A mechanistic hydro-epidemiological model of liver fluke risk

Published in:
JOURNAL OF THE ROYAL SOCIETY INTERFACE

Document Version:
Publisher's PDF, also known as Version of record

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

THE ROYAL SOCIETY
PUBLISHING

A mechanistic hydro-epidemiological model of liver fluke risk

Ludovica Beltrame[1], Toby Dunne[1], Hannah Rose Vineer[2,3,4], Josephine G. Walker[2,4,5], Eric R. Morgan[4,6], Peter Vickerman[5], Catherine M. McCann[7], Diana J. L. Williams[8] and Thorsten Wagener[1,4]

[1]Department of Civil Engineering, [2]School of Biological Sciences, [3]Bristol Veterinary School, [4]Cabot Institute, and [5]Bristol Medical School, University of Bristol, Bristol, UK
[6]School of Biological Sciences, Queen's University Belfast, Belfast, UK
[7]Epidemiology Research Unit, Scotland's Rural College, Inverness, UK
[8]Institute of Infection and Global Health, University of Liverpool, Liverpool, UK

LB, 0000-0002-8078-3871; HRV, 0000-0002-1488-0315; JGW, 0000-0002-9732-5738; ERM, 0000-0002-5999-7728; CMM, 0000-0002-5388-6031; DJLW, 0000-0001-8186-7236; TW, 0000-0003-3881-5849

The majority of existing models for predicting disease risk in response to climate change are empirical. These models exploit correlations between historical data, rather than explicitly describing relationships between cause and response variables. Therefore, they are unsuitable for capturing impacts beyond historically observed variability and have limited ability to guide interventions. In this study, we integrate environmental and epidemiological processes into a new mechanistic model, taking the widespread parasitic disease of fasciolosis as an example. The model simulates environmental suitability for disease transmission at a daily time step and 25 m resolution, explicitly linking the parasite life cycle to key weather–water–environment conditions. Using epidemiological data, we show that the model can reproduce observed infection levels in time and space for two case studies in the UK. To overcome data limitations, we propose a calibration approach combining Monte Carlo sampling and expert opinion, which allows constraint of the model in a process-based way, including a quantification of uncertainty. The simulated disease dynamics agree with information from the literature, and comparison with a widely used empirical risk index shows that the new model provides better insight into the time–space patterns of infection, which will be valuable for decision support.

1. Introduction

The transmission of several highly pathogenic infectious diseases is closely linked to weather and environmental conditions [1]. Waterborne diseases, like cholera, are directly affected by hydro-meteorological factors such as rainfall, through transport and dissemination of the pathogens, and water temperature, through their development and survival rates. Diseases involving a vector or intermediate host as part of their life cycle, such as schistosomiasis, are also indirectly controlled by characteristics of the water environment and land surface, through their influence on the vector or host [2,3].

Our environment is changing at unprecedented rates due to climate change and direct human activities [4,5], with implications for the behaviour, seasonality and distribution of many diseases and their carriers [6,7]. Evidence of climate and environment-driven changes in the phenology of pathogens and incidence of diseases already exists. The increase in frequency and intensity of extreme weather events is altering the occurrence of floods and droughts, changing the concentration of infectious agents such as *Vibrio cholerae* in the water environment and human exposure to infection [3]. Similarly, changes

in the prevalence of schistosomiasis have been observed due to the expansion of the snail intermediate host habitat, following the construction of dams and implementation of irrigation schemes to meet demands for food and energy from increasing numbers of people [8].

As climate change accelerates and other human-caused disturbances increase, it is urgent to assess impacts on disease transmission, to guide interventions that can reduce and/or mitigate risk [9]. To this end, we need to: (i) understand the mechanisms by which the environment affects epidemiological processes, addressing the system as a whole, (ii) represent these processes with models that are explicit in space and time, to more reliably simulate conditions beyond historically observed variability, and (iii) test these models in new ways, since simply reproducing past observations may no longer be sufficient to justify their use for decision support [1,3,7,10–12].

However, most current models that predict changes to disease patterns in response to climate change are empirical [7,13,14]. This means they do not explicitly represent mechanisms, but are based on statistical correlations between historical data, thus becoming unreliable when extrapolated to novel conditions, e.g. into different regions or future climates [15]. Moreover, empirical models do not allow for what-if analyses, i.e. they cannot be used to test the effect of interventions on disease incidence, which would be valuable for guiding decision-making [10,16].

In this paper, we incorporate knowledge of environmental and epidemiological processes into a new integrated mechanistic model, using fasciolosis as an example. This is a globally distributed parasitic disease of livestock and zoonosis, whose most widespread agent is *Fasciola hepatica*, the common liver fluke [17]. Clinical signs of disease in animals include weight loss, anaemia and sudden death, while sub-clinical infections result in lowered productivity and are estimated to cost the livestock industry $3 billion per year, globally [18,19]. Risk of infection with liver fluke is strongly influenced by weather and environmental conditions, especially temperature and soil moisture, as the parasite has an indirect life cycle involving an intermediate host (in the case of *F. hepatica*, the amphibious mud snail *Galba truncatula*) and free-living stages, which grow and develop in the environment [20–22].

Addressing fasciolosis is urgent for a number of reasons. First, resistance to available antiparasitic drugs is on the rise worldwide, making disease control challenging [23]. Second, increases in disease prevalence, expansions into new areas and shifts in its seasonality have been observed in recent years and attributed to altered temperature and rainfall patterns, raising concerns about the effects of climate change in the future [23,24]. Finally, fasciolosis is emerging as a major disease in humans, with at least 2.4 million people infected around the world, and human treatment relying on the same veterinary drug to which resistance is increasing [25]. Climate-based fluke risk models have been developed since the 1950s [20,26–28]. The Ollerenshaw index is the best-known example, which is still actively used to predict disease severity in Europe [20,29,30]. However, these models are empirical in nature and therefore of little use for assessing risk under changing conditions. On the other hand, previous attempts to model fasciolosis mechanistically, in connection with climate, neglect the role of soil moisture dynamics in driving infection and do not account for the spatial aspect of the disease (e.g. [19,31]).

Therefore, in this study, we introduce a new mechanistic coupled hydro-epidemiological model for liver fluke, which explicitly represents the parasite life cycle in time and space, linked with key environmental conditions. We then parametrize the model for two case studies in the UK and assess whether it can replicate temporal and spatial variability of observed infection levels. To overcome limitations of available epidemiological data, we propose a calibration approach that combines observations and expert knowledge. Finally, we further evaluate the model by comparing it with the widely used empirical Ollerenshaw index.

2. The Hydro-Epidemiological model for Liver Fluke

The Hydro-Epidemiological model for Liver Fluke (HELF) quantitatively captures the mechanisms underlying transmission of fasciolosis, describing the causal relationships between hydro-meteorological factors and biological processes, instead of relying on correlation. To this end, HELF integrates TOPMODEL [32,33], an existing hydrological model which we use to simulate soil moisture dynamics, and a novel epidemiological model, which represents the parasite life cycle. TOPMODEL is chosen because its underlying assumptions are physically realistic for humid–temperate catchments, such as UK catchments, where the dominant mechanism of run-off generation is soil saturation [32]. The epidemiological model is developed based on current understanding of the life cycle of *F. hepatica* and its dependence upon soil moisture and air temperature [20–22].

2.1. Hydrological component

TOPMODEL is a catchment-scale rainfall-run-off model, which was developed for hydrological predictions and has been extensively used for different water resources applications (e.g. references in [34]). The model uses air temperature, rainfall and Digital Elevation Model (DEM) data to estimate, at each time step, spatially distributed soil moisture over the catchment (calculated as a saturation deficit), as well as streamflow at the catchment outlet. The model we use is based on the version explained in [33] and has seven parameters (table 1).

In TOPMODEL, hydrological processes are represented using a sequence of conceptual stores for which the model estimates and tracks water balances. An interception store, representing vegetation cover, must be filled by rainfall before infiltration into the soil can occur. When water infiltrates into the soil, it first enters the root zone, from which it evaporates as a function of potential evapotranspiration, maximum capacity of the store, and its actual water content. Water that is not evaporated or retained by the soil percolates to the saturated zone (i.e. the groundwater), which contributes to the channel network through subsurface flow.

To simulate the spatial distribution of soil water content over the catchment, this water balance accounting routine, which is lumped at the catchment scale, is integrated with spatially distributed topographic information derived from DEM data. The effect of topography is captured, for each grid cell in the catchment, through calculation of a topographic index (TI):

$$TI = \ln\left(\frac{a}{\tan(\beta)}\right),$$

Table 1. Hydrological model parameters and initial ranges.

| no.	parameter	description	unit	min.	max.
1	LnTe	log of transmissivity of soil when saturated to the surface	mm^2 d^{-1}	−5	10
2	m	soil parameter controlling decline of transmissivity as saturation deficit increases	mm	1	200
3	Srz_init	initial saturation deficit in root zone	mm	0	300
4	Srz_max	maximum allowable saturation deficit in root zone	mm	5	300
5	td	time delay from unsaturated to saturated zone	day mm^{-1}	0	0.9
6	a	shape parameter for gamma distribution used for routing	—	0.01	5
7	b	scale parameter for gamma distribution used for routing	—	0.01	5

where a is the upslope contributing area and $\tan(\beta)$ the local slope. TI is used as a measure of the likelihood that a grid cell becomes saturated by downslope accumulation: high values occur over flat areas in valleys, which tend to saturate first, whereas low values are associated with areas at the top of hills, where there is little upslope area and slopes are steep (figure 1). The model assumes that all points with the same index value will respond similarly, hydrologically. For computational efficiency, the distribution of TI values is then discretized into classes, so that computations are performed for each class instead of for each grid cell.

Therefore, a saturation deficit for each TI class is calculated as a function of the catchment average saturation deficit, updated at each time step by water balance calculation, and the spatial distribution of the TIs. Rainfall that falls on saturated areas (i.e. where deficit is less than or equal to zero) cannot infiltrate into the soil and generates saturation-excess overland flow. Finally, total streamflow is calculated as the integrated subsurface flow and saturation-excess overland flow, and a gamma distribution is used to model the time delay in discharge generation at the catchment outlet, due to water moving through the river network of the catchment.

2.2. Epidemiological component

The epidemiological component of HELF represents the stages of the liver fluke life cycle that live on pasture: eggs, miracidia, snail infections and metacercariae (figure 2). Development and survival of these, as well as the presence of mud snails, require particular temperature conditions and wet soil. Therefore, the model takes as input variables temperature and soil moisture, as well as an egg scenario (i.e. number of embryonic eggs we assume are deposited on each TI class at each time step by infected animals), to calculate the abundance of individuals in each life-cycle stage.

Once passed out on pasture in the faeces of infected animals, eggs (E) develop at a temperature-dependent rate, and hatch into miracidia when both temperature and soil moisture conditions are suitable [35]. Miracidia (Mi) are short lived: either they find a snail host or die within 24 h [35,36]. Therefore, progression from miracidium to the next stage is calculated as the probability of finding a snail. This is assumed to depend on soil moisture levels and temperature, as *G. truncatula* snails are only found in poorly drained areas and are known to hibernate with cold weather and aestivate during hot dry periods [35]. Snail infections (SI) also develop in the model as a function of both temperature and soil moisture, as development within the snail may be
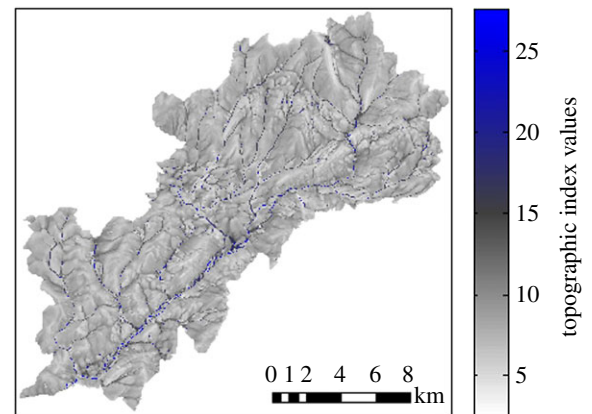
Figure 1. Spatial pattern of topographic index values for the River Tawe Catchment (UK).

halted due to hibernation and aestivation [21], until parasites emerge from snails in the form of cercariae. Once attached to grass as metacercariae (Me), these survive on pasture and retain infectivity based on temperature, with moderate weather being most favourable [35].

Each stage, except for miracidia that only have a lifespan of 1 day, is represented as a pool of developing cohorts of individuals to capture maturation progress in a more realistic way. Individuals in different cohorts are exposed to different environmental conditions, and therefore will develop at different times [35,36]. We account for this by using two state variables for each cohort within each stage: number of individuals and 'age' of the cohort. The rationale is that each cohort has a certain age, which increases with the number of days that have suitable environmental conditions, until the cohort eventually matures to the next life-cycle stage. Output from a stage is then the sum of cohorts per unit area which mature to the next one.

At each time step, development and/or survival rates for a stage are calculated based on the value of the relevant environmental conditions for that stage at that time step, and on the stage-specific requirements for development/survival, which are defined through model parameters (table 2). The technique employed to build the functions to calculate these rates has previously been used for modelling both liver fluke and other parasites (e.g. [19,31]). For temperature-dependent rates, we use information in the literature from laboratory experiments or controlled micro-environment studies that examine the time to development or death at a range of constant temperatures. First, rates are calculated for each constant temperature from the reported e.g. time to
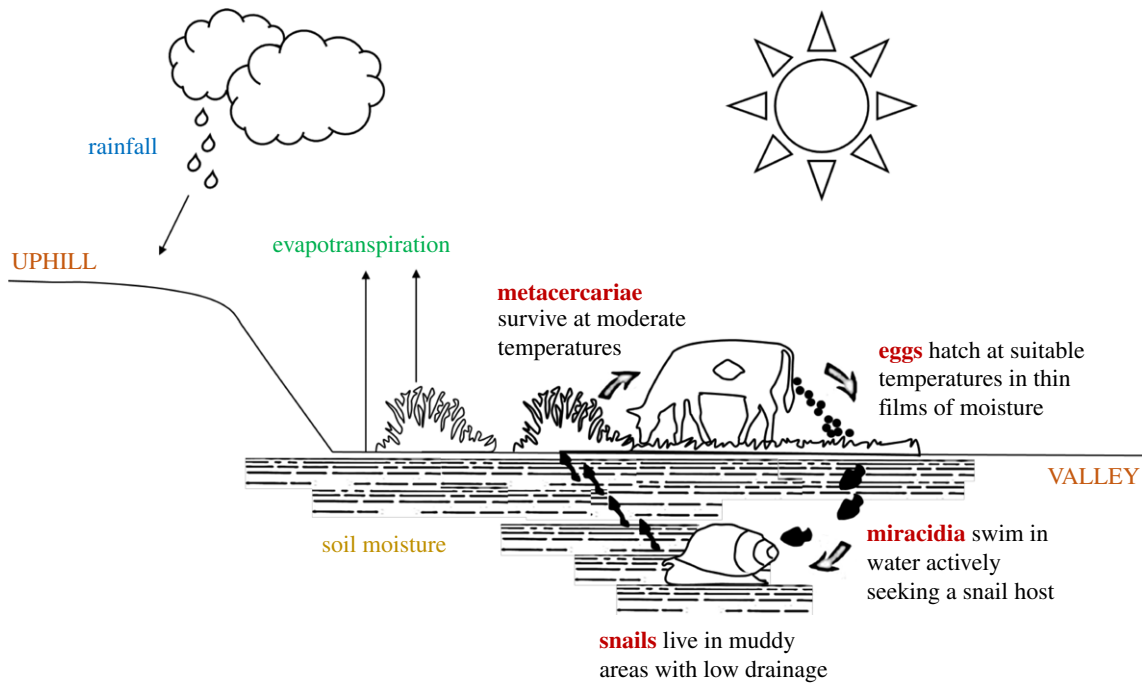
Figure 2. Representation of the liver fluke life cycle, with an amphibious mud snail serving as the intermediate host. (Online version in colour.)

development (i.e. rate = 1/time to development); then piece-wise linear models are fitted to these rates, yielding a regression equation which can be used to estimate the daily rates based on a time series of observed temperature. For soil moisture, we adopt the same approach, assuming that development is fastest when the soil is fully saturated (i.e. when deficit = 0) and that there is no development above a certain maximum deficit [20,35]. For stages with both temperature and soil moisture requirements, we allow for development to progress as a function of both (figure 3).

2.3. Coupled model

The coupled hydro-epidemiological model runs at a daily time step and has a total of 29 parameters. For each day, HELF calculates the catchment average saturation deficit based on rainfall and temperature, and derives the saturation deficit for each of 25 TI classes, as a function of this and the TI value for the class. Then, for each class and life-cycle stage, the model calculates the relevant development and/or survival rates, based on environmental conditions. The age of each cohort is updated based on the development rate, and, given an egg scenario, the model finally computes the number of individuals in the stage as a function of the number from the previous time step, plus the sum of the cohorts developed from the previous stage, minus those that die (figure 4). Therefore, the model outputs are the abundances of developed eggs, snails located and infected by miracidia, developed snail infections, and infective metacercariae surviving on pasture, which represents the environmental suitability for disease transmission to grazing livestock. These variables, calculated for each TI class, are then mapped back onto each grid cell in the catchment.

Regarding the egg scenario, the current assumption is that 100 embryonic eggs are introduced on each TI class daily, over the whole simulation period. This means we are considering a scenario of continuous livestock grazing and no disease management over the catchment. However, this assumption can be easily changed. The fact that the egg

scenario is a model input gives the model-user the possibility to estimate how the environmental suitability for disease transmission translates into risk of infection, based on local farm management factors such as grazing season length or disease control strategy.

3. Study sites and data

We test HELF in two UK catchments, located in South Wales and north-west Midlands (England), respectively. The datasets employed include both hydro-meteorological and epidemiological data.

3.1. The Tawe and Severn Catchments

The River Tawe flows approximately 50 km south-westwards from its source in the Brecon Beacons to the Bristol Channel at Swansea. The catchment is about 240 km² in size, with elevation ranging from about 10 to 800 m.a.s.l., and most of the area characterized by a relatively impermeable bedrock. The River Severn rises in mid Wales and flows through Shropshire, Worcestershire and Gloucestershire, before also discharging into the Bristol Channel. The catchment, gauged at Upton-on-Severn, is about 6850 km² in area, with elevation range and geological characteristics similar to the Tawe [37]. Both catchments have grassland as the dominant land cover (figure 5), which is extensively used for livestock farming, and are located in known fluke endemic areas [38]. Moreover, these areas are predicted to become increasingly warmer and wetter on average [39], which suggests they will become even more favourable for liver fluke transmission in the future.

3.2. Hydro-meteorological and epidemiological data

The hydro-meteorological dataset includes daily observations of rainfall, temperature and discharge. Gridded time series of rainfall and temperature are obtained from CEH-GEAR and the UK Met Office, respectively. For both case studies, to run HELF, we take the average over the grid cells within the

Table 2. Epidemiological model parameters and initial ranges.

no.	parameter	description	unit	min.	max.	references
1	S1MinTemp	min temp for egg development	°C	5	15	[20,21,35]
2	S1PeakTemp	optimal temp for egg development	°C	15	27	[35,36]
3	S1PeakRate	egg development rate at optimal temp	day^{-1}	0.025	0.5	[35,36]
4	S1MortRate	egg mortality rate	day^{-1}	0	0.0693	[35]
5	S1MaxTemp	max temp for egg development	°C	27	40	[35]
6	S1TrigSatDef	max saturation deficit for egg hatching	mm	1	100	[21,35]
7	S1TrigMinTemp	min temp for egg hatching	°C	5	15	[21]
8	S2SatDef	max saturation deficit for miracidia finding snails	mm	1	100	[20,35,36]
9	S2SatSlope	slope defining optimal saturation deficit for miracidia finding snails	—	0.01	0.5	
10	S2MinTemp	min temp for miracidia finding snails	°C	5	15	[35]
11	S2PeakTemp	optimal temp for miracidia finding snails	°C	15	27	[35]
12	S2MaxTemp	max temp for miracidia finding snails	°C	27	40	[21]
13	S3SatDef	max saturation deficit for snail infections development	mm	1	100	[20,35,36]
14	S3SatSlope	slope defining optimal saturation deficit for snail infections development	—	0.01	0.5	
15	S3MinTemp	min temp for snail infections development	°C	5	15	[20,21,30]
16	S3PeakRate	snail infections development rate at optimal temp	day^{-1}	0.0204	0.0357	[35]
17	S3MortRate	snail infections mortality rate	day^{-1}	0	0.0693	—
18	S4MinTemp	min temp for metacercariae survival	°C	−20	10	([35] and therein)
19	S4PeakTemp	optimal temp for metacercariae survival	°C	10	15	([35] and therein)
20	S4MaxTemp	max temp for metacercariae survival	°C	15	40	([35] and therein)
21	S4PeakRate	metacercariae survival rate at optimal temp	day^{-1}	0.0027	0.0333	([35] and therein)
22	S4MinRate	metacercariae survival rate at min/max temp	day^{-1}	0.0333	0.5	([35] and therein)

catchment. For the Tawe we use these time series for a 12-year period (1999–2010), whereas for the Severn we use 2 years of data (2013–2014), in line with the available epidemiological data periods. Observed discharge, at Ynystanglws for the Tawe and Upton-on-Severn for the Severn, is derived from [37]. DEM data for both catchments are obtained from Next-Map with spatial resolution of 5 m, then aggregated to 25 m.

The epidemiological dataset consists of a time series from the Veterinary Investigation Diagnostic Analysis (VIDA) database for the Tawe and a spatial dataset based on faecal egg counts (FECs) for the Severn. The VIDA database, compiled from reports from the UK Government's Animal and Plant Health Agency regional laboratories, provides diagnoses of fasciolosis made from ill or dead animals. The time series we use is the monthly number of sheep diagnosed with acute fasciolosis from the postcode district areas within the Tawe Catchment over 1999–2010. These data are believed to reflect well the temporal dynamics of within-year infection levels but may not always reflect the magnitude of infection in the field, as the rate of submission of animals to the laboratories is potentially influenced by multiple factors [40]. In our series, no cases are reported for 2001 and values over the following years are low, which may have been affected by the 2001 UK foot-and-mouth outbreak, which killed over 10 million animals, affecting submissions to the veterinary laboratories. On the other hand, the spatial dataset for the

Severn Catchment consists of 174 cattle herds, from farms within a 60 × 75 km area in Shropshire, that have been classified into infected and non-infected based on FECs collected over October 2014–April 2015. Unlike VIDA, these are active surveillance data, and thus more likely to reflect true levels of infection. However, rather than a continuous/quantitative measure of the magnitude of infection, this dataset only provides a binary classification into positive–negative farms, at one moment in time and at a limited number of points within the catchment.

4. Model calibration and testing

HELF comprises parameters related to the environment and parameters related to the phenology of the parasite (tables 1 and 2). Usually, more or less well-defined ranges of values can be found in the literature for these, rather than point estimates, partly because of their associated natural variability and partly due to uncertainty and poor understanding. Different parameter sets, selected from these ranges, often provide equally good representations of system behaviour, with implications in terms of predictive uncertainty and limitations for the applicability of the model [15,34]. This type of parameter uncertainty can be reduced through a calibration or constraining process.
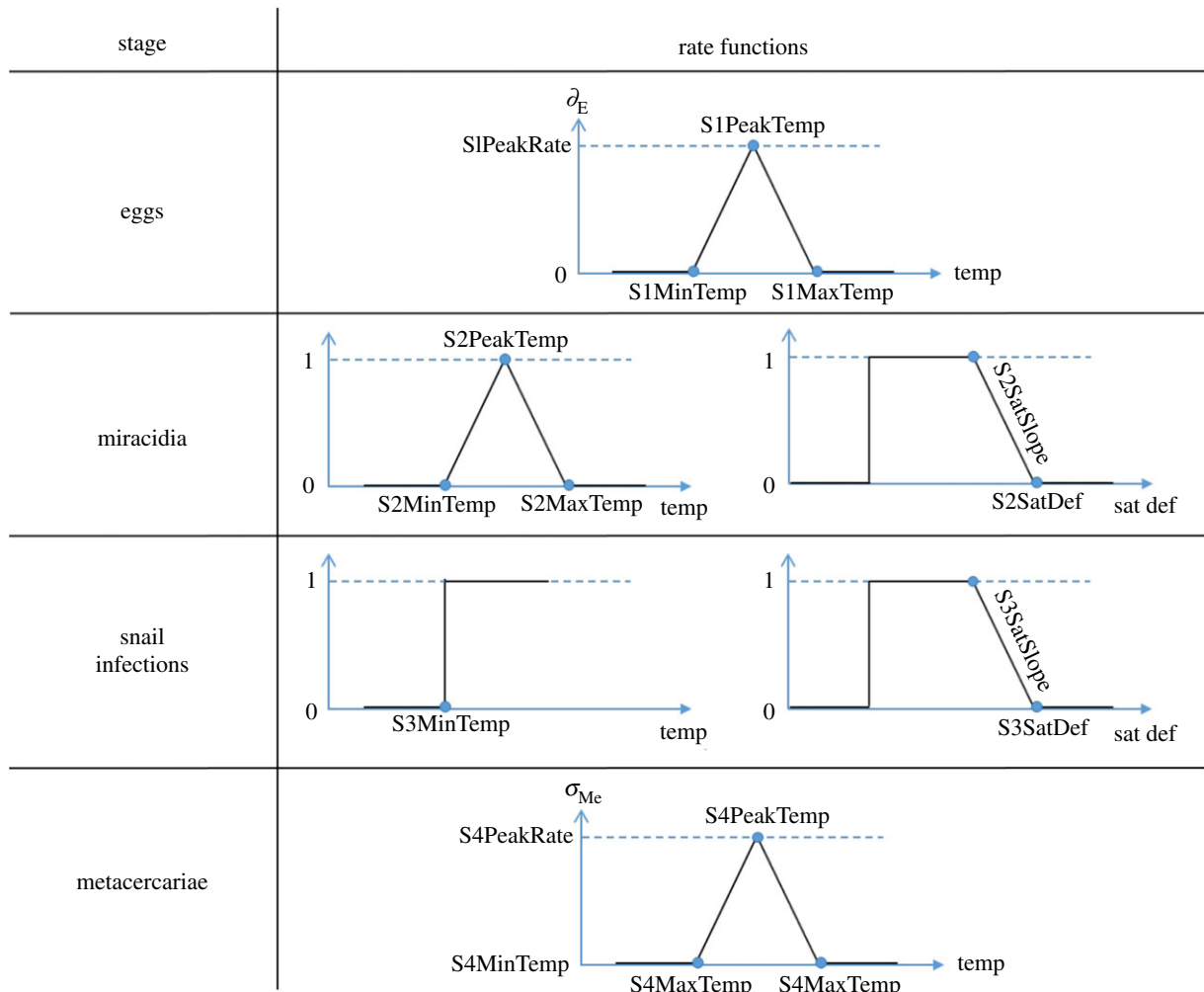
Figure 3. Functions used in HELF to calculate temperature and soil moisture-dependent development and survival rates. (Online version in colour.)
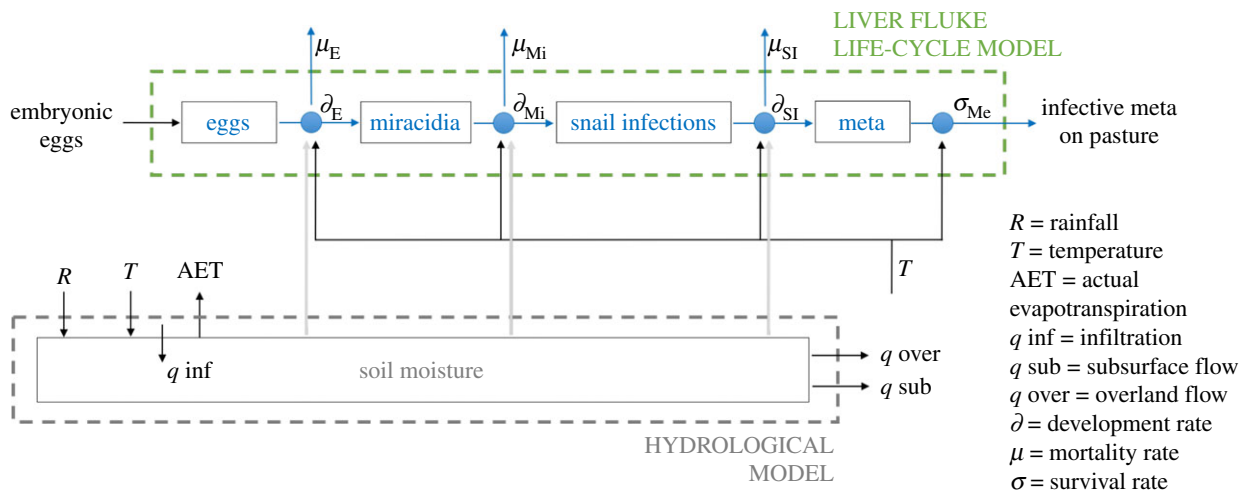

Figure 4. Simplified flow diagram of HELF, which integrates a hydrological and a liver fluke life-cycle component, to simulate the abundance of infective metacercariae (meta) on pasture. (Online version in colour.)

Usually models are calibrated and validated using historic records, assuming that the data available reflect the underlying system, and that conditions in the period considered are similar to those under which the model will be used. However, this may not be sufficient if data are disinformative in some respects and/or if the purpose of the model is to simulate conditions that are significantly different to those previously observed [11,41].

Our calibration strategy involves multiple datasets and methods. On one hand, we have high quality continuous data for both the meteorology and the hydrology. Therefore, we calibrate and validate the hydrological component of HELF by adopting a standard split-sampling approach [41]. On the other hand, given the epidemiological data limitations mentioned in §3.2, our approach for constraining the epidemiological model component
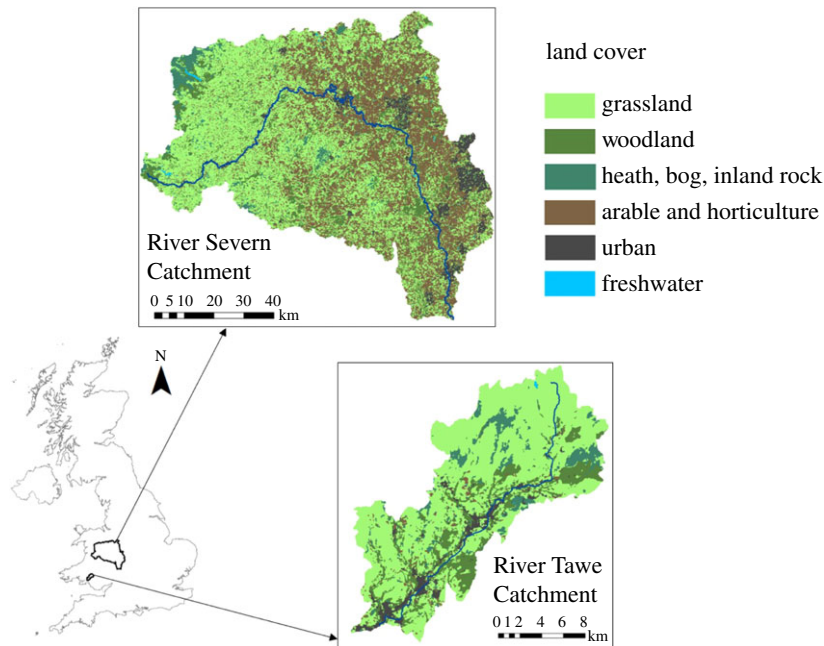
Figure 5. Location and land cover map of the Tawe and Severn Catchments [37].

not only uses past observations, but also expert-driven rules.

4.1. Calibration and testing of the hydrological component

To estimate TOPMODEL parameter values and evaluate its prediction capabilities, we perform a split-sample test using streamflow observations (years 2000–2006 for calibration and 2007–2010 for validation, with 1999 as warm-up period) [41]. The shuffled complex evolution (SCE-UA) optimization method is employed to find the parameter set which maximizes the coefficient of determination (R^2) between simulations and observations on our catchments [42]. The algorithm samples an initial population of parameter sets from *a priori* defined ranges (table 1) and then evolves this population of sets to find the best performing one with respect to R^2.

4.2. Calibration and testing of the epidemiological component

Using the best performing parametrization obtained for TOPMODEL (and therefore for now neglecting the uncertainty in representing the hydrology), first, we fit the fluke component of HELF to the two epidemiological datasets and assess whether we can reproduce the observed patterns of infection, ignoring the data limitations discussed. Second, under the assumption that these data may be disinformative, and given that we ultimately want to use HELF to simulate fluke risk under changing conditions, we propose an alternative calibration approach based on Monte Carlo sampling and expert knowledge. Finally, we evaluate the model by comparing results to information from previous studies and to the commonly used Ollerenshaw index.

4.2.1. Single-objective approach using epidemiological data

To estimate parameters of the epidemiological model for the Tawe Catchment, we fit HELF to the VIDA time series by using SCE-UA to maximize the Pearson coefficient of correlation (r) between simulated abundance of infective metacercariae and observed number of sheep infections. As the VIDA dataset only provides a single time series for the Tawe, we aggregate the simulated abundance of metacercariae over the catchment by taking the average across TI classes. Moreover, to account for the delay between the variable we simulate and the observations, a lag parameter is included in the optimization process, which is allowed to vary between 0 (no delay) and +5 months [18].

Similarly, to estimate parameters for the Severn Catchment, we fit HELF to the FEC-based spatial dataset. First, we divide the area over which we have observations into sub-areas with a minimum of 15 data points each. Second, we use SCE-UA to find the parameter set which maximizes r between the simulated percentage of grid cells at risk of infection and the observed percentage of herds infected, over each sub-area. To this end, for each parameter set, we aggregate the simulated abundance of metacercariae over months July–December 2014, assuming that pasture contamination over this period will be responsible for the observed infection levels [38]. Then, we classify the simulated abundance of metacercariae in each grid cell into two classes (no-risk and risk) by setting a threshold based on the overall observed percentage of infection.

4.2.2. Monte Carlo sampling-based approach using expert opinion

Given the limitations of the epidemiological datasets, we believe that simply fitting these may not be sufficient to guarantee reliability of our new model. Moreover, if HELF is to be used to assess future disease risk, its credibility should be assessed via more in-depth evaluation of the consistency with the real-world system, instead of just comparison against historical data [11]. To this end, we collect information from the literature (e.g. [24,27,30]) and use our perceptions to characterize the seasonality of the liver fluke life-cycle stages in the UK over years 2000–2010. This currently includes shifts in seasonality experienced over this
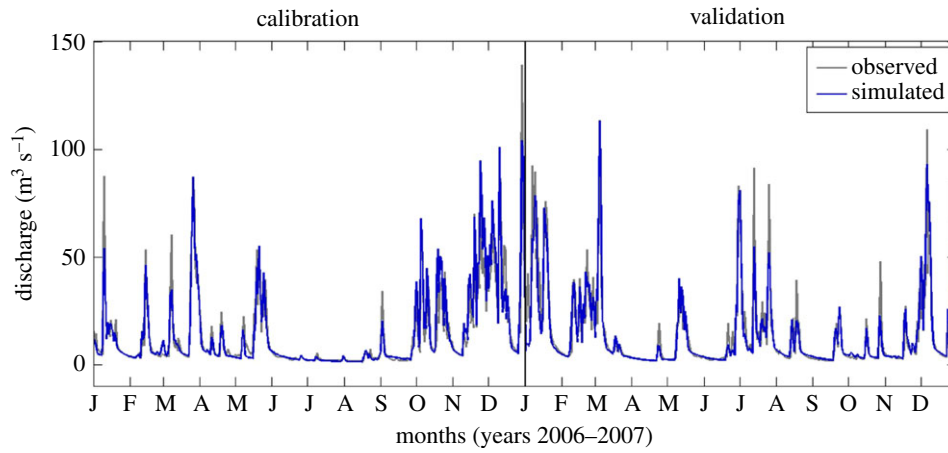
Figure 6. Extract of the calibration and validation periods using daily streamflow data for the River Tawe Catchment (total period is 2000 – 2006 for calibration and 2007 – 2010 for validation).

period, compared to what has been traditionally observed, driven by altered temperature and rainfall patterns, but could be adjusted to account for further changes and shifts, going forwards. Then, we formalize this knowledge into a set of rules:

— *Rule 1*. Retain parameter sets that every year predict the first month of snail presence (i.e. with positive number of snails) to happen earlier than average, if temperature is above average over January–March.
— *Rule 2*: Retain sets that every year produce higher mean number of snail infections over summer (June–August), if number of rain days over May–July is above average.
— *Rule 3*: Retain sets that every year produce higher mean number of metacercariae over autumn (August–October), if rainfall is above average and the number of days $>20°C$ is below average over May–August.
— *Rule 4*: Retain sets that every year produce higher mean number of metacercariae over winter (January–February), if total number of days $>10°C$ is above average over Jan–Feb.

We randomly sample 8000 parameter sets using uniform distributions from ranges in table 2, and reject all sets producing model outputs that are inconsistent with these rules.

4.2.3. Comparison with the Ollerenshaw index

To further evaluate HELF, we use the behavioural parametrizations, i.e. those retained from sequential application of the rules, and compare disease risk simulated using these with the Ollerenshaw index. This, calculated at the monthly scale based on rainfall and temperature characteristics as explained in [29], is the current standard for providing liver fluke forecasts in the UK, where it is used by the National Animal Disease Information Service to warn farmers about high risk years [30].

5. Results

5.1. Performance of the hydrological model

Comparison of simulated and observed daily streamflow shows that TOPMODEL is capable of reproducing the temporal dynamics of observations well, including the peaks and recession periods of the hydrograph. The model achieves

an $R^2 = 0.87$ during calibration and 0.84 in the validation phase (figure 6).

5.2. Performance of the epidemiological model
5.2.1. Fit to epidemiological data

A delay is evident between simulated catchment average number of metacercariae and reported number of sheep diagnosed with fasciolosis from the Tawe Catchment (figure 7). This is due to the time-lag between pasture contamination, which HELF simulates, and infection diagnosed in the animal, which the VIDA dataset reports. Except for the year 2000, for which the model predicts risk of infection that is not reflected in the VIDA numbers over 2001, HELF seems to adequately predict the observed temporal dynamics of infection. It simulates low pasture contamination for most of the period and captures the higher peaks over winters 2008–2009 and 2009–2010, driven by the preceding exceptionally wet summers and rainy autumns. The highest correlation between the two series ($r = 0.62$) is found at a lag of three months, which corresponds to the prepatent period of fasciolosis reported in the literature [18]. If, instead of using the whole dataset for calibration, we perform a five-fold cross-validation, mean correlation results are 0.52 in calibration and 0.41 in validation.

Division of the area for which we have observations within the Severn Catchment into sub-areas with at least 15 data points each, results in nine sub-areas (figure 8). When we compare the simulated percentage of grid cells at risk of infection with the observed percentage of infected herds, in each of the sub-areas, the two are in good agreement ($r = 0.83$), suggesting that the model can replicate the observed spatial pattern (here, performing a leave-one-out cross-validation results in a mean absolute error of 0.1). Risk of infection seems overestimated in sub-areas A2 and A5. However, these areas were significantly drier than the others in 2014 (electronic supplementary material, figure S1) and have a lower percentage of area suitable for snail hosts in terms of soil pH (electronic supplementary material, figure S2), which HELF currently does not account for.

5.2.2. Results of the expert-driven approach

Sequential application of the expert-driven rules reduces the initial sample of 8000 parameter sets to 14 behavioural
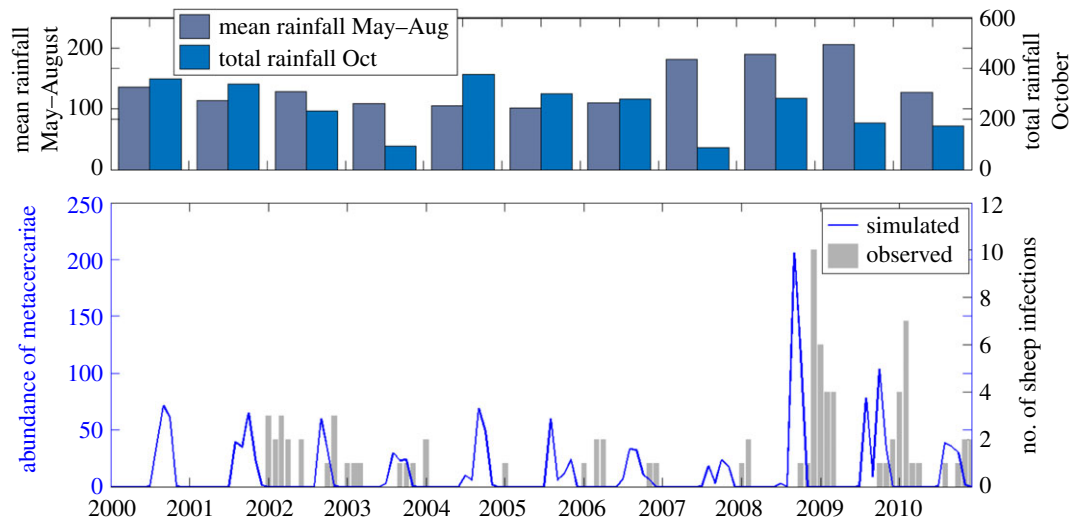
Figure 7. (bottom) Monthly comparison of simulated catchment average number of metacercariae and observed number of infections (VIDA data) over 2000–2010 for the Tawe Catchment. (top) Years 2008 and 2009 have the highest mean summer rainfall within the simulation period, as well as a sufficiently wet autumn, resulting in high suitability for disease transmission.
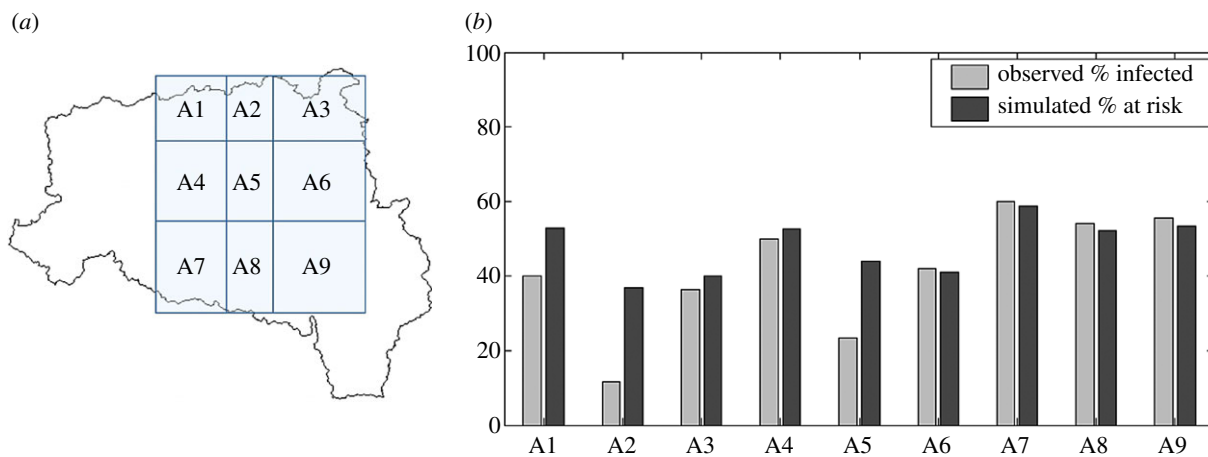

Figure 8. (*a*) Sub-areas within the Severn Catchment for which we have data points (i.e. cattle herds classified into infected and not-infected). (*b*) Comparison of simulated percentage of grid cells at risk of infection and observed percentage of infected herds for each sub-area.
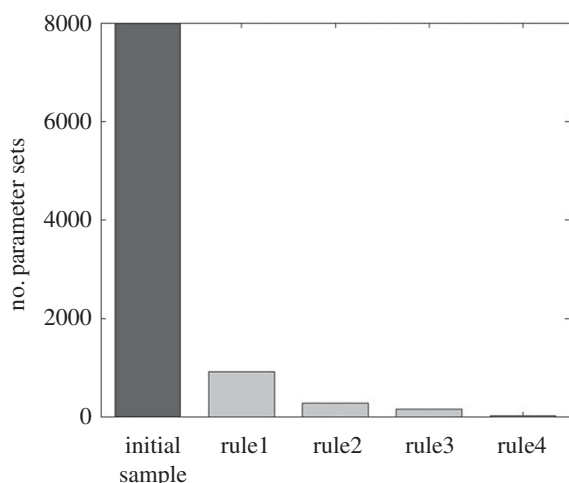

Figure 9. Evolution of the initial sample of parametrizations (each including the 22 epidemiological model parameters sampled from within their initial ranges) along the four confinement steps.

parametrizations (figure 9). The resulting simulated abundance of developed eggs on pasture seems to increase in March, as the weather warms up, before decreasing gradually over the following months, as hatching into miracidia begins (figure 10). Snail activity, and therefore infection of snails by miracidia, also starts in spring and carries on until November, when frosts may send snails back into hibernation. Development of intra-molluscan infections peaks around August, leading to high numbers of infective metacercariae on pasture in Autumn. Finally, if we compare the abundance of metacercariae—this time obtained using the whole set of behavioural parametrizations—with the VIDA time series, first, we still see the expected delay between simulations and observations (figure 11). Second, we note that, while uncertainty is still large in terms of magnitude of the yearly peak of infection, bounds are narrower in terms of timing and duration of the outbreaks, with the number of infective metacercariae on pasture beginning to increase in July, reaching a peak in September, before decreasing again in December, on average.

5.2.3. Comparison with the Ollerenshaw index

Temporal comparison of the suitability for disease transmission simulated by HELF, constrained using the rules, with the Ollerenshaw index, shows a time-lag of one month
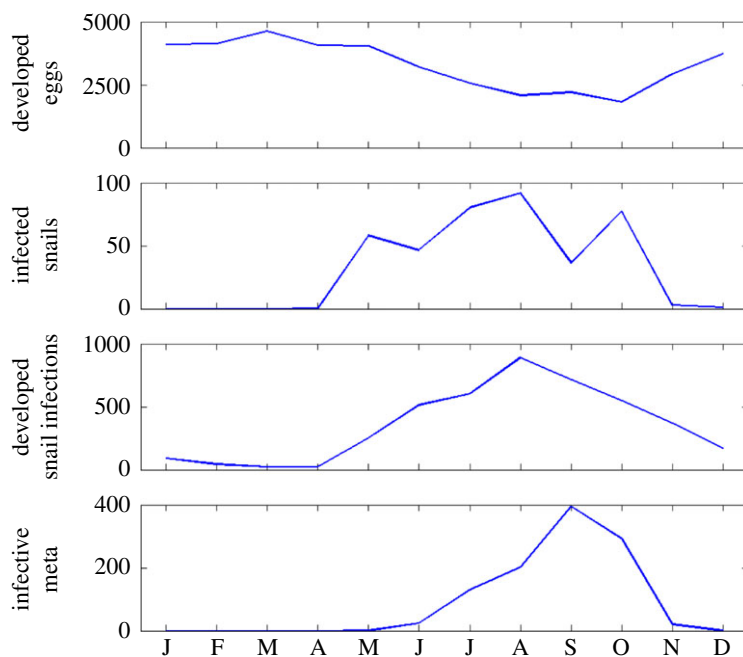
Figure 10. Monthly behaviour of the parasite life-cycle stages simulated with HELF for 2001, as an example (median of the behavioural simulations). (Online version in colour.)
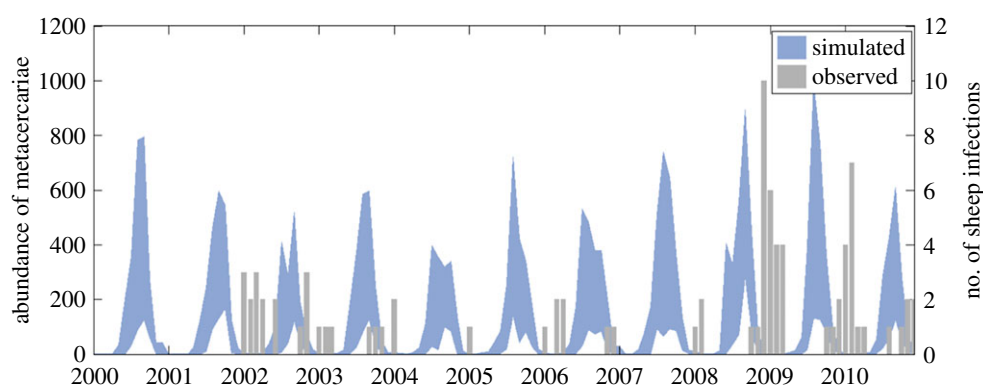

Figure 11. Monthly comparison of simulated catchment average number of metacercariae, obtained using the behavioural parameter sets (90% bounds), and observed number of infections (VIDA data) over 2000–2010 for the Tawe Catchment.

between the two series (figure 12*a* and electronic supplementary material, figure S3). This is due to the two models representing different things: a risk index based on monthly temperature and rainfall in the case of Ollerenshaw, and the abundance of metacercariae, based on soil moisture and accounting for the delays in the parasite life cycle, in the case of HELF. Moreover, we see that, while matching the empirical index on interannual variation (at lag of one month, r = 0.73), the two models' responses may differ at higher temporal resolution. For example, the Ollerenshaw index reaches the same peak value in years 2007 and 2008, but risk of infection in 2007 seems lower than the following year according to HELF. Comparison of the two models in space, presented in figure 12*b* for August 2006 as an example, shows the presence of high risk areas in the Tawe Catchment according to both models. However, when using the Ollerenshaw index, no proportion of the catchment seems risk-free and risk of infection is highest in the north-east where rainfall is highest [37]. In contrast, for the same month, assuming an area is at risk if its number of metacercariae is positive, HELF estimates that 17.3% of the catchment is risk-free, and

that there are 134 patches at risk, spread throughout the catchment, with mean size of 1.6 km².

6. Discussion

In this study, we developed the first mechanistic model which explicitly simulates the risk of infection with *F. hepatica* in time and space, driven by temperature and soil moisture dynamics. The novelty of our work lies in the description of the bio-physical processes underlying transmission of fasciolosis, advancing the study of the disease beyond empirical associations of infection levels with temperature and rainfall. Despite current forecasting models calculating fluke risk based on these meteorological variables [20,29,30], soil moisture has always been recognized as the critical driver of disease transmission for its role on development of the free-living stages and presence of the snail intermediate hosts [20]. Here we included it using an existing hydrologic model, which is based on spatially distributed topographic information, also known as an important fluke risk factor
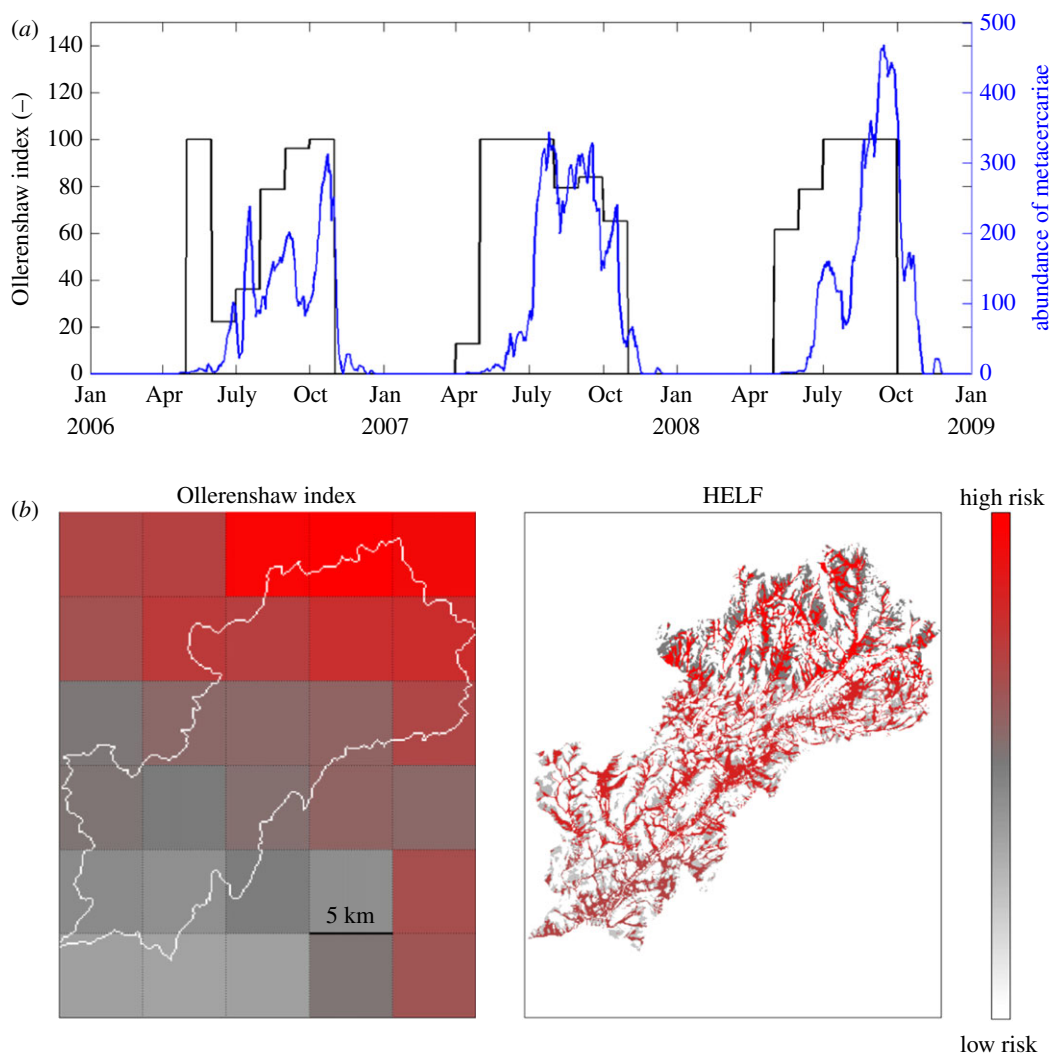
Figure 12. Comparison between the Ollerenshaw index and HELF for the Tawe Catchment: (*a*) in time, for an extract of the simulation period (median of the behavioural parametrizations in blue); and (*b*) in space, for August 2006, as an example.

[27]. Moreover, collaboration across the physical and biological sciences was necessary to analyse the effect of both soil moisture and temperature on the multiple parasite life-cycle stages (figure 3), and translate the mechanistic understanding of the system into an integrated model (figure 4).

By simulating the system at 25 m with a daily time step, HELF provides new insight into the space–time patterns of disease risk, which will be valuable for decision support. Compared to the Ollerenshaw index, which considers each month independently from every other, HELF is dynamic. Therefore, high rainfall may result in high risk of infection depending on the antecedent moisture conditions of the soil and their effect on the life-cycle progress (figure 12*a*). Moreover, by providing greater temporal resolution, HELF allows capturing the impact of short-term weather events, such as extremely warm days or intense concentrated rainfall, which are believed to be particularly relevant for the biological system [13–15]. Combined with the fact that HELF can identify hotspots of transmission potential (figure 12*b*), this means it may be possible for farmers to control the magnitude of exposure to fluke in the field, e.g. by altering management practices to avoid livestock grazing in high risk areas during peak metacercarial abundance. Finally, the stages included in HELF represent the part of the life cycle which is missing in the model of fluke dynamics within the final host developed in [19]. Integration of the two would

allow a mechanistic description of the whole cycle, thus providing the opportunity to assess e.g. the impact of vaccines on infection levels.

In addition to aiding the management of fasciolosis, HELF could also benefit the study of other diseases. A similar model could be useful for rumen fluke, which is on the rise in British and Irish livestock and has a similar life cycle to liver fluke, sharing the same intermediate host [43]. On the other hand, a different hydrological model component could be employed instead of TOPMODEL, depending on the hydro-environmental drivers relevant for the disease considered [3]. For example, a model describing freshwater connectivity would be needed for diseases involving aquatic intermediate hosts, such as freshwater snails in the case of schistosomiasis [2].

Several assumptions are embedded in HELF. Notably, to account for seasonality and distribution of the disease, we assumed the parasite life cycle is entirely driven by environmental conditions, simplifying the mechanisms related to the intra-molluscan stage and neglecting density-dependent processes. Even with regard to environmental factors, characteristics such as soil pH and texture have been described as potentially relevant for the suitability of snail habitats [27], but have not been included in our model, yet. Similarly, surplus run-off water may have a role in the infection transmission pathway, contributing to the dispersal of snails and metacercariae down water courses [44]. However, HELF

could be expanded to incorporate these, as well as additional spatial data, including remote sensing information.

To address common disease data limitations, we proposed an approach that includes the use of expert knowledge to constrain and evaluate our new model. Fitting observations is standard practice for calibration of hydrologic models, when there is a gauging station providing data to compare simulations against (figure 6). Distributed soil moisture observations were not available for our case studies, but previous works have shown that TOPMODEL can provide good representation of the spatial pattern of saturated areas [45]. Less frequently, when data are available, calibration is performed to parametrize epidemiological models (e.g. [10,16,46]). Our results show that HELF is flexible enough to replicate the observed time–space patterns of infection over two case study catchments (figures 7 and 8). We speculate remaining mismatches when we fit the two datasets are not necessarily due to aspects not yet included in the model only, but may also be related to data issues. The absence of reported cases for 2001 from the Tawe Catchment is believed to have been influenced by the UK outbreak of foot-and-mouth in the same year. Similarly, discrepancies in some sub-areas of the Severn Catchment may also be due to our underlying assumption of uniform distribution of farms per sub-area, which may not reflect the real-world system. Mis-reporting and low space–time resolution of data are common issues for many diseases and have often been recognized as a bottleneck to developing models providing meaningful predictions of disease risk [12,14,16]. Recent correlative fluke studies (e.g. [47]), have used geo-referenced data from abattoir liver condemnations, which, if routinely collated and made available, may benefit testing of models such as HELF across wider areas. However, even if larger, potentially more reliable epidemiological data were available, they would still reflect historical conditions, which may not necessarily be relevant for the future [11,15]. Our calibration strategy includes the use of expert-driven rules to overcome these issues. The rules represent mechanistic knowledge of the system translated into prior information about the output variables. By using these, we can constrain aspects of the model for which no hard data are available in a process-based manner, without biasing the parameters towards external drivers not included in the model. The current formulation reflects changes in seasonality experienced over our simulation period. However, going forward, this can be adjusted to account for further changes, in order to reliably assess the impact on disease risk of conditions beyond the range of historical variability [48]. Our results show there are parametrizations satisfying all four our rules (figure 9), and that the behaviour of the simulated stages and the lags between them (figure 10) agree with what is reported in the literature [20,24]. This suggests that HELF reflects well (our current knowledge of) the real-world system. The fact that simulations are rejected from the initial sample suggests that our parameter confinement strategy is effective, which is crucial as the inability to identify behavioural parametrizations may result in significant predictive uncertainty when using the model under changing conditions [15,34]. Moreover, using HELF with Monte Carlo sampling allows explicit consideration of uncertainty, by propagating it from the parameter ranges to the model simulations. This means we can provide decision-makers with a degree of confidence attributed to the model results. The reason why uncertainty in the simulated risk of infection still seems high in terms of magnitude (figure 11) is that the rules are currently based on information about the seasonality of the disease only, driven by our aim of providing a model that is generally applicable across the UK. However, if reliable local data were available, the rules could be modified or increased in number to make the model more accurate locally (see [16,49]). Instead, the fact that uncertainty bounds are narrow in terms of timing and duration of the outbreaks is particularly useful to inform farmers' decisions about e.g. when to allow grazing of animals or when to treat them.

7. Conclusion

We developed and tested a new mechanistic hydro-epidemiological model to simulate the risk of liver fluke infection linked to key weather–water–environmental processes (HELF). The fact that, unlike previous models, HELF explicitly describes processes, rather than relying on correlation, makes it better suited for capturing the impact of 'new' conditions on disease risk. We showed that the model is sufficiently flexible to fit observations for two UK case studies, but also introduced an expert-driven calibration strategy to make the model more robust to data with limited reliability and in the presence of climate change. Finally, comparison with a widely used empirical model of fluke risk showed that, while matching the existing index on interannual variation, HELF provides better insight into the time–space patterns of disease, which will be valuable for decision support. Driving the model with climate and management scenarios will enable assessment of future risk of infection and evaluation of control options to reduce and/or mitigate disease burden. This is urgent, given the widespread increase in drug resistance and threat of altered patterns of transmission due to climate-environmental change. Through the example of fasciolosis, we demonstrated (i) that sufficient mechanistic understanding of the bio-physical system may be available to develop and test a process-based model for an environment-driven disease, without having to rely only on limited and potentially disinformative data, and (ii) how accounting for the critical hydro-environmental controls underlying transmission can be valuable to better understand seasonality and spread of emerging or re-emerging threatening diseases.

References

1. Altizer S, Dobson A, Hosseini P, Hudson P, Pascual M, Rohani P. 2006 Seasonality and the dynamics of infectious diseases. *Ecol. Lett.* **9**, 467–484. (doi:10.1111/j.1461-0248.2005.00879.x)

2. Perez-Saez J, Mande T, Ceperley N, Bertuzzo E, Mari L, Gatto M, Rinaldo A. 2016 Hydrology and density feedbacks control the ecology of intermediate hosts of schistosomiasis across habitats in seasonal climates. *Proc. Natl Acad. Sci. USA* **113**, 6427–6432. (doi:10.1073/pnas.1602251113)

3. Rinaldo A, Gatto M, Rodriguez-Iturbe I. 2018 River networks as ecological corridors: a coherent ecohydrological perspective. *Adv. Water Resour.* **112**, 27–58. (doi:10.1016/j.advwatres.2017.10.005)

4. Coumou D, Rahmstorf S. 2012 A decade of weather extremes. *Nat. Clim. Chang.* **2**, 491–496. (doi:10.1038/NCLIMATE1452)

5. Van Loon AF *et al.* 2016 Drought in the Anthropocene. *Nat. Geosci.* **9**, 89–91. (doi:10.1038/ngeo2646)

6. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, Daszak P. 2008 Global trends in emerging infectious diseases. *Nature* **451**, 990–993. (doi:10.1038/nature06536)

7. Wu X, Lu Y, Zhou S, Chen L, Xu B. 2016 Impact of climate change on human infectious diseases: empirical evidence and human adaptation. *Environ. Int.* **86**, 14–23. (doi:10.1016/j.envint.2015.09.007)

8. Steinmann P, Keiser J, Bos R, Tanner M, Utzinger J. 2006 Schistosomiasis and water resources development: systematic review, meta-analysis, and estimates of people at risk. *Lancet Infect. Dis.* **6**, 411–425. (doi:10.1016/S1473-3099(06)70521-7)

9. World Bank. 2014 Reducing climate-sensitive disease risks. Agriculture and environmental services discussion paper no.7. Washington, DC.

10. Lloyd-Smith JO, George D, Pepin KM, Pitzer VE, Pulliam JRC, Dobson AP, Hudson PJ, Grenfell BT. 2009 Epidemic dynamics at the human–animal interface. *Science* **326**, 1362–1367. (doi:10.1126/science.1177345)

11. Wagener T *et al.* 2010 The future of hydrology: an evolving science for a changing world. *Water Resour. Res.* **46**, W05301. (doi:10.1029/2009WR008906)

12. Fox NJ, Marion G, Davidson RS, White PCL, Hutchings MR. 2012 Livestock helminths in a changing climate: approaches and restrictions to meaningful predictions. *Animals* **2**, 93–107. (doi:10.3390/ani2010093)

13. Lo Iacono G, Armstrong B, Fleming LE, Elson R, Kovats S, Vardoulakis S, Nichols GL, Remais JV. 2017 Challenges in developing methods for quantifying the effects of weather and climate on water-associated diseases: a systematic review. *PLoS Negl. Trop. Dis.* **11**, e0005659. (doi:10.1371/journal.pntd.0005659)

14. Urban MC *et al.* 2016 Improving the forecast for biodiversity under climate change. *Science* **353**, aad8466. (doi:10.1126/science.aad8466)

15. Dormann CF *et al.* 2012 Correlation and process in species distribution models: bridging a dichotomy. *J. Biogeogr.* **39**, 2119–2131. (doi:10.1111/j.1365-2699.2011.02659.x)

16. Eisenberg JNS, Brookhart MA, Rice G, Brown M, Colford JM. 2002 Disease transmission models for public health decision making: analysis of epidemic and endemic conditions caused by waterborne pathogens. *Environ. Health Perspect.* **110**, 783–790. (doi:10.1289/ehp.02110783)

17. Mas-Coma S, Bargues MD, Valero MA. 2005 Fascioliasis and other plant-borne trematode zoonoses. *Int. J. Parasitol.* **35**, 1255–1278. (doi:10.1016/j.ijpara.2005.07.010)

18. Kaplan RM. 2001 *Fasciola hepatica:* a review of the economic impact in cattle and considerations for control. *Vet. Ther.* **2**, 40–50.

19. Turner J, Howell A, McCann C, Caminade C, Bowers RG, Williams D, Baylis M. 2016 A model to assess the efficacy of vaccines for control of liver fluke infection. *Sci. Rep.* **6**, 23345. (doi:10.1038/srep23345)

20. Ollerenshaw CB, Rowlands WT. 1959 A method of forecasting the incidence of fascioliasis in Anglesey. *Vet. Rec.* **71**, 591–598.

21. Pantelouris EM. 1963 Environmental influences on the life-cycle of the liver-fluke, *Fasciola hepatica* L. *Ir. Nat. J.* **14**, 94–97.

22. Van Dijk J, Sargison ND, Kenyon F, Skuce PJ. 2010 Climate change and infectious disease: helminthological challenges to farmed ruminants in temperate regions. *Animal* **4**, 377–392. (doi:10.1017/s1751731109990991)

23. Charlier J, Vercruysse J, Morgan E, van Dijk J, Williams DJL. 2014 Recent advances in the diagnosis, impact on production and prediction of *Fasciola hepatica* in cattle. *Parasitology* **141**, 326–335. (doi:10.1017/S0031182013001662)

24. Relf V, Good B, Hanrahan JP, McCarthy E, Forbes AB, deWaal T. 2011 Temporal studies on *Fasciola hepatica* in *Galba truncatula* in the west of Ireland. *Vet. Parasitol.* **175**, 287–292. (doi:10.1016/j.vetpar.2010.10.010)

25. WHO. 2017 See http://www.who.int/foodborne_trematode_infections/fascioliasis/en/.

26. Malone JB, Williams TE, Muller RA, Geaghan JP, Loyacano AF. 1987 Fasciolosis in cattle in Louisiana: development of a system to predict disease risk by climate, using the Thornthwaite water budget. *Am. J. Vet. Res.* **48**, 1167–1170.

27. McCann CM, Baylis M, Williams DJL. 2010 The development of linear regression models using environmental variables to explain the spatial distribution of *Fasciola hepatica* infection in dairy herds in England and Wales. *Int. J. Parasitol.* **40**, 1021–1028. (doi:10.1016/j.ijpara.2010.02.009)

28. Selemetas N, Phelan P, O'Kiely P, Waal T. 2014 Weather and soil type affect incidence of fasciolosis in dairy cow herds. *Vet. Rec.* **175**, 371. (doi:10.1136/vr.102437)

29. Fox N *et al.* 2011 Predicting impacts of climate change on *Fasciola hepatica* risk. *PLoS ONE* **6**, e16126. (doi:10.1371/journal.pone.0016126)

30. NADIS. 2017 See http://www.nadis.org.uk/bulletins/liver-fluke-control-in-sheep.aspx.

31. Nice NG, Wilson RA. 1974 A study of the effect of temperature on the growth of *Fasciola hepatica* in *Lymnaea truncatula*. *Parasitology* **68**, 47–56. (doi:10.1017/S0031182000045364)

32. Beven KJ, Kirkby MJ. 1979 A physically based, variable contributing area model of basin hydrology. *Hydrol. Sci. Bull.* **24**, 43–69. (doi:10.1080/02626667909491834)

33. Beven K *et al.* 1995 TOPMODEL. In *Computer models of watershed hydrology* (ed. VP Singh), pp. 627–668. Colorado: Water Resour. Publ.

34. Beven K. 1997 TOPMODEL: a critique. *Hydrol. Process.* **11**, 1069–1085.

35. Andrews SJ. 1999 The life cycle of *Fasciola hepatica*. In *Fasciolosis* (ed. JP Dalton), pp. 1–29. Wallingford, UK: CABI Publishing.

36. Thomas AP. 1883 The life history of the liver-fluke (*Fasciola hepatica*). *Q. J. Microsc. Sci.* **23**, 99–133.

37. NRFA. 2017 See http://nrfa.ceh.ac.uk/.

38. Williams DJL, Hodgkinson JE. 2014 Liver fluke—an overview for practitioners. *Cattle Pract.* **22**, 238–244.

39. Murphy DM *et al.* 2009 UK climate projections science report: climate change projections. Exeter, UK: Met Office Hadley Centre.

40. Van Dijk J, David GP, Baird G, Morgan ER. 2008 Back to the future: developing hypotheses on the effects of climate change on ovine parasitic gastroenteritis from historical data. *Vet. Parasitol.* **158**, 73–84. (doi:10.1016/j.vetpar.2008.08.006)

41. Klemes V. 1986 Operational testing of hydrological simulation models. *Hydrol. Sci. J.* **31**, 13–24. (doi:10.1080/02626668609491024)

42. Duan Q, Sorooshian S, Gupta V. 1992 Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resour. Res.* **28**, 1015–1031. (doi:10.1029/91WR02985)

43. Huson KM, Oliver NAM, Robinson MW. 2017 Paramphistomosis of ruminants: an emerging parasitic disease in Europe. *Trends Parasitol.* **33**, 836–844. (doi:10.1016/j.pt.2017.07.002)

44. Mas-Coma S, Valero MA, Bargues MD. 2009 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)

45. Güntner A, Seibert J, Uhlenbrook S. 2004 Modeling spatial patterns of saturated areas: an evaluation of different terrain indices.

Water Resour. Res. **40**, W05114. (doi:10.1029/2003WR002864)

46. Bertuzzo E, Finger F, Mari L, Gatto M, Rinaldo A. 2016 On the probability of extinction of the Haiti cholera epidemic. *Stoch. Environ. Res. Risk Assess.* **30**, 2043 – 2055. (doi:10.1007/s00477-014-0906-3)

47. Byrne AW *et al.* 2016 Liver fluke (*Fasciola hepatica*) infection in cattle in Northern Ireland: a large-scale epidemiological investigation utilising surveillance data. *Parasit. Vectors* **9**, 209. (doi:10.1186/s13071-016-1489-2)

48. Singh R, Wagener T, van Werkhoven K, Mann ME, Crane R. 2011 Trading space for time: a nonstationary uncertainty framework for hydrologic modeling in a changing climate. *Hydrol. Earth Syst. Sci.* **15**, 3591 – 3603. (doi:10.5194/hess-15-3591-2011)

49. Liang S, Spear RC, Seto E, Hubbard A, Qiu D. 2005 A multi-group model of *Schistosoma japonicum* transmission dynamics and control: model calibration and control prediction. *Trop. Med. Int. Health* **10**, 263 – 278. (doi:10.1111/j.1365-3156.2005.01386.x)