapes predicted habitat suitability for the threatened mountain skink *Liopholis montana*. *Pacific Conservation Biology* 32, PC26015. <https://doi.org/10.1071/PC26015>
- Fielding, A.H., Bell, J.F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24(1):38–49. doi:10.1017/S0376892997000088
- GBIF.org [19 September 2026] GBIF Occurrence Download <https://doi.org/10.15468/dl.m5b2er>
- Guisan, A., Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135(2–3), 147–186. [https://doi.org/10.1016/s0304-3800\(00\)00354-9](https://doi.org/10.1016/s0304-3800(00)00354-9)
- Hao, T., Elith, J., Guillera-Aroita, G., et al. (2019). A review of evidence about use and performance of species distribution modelling ensembles like BIOMOD. *Divers Distrib.*, 25: 839–852. <https://doi.org/10.1111/ddi.12892>
- Hirzel, A. H., Le Lay, G., Helfer, V., et al. (2006). Evaluating the ability of habitat suitability models to predict species presences. *Ecological modelling*, 199(2), 142–152. DOI: 10.1016/j.ecolmodel.2006.05.017
- Honcharov, S. L., Soroka, N. M., Derkach, I. M., et al. (2024). Current status of the prevalence of fascioliasis in cattle in Ukraine. *Animal Health*, 26(114), 132–137. <https://doi.org/10.32718/nvlvet11419>

- Kendall, S.B., McCullough, F.S. (1951). The emergence of the cercariae of *Fasciola hepatica* from the snail *Limnaea truncatula*. *Journal of Helminthology*, 25, 77–92. doi:10.1017/S0022149X0001899X
- Kriticos, D.J., Webber, B.L., Leriche, A., et al. (2012). *CliMond*: global high-resolution historical and future scenario climate surfaces for bioclimatic modelling. *Methods in Ecology and Evolution*, 3: 53-64. <https://doi.org/10.1111/j.2041-210X.2011.00134.x>
- Lan, Z., Zhang, X.H., Xing, J.L., et al. (2024). Global prevalence of liver disease in human and domestic animals caused by *Fasciola*: A systematic review and meta-analysis. *J. Glob. Health*, Sep 19;14:04223. doi: 10.7189/jogh.14.04223
- Lytvynenko, O. P., Miroshnichenko, O. I., Pishchanskyi, et al. (2024). Issues regarding the prevalence of fascioliasis in cattle in Ukraine (2021–2023). *One Health Journal*, 2(III), 32–38. <https://doi.org/10.31073/onehealthjournal2024-III-04>
- Lundberg, S. M., B. Nair, M. S. Vavilala, et al. 2018. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nature Biomedical Engineering*, 2(10): 749–760. DOI:10.1038/s41551-018-0304-0
- Malone, J.B., Williams, T.E., Muller, R.A., et al. (1987). *Fasciolosis* in cattle in Louisiana, development of a system to predict disease risk climate, using the Thornthwaite water budget. *Am. J. Vet. Res.*, 48, 1167-1170. <https://pubmed.ncbi.nlm.nih.gov/3631704/>
- Marynych, A.M., Pashchenko, V.M. & Shishchenko P.G. (1985). The nature of the Ukrainian SSR. Landscapes and physical-geographical zoning (in Russian). Kyiv: Naukova Dumka
- Marynych, O. M., Shyshchenko, P. H. *Physical Geography of Ukraine: Textbook*. Kyiv: Znannia, 2005. 128 p.
- Mas-Coma, S., Valero, M.A., Bargues, M.D. (2009a). Climate change effects on trematodiasis, with emphasis on zoonotic fascioliasis and schistosomiasis. *Vet. Parasitol.*, 163, 264-280. <https://doi.org/10.1016/j.vetpar.2009.03.024>
- Mas-Coma, S., Valero, M.A., Bargues, M.D. (2009b). Chapter 2. *Fasciola*, lymnaeids and human fascioliasis, with a global overview on disease transmission, epidemiology, evolutionary genetics, molecular epidemiology and control. *Adv. Parasitol.*; 69:41-146. doi: 10.1016/S0065-308X(09)69002-3
- Mas-Coma, S., Valero, M.A., Bargues, M.D. (2022). Human and Animal Fascioliasis: Origins and Worldwide Evolving Scenario. *Clin Microbiol Rev* 35:e00088-19. <https://doi.org/10.1128/cmr.00088-19>
- Mas-Coma, S. (2005). Epidemiology of fascioliasis in human endemic areas. *J. Helminthol.*, Sep; 79(3):207-16. doi: 10.1079/joh2005296
- Mayer, M. (2026). shapviz: SHAP Visualizations. R package version 0.10.4, <https://github.com/modeloriented/shapviz>
- Nikraftar, Z., Parizi, E., Saber, M., et al. (2025). An Interpretable Machine Learning Framework for Unraveling the Dynamics of Surface Soil Moisture Drivers. *Remote Sensing*, 17(14),2505.<https://doi.org/10.3390/rs17142505>
- Ollerenshaw, C. B., Rowlands, W. T. (1959). A method of forecasting the incidence of fascioliasis in Anglesey. *Veterinary Record*, Vol. 71, No. 29, 591-598.

- Phillips, S. J., R. P. Anderson, R. E. Schapire. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190 (3–4): 231–259. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>
- Poulin, R. (2006). Global warming and temperature-mediated increases in cercarial emergence in trematode parasites. *Parasitology*, 132(1):143-151. doi:10.1017/S0031182005008693
- Qin, H., Gao, X., Wang, H., et al. (2016). Relative importance of meteorological and geographical factors in the distribution of *Fasciola hepatica* infestation in farmed sheep in Qinghai province, China. *Parasite*, 23:59. doi: 10.1051/parasite/2016070
- Radosavljevic, A., Anderson, R. P. (2014). Making better Maxent models of species distributions: complexity, overfitting and evaluation. *Journal of Biogeography*, 41(4), 629-643. <https://doi.org/10.1111/jbi.12227>
- Rapsch, C., Dahinden, T., Heinzmann, D., et al. (2008). An interactive map to assess the potential spread of *Lymnaea truncatula* and the free-living stages of *Fasciola hepatica* in Switzerland. *Vet. Parasitol.*, 154:242-9. doi: 10.1016/j.vetpar.2008.03.030
- Roberts, D. R., Bahn, V., Ciuti, S., et al. (2017). Cross validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8), 913-929. <https://doi.org/10.1111/ecog.02881>
- Segurado, P., Araújo, M.B. Kunin, W.E. (2006), Consequences of spatial autocorrelation for niche-based models. *Journal of Applied Ecology*, 43: 433-444. <https://doi.org/10.1111/j.1365-2664.2006.01162.x>
- Selemetas, N., Ducheyne, E., Phelan, P., et al. (2015). Spatial analysis and risk mapping of *Fasciola hepatica* infection in dairy herds in Ireland. *Geospatial Health*, 9(2), 281-291. <https://doi.org/10.4081/gh.2015.350>
- Skorobogatov V, Vasyliuk O (2023). Records of parasites of domestic animals in Ukraine (according to the State Service of Ukraine for Food Safety and Consumer Protection). Version 1.2. Ukrainian Nature Conservation Group (NGO). Occurrence dataset <https://doi.org/10.15468/k5djvr> accessed via GBIF.org on 2026-03-14.
- Valencia-López, N., Malone, J.B., Carmona, C.G., et al. (2012). Climate-based risk models for *Fasciola hepatica* in Colombia. *Geospat Health*, Sep; 6(3):S67-85. doi: 10.4081/gh.2012.125
- Velazco, S.J.E., Rose, M.B., Andrade, A.F.A. et al. (2022). flexsdm: An R package for supporting a comprehensive and flexible species distribution modelling workflow. *Methods in Ecology and Evolution*, 13, 8. P. 1661–1669. DOI: <https://doi.org/10.1111/2041-210X.13874>
- Zastavnyi, F.D. (1994). Geography of Ukraine: in 2 books: a study guide for students of geography faculties and economics specialties at higher education institutions. Lviv: Svit,. 470 pp. <https://history.org.ua/LiberUA/5-7773-0043-8/5-7773-0043-8.pdf>
