

1 Hydrology-informed metapopulation modeling of liver fluke
2 transmission in the Lawa Lake complex of northeast Thailand

3

4

5

6

7

8

9

10 Tomás M. León^{1,3*}, Vichian Plermkamon², Kittiwet Kuntiyawichai², Banchob Sripa³,
11 Robert C. Spear¹

12 ¹School of Public Health, University of California, Berkeley, USA

13 ²Faculty of Engineering, Khon Kaen University, Khon Kaen, Thailand

14 ³Tropical Disease Research Center, Faculty of Medicine, Khon Kaen University, Khon
15 Kaen, Thailand

16

17 * Corresponding author

18 Email: tomas.leon@berkeley.edu (TML)

19

20 **Abstract:**

21 A metapopulation model linking local hydrology with transmission of the liver fluke
22 *Opisthorchis viverrini* in a lake system in northeast Thailand was developed and
23 parameterized using infection data from 2008-2016. A rainfall-runoff model and other
24 hydrologic data were used to assess level of connectivity between villages and the
25 influence of upstream communities on parasite distribution in the study area, while the
26 disease transmission model was expanded from a single-village model for *O. viverrini*
27 transmission. Connectivity between villages and hydrologic variables was assessed
28 monthly and showed strong seasonality trends. The metapopulation model improved
29 upon the single-village model in its fit to historical data patterns for the six village
30 clusters with the introduction of the new time-variable parameters. Results suggest
31 there are three unique hydrologic-epidemiologic regimes within the Lawa Lake system
32 in response to upstream watersheds and risk of overland flooding that contribute to risk
33 for *O. viverrini* infection. Similar approaches using a hydrologic submodel to inform a
34 mechanistic disease transmission model could be applied across many water-related
35 disease systems.

36
37 **Author Summary:**

38 While hydrology is intuitively understood to influence transmission dynamics of water-
39 related diseases, limited research exists that explicitly links hydrologic and infectious
40 disease data. In this work, we use an approach that leverages a rainfall-runoff model to
41 better understand water movement into, out of, and around Lawa Lake in northeast
42 Thailand and how that affects fate and transport of the multiple waterborne life stages of
43 *Opisthorchis viverrini*. To model disease transmission, we represent six village clusters

44 around the lake using known infection prevalence data of humans, cats and dogs, snails,
45 and fish to parameterize and fit a metapopulation model. The connectivity between
46 village clusters and external inputs of parasites are derived from the hydrology data and
47 the rainfall-runoff model. Results suggest three unique hydrologic regimes that also
48 reflect unique patterns in disease prevalence among the different hosts. Other water-
49 related disease systems can use similar approaches to assess the impacts of water on
50 pathogen transmission dynamics.

51

52 **Introduction**

53 Opisthorchiasis, infection with liver flukes of the genus *Opisthorchis*, is a disease
54 whose transmission and distribution are largely determined by hydrology. The
55 parasites' egg and cercarial forms require sufficient water and transport to the next
56 intermediate host for the transmission cycle to be sustained. Eggs are excreted in the
57 feces of infected final hosts (humans, and reservoir cats, dogs, and other mammals to a
58 limited extent); if not safely treated or contained, these eggs reach the amphibious
59 habitat of intermediate host snails, which consume the eggs and enable maturation to
60 the cercarial stage. The cercariae are then released back into water, where they swim
61 and seek out the second intermediate host cyprinid fish. They encyst in the fish, which
62 if then consumed raw or undercooked by humans or certain other mammals can
63 migrate to the bile ducts and mature into adult worms. Water is a sustaining force for
64 this parasitic life cycle, and its movement permits viable infection at each successive
65 host stage.

66

67 The major liver fluke of interest in Thailand is *Opisthorchis viverrini*. Given the

68 known disease burden in Thailand, *O. viverrini* has been a public health priority there,
69 where it is transmitted to humans via the consumption of popular local raw and
70 fermented fish dishes such as *koi pla* and *pla som*. The highest prevalence of
71 opisthorchiasis and cholangiocarcinoma (CCA), the fatal bile duct cancer associated
72 with *O. viverrini* infection, are found in northeast Thailand in the region surrounding
73 Khon Kaen [1]. Historical hotspots of opisthorchiasis and CCA were in the villages
74 around Lawa Lake. While much research has been conducted on the pathology of
75 opisthorchiasis and CCA, there is limited literature addressing the ecological and
76 hydrological aspects of parasite transmission in the environment [2,3].

77

78 Lawa Lake is an approximately 4000-acre body of water that is highly vegetated
79 and subject to significant hydrologic changes caused by seasonal variation in northeast
80 Thailand. A peak in liver fluke infections is seen with lag following the rainy season in
81 Thailand, as flooding facilitates the spread of fecal contamination and coincides with
82 the rapid increase in snail populations [1]. Since several weeks are required for the
83 parasite to mature through its life stages, high infection rates in fish are seen in the late
84 rainy season and summer (July-January). Low infection risks occur in the dry season
85 and summer (March-June). A primary industry on the lake is fishing, which contributes
86 to the environment mediating ongoing liver fluke transmission in the local area [4]. The
87 hydrology of the Lawa Lake region is exceedingly complex and disturbed, as significant
88 changes have occurred in recent years due to construction of new irrigation canals and
89 ditches, new culverts and spillways that are opened and closed in the flooding season,
90 and fish ponds that have become increasingly popular as a source of food and revenue.

91

92 Metapopulation modeling is commonly used to better understand the
93 connectivity and influence of discrete human populations and environmental patches
94 on each other. Metapopulation modeling is especially powerful in understanding
95 pathogen transmission in complex, interrelated environments, where contiguous areas
96 and their inhabitants influence each other. Connectedness between these environmental
97 and host patches can occur in multiple ways, including migration of humans between
98 villages and movement of a waterborne pathogen from a section of river or lake
99 adjacent to one population to another section adjacent to a separate population. The
100 second example demonstrates hydrological influence on a disease transmission system,
101 as waterborne diseases as diverse as cholera and schistosomiasis rely on advective
102 transport to expose new susceptible individuals with pathogens excreted or shed by
103 infected individuals. Hydrologic patterns are time-varying and markedly local in
104 nature. Given this environmental complexity, hydrologic fate and transport of
105 pathogens are difficult to study. In addition, motile waterborne parasitic forms, such as
106 liver flukes and schistosome cercariae, have independent mobility behaviors, making
107 hydrologic flows not entirely representative of how these parasites are dispersed [5,6].

108

109 Research connecting hydrology with waterborne disease transmission is an
110 emerging field with recent work on cholera and schistosomiasis [7,8,9]. Tracing the
111 spread of pathogens in the environment is challenging, and countervailing forces make
112 it difficult to ascribe an increase or decrease in human infection to trends in
113 meteorology and climate or consequent hydrology. Long-term studies of climatic
114 changes in rainfall patterns or the influence of dams allow more definitive statements
115 about impacts on disease transmission, though these are also complicated by shifts in

116 host and vector habitat and in seasonal patterns that may disrupt or exacerbate host and
117 vector growth and reproduction [10,11].

118

119 In this work, a metapopulation disease transmission model is developed and
120 parameterized to assess hydrologic connectivity and *O. viverrini* parasite movement
121 between six village clusters around Lawa Lake in Khon Kaen Province, Thailand, and
122 how that is reflected by opisthorchiasis prevalence in hosts. Understanding liver fluke
123 transmission in this seasonal, hydrologically connected environment with modeling can
124 help define the scale of transmission processes and thereby optimize environmental
125 control and treatment to have maximum impact on reducing disease transmission in
126 this setting and others.

127

128 **Methods**

129 The model structure is an extension of the modeling framework presented in
130 [12]. The six village clusters studied are now connected in a metapopulation framework
131 to account for exchange of parasites and hosts between village clusters and their
132 associated environments. This enables the model to include the influence of population-
133 level factors, spatial heterogeneity, and degrees of connectedness between patches. This
134 metapopulation model leverages information about hydrologic connectivity between
135 village and host clusters to understand the movement of the liver fluke parasite's
136 various forms in the environment as mediated by water. To consider local hydrologic
137 impacts on the liver fluke transmission cycle, five main factors are included: 1) egg
138 inputs into the system from upstream watersheds; 2) egg inputs into the system from
139 overland flooding; 3) snail and fish mobility due to hydrologic connectivity; 4) snail and
140 fish available habitat; and 5) hotspots where infectious snails come into contact with

141 susceptible fish. These factors are modeled monthly with seasonality to account for
142 changing patterns throughout the year.

143

144 The metapopulation model connects the six villages or geographically proximate
145 village clusters shown in Fig 1 around Lawa Lake in Khon Kaen Province of northeast
146 Thailand. The villages or village clusters described here are CCK, Lawa, BT (cluster of 5
147 villages), NNK, KSR, and DPD (cluster of 2 villages), which were chosen and clustered
148 based on geographical location and how historical human infection survey data was
149 collected. The six clusters further sort into three groups based on impact or lack thereof
150 of flooding and upstream watersheds. The two upstream watersheds (Fig 2) flow into
151 Lawa Lake at locations adjacent to NNK and BT and contribute to egg input there.
152 Villages CCK and Lawa are in proximity to the Chi River and are most susceptible to
153 seasonal overland flooding. KSR and DPD are the villages most “downstream” and are
154 relatively isolated from major flooding or upstream drainage impacts. Villages within
155 the upstream watershed have not had as significant treatment and control programs as
156 the villages around Lawa Lake, and reports suggest that upstream villages still have
157 high infection prevalence values over 30%.

158

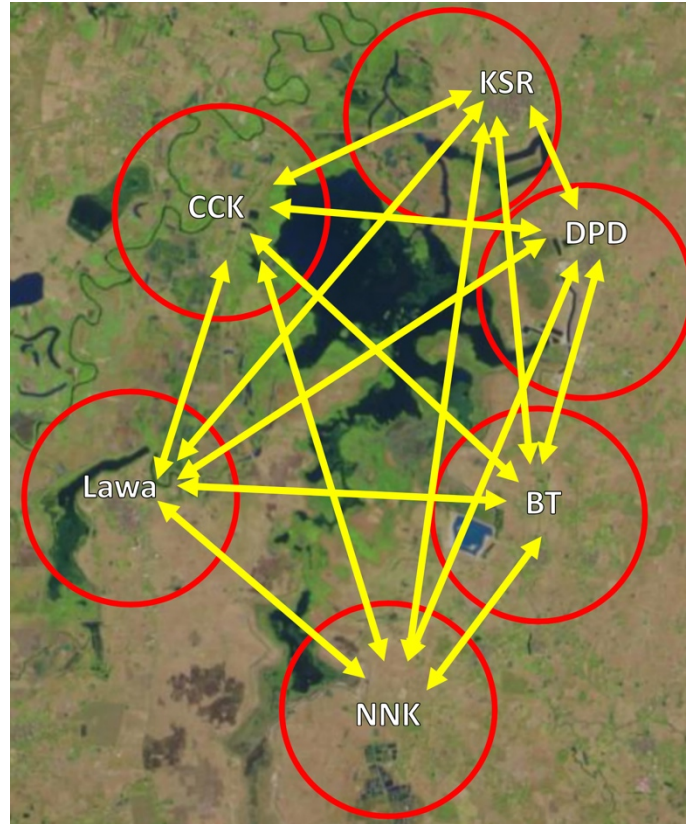


Fig 1: Connectivity between six village clusters around Lawa Lake (Map data: Landsat, USGS).

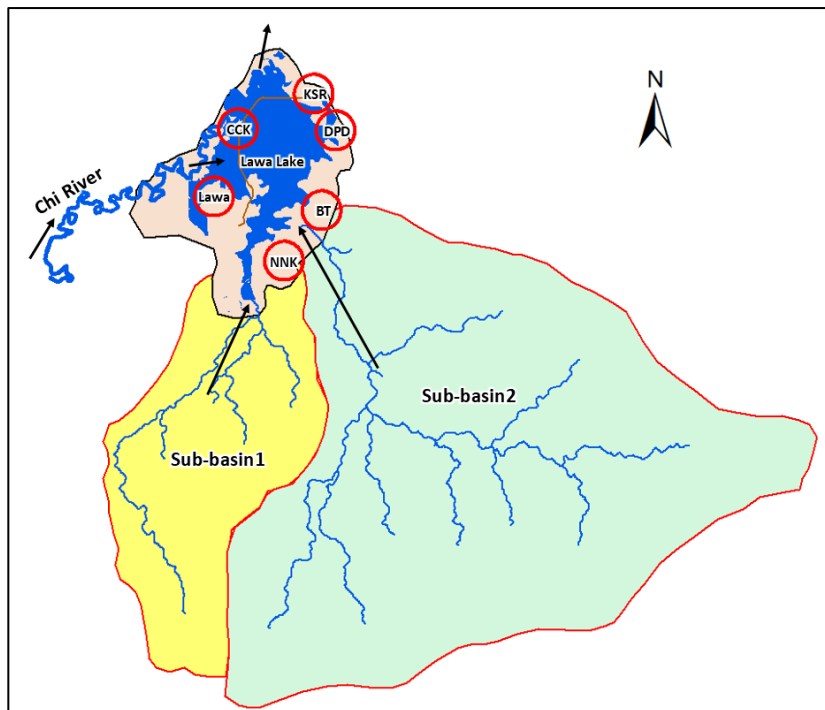


Fig 2: Two upstream sub-basins flow into Lawa Lake near NNK and BT, respectively (QGIS).

160 These hydrologic impacts seem to mirror trends in epidemiological patterns.
161 Previous research highlighted the north-to-south gradient of nitrogen and salinity in the
162 water that predicted higher snail abundance [13]. NNK, the southernmost village
163 cluster, had the highest infection prevalence at the time points when it was studied
164 (Table 1). KSR and DPD, the northernmost villages and the farthest from the Chi River,
165 had the lowest baseline prevalence of *O. viverrini* infection before the control program
166 started. In the single-village model, varying transmission parameters by village did not
167 fully capture the different patterns that occurred in the six village clusters when
168 connectivity was not included. Therefore, the role of hydrology and connectivity
169 between villages needs to be considered in the disease transmission model to better
170 account for the patterns observed.

Table 1: Infection prevalence (%) and mean intensity in positive individuals (EPG) for six village clusters around Lawa Lake. * indicates less sensitive diagnostic method (Kato-Katz or Kato thick smear).

Village Cluster	2008	2010	2011	2012	2014	2015
CCK	54.9% (593)		33.0%*	44.3% (139)		
Lawa	67.1% (501)	63.1% (108)	19.0%*	(16)		8.7%*
BT	61.9% (346)	37.2% (131)		35.0% (136)	9.0%*	14.2%*
NNK	74.1% (499)			50.0% (61)		
KSR	16.4% (101)					
DPD	22.1% (112)	36.5% (82)				14.6%*

171
172 To understand the effects of complex hydrologic factors on snail habitat, fish
173 access to these habitats, and the pathways of parasite transmission, a hydrologic model
174 of the Lawa Lake system was utilized to simulate flow patterns and changes in water
175 levels over time. This model is a rainfall/runoff model common in hydrology that

176 considers the transport of water through a system originating from upstream in the
177 catchment basin or from precipitation. It uses the Soil & Water Assessment Tool
178 (SWAT) model to generate runoff for the PCSWMM hydraulic model to determine
179 hydraulic parameters of Lawa Lake such as flood depth and extent, flow velocity, and
180 travel time [14]. The inputs for the SWAT model include meteorological data (rainfall,
181 temperature, relative humidity, and windspeed) from the Thai Meteorological
182 Department, soil type, land use, and a digital elevation model (DEM) generated from
183 satellite imagery. For the hydraulic model, hydrologic structures and key parts of Lawa
184 Lake were surveyed for elevation at 1m x 1m resolution using drone data, and a 2D
185 model integrating runoff, the improved DEM, and meteorological data from 2008 to
186 2016 was developed using PCSWMM. Outputs include water level and flow vectors for
187 the time points modeled between 2008 and 2016. Calibration was conducted with
188 available precipitation and gauge data in the area from the Bureau of Water
189 Management and Hydrology, Royal Irrigation Department, in Thailand. Fig 3
190 demonstrates an example of the variation in flows predicted by the model over the
191 course of a calendar year encompassing the rainy and dry seasons; arrow direction and
192 thickness represent the relative change in flow vectors. While all of the villages are
193 adjacent to Lawa Lake, inflows and outflows as well as the relevant snail and fish
194 populations are not static, homogeneously distributed, or relevant to each village. The
195 rainy season is characterized by high and active flows that generally peak in October
196 with flooding from the Chi River varying from year to year. The dry season has
197 relatively little hydrologic activity establishing connectivity between village clusters.
198
199

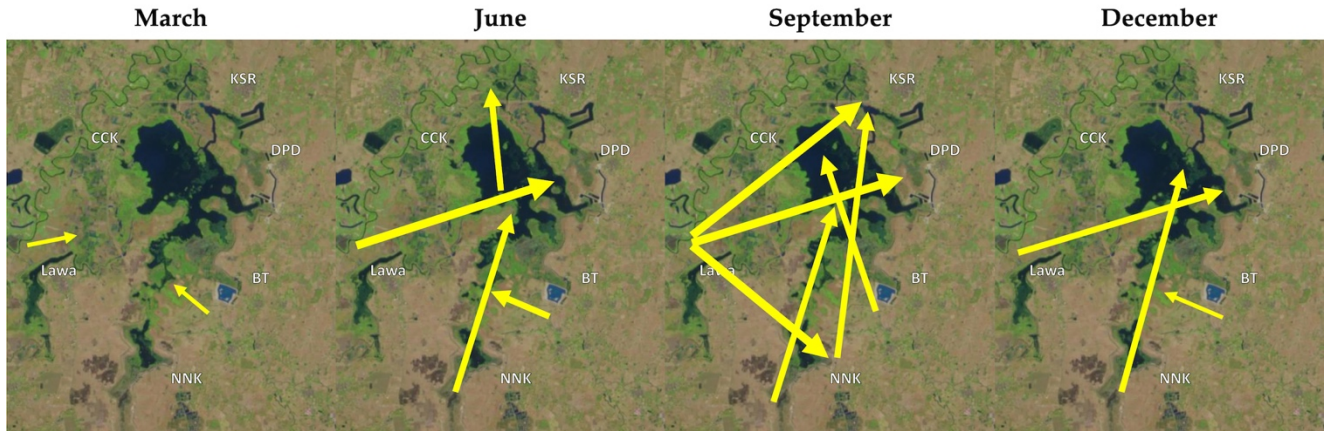


Fig 3: Hydrologic flows change dynamically throughout the year (semiquantitative interpretation of hydrology model results for an annual cycle) (Map data: Landsat, USGS).

200 Because gauge data was not available for the sub-basins upstream of Lawa Lake,
201 a Soil and Water Assessment Tool – Calibration and Uncertainty Procedures (SWAT-
202 CUP) model was used to calibrate and validate runoff into the Lawa Lake system from
203 these sources generated by the SWAT model. Inputs for the upstream sub-basins
204 included a 30m x 30m DEM, land use data, and soil type data from the Land
205 Development Department of Thailand. To map the hydrologic features of Lawa Lake
206 and finer scale structures, a drone was used to chart these areas in greater detail and
207 determine elevations where water was flowing into or out of Lawa Lake. Sub-basin
208 calibration and validation graphs are shown in Fig 4; 2006-2010 data was used for
209 calibration, and 2011-2013 data was used for validation. The R^2 values range from 0.61 to
210 0.81; both the calibration and validation models miss late peaks in their runs. In the case
211 of calibration, fitting the less extreme peaks may have disadvantaged the model from
212 predicting the major discharge in 2010. In the case of validation, the 2011 peak
213 discharge was better modeled, but 2013 was missed by a large margin for reasons that
214 are not entirely clear but may be related to the different timing of precipitation-driven
215 flooding in 2013 compared with other years.

216

217 Based on these data and tools, the presence/absence of connections between
218 parcels of water associated with different village clusters were assessed. For example,
219 year-round one would not expect KSR to contribute to transmission in Lawa or NNK
220 but would expect the reverse to be true. In March, when less rainfall and flooding occur,
221 each village cluster is relatively isolated, with the exception of the relatively weak
222 influence of near neighbors (Lawa to CCK or BT to DPD). The origin of flows is also
223 subject to a differential dilution effect; contaminated waste from the upstream
224 watershed (Fig 2) would more strongly impact NNK than DPD or KSR and would
225 experience the effects sooner.

226

227 The mathematical model of disease transmission (which incorporates data from
228 the hydrologic model as variables and parameters) is an expansion of the single-village
229 model described in [12] to connect the six village clusters and uses as its state variables
230 the infection prevalence in humans, reservoir hosts (cats and dogs), snails, and fish. Base
231 equations are included below in Equations (1) – (5). Initial values were set from the
232 baseline surveys in 2008, and base transmission parameters are carried over from the
233 single-village model found using Markov chain Monte Carlo (MCMC) methods to fit the
234 model onto known infection prevalence data as described below. The infection
235 prevalence data includes surveys using two different methods: formalin-ether
236 concentration technique (FECT) and Kato-Katz (or Kato thick smear). In Thailand, FECT
237 has been regarded as the significantly more sensitive method to detect *O. viverrini*
238 infection based on available data and because the protocol is intended to make

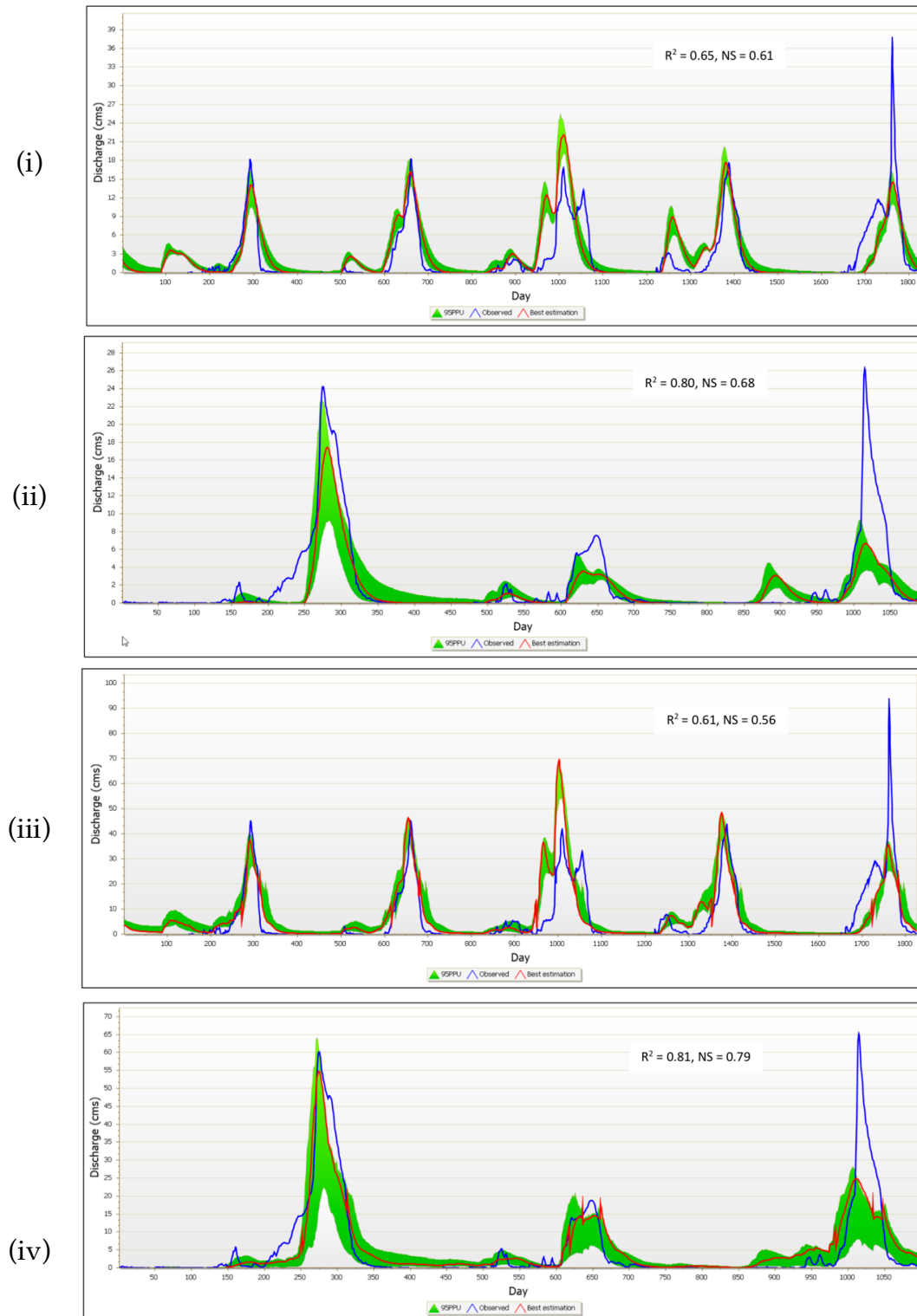


Fig 4: (i) Calibration curve for sub-basin 1 (2006-2010); (ii) validation curve for sub-basin 1 (2011-2013); (iii) calibration curve for sub-basin 2 (2006-2010); (iv) validation curve for sub-basin 2 (2011-2013); green is prediction uncertainty, blue is observed discharge, red is estimated discharge.

239 microscopic examination easier [15,16]. Within the fish state variable is a fish
240 demography model that captures the small window of time in the first few months of a
241 fish's life when it is susceptible to cercarial infection (before its scales harden and it
242 becomes more resistant). This model assumes a maximum fish lifespan of 4 years before
243 either being caught or natural death.

244

245 Using this approach, the fish (F) differential equations (4) and (5) solve for the
246 number of infected fish in a cluster, whereas the human (H , equation 1), reservoir host
247 (R , equation 2), and snail (S , equation 3) differential equations solve for infection
248 prevalence as in the single-village model. The egg inputs come from the Chi River and
249 the two upstream watersheds and affect the snail state variable, contributing to the
250 force of infection in that linkage; these time-varying parameters are derived on a
251 monthly basis from the rainfall-runoff model. Other egg inputs from open defecation
252 and disposal of septic tank sludge are not modeled due to lack of information about
253 where and when they occur. The egg inputs from overland flooding of the Chi River
254 were assumed to affect CCK and Lawa villages equally and were calculated by using
255 flow measurements from the river and multiplying by a scalar to relate the impact of
256 that water source with the upstream sub-basins. The first and second upstream sub-
257 basins' outflow were modeled to contribute eggs to the systems in BT and NNK
258 exclusively and multiplied by their own scalars to translate those flows into
259 contributions to human and reservoir host infection.

260

261

$$\frac{dI_{H,i}}{dt} = \beta_{FH,i} (1 - I_{H,i}(t)) \sum_{j=1}^n r_{FH_{i,j}}(t) * \gamma \sum_{k=1}^3 I_{F_{j,k}}(t) - (\mu_H + \alpha_{PZQ}(t)) I_{H,i}(t) \quad (1)$$

$$\frac{dI_{R,i}}{dt} = \beta_{FR,i} (1 - I_{R,i}(t)) \sum_{j=1}^n r_{FR_{i,j}}(t) * \gamma \sum_{k=1}^3 I_{F_{j,k}}(t) - (\mu_R + \alpha_{PZQ,R}(t)) I_{R,i}(t) \quad (2)$$

$$\frac{dI_{S,i}}{dt} = \beta_{HS,i} (1 - I_{S,i}(t)) \left(\sum_{j=1}^n r_{HS_{i,j}} I_{H,j}(t) N_{H,j}(t) + \sum_{j=1}^n r_{RS_{i,j}} I_{R,j}(t) N_{R,j}(t) \right) - \mu_S I_{S,i}(t) \quad (3)$$

$$\frac{dI_{F,i}^0}{dt} = \beta_{SF,i} (N_{F,i}^0(t) - I_{F,i}^0(t)) \lambda(t) \sum_{j=1}^n r_{SF_{i,j}} I_{S,j}(t) N_{S,j}(t) - \mu_F I_{F,i}^0(t) \quad (4)$$

$$\frac{dI_{F,i}^{1-3}}{dt} = -(\gamma + \mu_F) I_{F,i}^{1-3}(t) \quad (5)$$

262 Connectivity rates between patches were varied on a monthly basis between
 263 2008 and 2016 based on the hydrology model results to account for snail and fish
 264 mobility due to hydrologic connectivity. These connectivity parameters were derived
 265 by assessing the fate and transport of parcels of water in a village cluster's area and
 266 what proportions reached other village clusters in the Lawa Lake system. These r
 267 parameters are unitless and vary from 0 to 1, describing the proportion of each village
 268 cluster's force of infection for that host stage that affects each other village. Fig 5 shows
 269 examples of how these connectivity parameters varied by village pairing, month, and
 270 year (Figs 5-8 were produced in R using the ggplot2 package [17,18]).

271

272 The parameters carried over from the single-village model are listed in Table S1.
273 The β values are transmission parameters and are village cluster-specific (Table S2). The
274 r connectivity parameters are specific to each patch-to-patch relationship (Fig 1), γ is the
275 fish catch rate describing the fraction of the total fish population caught at each time
276 step, $\lambda(t)$ is a gating function to control fish birth, death, and aging processes, μ are
277 mortality rates, and α_{PZQ} are PZQ treatment events.
278

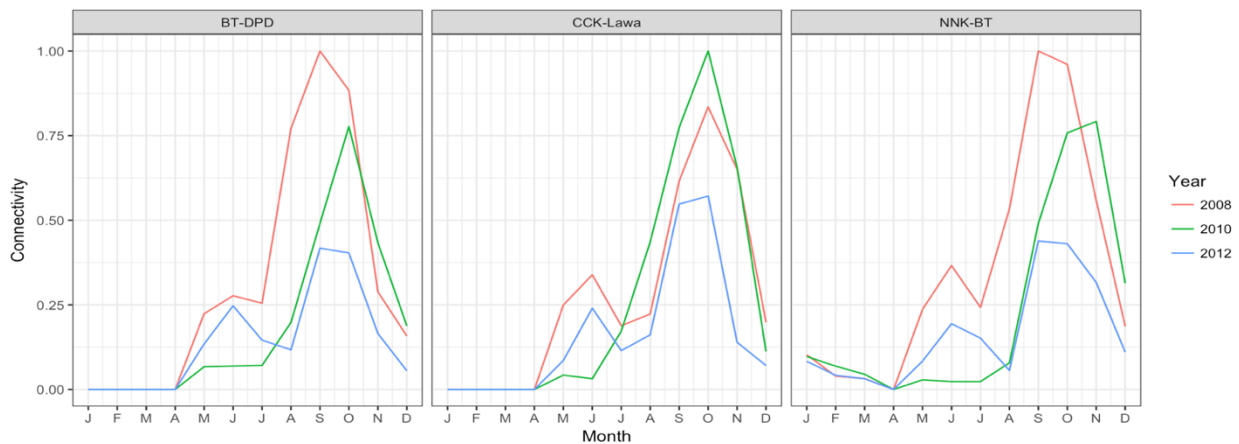


Fig 5: Village connectivity for BT-DPD, CCK-Lawa, and NNK-BT in 2008, 2010, and 2012

279 While a daily time scale could be preferable for assessing hydrologic impact,
280 historical data only captured month-to-month variability. Because human infection
281 survey data only measures prevalence and not incidence, this time scale is reasonable
282 for this study. From the hydrology model, the extent of water surface area at a suitable
283 depth (under 0.3m) for the contact events between juvenile cyprinid fish and the aquatic
284 snail intermediate hosts (“hotspots” for infection transfer) were used and to estimate N_s
285 and N_r . The transport time for a parcel of water between each village cluster was
286 employed to estimate the time scale of movement between the locations, but these
287 interactions happened on the order of days and not months and therefore the model’s

288 time scale was not granular enough to introduce a time lag component. There was
289 insufficient data to track fish mobility between patches, so fishermen's movement data
290 from [19] was used to inform the exchange of fish in inter-village commerce as
291 contributing to human infection from other village clusters. Table 2 summarizes these
292 linkages and use of hydrology-related data in the disease transmission model. Fig 5
293 shows examples of the variability between village connectivity across months and years
294 for the human/reservoir-to-snail and snail-to-fish transmission processes (the
295 remaining connectivity parameter graphs are shown in the Appendix). The general
296 trends persist from year to year, but the timing of peaks differ and affect village
297 connections differently. The year 2008 produced stronger connectivity for BT-DPD and
298 NNK-BT, while 2010 had stronger connectivity for CCK-Lawa.

299

Table 2: Description of linkages between hydrology model and disease transmission model

Hydrology-related transmission impact	Quantification method	Retained in model?
Egg inputs from overland flooding	Rainfall-runoff model output from Chi River summarized on monthly basis	Yes
Egg inputs from upstream water basins	Sub-basin model output summarized on monthly basis	Yes
Snail and fish mobility	Snail: Patch connectivity from rainfall-runoff model; Fish: fishermen catch data	Yes
Snail and fish available habitat	Snail: Rainfall-runoff model output and GIS analysis; Fish: N/A	Yes
Hotspots for snail-to-fish contact	Rainfall-runoff model output and GIS analysis	No

300

301

302 Results

303 Fig 6 shows the metapopulation model results for the six village clusters in the
304 base scenario with the metapopulation model establishing relative connectivity between
305 villages. Prevalence data points are included, distinguishing between the more sensitive
306 FECT surveys and the less sensitive Kato method surveys. The heterogeneity of
307 outcomes reflects the data: some villages saw reductions in infection prevalence to less
308 than 10% (KSR, DPD), yet a few villages continued to have predicted prevalence values
309 greater than 20% (BT and NNK). The steep drops in the graph were treatment events,
310 when a subgroup of villagers was tested for infection and given praziquantel if they
311 tested positive (the model assumes 100% drug efficacy). Model simulations were run for
312 eight years between 2008 and 2016.

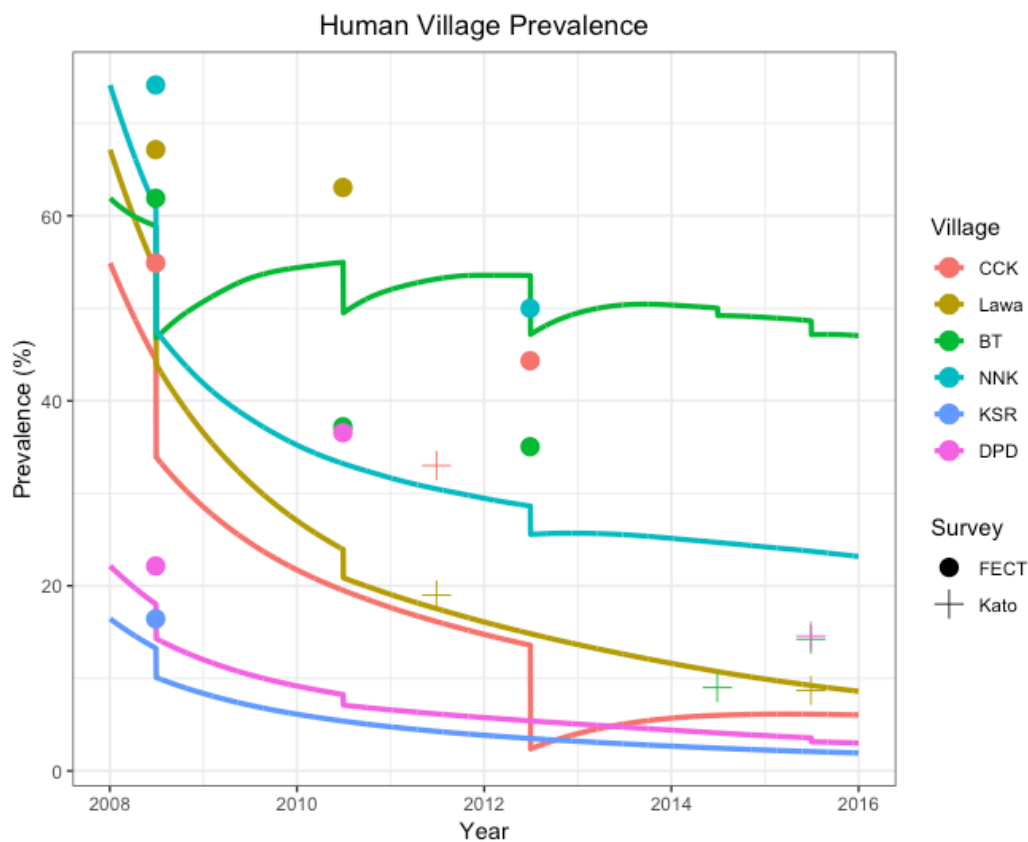


Fig 6: Metapopulation model run for human infection prevalence in villages around Lawa Lake

313 Figs 7 and 8 show infection prevalence values for intermediate snail and fish
314 hosts. In snails, the prevalence cycles seasonally with most patch populations
315 decreasing to 0.1% from initial values of 0.2% (with the exception of Lawa where
316 prevalence approaches 0%). NNK has the highest final snail prevalence value at 0.18%,
317 which is well within the range of what would be expected in this type of environment.
318 For fish prevalence, because the initial conditions are disparate and based on baseline
319 survey data, the model behavior is quite different. There is a seasonal aspect to their
320 dynamics though this is dampened for most villages except NNK, where it is readily
321 apparent. The end prevalence values range from 8-41%, with CCK, Lawa, BT, and DPD
322 having the lowest values and NNK having the highest.

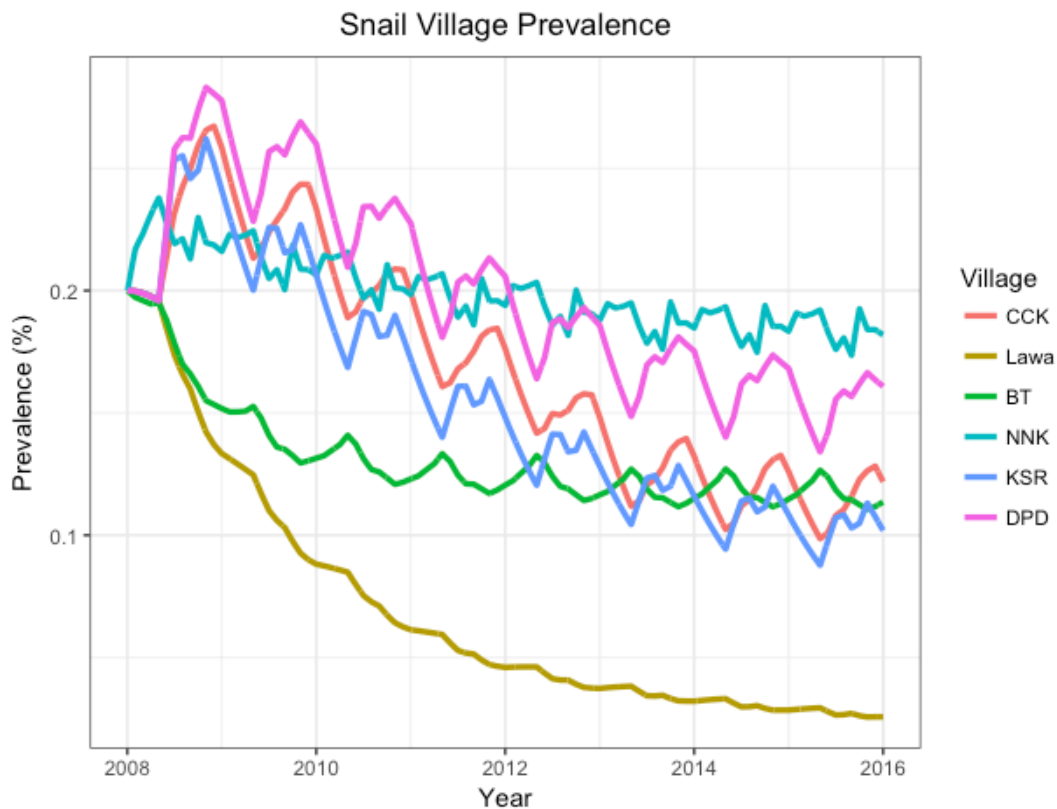


Fig 7: Snail prevalence values for metapopulation model

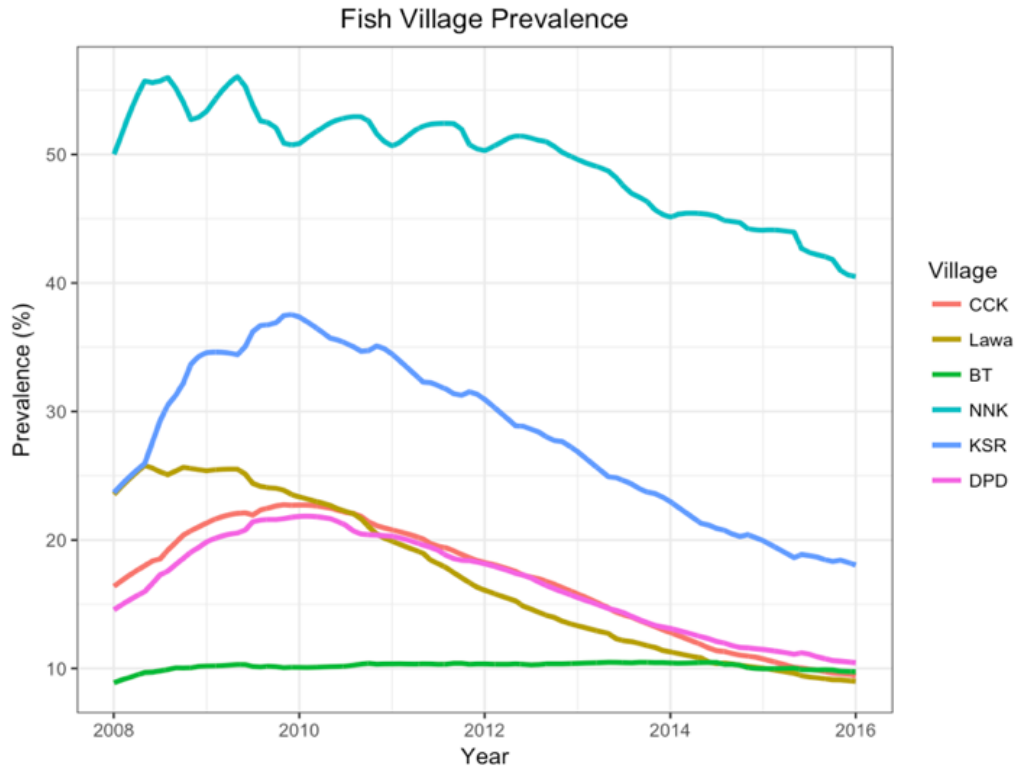


Fig 8: Fish prevalence values for metapopulation model

324

325 Discussion

326 Compared with the single-village model, the metapopulation model no longer
327 over-predicts final infection prevalence for the villages in 2016. Instead, the model now
328 occasionally under-predicts prevalence for some data points, notably, CCK in 2012. This
329 specific issue can be partially accounted for by the number of treatment-naïve
330 individuals included in the 2012 infection surveys. Interpretation of the data leads to
331 some speculation about the meaning of the surveys and their different diagnostic
332 methods. BT stands out as the modeled village with the least improvement (though it
333 overpredicts the 2012 and 2014 data). According to control program managers, BT
334 villagers were the least participatory in the Lawa Project and may therefore have reaped
335 the least benefit from the control program. Given that this mathematical model is

336 primarily concerned with infectious individuals actively shedding parasites into the
337 environment rather than asymptomatic cases, the Kato-based surveys from 2011, 2014,
338 and 2015 may actually capture the most infectious and relevant individuals in the
339 system and therefore be useful for thinking about infection prevalence patterns.
340 However, fitting all of the data points accurately would be nearly impossible since most
341 villages show non-monotonic patterns, and infection burden builds up over time.

342

343 Two major questions are, given the discrepancy between model fits and the data,
344 whether the data accurately reflects the reality of the disease transmission situation and
345 should the model be tuned with yet more parameters to get a close fit to the data. While
346 this dataset is relatively complete and informative by the standards of NTDs, it still
347 lacks enough time points, consistency in collection methods, and large enough sample
348 proportions of the populations to give a detailed picture of the *O. viverrini* infection
349 situation around Lawa Lake. The data (Table 1) show major swings across time points
350 and discrepancies between the survey methods. Therefore, while the FECT data was
351 used to fit via MCMC the transmission parameters in the single-village model, no
352 parameters were fit for the metapopulation model because of the increase in model
353 complexity and number of parameters (eighteen transmission parameters alone
354 compared with three per single-village model, plus seventy-two time-varying
355 connectivity parameters between the six village clusters). For this reason, the
356 connectivity parameters were derived exclusively from hydrology submodel results.
357 The metapopulation model is more believable than the single-village model in part
358 because of its inclusion of external hydrologic influences and connectivity and in part
359 because the modeled behavior is more nuanced in the clustered patterns of village
360 prevalence change it exhibits. The lack of parameters fitted to this model structure and

361 the higher quality data that informed the hydrology model lend a realism to the
362 underlying mechanics of the metapopulation model that improves upon the relatively
363 straightforward transmission framework of the single-village model. The
364 metapopulation model is meant to capture patterns of transmission rather than exact
365 fits to the data. Nonetheless, the lack of information about differences in snail infection
366 and raw fish eating patterns between villages remains a weakness, as they could not be
367 incorporated into the model. Better data on these aspects of the transmission cycle
368 would further strengthen the modeling framework and bring it into greater alignment
369 with reality.

370

371 The patterns observed in these results support the sorting of the six village
372 clusters into three geographical clusters that exhibit different patterns and trends based
373 on human prevalence values. The first cluster, consisting of CCK and Lawa, is in close
374 proximity to the Chi River and is most susceptible to overland flooding experienced
375 during the rainy season. Its villages had high human prevalence values at the start of
376 the control program, which decreased sharply during the period of treatment and
377 control activity. These villages were the headquarters and major focal area of Lawa
378 Project activities, suggesting that they benefited the most from health education and
379 health volunteer engagement. The model is able to accurately account for the decrease
380 in prevalence without making any assumptions about reinfection. Additionally, these
381 villages are in close proximity to heavily fished waters in the lake, as supported by GPS
382 evidence and interviews described in [19].

383

384 The second cluster consists of BT and NNK, the villages to the south and
385 southeast of the lake. These village also had high prevalence values at the start of the

386 treatment and control program but experienced much more modest decreases when
387 comparing data from later infection surveys. These villages are most impacted by
388 upstream watersheds draining into Lawa Lake (Fig 2.2), where some villages still have
389 over 50% *O. viverrini* infection prevalence (B. Sripa, unpublished data). Consequently, if
390 open defecation or unsafe disposal of human or reservoir waste is occurring in these
391 watersheds, the runoff will disperse parasite eggs into the canals, ponds, and sections of
392 the lake in close proximity to the second cluster's villages. These villages are also
393 adjacent to the highest concentration of fish ponds in the system and were not a focus of
394 major emphasis for Lawa Project activities.

395

396 The third cluster is KSR and DPD, which are located to the northeast of Lawa
397 Lake. At the start of the control program, these villages had significantly lower
398 prevalence values, which may be attributable to a lower degree of baseline
399 environmental contamination. These villages were significantly affected by neither the
400 Chi River nor the upstream watersheds, so they experienced fewer external inputs of
401 infectious individuals or waterborne forms of the parasite into their local systems.
402 These villages were not adjacent to high concentrations of fish ponds or fishing activity
403 in their local waters and received less emphasis by the Lawa Project compared to the
404 first cluster.

405

406 Because of the low prevalence of snail infection, the lack of field survey data, and
407 the difficulty measuring snail prevalence precisely, strong claims cannot be made about
408 the model results for the snail intermediate host. However, most field surveys indicate
409 that snail prevalence in this region ranges between 0 and 0.2%, agreeing with the model
410 results [13]. With current diagnostic methods, differentiating between the clusters

411 would require surveys of tens of thousands of snails at least. In recent years, the
412 highest snail prevalence values found are still less than 10% [20]. Much like other snail-
413 borne diseases such as schistosomiasis though, only a few snails are required to
414 maintain transmission in an area because of the high number of cercariae they shed into
415 water bodies. Further understanding of where snails are most likely to be infected will
416 help with environmental measurement and control. Bottlenecks of water flow, such as
417 certain ditches and culverts, would concentrate fecal waste and parasite eggs and be
418 zones of likely contact with susceptible snail hosts. Eliminating or protecting these areas
419 could be an effective method of environmental control subject to proper coverage in the
420 environment of interest and patch/cluster connectivity.

421

422 Considering fish, infection prevalence is much higher and therefore it is easier to
423 discern differences between clusters. Because transmission is foodborne, we are
424 interested in the supply chain of food to consumers, which is not necessarily related to
425 proximity between where fish lived and where they are eaten. Fishermen, middlemen,
426 and merchants are all mobile and may choose to sell and distribute fish to other villages
427 to expand their market. In the model simulation, the relative ordering of low to high
428 fish prevalence values in fish hosts differ considerably from the results in humans, with
429 NNK and KSR having the highest infection prevalence. NNK's is driven largely by the
430 initial value, but KSR's is driven by dynamics, as its outcome is quite different from
431 Lawa's, which started with a similar prevalence level. Fish are infected by having
432 infectious snails releasing cercariae into areas with juvenile fish, so KSR is the cluster
433 with the greatest magnitude of this process taking place. Paying attention to fish
434 prevalence results and how they interact with patterns of fish commerce can help
435 identify where to target health education interventions related to cooking.

436

437 The model and the data that inform the model have limitations due to their
438 fragmentary nature. Other model assumptions ignore the impact of different parasite
439 burden levels in hosts and the age structure of human populations. The hydrology
440 model was calibrated and validated against available data, leading to a plausible
441 interpretation of the observed hydrologic behavior of the system. However, it could not
442 account for very local effects that eluded its time and spatial scale and may have
443 relevance for the points of contact between stages of the life cycle. The infection survey
444 data may not be representative of the host populations because of sampling bias, but its
445 overall spatial and temporal patterns align with local understanding and experience.

446

447 **Conclusion**

448 This work highlights a major driver of persistent *O. viverrini* infection in
449 northeast Thailand: a disturbed and dynamic hydrologic environment that mediates
450 parasite transmission between connected village clusters and environments. This was
451 accomplished by discussing and modeling five different means of hydrology
452 influencing parasite transmission and showing that its effects are significant and
453 complex, acting heterogeneously across the Lawa Lake system. While local infection
454 and contamination may be the main factor driving transmission at higher prevalence
455 levels, as prevalence declines and villages move toward elimination connectivity will
456 play a bigger role in maintaining the transmission cycle and preventing local
457 elimination of the parasite.

458

459 The connectivity between water bodies and villages ensures that elimination of
460 local infection is not possible without addressing upstream and adjacent environments.

461 If infectious fecal waste from human and reservoir hosts is continually flushed
462 downstream and the source is not treated, transmission will be restarted even if
463 previously disrupted. This argument supports an approach that first targets villages
464 and populations near headwaters and then proceeds further downstream while still
465 accounting for human mobility and migration that could reintroduce infection into a
466 previously cleared environment. Flood pulses and overland flooding also risk moving
467 parasites into isolated and disconnected environmental patches on an annual basis,
468 which requires constant treatment and attention to infection status of individuals in
469 those patches. Snail and fish mobility remain little understood but have major relevance
470 for *O. viverrini*'s life cycle, specifically how infection propagates in the environment.
471 Targeting the locations where snails and juvenile fish come into close proximity with
472 each other could be a promising environmental control technique but requires greater
473 knowledge about the conditions that enable this transmission process.

474

475 A linked disease transmission-hydrologic modeling approach was employed
476 here that uses hydrology model outputs as time-varying inputs in the disease
477 transmission model to account for seasonal effects of flooding and water movement
478 relevant to the intermediate hosts and waterborne forms of *O. viverrini*. Based on model
479 results, village clusters were grouped into three disease prevalence curve patterns based
480 on presence/absence of upstream and flooding impacts and history of control program
481 intensiveness. Considering these findings, we argue for the use of this modeling
482 approach and its results to inform environmental control of *O. viverrini* and for the need
483 for environmental surveillance. While the specifics of the hydrology, population
484 structure, and pathogen transmission cycle are local and specific in nature, this

485 approach can be replicated across a variety of disease systems that are impacted by
486 seasonally dynamic water movement.

487 **References**

- 488 [1] Sithithaworn P, Pipitgool V, Srisawangwong T, Elkins DB, Haswell-Elkins MR.
489 Seasonal variation of *Opisthorchis viverrini* infection in cyprinoid fish in north-east
490 Thailand: implications for parasite control and food safety. *Bulletin of the World Health*
491 *Organization*. 1997;75(2):125.
- 492 [2] Wang YC, Yuen R, Feng CC, Sithithaworn P, Kim IH. Assessing the role of
493 landscape connectivity on *Opisthorchis viverrini* transmission dynamics. *Parasitology*
494 *international*. 2017 Aug 1;66(4):402-12.
- 495 [3] Grundy-Warr C, Andrews RH, Sithithaworn P, Petney TN, Sripa B, Laithavewat L,
496 Ziegler AD. Raw attitudes, wetland cultures, life-cycles: socio-cultural dynamics
497 relating to *Opisthorchis viverrini* in the Mekong Basin. *Parasitology International*. 2012
498 Mar 1;61(1):65-70.
- 499 [4] Aunpromma S, Tangkawattana P, Papirom P, Kanjampa P, Tesana S, Sripa B,
500 Tangkawattana S. High prevalence of *Opisthorchis viverrini* infection in reservoir hosts
501 in four districts of Khon Kaen Province, an opisthorchiasis endemic area of Thailand.
502 *Parasitology international*. 2012 Mar 1;61(1):60-4.
- 503 [5] Haas W, Granzer M, Brockelman CR. *Opisthorchis viverrini*: finding and recognition
504 of the fish host by the cercariae. *Experimental Parasitology*. 1990 Nov 1;71(4):422-31.
- 505 [6] Krishnamurthy D, Katsikis G, Bhargava A, Prakash M. *Schistosoma mansoni*
506 cercariae swim efficiently by exploiting an elasto-hydrodynamic coupling. *Nature*
507 *Physics*. 2017 Mar;13(3):266.
- 508 [7] Remais J, Liang S, Spear RC. Coupling hydrologic and infectious disease models to
509 explain regional differences in schistosomiasis transmission in southwestern China.
510 *Environmental Science & Technology*. 2008 Feb 20;42(7):2643-9.

- 511 [8] Rinaldo A, Bertuzzo E, Mari L, Righetto L, Blokesch M, Gatto M, Casagrandi R,
512 Murray M, Vesenbeckh SM, Rodriguez-Iturbe I. Reassessment of the 2010–2011 Haiti
513 cholera outbreak and rainfall-driven multiseason projections. PNAS. 2012 Apr
514 24;109(17):6602-7.
- 515 [9] Perez-Saez J, Mande T, Ceperley N, Bertuzzo E, Mari L, Gatto M, Rinaldo A.
516 Hydrology and density feedbacks control the ecology of intermediate hosts of
517 schistosomiasis across habitats in seasonal climates. PNAS. 2016 May 5:201602251.
- 518 [10] Tompkins AM, Ermert V. A regional-scale, high resolution dynamical malaria
519 model that accounts for population density, climate and surface hydrology. Malaria
520 Journal. 2013 Dec;12(1):65.
- 521 [11] Ziegler AD, Petney TN, Grundy-Warr C, Andrews RH, Baird IG, Wasson RJ,
522 Sithithaworn P. Dams and disease triggers on the lower Mekong river. PLoS NTDs.
523 2013 Jun 13;7(6):e2166.
- 524 [12] León TM, Porco TC, Kim CS, Kaewkes S, Kaewkes W, Sripa B, Spear RC. Modeling
525 liver fluke transmission in northeast Thailand: Impacts of development, hydrology, and
526 control. Acta Tropica. 2018 Dec;188:101-7.
- 527 [13] Kim CS, Echaubard P, Suwannatrai A, Kaewkes S, Wilcox BA, Sripa B. Seasonal
528 and spatial environmental influence on *Opisthorchis viverrini* intermediate hosts,
529 abundance, and distribution: insights on transmission dynamics and sustainable
530 control. PLoS NTDs. 2016 Nov 23;10(11):e0005121.
- 531 [14] Computational Hydraulics International [CHI] (2018). PCSWMM Modeling
532 Software. Available at: <https://www.pcswmm.com>
- 533 [15] Laoprom N, Laithavewat L, Kopolrat K, Kiatsopit N, Kaewkes S, Chantalux S,
534 Mongkolsin C, Chanmaha B, Andrews RH, Petney TN, Sithithaworn P. Evaluation of a

535 commercial stool concentrator kit compared to direct smear and formalin-ethyl acetate
536 concentration methods for diagnosis of parasitic infection with special reference to
537 *Opisthorchis viverrini sensu lato* in Thailand. The Southeast Asian Journal of Tropical
538 Medicine and Public Health. 2016 Sep;47(5):890-900.

539 [16] Saijuntha W, Duengngai K, Tangkawattana S, Petney TN, Andrews RH, Sithithaworn
540 P. Recent Advances in the Diagnosis and Detection of *Opisthorchis viverrini* Sensu Lato
541 in Human and Intermediate Hosts for Use in Control and Elimination Programs. In
542 Advances in Parasitology. 2018 (Vol. 101, pp. 177-214).

543 [17] R Core Team. R: A language and environment for statistical computing. R
544 Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
545 2018.

546 [18] Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer; 2016.

547 [19] Kim CS, Smith JF, Suwannatrai A, Echaubard P, Wilcox B, Kaewkes S, Sithithaworn
548 P, Sripa B. Role of socio-cultural and economic factors in cyprinid fish distribution
549 networks and consumption in Lawa Lake region, Northeast Thailand: Novel
550 perspectives on *Opisthorchis viverrini* transmission dynamics. Acta Tropica. 2017
551 Jun;170:85-94.

552 [20] Kiatsopit N, Sithithaworn P, Saijuntha W, Boonmars T, Tesana S, Sithithaworn J,
553 Petney TN, Andrews RH. Exceptionally high prevalence of infection of *Bithynia*
554 *siamensis* goniomphalos with *Opisthorchis viverrini* cercariae in different wetlands in
555 Thailand and Lao PDR. The American Journal of Tropical Medicine and Hygiene. 2012
556 Mar;86(3):464-9.

557 **Supporting information**

Table S1: Parameter values for single-village model

Parameter	Value	Units	Source	Symbol
Natural mortality of snails	1.37E-03	per day	Kruatrachue et al. 1982	μ_S
Parasite dependent mortality of snails	1.37E-03	per day	Chanawong & Waikagul 1991	α_S
Mortality of fish	6.85E-04	per day	Suvarnaraksha et al. 2011	μ_F
Parasite dependent mortality of fish	0	per day	Assumption (unstudied)	α_F
Natural mortality of humans	3.69E-05	per day	CIA 2015 (Factbook)	μ_H
Human infection clearance by praziquantel	Variable	Episodic	Treatment data from clinics	α_{PZQ}
Mortality of reservoir host	2.74E-04	per day	Local interview data	μ_R
Reservoir infection clearance by praziquantel	Variable	Episodic	Cat/dog treatment data from veterinarians	$\alpha_{PZQ,R}$
Transmission parameters (fish-to-human, fish-to-reservoir, human-to-snail, snail-to-fish)	See Table 3.5.1	per day per infectious host/worm	Equilibrium conditions and MCMC	$\beta_{FH}\beta_{FR}, \beta_{HS}, \beta_{SF}$
Fish population	3000	fish	Estimate	N_F
Snail population	30000	snails	Estimate	N_S
Human population	Variable	humans	Village censuses	N_H
Cat and dog population	100	reservoir hosts	Estimate from village censuses	N_R
Mean human worm count	Variable	worms	Infection survey data	W_H
Mean reservoir worm count	Variable	worms	Infection survey data	W_R

558

Table S2: Beta transmission parameters for single village model

Equilibrium	CCK	Lawa	BT	NNK	KSR	DPD
Fish to Human	2.50E-10	7.24E-11	2.53E-10	1.81E-11	3.32E-07	2.68E-07
Human to Snail	8.90E-09	1.01E-08	4.92E-09	8.92E-09	6.56E-12	1.18E-11
Snail to Fish	7.34E-06	7.10E-06	3.40E-06	8.96E-06	1.67E-05	1.35E-05
MCMC	CCK	Lawa	BT	NNK	KSR	DPD
Fish to Human	2.95E-07	3.45E-07	7.26E-07	2.28E-07	3.30E-08	7.75E-08
Human to Snail	1.25E-08	2.53E-09	3.45E-09	2.26E-08	1.28E-08	1.48E-08
Snail to Fish	4.48E-06	7.03E-06	2.23E-06	2.28E-05	7.08E-06	3.89E-06

559