Environmental determinants of *Oncomelania hupensis* snail occurrence along irrigation ditch networks in Sichuan Province, China

Eyal Matalon

Abstract

The *Oncomelania hupensis* snail transmits the infectious parasite *Schistosoma japonicum,* which, at the end of the 20th century, was estimated to infect 865,000 total people along the Yangtze River Basin. Whereas Geographic Information Systems and Remote Sensing have been indispensable tools for modeling the habitat of this disease vector, most studies are limited by the low spatial resolution of remotely-sensed environmental data and have focused on identifying potential snail habitat on a regional scale. There is limited understanding of what influences snail occurrence at the level of human communities. In this paper, the association between snail occurrence and landscape features at varying spatial scales derived from field measurements and remote sensing was explored using logistic and negative binomial regression. Water conditions (particularly the flow rate and the existence of a thin film of water) were the most important small-scale ecological variables affecting snail occurrence. Percent forest cover and crop cover were the most significant macro-scale landscape features associated with snail occurrence. This paper discusses the possible underlying ecological mechanisms for these associations and suggests directions for future study at the village level.

Introduction

The *Oncomelania hupensis* snail transmits the parasite responsible for schistosomasis in China, *Schistosoma japonicum.* At the end of the 20^{th} century, schistosomiasis was estimated to infect 865,000 people along the Yangtze River Basin (Chen and Feng 1999). Surveys taken in the early 1990s found infected human subjects and *O. hupensis* snails in approximately 400 counties or cities in the region (Schistosomiasis Expert Advisory Committee 1993). The parasite is water-borne and endemic to plains regions, swamp and lake regions, and hilly and mountainous regions (Chen and Feng 1999). Because individuals in agricultural, rural areas have high levels of water contact (working along irrigation ditches) they are most susceptible to infection (Yang *et al.* 2008). It is estimated that 40 million people in the People's Republic of China are at risk (Chen and Feng 1999).

Propagation of the parasite depends on both snail and mammalian hosts. In the Yangtze River, two subspecies of the amphibious snail *O. hupensis* transmit the schistosome parasite (Davis *et al.* 1999). A free-swimming form of the parasite (miracidia) infects snails living in agricultural ditches. The infected snails release another free-swimming form of the parasite (cercariae) into the water network. Humans or domesticated animals then become infected through skin-water contact in agricultural ditches. Once inside the body, juvenile worms mature and lay eggs in the liver, causing a severe immune reaction in the host. Symptoms include a "progressive enlargement of the liver and spleen, intestinal damage, and hypertension of abdominal blood vessels" (WHO 2008). Schistosome eggs leave the human body in the stool, and the untreated waste is, in turn, used to fertilize agriculture fields through the ditch networks. Inside the ditches, schistosome eggs hatch into the miracidia that then infect the *Oncomelania* snails living along the ditches, completing the life cycle (Xu *et al.* 2006).

Because propagation of the disease is completely dependent on the presence of *O. hupensis*, it is important to understand the ecology of this intermediate host. As amphibious snails, adult *O. hupensis* are found at the interface between soil and water along the banks of irrigation ditches and swamps (Davis *et al.* 1999). Generally, *Oncomelania spp.* distributions are associated with agricultural land-uses, heavily vegetated soils rich in organic matter, and a wide range of temperatures and moisture levels (Seto *et al.* 2002). Understanding the relative importance of these ecological factors in the abundance and distribution of *O. hupensis* is critical to schistosomiasis control efforts, that can include focal mollusciciding to reduce snail densities

(Bergquist 2001). This strategy relies upon the ability to identify areas of high snail density in order to maximize the effectiveness of snail control while minimizing the environmental impacts of molluscicides (e.g., toxicity to fish that live in the ditches and in the fishponds that are connected to the ditch network) (Liang et al 2002). Because snail distributions are wide-spread across various habitats, detailed surveys to identify hot spots are labor-intensive and expensive. Therefore, spatial analysis is needed to further describe ecological determinants and predict the occurrence of *O. hupensis* at the relatively small scales needed to inform focal mollusciciding efforts.

Geographic Information Systems (GIS) and Remote Sensing (RS), two tools that allow for remote analysis of environmental and epidemiological factors, have been indispensable to understanding the ecology of disease (Clarke *et al.* 1996). RS tools have generated indices for vegetation, surface temperature, soil moisture, and precipitation, all of which have been useful for exploring spatial association between epidemiological and environmental data (Cringoli *et al.* 2005). Further, the development of GIS-based analytical regression techniques in recent years has been instrumental in modeling disease transmission (Jerrett *et al.* 2003).

Such methods have proved especially useful for modeling vector-borne disease transmission (Malone 2005). For example, epidemiologists have studied the associations between remotely sensed land-use changes and distributions of tsetse flies that serve as vectors for trypanosomiasis, a disease in vertebrates caused by parasitic trypanosomes (de la Rocque *et al.* 2005). Malone (2005) has integrated aerial photographs, soil maps, farm boundary maps and snail habitat survey data from southwest coastal Louisiana to describe relationships between soil temperature, water, and the transmission of fasciolosis in cattle by freshwater snail hosts. Ceccato et al. (2005) discuss the potential of similar integrated GIS-RS models to produce an early warning system for mosquito-borne malaria. Daniel et al. (2004) describe how mapping microclimatic factors such as temperature and humidity can, in turn, help map tick distributions and tick-borne disease occurrence. Because all of these diseases are associated with specific vectors or intermediate hosts, their epidemiologies lend themselves to georeferenced population surveys, subsequent GIS spatial regression methods, and modeling of vector habitats.

In the last decade, researchers have used GIS and RS to understand the ecology of *O. hupensis* and predict its occurrence along the Yangtze River basin. For example, Seto *et al*. (2002) used Landsat imagery of habitat and non-habitat sites to generate a predictive habitatranking statistic for different agricultural landscapes in the Anning River Valley of Sichuan Province. In a later study, Zhang *et al*. (2004) found that values of three RS indices (modified soil-adjusted vegetation, wetness, and land surface temperature) had strong associations with high snail densities. Similar indices were used in a later model that estimated potential *O. hupensis* habitats in the marshland around Poyang Lake (Guo *et al.* 2005). While these studies have been formative in modeling snail habitats at the macro-scale (national, provincial, or county level) Yang *et al.* (2008) argue that they are limited due to the low spatial resolution of remotely-sensed environmental data.

Because most prior studies have focused on describing snail habitat on a larger scale, there is limited understanding of what influences snail distributions at the level of human communities. Along the agricultural Yangtze River Basin, these are rural Chinese production groups–geographic entities that correspond to approximately 250 individuals who collectively farm an area of land approximately 0.25 km^2 in size. To improve finer-scale description of snail distribution, Seto *et al*. (2001) designed a protocol for geographically randomized snail surveys at the production group level. In a subsequent study, Xu *et al*. (2004) integrated snail survey data with land-cover, land-use, and elevation data from high-resolution IKONOS and ASTER satellite images to generate a predictive model of snail abundance for a 30-meter resolution grid. A later study by Spear *et al*. (2004) implemented the protocol in 20 production groups in Sichuan Province, China and showed that snails exhibit an overdispersed, clumped distribution along agricultural ditches. The authors suggest that this distribution is due to underlying ecological features of the irrigation ditch network. Descriptions of these features of *O. hupensis* habitat have been largely anecdotal (Davis 2008, Seto 2009 pers. comm.). It is generally known that snails are mainly present in areas where mud surfaces and emergent vegetation are covered by a film of slow-flowing water (Sturrock 2001).

Given the importance of small-scale factors in determining village-level *O. hupensis* distribution, it is necessary to verify this anecdotal evidence with empirical research. Further, it is important to explore the utility of RS/GIS-based techniques for modeling snail habitats at this scale by examining the relative importance of and relationships between remotely-sensed landscape features and smaller scale ecological conditions on the ground. Some RS/GIS-based studies at larger scales have considered associations between variables at different scales while others have neglected to mention them entirely. For example, Seto *et al.* (2002) suggest that the suitability of snail habitats within different land cover categories may have more to do with factors such as soil composition rather than variables that can be described using RS. Guo *et al.* (2005), on the other hand, forward three RS indices (green vegetation, elevation, and soil moisture) as proxies for suitable *O. hupensis* habitats without discussing any underlying ecological reasons or mechanisms. Whereas strictly GIS/RS-based studies generally preclude the examination of small-scale ecological variables in favor of lower-resolution environmental variables, an integration of ground survey data with remotely-sensed data can establish important links between environmental factors at larger and smaller scales (Seto 2009, pers. comm.). These links are especially important for the continued utility of RS/GIS in mapping and predicting *O. hupensis* distribution at the production group level.

This study assesses the relative importance of remotely sensed landscape features and small-scale ecological conditions in the distribution of *Oncomelania hupensis* at the production group scale. Specifically, given the overdispersed distribution of *O. hupensis* along agricultural ditch networks, the study determines (1) the association between conditions at the ditch and *O. hupensis* densities, (2) the association between remotely-sensed landscape features and *O. hupensis* occurrence, and (3) whether remotely sensed landscape features are useful proxies for suitable small-scale ecological conditions at the production group level. The hypotheses that higher-density sites along the ditch are associated with three ditch conditions: visible, low-flowing water, thick grass, and soil ditch construction; and three remotely-sensed landscape features: percent crop land cover, a wetness index, and slope will be tested. Furthermore, the hypothesis that crop land cover, built infrastructure land cover, the wetness index, and slope will serve as useful proxies for the grass thickness, concretized ditch construction, water level, and water flow at the ditch, respectively will be explored.

Methods

An observational study was conducted to understand how small-scale ecological conditions and large-scale landscape features affect snail distributions along irrigation ditch networks at the village level.

Study Region and Study Specimen

Snail density and satellite imagery data were collected for 13 production groups in Jinyang, county in Sichuan Province in the fall of 2007 (see Fig. 1). Land cover in production groups is primarily agricultural fields but manmade ponds, forests, and built infrastructure are also present within village boundaries. Each production group surveyed contains a network of irrigation ditches that transport water and nutrients to agricultural fields (see Fig. 2). This study investigated distributions of the snail subspecies *Oncomelania hupensis robertsoni* (Davis 1999). The two known subspecies, *O. h. hupensis* and *O. h. robertsoni*, are geographically separated by the Three Gorges along the Yangtze River and are therefore highly genetically and ecologically divergent (Wilke 2000). Whereas *O. h. hupensis* is primarily found along low-lying reservoir banks (Guo *et al.* 2005), *O. h. robertsoni* lives in the high plateaus, hills, and mountains of Yunnan and Sichuan Provinces (Davis 1999). As amphibious snails, *O. h. robertsoni* adults are found along the banks of irrigation ditch networks (Seto *et al.* 2002).

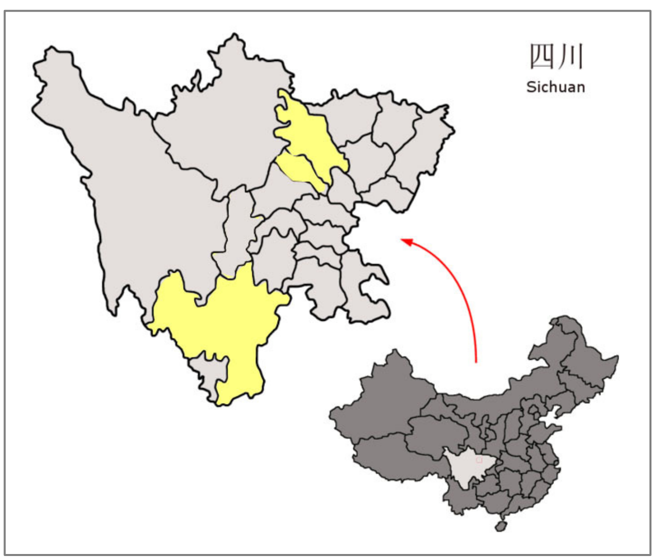

Figure 1: Jinyang counties in Sichuan Province, China (Courtesy of commons.wikimedia.org)

Figure 2: Irrigation Ditches in a Sichuan Production Group (Courtesy of ehs.sph.berkeley.edu/china)

Data Collection

Snail surveys were loosely based on the protocol outlined in Seto *et al.* (2001). Snails were collected and analyzed at 10-meter intervals along the entire ditch network of each production group by field staff of the Sichuan Institute of Parasitic Diseases. At each site, a *kuang* (square frame, 0.11 m^2) was placed on the bank of the ditch. After GPS coordinates were taken at the site (Trimble GeoExplorer, Trimble Navigation Limited, Sunnyvale, CA, USA), all adult snails within the *kuang* were collected into envelopes and labeled with a location ID. Various characteristics of the ditch were also observed and recorded at each site by trained observers: water conditions (wet, dry, or visible water), water flow (< 0.15 m/s or > 0.15 m/s), grass conditions (no grass, thin grass, or thick grass) and ditch construction (soil, brick, or concrete). Collected snails were crushed under a microscope to check for schsitosome infection, indicated by the presence of schistosome cercariae.

Two IKONOS images of the study site were obtained from November 30, 2002, a 1-meter resolution panchromatic image and a 4-meter true color image. These images were merged using the Pan-sharpening tool in ArcGIS (Version 9.0, ESRI, Redlands, CA, USA) to create a single 1-meter resolution true color image.

A supervised maximum likelihood classification was performed to classify the image into six land cover categories based on remotely-sensed spectral signatures. Based on visual interpretation done in collaboration with Dr. Edmund Seto, who has familiarity with the ground land cover in this region, 554 pixels were selected and identified as one of six land cover categories: built infrastructure (48 pixels), crop (65 pixels), bare soil/dry field (74 pixels), river (67 pixels), forest (258 pixels), and pond (52 pixels). The Maximum Likelihood Classification tool in the Spatial Analyst Extension of ArcMap (Version 9.0, ESRI, Redlands, CA, USA) classified all pixels in the image into the six land cover categories based on this 554-pixel training set. The same classification algorithm was applied to a validation set. For each category, pixels were selected (but not identified in advance) and input into the classification model (552 pixels total): built infrastructure (51 pixels), crop (47 pixels), bare soil/dry field (62 pixels), river (78 pixels), forest (284 pixels), and pond (30 pixels). Subsequent accuracy assessment was performed to generate the overall accuracy and Kappa coefficient of the Maximum Likelihood Classification.

Elevation data was obtained through NASA's Shuttle Range Topographic Mission (SRTM) at 90-meter resolution. Degrees slope were calculated using the Slope Tool in Spatial Analyst in ArcMap (Version, 9.0, ESRI, Redlands, CA, USA). Topographic wetness index was also derived from the SRTM image by using the Flow Direction and Flow Accumulation Tools from the Spatial Analyst toolbox and inputting the resulting raster file into the Raster Calculator of ArcMap (Version, 9.0, ESRI, Redlands, CA, USA):

$$\frac{Flow\ Accumulation\ (upstream\ pixels)}{Slope\ (degrees)} = Wetness\ (upstream\ pixels\ per\ degree\ slope)$$

To account for autocorrelation, the variable, Distance to Nearest Presence Site, was generated by performing a spatial join between all sampling sites and only sites in which snails were present using ArcGIS (Version, 9.0, ESRI, Redlands, CA, USA).

Lastly, a Shannon's evenness index was calculated using the results of Zonal Statistics analysis for the six land cover categories (see next section).

Summarizing Environmental Data in Proximity to Sampling Sites

To understand whether snail densities might be associated with the above remotely-sensed landscape features, 50-meter buffers were generated around all snail sampling sites using the Buffer tool in Analysis Tools in ArcMap (ESRI, Redlands, CA) (see Figure 3). Land cover, elevation, wetness, slope, and elevation data were then imported into ArcMap and projected onto a single coordinate system (WGS84 UTM Zone 48N).

Zonal Statistics ++ in HawthsTools (Beyer 2004) were used to summarize the SRTM elevation, slope, and wetness raster data within each snail-site buffer. Average elevation meters, slope degrees, wetness index values were calculated for each snail buffer. The single land cover raster layer was converted into six distinct raster layers using Raster Calculator in the Spatial Analyst Extension of ArcMap. For each new raster layer, ones and zeros indicated whether given land cover types were present or absent, respectively. Zonal Statistics ++ in HawthsTools were then used to summarize land cover data within the snail-site buffers. The number of pixels of each land cover category was calculated for each snail buffer.

To calculate the evenness index for each snail-site buffer, the results of the Zonal Statistics analysis for land cover were input into the following equation using Microsoft Excel 2007 (Microsoft Corporation, Seattle, WA, USA)

$$H' = -\sum_{i=1}^{S} p_i \ln p_i, \text{ where } p_i = \frac{n_i}{N}$$

Where:

p_i is the proportion of pixels of land cover i within a snail-site buffer.

n_i is the number of land cover pixels of class i in a snail-site buffer.

N is the total number of pixels in a snail-site buffer.

S is the total number of land cover categories (in this case, six)

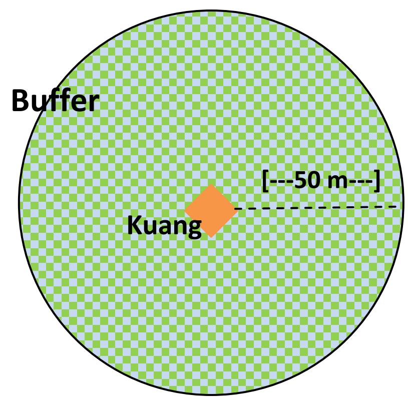

Figure 3: Diagram of a 50-meter buffer around a snail sampling site (kuang). Landscape features such as elevation, slope, wetness, and land cover were summarized within 50 meters of each snail sampling sites using the Zonal Statistics tool in ArcGIS (ESRI, Redlands, CA)

Statistical Data Analysis

Because snail count data gathered in this study was zero-inflated (i.e. has a modal value of zero), inputting data into a standard linear regression model resulted in residuals that were neither normally distributed nor homogenous in variance (two important assumptions of linear regression). Previous studies have stressed that zero-inflated data is better fit by negative binomial distributions than by normal distributions (White and Bennetts 1996). Although there are numerous methods for modeling associations between zero-inflated count data and environmental data (see, for example, Agarwal *et al.* 2002, Cunningham and Lindenmayer 2005, Sileshi 2008) a review by Martin *et al.* (2005) stresses that negative binomial regression, available in most statistical software packages, models such data with sufficient accuracy. Small- and large-scale variables were regressed against continuous snail count data using stepwise negative binomial regression models in Stata/IC (Version 8.1, Stata Corporation, College Station, TX, USA). Significant determinants of snail counts were identified by assessing the p-values of their regression coefficients in both univariate and two-variable models. The two-variable models with the least negative log likelihood values were identified as the most suitable models of snail counts for both scales of analysis.

Additionally, continuous snail count data was divided into binomial "presence" / "absence" data. Logistic regression was performed to test whether there was any significant difference presence and absence sites for any of the above small- and large-scale environmental variables. An odds ratio, or the ratio between the odds of snail presence and the odds of snail absence, was calculated for each environmental variable.

Lastly, pair-wise correlation coefficients between four large-scale and four small-scale variables were calculated to verify the hypothesis that certain landscape features serve as useful proxies for certain environmental conditions at the ditch.

Results

Data Collection

Out of a total of 2525 sampling sites, 48 sites had snails present. The mean number of snails per sampling site was 0.063 with a standard deviation of 0.878. Counts at sites in which snails were present ranged from 1 to 37 snails. For a statistical summary of snail count data for each production group, see Table 1.

Table 1: Summary statistics of stratified snail surveys for 13 villages in Jinyang County, Sichuan Province. Snail count data was zero-inflated (modal value of 0, = .063, SD = 0.833) and statistical analysis was performed on the entire data set.

Village	# sites	mean snail count	SD snail count	# presence sites
Gaohuai 3	227	0.211	0.810	22
Gaohuai 7	157	0.166	1.25	8
Gongqiao 2	280	0	0	0
Gongqiao 3	143	0.042	0.426	2
Shuiku 4	340	0	0	0
Guihua 8	233	0.004	0.065	1
Shiban 1	977	0.035	0.993	4
Longfeng 2	140	0	0	0
Longfeng 4	155	0	0	0
Longfeng 6	348	0.006	0.076	2
Dashu 4	357	0.036	0.310	6
Dashu 5	137	0.058	0.601	2
Dongchao 5	145	0.021	0.248	1
Total (within IKONOS image)	2525	0.063	0.878	48

Of the sampling sites for which snail and small-scale ecological data were collected, 1,114 corresponded to sites on terraces (away from irrigation ditches) or to sites that did not correspond to any remotely-sensed data when overlaid over the IKONOS imagery (see Fig. 4). These 1,114 sites were not included in the analysis.

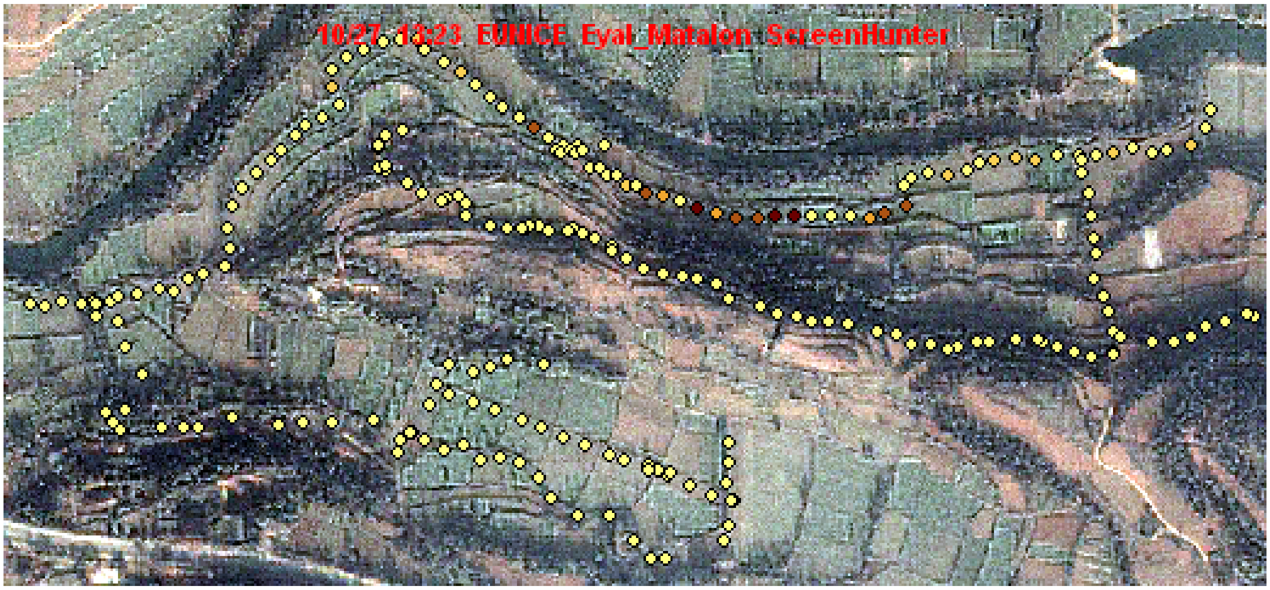

Figure 4: Georeferenced snail survey data overlaid on IKONOS imagery (November 30, 2002 1-m pan-sharpened resolution). Each point represents a sampling site along the irrigation ditch network in which snails were counted and collected and ditch conditions were recorded. Image is from 'Gaohuai7' production group in Jinyang County, Sichuan Province, China.

Land Cover Classification

The result of the land cover classification is illustrated in Figure 5. The overall accuracy of the land-cover categories for the MLC training set was 85% and the overall Kappa statistic was 0.798. The accuracies for built infrastructure, crop, bare soil / dry field, river, forest, and pond were 97.9%, 93.8%, 100%, 29.9%, 89.5%, and 90.5%, respectively (see Table 2). The overall accuracy of the land-cover categories for the MLC validation set was 83% and the overall Kappa statistic was 0.754. The accuracies for built infrastructure, crop, bare soil / dry field, river, forest, and pond were 86.3%, 91.5%, 96.8%, 28.2%, 91.2%, and 93.3%, respectively (see Table 3).

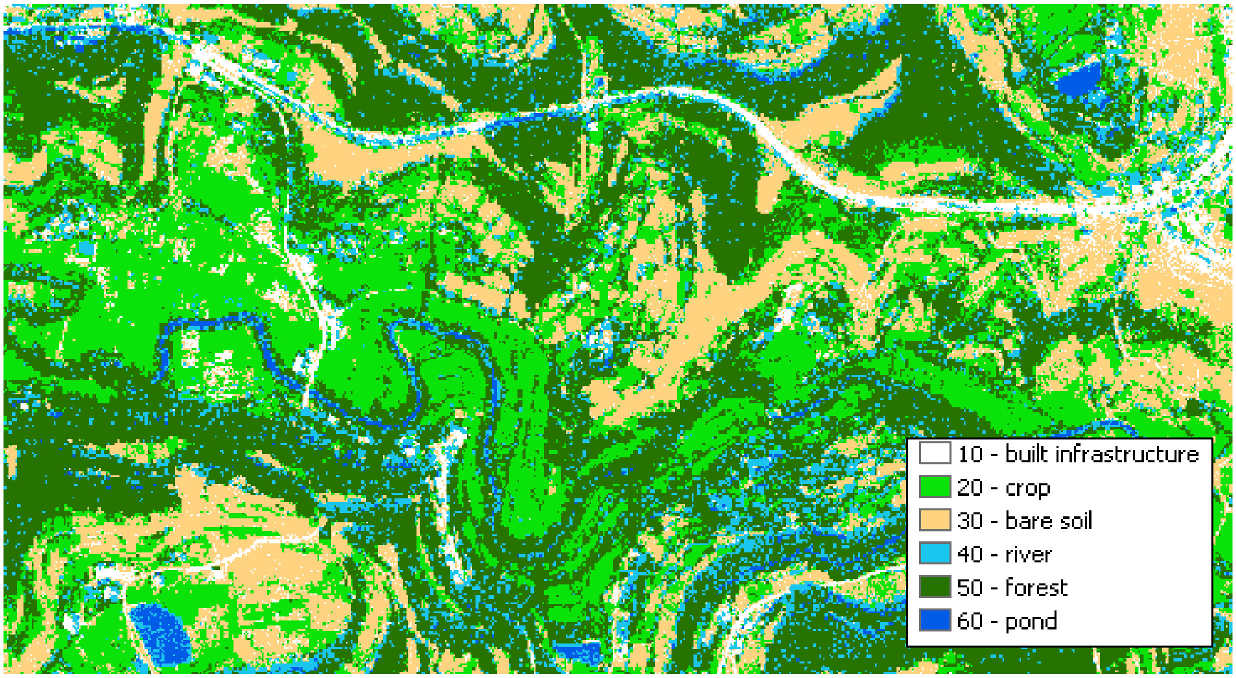

Figure 5: Land Cover Data based on Supervised Maximum Likelihood Classification of IKONOS imagery (November 30, 2002 1-m pan-sharpened resolution) Image is from 'Gaohuai7' production group in Jinyang County, Sichuan Province, China.

Table 2: Maximum Likelihood Classification Accuracy Matrix for a training set of 554 pixels. MLC performed on IKONOS image from Jinyang County, Sichuan Province (November 30, 2002; 1-m pan-sharpened resolution).

Reference	**Classification**						
	Built	Crop	Baresoil	River	Forest	Pond	**Total**
Built	47	0	1	0	0	0	48
Crop	0	61	2	2	0	0	65
Baresoil	0	0	74	0	0	0	74
River	1	3	0	20	2	41	67
Forest	0	8	1	18	231	0	258
Pond	0	0	0	3	1	38	42
Total	48	72	78	43	234	79	554

Kappa Statistic: 0.798 **Overall Accuracy:** 85%

Table 3: Maximum Likelihood Classification Accuracy Matrix for the validation set of 552 pixels. From IKONOS satellite image of Jinyang County, Sichuan Province (November 30, 2002; 1-m pan-sharpened resolution)

Reference	**Classification**						
	Built	Crop	Baresoil	River	Forest	Pond	**Total**
Built	44	1	2	3	0	1	51
Crop	0	43	1	2	1	0	47
Baresoil	2	0	60	0	0	0	62
River	3	0	0	22	1	52	78
Forest	0	8	0	17	259	0	284
Pond	0	0	0	2	0	28	30
Total	49	52	63	46	261	81	552

Kappa Statistic: .754 **Overall Accuracy:** 83%

Small-Scale: Ditch Conditions

Three ditch conditions were significantly associated with snail presence sites (i.e. odds ratio > 1 for univariate logistic regression against a snail presence indicator variable): thin grass, visible water, and slow water flow (< 0.15 m/s). Three ditch conditions were significantly associated with snail absence sites (i.e. odds ratio < 1): thick grass, dry conditions, and fast water flow (> 0.15 m/s). Visible water and slow water flow were significant predictors of snail counts in univariate negative binomial (NB) models at $p \leq 0.05$ (see Table 4). Using the log likelihood value as a measure of best fit, the NB regression model predicting snail count from visible water (coefficient = 2.18, p = 0.0001) and slow water flow (coefficient = 3.40, p = 0.0001) was the best two-variable micro-scale model (chi-squared = 24.8, df = 2, p = 0.0001, log likelihood = –320.4). The NB regression model predicting snail count from fast water (coefficient = –2.51, p = 0.0001) and dry conditions (coefficient = –1.21, p = 0.011) was also statistically significant (chi-squared = 14.9, df = 2, p = 0.0006, log likelihood = –325.3). Thin grass and soil ditch construction, although not significant predictors of snail counts in univariate models, were positively correlated with snail counts in two-variable models at significance, $p \leq 0.05$. Thick grass, concrete ditch construction, wet conditions without visible water, and dry conditions were negatively correlated with snail counts in two-variable models at significance, $p \leq 0.05$ (see Table 5).

Table 4: Univariate associations between small-scale ditch conditions and snail presence (logistic model) and snail counts (negative binomial [NB] model). An odds ratio > 1 indicates that a given variable is associated with snail presence sites; an odds ratio < 1 indicate that a given variable is associated with snail absence sites. *denotes significance at $p \leq 0.05$

		Logistic Model (Snail Presence)	NB Model (Snail Counts)
Variable		**Odds Ratio**	**Coefficient**
GRASS	None	.628	.020
	Thin	2.41*	.569
	Thick	.403*	-.961
WATER LEVEL	Dry	.365*	-.608
	Wet	.964	-.939
	Visible	2.79*	.977*
WATER VELOCITY	Fast	.144*	-1.83*
	Slow	6.96*	1.83*
DITCH CONSTR.	Concrete	.579	-.642
	Brick	1.18	-1.02
	Soil	1.69	.681

Table 5: Two-variable negative binomial regression models associating small-scale ditch conditions and snail counts. Regression coefficients and their corresponding p-values are indicated in parenthesis next to each variable. Log-likelihood is a measure of goodness of fit such that the higher (more positive) the log-likelihood value is, the more the model explains the data's variability.

2-variable NB Models	**Log-Likelihood**
Thick Grass (-1.23, p=.023) Dry Conditions (-.884, p=.051)	-329.5
Thick Grass (-1.19, p=.024) Visible Water (1.15, p=.011)	-328.0
Thick Grass (-1.21, p=.024) Concrete Constr. (-.924, p=.052)	-329.6
Thick Grass (-1.25, p=.021) Soil Constr. (.978, p=.039)	-329.4
Dry Conditions (-.887, p=.056) Wet Conditions (-1.41, p=.043)	-330.0
Dry Conditions (-1.21, p=.011) SlowWater (2.51, p=.001)	-325.3
Wet Conditions (-1.20, p=.057) Fast Water (-1.98, p=.001)	-327.2
Visible Water (2.18, p=.0001) Slow Water (3.40, p=.0001)	-320.4

Large-Scale: Remotely Sensed Landscape Features

Two remotely sensed landscape features were significantly associated with snail presence sites (i.e. odds ratio > 1 for univariate logistic regression against a snail presence indicator variable): percent forest cover and percent pond cover. Three landscape features were significantly associated with snail absence sites (i.e. odds ratio < 1): elevation, % built infrastructure land cover, and percent crop field land cover. Percent forest cover was a significant predictor of snail counts in univariate negative binomial (NB) models at significance, $p \leq 0.05$. Elevation, percent built infrastructure land cover, and percent river land cover were negatively associated with snail counts in univariate NB models at significance, $p \leq 0.05$. Univariate models regressing the autocorrelation variable, distance between sites, did not converge (see Table 6).

Using the log likelihood value as a measure of best fit, the NB regression model predicting snail count from percent forest cover (coefficient = 6.08, p = 0.001) and percent crop field cover (coefficient = 3.38, p= .038) was the best two-variable micro-scale model (chi-squared = 12.41, df = 2, p<.002, log likelihood = -326.6). All other significant two-variable macro-scale models included the variable % river cover and were discounted due to this land category's low classification accuracy (see Tables 1 and 2). The autocorrelation variable was not a significant predictor of snail counts for any two-variable NB model and was therefore not included in the analysis (see Table 7).

Table 6: Univariate associations between remotely sensed landscape features and snail presence (logistic model) and snail counts (negative binomial [NB] model). An odds ratio > 1 indicates that a given variable is associated with snail presence sites; an odds ratio < 1 indicate that a given variable is associated with snail absence sites. *****denotes significance at $p \leq 0.05$

	Logistic Model (Snail Presence)	NB Model (Snail Counts)
Variable	**Odds Ratio**	**Coefficient**
Elevation	.968*	-.028*
Slope	1.04	.032
% Built	9.21×10^{-8}*	-12.4*
% Crop	.127*	-1.29
% Bare Soil	2.28	-.387
% River	.057	-10.1*
% Forest	22.2*	2.93*
% Pond	4260*	12.9
Wetness	.991	-.011
Shannon	2.41	.621
Autocorrelation	Non-converge	Non-converge

Table 7: Two-variable negative binomial regression models associating remotely sensed landscape features and snail counts. Regression coefficients and their corresponding p-values are indicated in parenthesis next to each variable. Log-likelihood is a measure of goodness of fit such that the higher (more positive) the log-likelihood value is, the more the model explains the data's variability.

2-variable NB Models	Log-Likelihood
Elevation (-.042, p=.005) % River (-122.5, p=.001)	-324.8
% Road (-9.884, p=.056) % River (-8.296, p=.021)	-326.3
% Crop (-2.70, p=.007) % River (-14.3, p=.001)	-324.7
% Crop (3.38, p=.038) % Forest (6.08, p=.001)	-326.6
% River (-12.4, p=.001) % Forest (3.51, p=.001)	-322.4
% River (-12.6, p=.001) % Pond (25.6, p=.035)	-325.7

Remotely-Sensed Proxies for Small-Scale Environmental Conditions

Crop land cover and the topographic wetness index serve as useful proxies for thick grass and visible water, respectively. Built infrastructure shows no clear relationship with concretized ditch construction while degree slope shows an inverse correlation with fast water velocity (see Table 8).

Table 8: Pair-wise correlations between large-scale remotely sensed features and small-scale conditions at the ditch using the information at or within 50-meters of snail sampling sites. (n = 2525).

Large-Scale Variable	Small-Scale Variable	r^2	p-value
Crop Land Cover	Thick Grass	**0.226**	.0001
Built Infrastructure	Concrete Ditch Constr.	0.002	.9102
Topographic Wetness Index	Visible Water	**0.165**	.0001
Degree Slope	Fast Water	**-0.3368**	.0001

Discussion

This study integrated georeferenced snail survey data and remotely-sensed environmental data in order to better understand *O. hupensis* snail ecology and more accurately predict *O. hupensis* habitats at the production group level. Below, the important small-scale ecological conditions and large-scale remotely-sensed landscape features associated with *O. hupensis* snail occurrence along irrigation ditch networks are described.

Small-Scale Environmental Conditions

Water conditions at the ditch scale were significant ecological determinants of snail occurrence (both presence and counts) along irrigation ditch networks. Low-flowing water was positively correlated with snail presence sites as well as snail counts at those sites. This trend corroborates common anecdotal accounts of *O. hupensis* microhabitat (Sturrock 2001). It should be noted that the indicator variable, *slow water*, also applies to dry sampling sites without any water in the ditch, which were associated with snail absence sites (see Table 4). Therefore, the fact that snail counts are predicted by both variables *slow water* and *visible water* (see Table 5) suggests that snails prefer microhabitats in which water is present but flowing at less than 0.15 meters/second. Conversely, sites with fast-flowing water were correlated with snail absence sites and lower snail counts in general. This trend suggests that snails cannot attach to sites in which water flows at faster than 0.15 meters/second and are flushed downstream to areas with more suitable habitat. Although prior studies have shown that snails in the mountainous regions of Sichuan and Yunnan are less affected by flooding than their downstream counterparts below the Three Gorges Dam, faster water flow may nevertheless preclude snail occurrence along the irrigation ditch networks (Davis *et al.* 2001).

The thickness of grass is predictive of snail presence and may be predictive of snail counts. More *thin grass* sampling sites had snails present, and conversely, fewer *thick grass* sampling sites had snails present than would be expected by chance (see Table 4). This corroborates prior anecdotal accounts describing *O. hupensis* as favoring sites with emergent vegetation (Sturrock 2001). Field observations have described snails clinging on to the base of grasses, which provides an "environmental buffer, protecting them from direct light, rain, and extreme fluctuations in temperature and moistures" (Sturrock 2001). However, counter to this study, which found associations between snail presence and emergent vegetation, prior research has found a positive correlation between grass density and snail density (Guo 1990). In fact, the existence of grass has traditionally been used as an indicator of snail habitat in China's national schistosomiasis control program (Yang *et al.* 2006). However, it is possible that the effects of grass density on snail occurrence are confounded by the effects of soil composition. A study by Seto *et al.* (2002b) revealed that the lack of certain soil conditions potentially excluded *O. hupensis robertsoni* from some sites in the Anning River Valley.

Ditch construction seems to play a far less significant role in the distribution of *O. hupensis* than previously thought (Seto *et al.* 2002a). Although soil and concrete ditch construction were positively and negatively correlated with snail counts, respectively, in some two-variable NB regression models, these variables were not recurring predictors of snail counts. The lack of any obvious association between ditch construction and snail occurrence (see Table 4), suggests that, contrary to anecdotal accounts, concretized ditches do not necessarily exclude *O. hupensis* (Seto 2008, pers. comm.).

Remotely-Sensed Landscape Features

Three variables were derived from remotely-sensed elevation data at 90-meter resolution. Elevation was a significant predictor of snail absence sites and negatively correlated with snail counts in univariate regression models (see Table 6). Elevation in this study ranged from 490m to 517m. However, Davis *et al.* (2001) has described the subspecies *O. hupensis robertsoni* as occurring at elevations ranging from 500m to 2000m, with several populations living on lower plateaus or basins at 200-500m. This wide range of elevation for *O. h. robertsoni* has been confirmed by other studies (Xu *et al.* 2004, Yang *et al.* 2008). Therefore, if there is an inverse relationship between snail occurrence and elevation it is only within the elevation range relevant to this study. However, it is unclear why elevation as a single variable at such a narrow range should affect snail habitat. The disparate resolution of snail survey data and elevation data may explain this trend. Whereas snail counts were surveyed every 10 meters along the irrigation ditch, elevation data for the study was acquired at 90-meter resolution. The analysis in this study did not consider the topographical heterogeneity of the several sampling sites that existed within each 90m x 90m elevation pixel.

The other two features derived from elevation data, slope and a wetness index, did not show any significant association with snail occurrence (both presence and counts, see Table 5). The relationship between slope and water velocity, which was predictive of snail counts, was the reverse of what would be expected: lower-sloped sites were generally correlated with faster water flows (see Table 8). This is likely due to the relatively flat terrain in this study (mean slope value was 5.5 degrees, with a standard deviation of 4.9 degrees). The loose correlation between slope and water flow with regards to snail habitat has been noted anecdotally by Davis *et al.* (2001), who observed *Oncomelania hupensis* in "small, trickling perennial flows" at a slope of up to 25 degrees. Wetness, which has been a key indicator of snail habitat in lower-resolution studies (municipal or provincial level), was not a significant predictor of snail occurrence at the village level (Yang *et al.* 2006, Guo *et al.* 2005). This may be explained by the result that wetness is a poor proxy for water level at the ditch networks (see Table 8). As with the elevation data, it is also likely that slope and wetness data was too coarse to explain any variability in snail counts.

Three of the six land cover categories were predictive of snail counts. Percent forest cover was the most predictive land cover category as it was positively correlated with snail counts and presence sites in both univariate and multivariable regression models (see Tables 5 and 6). This positive association is likely due to the shading created by forest cover, which shields snails from intense UV radiation and prevents desiccation (Davis *et al.* 2001).

Percent built infrastructure was associated with snail absence sites. Given the low predictive value of ditch construction on snail occurrence, it is unclear why this might be the case. Built infrastructure was not associated with higher or lower snail counts in neither univariate nor two-variable regression models. It is possible that the absence of snails in areas in proximity to built infrastructure is due to the implementation of snail control methods in areas in which people work and reside. However, disease control information within the production group was not available for this study.

Percent crop cover was associated with snail absence sites but positively correlated with snail counts in a two-variable regression model with percent forest cover. This discrepancy is likely due to the heterogeneity of crop land cover in this study. Prior studies at the village level that have found correlations between snail occurrence and crop land cover have noted that snails are associated with some crop land covers and not others (Xu *et al.* 2004, Yang *et al.* 2008). The predictive strength of the two-variable model with land and forest cover suggests that crop cultivation in proximity to forested areas provides more ideal habitat for *O. hupensis,* likely due to suitable vegetation, soil composition, and shading in these areas.

Broader Implications, Limitations, and Directions for Future Study

The study of *O. hupensis* ecology and prediction of snail occurrence at the village scale constitutes a novel and important research objective (Yang *et al.* 2008). The purpose of these studies is to identify environmental determinants of snail habitat at a higher resolution. Remote sensing and geographic information systems have been indispensible tools for identifying landscape features associated with snail habitat at a municipal or provincial level. However, the results of this analysis suggests that a different set of remotely-sensed variables that play an important role in the occurrence of *O. hupensis* at the village level, namely, land cover classes, derived from IKONOS imagery at 1-meter resolution. This study also illustrates that additional explanatory power for snail occurrence can be attained by examining smaller-scale variables that cannot be remotely sensed, namely water conditions and grass density within irrigation ditches. Identification of these small-scale environmental determinants of *O. hupensis* in this analysis corroborates descriptions of snail habitats that have largely been anecdotal (Sturrock 2001). Lastly, exploring associations between variables at the two different scales, allows for a better understanding of the underlying ecological mechanisms of associations between remotely sensed features and snail occurrence.

Despite the novelty of this approach, this study had several inherent limitations. First, this study did not take into account temporal fluctuations in neither the environmental variables examined nor snail distribution and density. Snail surveys were carried out once in the fall of 2007. Prior studies have shown that snail counts vary widely throughout the year and that various seasonal cues trigger key population processes, such as recruitment and mortality (Remais *et al.* 2006). Further studies should conduct snail surveys multiple times throughout the year to understand how the association between ecological determinants and snail occurrence evolves temporally. However, higher-resolution remotely sensed data is difficult to obtain with regularity. IKONOS imagery at 1-meter resolution for the study site could only be obtained for the fall of 2002, five years prior to the time of the snail survey. During that time, it is likely that land cover data has changed, particularly in light of demographic shifts to urban centers and national afforestation programs in rural areas (Shen *et al.* 2005, Démurger and Yang 2006). This is a general limitation of high-resolution IKONOS imagery. Unlike lower-resolution satellites (e.g. Landsat), IKONOS satellites are not scheduled for regular acquisitions over the entire surface of the earth. Therefore, in order to understand seasonal variability in the relation between ecological determinants and snail densities, there would need to be a larger reliance on smaller-scale field observations.

Despite these limitations, this study constitutes an important step in predicting *O. hupensis* habitat at multiple scales. More importantly, it reaffirms the utility of remote sensing techniques for studying the ecology of disease vectors at higher resolutions. It is hoped that this general approach will inform the implementation of focal mollusciciding and other snail control strategies at the production group level.

Acknowledgements

The author thanks his mentor Edmund Seto of the Center for Occupational and Environmental Health in the University of California, Berkeley for all of his guidance and support during this project. The author also thanks instructors Shelly Cole and Tim De Chant for guiding the translation of this research into an increasingly coherent academic work. The author thanks Maureen Lahiff for her input regarding statistical methods applicable to zero-inflated count data. Finally, the author thanks the Sichuan Institute of Parasitic Disease for collecting the field data for this study.

Works Cited

Agarwal, D. K., A. E. Gelfand, and S. Citron-Pousty. 2002. Zero-inflated models with application to spatial count data. Environmental and Ecological Statistics 9:341-355.

Beyer, H. L. 2004. Hawth's Analysis Tools for ArcGIS. Available at http://www.spatialecology.com/htools, accessed January 28, 2009.

Bergquist, R. 2001. Strategies for control of infection and disease: current practice and future potential. Pp. 413-467 in Adel A. F. Mahmoud, editor. Schistosomiasis. Imperial College Press. London, UK

Chen M. and Z. Feng. 1999. Schistosomiasis control in China. Parasitology International, 48:11-19.

Cunningham, R. B. and D. B. Lindenmayer. 2005. Modeling Count Data of Rare Species: Some Statistical Issues. Ecology 86:1135-1142.

Ceccato P., S. J. Connor, I. Jeanne, and M. C. Thomson. 2005. Application of Geographical Information Systems and Remote Sensing technologies for assessing and monitoring malaria risk. Parassitologia (Rome) 47:81-96.

Clarke K. C., S. L. McLafferty, and B. J. Tempalski. 1996. On epidemiology and geographic information systems: A review and discussion of future directions. Emerging Infectious Diseases 2:85-92.

Clennon J. A., C. H. King, E. M. Muchiri, and U. Kitron. 2007. Hydrological modelling of snail dispersal patterns in Msambweni, Kenya and potential resurgence of Schistosoma haematobium transmission. Parasitology 134:683-693.

Cringoli G., A. Ippolito, and R. Taddei. 2005. Advances in satellite Remote Sensing of pheno-climatic features for epidemiological applications. Parassitologia (Rome) 47:51-62.

Daniel M., J. Kolar, and P. Zeman. 2004. GIS tools for tick and tick-borne disease occurrence. Parasitology 129:S329-S352.

Davis, G. M., T. Wilke, Z. Yi, Z. Xu, C. Qiu, C. Spolsky, D. Qiu, Y. Li, M. Xia, and Z. Feng. 1999. Snail-Schistosoma, Paragonimus interactions in China: population ecology, genetic diversity, coevolution, and emerging diseases. Malacologia. 41(2): 355-377.

Davis, George M. Department of Malacology in the Academy of Natural Sciences, Philadelphia, Pennsylvania. 2008, personal communication.

de La Rocque S., J. F. Michel, J. Bouyer, G. De Wispelaere, and D. Cuisance. 2005. Geographical Information Systems in parasitology: a review of potential applications using the example of animal trypanosomosis in West Africa. Parassitologia (Rome) 47:97-104.

Démruger S. and W. Yang. 2006. Economic changes and afforestation incentives in rural China. Environment and Development Economics 11:629-649.

Guo J., V. Penelope, C. Cao, U. Jürg, H. Zhu, A. Daniel, R. Zhu, Z. He, D. Li, F. Hu, M. Chen, and T. Marcel. 2005. A geographic information and remote sensing based model for prediction of Oncomelania hupensis habitats in the Poyang Lake area, China. Acta Tropica, 96:213-222.

Jerrett M., R. T. Burnett, M. S. Goldberg, M. Sears, D. Krewski, R. Catalan, P. Kanaroglou, C. Giovis, and N. Finkelstein. 2003. Spatial analysis for environmental health research: Concepts, methods, and examples. Journal of Toxicology and Environmental Health Part A 66:1783-1810.

Liang S., D. Maszle, and R. C. Spear. 2002. A quantitative framework for a multi-group model of Schistosomiasis japonicum transmission dynamics and control in Sichuan, China. Acta Tropica, 82:263-277.

Maszle D. R., P. G. Whitehead, R. C. Johnson, and R. C. Spear. 1998. Hydrological studies of schistosomiasis transport in Sichuan Province, China. The Science of the Total Environment, 216:193-203.

Schistosomiasis Expert Advisory Committee, Ministry of Health, Analysis on the epidemiological situation of schistosomiasis in 1992 in China. *Chin J Schisto Contr* 5 (1993), pp. 257–259.

Martin T. G., B. A. Wintle, J. R. Rhodes, P. M. Kuhnert, S. A. Field, S. J. Low-Choy, A. J. Tyre, and H. P. Possingham. 2005. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. Ecology Letters 8:1235-1246.

Seto E., S. Liang, D. Qiu, X. Gu, and R. C. Spear. 2001. A protocol for geographically randomized snail surveys in schistosomiasis fieldwork using the global positioning system. American Journal of Tropical Medicine and Hygiene 64:98-99.

Seto E., B. Xu, S. Liang, P. Gong, W. P. Wu, G. Davis, D. C. Qiu, X. G. Gu, and R. Spear. 2002a. The use of remote sensing for predictive modeling of schistosomiasis in China. Photogrammetric Engineering and Remote Sensing 68:167-174.

Seto E. Y. W., W. P. Wu, D. C. Qiu, H. Y. Liu, X. G. Gu, H. G. Chen, R. C. Spear, and G. M. Davis. 2002b. Impact of soil chemistry on the distribution of Oncomelania hupensis (Gastropoda : Pomatiopsidae) in China. Malacologia 44:259-272.

Seto, M. Center for Occupational and Environmental Health, School of Public Health, University of California, Berkeley. 2008, personal communication.

Sileshi, G. 2008. The excess-zero problem in soil animal count data and choice of appropriate models for statistical inference. Pedobiologia 52:1-17.

Shen L., S. Cheng, A. J. Gunson, and H. Wan. 2005. Urbanization, sustainability and the utilization of energy and mineral resources in China. Cities 22:287-302.

Spear, Robert C., Edmund Seto, Song Liang, Merril Birkner, Alan Hubbard, Dongchuan Qiu, Changhong Yang, Bo Zhong, Fashen Xu, Xueguang Gu, and George M. Davis. 2004. Factors influencing the transmission of *Schistosoma japonicum* in the mountains of Sichuan Province of China. The American Journal of Tropical Medicine and Hygiene. 70(1): 48-56.

Sturrock, R. 2001. Schistosomes and their intermediate hosts. Pp. 7-83 in Adel A. F. Mahmoud, editor. Schistosomiasis. Imperial College Press. London, UK

White, G. C. and R. E. Bennetts. 1996. Analysis of Frequency Count Data Using the Negative Binomial Distribution. Ecology 77:2549-2557.

Wilke T., G. M. Davis, C. Cui-E, Z. Xiao-Nung, Z. Xiao Peng, Z. Yi, and C. M. Spolsky. 2000. Oncomelania hupensis (Gastropoda: Rissooidea) in eastern China: molecular phylogeny, population structure, and ecology. Acta Tropica, 77:215-227.

World Health Organization. 2008. Health Topics: Schistosomiasis. <www.who.int/topics/schistosomiasis/en/>

Xu B., P. Gong, G. Biging, S. Liang, E. Seto, and R. Spear. 2004. Snail density prediction for schistosomiasis control using ikonos and ASTER images. Photogrammetric Engineering and Remote Sensing 70:1285-1294

Xu B., P. Gong, E. Seto, S. Liang, C. Yang, S. Wen, D. Qiu, X. Gu, and R. Spear. 2006. A Spatial-Temporal Model for Assessing the Effects of Intervillage Connectivity in Schistosomiasis Transmission. Annals of the Association of American Geographers 96:31-46.

Yang G. J., P. Vounatsou, M. Tanner, X. N. Zhou, and J. Utzinger. 2006. Remote sensing for predicting potential habitats of Oncomelania hupensis in Hongze, Baima and Gaoyou lakes in Jiangsu province, China. Geospatial Health 1:85-92.

Yang, K. Xian-Hong Wang, Guo-Jing Yang, Xiao-Hua Wu, Yun-Liang Qi, Hong-Jun Li, and Xiao-Nong Zhou. 2008. An integrated approach to identify distribution of *Oncomelania hupensis*, the intermediate host of *Schistosoma japonicum*, in a mountainous region in China. International Journal for Parasitology. 38: 1007-1016

Zhou X. N., J. B. Malone, T. K. Kristensen, and N. R. Bergquist. 2001. Application of geographic information systems and remote sensing to schistosomiasis control in China. Acta Tropica, 79:97-106.