Econometric Evidence on Forest Ecosystem Services: Deforestation and Flooding in Malaysia

Jie-Sheng Tan-Soo, Norliyana Adnan, Ismariah Ahmad, Subhrendu K. Pattanayak & Jeffrey R. Vincent

Environmental and Resource Economics
The Official Journal of the European Association of Environmental and Resource Economists

ISSN 0924-6460

Environ Resource Econ
DOI 10.1007/s10640-014-9834-4

Springer

Environ Resource Econ
DOI 10.1007/s10640-014-9834-4

Econometric Evidence on Forest Ecosystem Services: Deforestation and Flooding in Malaysia

Jie-Sheng Tan-Soo · Norliyana Adnan · Ismariah Ahmad · Subhrendu K. Pattanayak · Jeffrey R. Vincent

Accepted: 20 September 2014

Abstract Governments around the world are increasingly invoking hydrological services, such as flood mitigation and water purification, as a justification for forest conservation programs in upstream areas. Yet, rigorous empirical evidence that these programs are actually delivering the intended services remains scant. We investigate the effect of deforestation on flood-mitigation services in Peninsular Malaysia during 1984–2000, a period when detailed data on both flood events and land-use change are available for 31 river basins. Floods are the most common natural disaster in tropical regions, but the ability of tropical forests to mitigate large-scale floods associated with heavy rainfall events remains disputed. We find that the conversion of inland tropical forests to oil palm and rubber plantations significantly increased the number of days flooded during the wettest months of the year. Our results demonstrate the importance of using disaggregated land-use data, controlling for potentially confounding factors, and applying appropriate estimators in econometric studies on forest ecosystem services.

Keywords Ecosystem service · Tropical forests · Floods · Oil palm · Rubber · Malaysia

1 Introduction

The Millennium Ecosystem Assessment (2005) highlighted several hydrological services that forests in upstream areas can supply to households and industries located downstream.

This paper has not been submitted elsewhere in identical or similar form, nor will it be during the first 3 months after its submission to the Publisher.

J.-S. Tan-Soo
Duke Global Health Institute, Nicholas School of the Environment and Sanford School of Public Policy, Duke University, Durham, NC 27708, USA

S. K. Pattanayak · J. R. Vincent (✉)
Nicholas School of the Environment and Sanford School of Public Policy, Duke University, Durham, NC 27708, USA
e-mail: Jeff.Vincent@duke.edu

N. Adnan · I. Ahmad
Forest Research Institute Malaysia, 52109 Kepong, Selangor, Malaysia

Published online: 07 October 2014

These services, which include flood mitigation and water purification, have increasingly been invoked as a justification for forest conservation programs. A good example is the proliferation of watershed-payment programs, which compensate landowners for refraining from harvesting or converting their forests (Brauman et al. 2007; Kinzig et al. 2011). As of 2011, more than 100 programs in developing countries were paying landowners to supply hydrological services, with another 50 programs in process (Bennett et al. 2013).

A growing body of economic research seeks to determine the causal impact of conservation programs on the delivery of these and other forest ecosystem services (Pattanayak et al. 2010; Ferraro et al. 2012). The chain of causality has two links: first, a program must induce forest owners to protect a larger area of forest from harvesting or conversion than they would have protected in the absence of the program; and second, the supply of the service must increase in response to the larger area protected. With few exceptions (e.g., Kramer et al. 1997, Pattanayak and Kramer 2001), most of the economic literature investigates only the first link. We investigate the second link, with a focus on the effect of tropical forest protection on flooding.

Floods are the most common natural disaster in tropical regions (Jonkman 2005). The ability of tropical forests to mitigate floods has been debated for almost a century (Bruijnzeel 2004). There is general agreement that, compared to other land uses, tropical forests reduce peak runoff from small catchments during small to moderate rainfall events (Bruijnzeel 2004). What is disputed is whether they mitigate large-scale floods associated with heavy rainfall events. For example, a well-known report by two prominent international agencies, the UN Food and Agriculture Organization and the Center for International Forestry Research, argued that large-scale floods in tropical regions are purely meteorological events that are not influenced much by land use (FAO and CIFOR 2005). It presented a table from Kiersch (2000) indicating that land use does not affect peak flows in river basins larger than 100 km^2 (FAO and CIFOR 2005, p. 8).[1]

Forest hydrologists have traditionally investigated the effects of forests on storm runoff by conducting experiments in matched treatment-control pairs of small catchments. Recent paired-catchment studies in temperate and tropical locations offer evidence that intact forests can reduce peak runoff even during heavy rainfall events (Alila et al. 2009; Ogden et al. 2013), but the catchments in these studies were small (less than 3 km^2). Alila et al. (2009, p. 18) acknowledged that their findings "may not be viewed by some as important, for instance, to the more challenging question of the effects of massive quasi-permanent deforestation on floods in large tropical mountain basins."

Research on the effects of deforestation on floods at larger, more policy-relevant scales has relied instead on statistical analysis of nonexperimental data from multiple locations (river basins or countries), multiple time periods, or both (panel data). Several recent studies, including ones by economists (e.g., Ferreira and Ghimire 2012; Ferreira et al. 2013), have analyzed data from the Global Active Archive of Large Flood Events administered by the Dartmouth Flood Observatory (DFO; http://www.dartmouth.edu/~floods/Archives/index.html).[2] The DFO is funded by NASA and the European Commission. As its name indicates, the DFO Global Archive records information on floods it defines as large, in the

[1] Ogden et al. (2013, p. 8444) note that the original table in Kiersch (2000) did not include any supporting references.

[2] A pioneering cross-country study on the economics of natural disasters by Kahn (2005) included an analysis of flood occurrence, but it focused on issues other than the effects of land use. It obtained flood data from the Emergency Events Database (EM-DAT; http://www.emdat.be), which is widely used but has shortcomings that include separate flood events being recorded as a single one and smaller flood events being underreported in developing countries (Jonkman 2005, p. 153).

sense of being associated with "significant damage to structures or agriculture, long (decades) reported intervals since the last similar event, and/or fatalities." An initial analysis of data from the DFO Global Archive reported that deforestation increased the number of large floods in developing countries (Bradshaw et al. 2007). Subsequent re-analyses found that this conclusion does not hold, however, when factors that confound the effect of deforestation are controlled more completely (Van Dijk et al. 2009; Ferreira and Ghimire 2012; Ferreira et al. 2013). Chief among these factors is population density, which can affect the likelihood that floods are reported.

We present econometric evidence that flood mitigation by tropical forests can occur at scales much larger than previously reported. Instead of relying on the DFO Global Archive, we drew on rich governmental data sources for Peninsular Malaysia, one of the three major divisions of Malaysia.[3] The Malaysian federal water agency, the Drainage and Irrigation Department (DID), divides Peninsular Malaysia into 41 river basin management units (RBMUs). Our chief data source on floods was a report that provided information on the number and duration of floods per month for nearly all RBMUs during 1980–2000 (KTA Tenaga 2003). That report also provided information on the number of deaths and evacuations associated with each flood. It is vastly more complete than the DFO Global Archive: while the latter lists only five floods in Peninsular Malaysia between 1985 (the start of the DFO data) and 2000, the DID report lists 587. The hundreds of additional floods in the DID report that the DFO excludes are not unimportant events, as they accounted for 70 % of the flood deaths and 92 % of the flood evacuees in Peninsular Malaysia during 1985–2000.

We obtained land-use data from Peninsula-wide surveys conducted by the Malaysian Department of Agriculture in 1984–1985, 1997–1998, and 2004–2005.[4] We classified these data into two forest types and four nonforest land uses, allowing us to study how the effect of deforestation on floods varied by the type of forest that was converted and the use to which it was converted. Our main finding is that conversion of inland forests to oil palm and rubber plantations significantly increased the number of days flooded during the wettest months of the year. We detected this effect even though we controlled more carefully for potentially confounding factors than recent cross-country studies have done.

The remainder of the paper is organized as follows. We describe our econometric model and data sources in the next section. After that, we present our estimation results, and we use those results to calculate the effects of conversion of inland forests to oil palm and rubber on flood-related deaths and evacuations. We conclude by recapping our findings and discussing caveats and areas for future research.

2 Materials and Methods

2.1 Econometric Approach

Our analysis aimed at identifying the mean marginal effect of deforestation on floods. We analyzed two measures of flood events: the monthly number of floods, and the monthly number of days flooded. Both were count variables. The standard regression models for analyzing count data are the Poisson and negative binomial models (Hilbe 2011; Cameron and Trivedi 2013). For reasons given below, we used the Poisson model.

[3] The other two are the states of Sabah and Sarawak on Borneo.

[4] These surveys do not cover Sabah and Sarawak.

The conditional expectation function for the Poisson model has an exponential form. It can be written as follows for the particular specification we estimated (Wooldridge 2002, Ch. 19):

$$\mathrm{E}\left(f_{rym}\left|\mathbf{L}_{ry}, \mathbf{X}_{rym}, c_r, \theta_y, \theta_m\right.\right) = \exp\left(\mathbf{L}_{ry}\boldsymbol{\beta} + \mathbf{X}_{rym}\boldsymbol{\gamma} + c_r + \theta_y + \theta_m\right) \tag{1}$$

f_{rym} is the monthly number of flood events (floods or days flooded, depending on the model), with r denoting RBMU, y denoting year, and m denoting month. As indicated by these subscripts, our flood-event data varied both cross-sectionally and longitudinally. "exp" is the exponential operator. $\mathbf{L}_{ry}$ is a matrix of land-use variables, which varied across years in an RMBU but not across months within a year, while $\mathbf{X}_{rym}$ is a matrix of other covariates, some of which varied by both year and month and others only by year. $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are parameter vectors for $\mathbf{L}$ and $\mathbf{X}$, respectively.

The panel structure of our data enabled us to include c_r, θ_y, and θ_m as fixed effects that controlled, respectively, for time-invariant RBMU characteristics (e.g., area, topography, soils, geology), RBMU-invariant annual characteristics (e.g., ENSO cycle, CO_2 concentrations, water management policies and procedures, land-use survey procedures, national political and economic conditions), and RBMU-invariant monthly characteristics (e.g., seasonality). The inclusion of fixed effects for RBMUs implies that our models used changes over time within RBMUs, not differences between RBMUs, to identify the effects of the land-use variables on flooding. Fixed effects were more appropriate than random effects because the sample included nearly all RBMUs in Peninsular Malaysia, not a random selection of them (Wooldridge 2002, pp. 250–252; Kennedy 2008, p. 291).

An individual variable L in $\mathbf{L}$ gave the percentage of the surface area of an RBMU that was in a particular land use in a particular year. Because the percentages summed to 100 % across all land uses in each year, one of the land-use variables had to be excluded during estimation. We excluded the variable for forestland, which enabled us to interpret the parameter estimate $\hat{\beta}$ on L as a relative measure of the mean marginal effect of converting forest to that land use. The mean marginal effect of L on flood events is given by differentiating Eq. (1) with respect to L,

$$\frac{\partial \mathrm{E}\left(f_{rym}\left|\mathbf{L}_{ry}, \mathbf{X}_{rym}, c_r, \theta_y, \theta_m\right.\right)}{\partial L} = \hat{\beta} \cdot \mathrm{E}\left(f_{rym}\left|\mathbf{L}_{ry}, \mathbf{X}_{rym}, c_r, \theta_y, \theta_m\right.\right). \tag{2}$$

This implies that $\hat{\beta}$ can be interpreted as a semi-elasticity (Wooldridge 2002, pp. 17–18): if 1 % of the area of an RBMU changes from forest to the nonforest land use represented by L, then the expected number of flood events changes by $\left(100 \cdot \hat{\beta}\right)$ %. An elasticity can be obtained by multiplying $\hat{\beta}$ by the mean of L: if L increases by 1 %, then the expected number of flood events changes by $\left(\hat{\beta} \cdot \bar{L}\right)$ %.

Given our objective to identify the effect of deforestation on flood events, our overriding concern was to minimize omitted-variables bias: the risk that the land-use variables were proxying for factors omitted from the model. Instead of using a goodness-of-fit statistic (e.g., AIC or BIC) to develop a model that parsimoniously explained flood variation, we therefore erred on the side of including a large number of controls in our model, through the covariates in $\mathbf{X}$ and especially the fixed effects.[5] This approach can cause overfitting in small samples,[6]

[5] If our objective had been to investigate underlying hydrological processes, then we would have needed to include potentially confounding factors directly as covariates instead of using fixed effects to sweep away their effects.

[6] One type of overfitting in panel models occurs when fixed effects for cross-sectional units are included at a level that is more finely disaggregated than necessary for identifying the effect of interest; it causes standard

but the number of degrees of freedom in our models was large both absolutely and relative to the number of observations: the smallest number of degrees of freedom in any of our models was 579 (91 % of the observations in the sample).[7]

Standard errors in Poisson models tend to be biased downward (i.e., underestimated) when the conditional variance exceeds the conditional mean (overdispersion; Wooldridge 2002, p. 651). The presence of overdispersion has sometimes led researchers to reject the Poisson model in favor of the negative binomial model (Hilbe 2011; Cameron and Trivedi 2013). Overdispersion does not affect the consistency of parameter estimates in the Poisson model, however. As an alternative to using the negative binomial model to address overdispersion, one can therefore retain the Poisson model but use robust standard errors (Huber 1967; White 1980) instead of conventional Poisson standard errors (Wooldridge 2002, pp. 649–651). We followed this approach, which has the added advantage of correcting the standard errors for heteroskedasticity.

The rationale for using the Poisson model instead of the negative binomial model is even stronger when the model includes fixed effects for the panels (RBMUs in our models). There are two variants of the fixed-effects (FE) negative binomial model: the unconditional model, which includes fixed effects directly as covariates (i.e., dummy variables), and the conditional model, which instead uses a sufficient statistic to condition them out of the likelihood function (Hilbe 2011, pp. 468–478). Both are inferior to the FE Poisson model, for different reasons. In the case of the unconditional model, parameter estimates are biased and inconsistent and standard errors are underestimated when the number of panels is large. Hilbe recommends using the conditional model instead of the unconditional model when the number of panels exceeds 20. As discussed in the next section, our dataset included 31 RBMUs.

Hilbe also cautions, however, that the conditional FE negative binomial model is not a true fixed-effects model that controls fully for unobserved time-invariant factors. In contrast, the unconditional and conditional FE Poisson models are statistically identical to each other, yield consistent parameter estimates when the number of panels is large, and are true fixed-effects models (Hilbe 2011, pp. 473–474). Although we therefore preferred the FE Poisson model, as a robustness check we compared results from it to results from the unconditional FE negative binomial model.

The Poisson and negative binomial models share a problem when they are applied to count-data variables that include more zeros than expected for the corresponding distributions. An "excess" number of zeros is a common cause of overdispersion, but it can have a more serious consequence than underestimated standard errors: it can also cause parameter estimates to be biased and inconsistent (Kennedy 2008, p. 254). This happens when the data-generating process that determines whether a nonzero count is observed differs from the process that determines the observed count structure for the nonzero values. This problem can be addressed by estimating a logit-Poisson hurdle model, which uses a logit model to describe the first process (with all nonzero observations coded as unity) and a zero-truncated

Footnote 6
errors to be underestimated but does not affect the consistency of parameter estimates (Ritschl 2009). In our models, this type of overfitting would occur if RBMUs could be classified into homogeneous groups, in which case fixed effects should be included for groups instead of individual RBMUs. The RBMUs do not fall into obvious groups, however, as they differ greatly in area, topography, and other fixed characteristics. Moreover, this type of overfitting occurs only if the number of fixed effects is large relative to the number of observations, which was not the case in our models (31 RBMU fixed effects, 636–716 observations).

[7] Including a large number of controls can also cause multicollinearity, but given our objective this was a lesser evil than omitted-variables bias. Multicollinearity inflates standard errors, but it does not bias parameter estimates (Kennedy 2002, pp. 193–194). Hence, it contributes to a conservative estimation strategy by reducing the risk of overestimating parameter significance.

Poisson model to describe the nonzero count structure in the second process (Mullahy 1986; Hilbe 2011, pp. 354–370). We examined the distributions of our two count-data variables, and we used deviance and Pearson goodness-of-fit statistics to test the FE Poisson models for overdispersion. If these investigations indicated an excess number of zeros, then we reestimated the models using a FE logit-Poisson hurdle model.

Another class of count-data models, zero-inflated models, can also be applied to data with excess zeros (Lambert 1992; Hilbe 2011, pp. 370–382). Like the hurdle models, these models assume two data-generating processes. A critical difference, however, is that the zero-inflated models impose more structure on the data by assigning some zeros to the first process and others to the second process. This approach makes more efficient use of the data when there are sound priors for distinguishing the two types of zeros, but otherwise it runs the risk of biased and inconsistent estimates of the parameters in the first process. We opted for the less restrictive hurdle model to avoid imposing additional assumptions on the data, but as a robustness check we compared results from it to results from a zero-inflated Poisson model.

2.2 Definitions of Variables and Data Sources

The regression sample spanned 1984–2000, which was determined by data availability. Each observation was a particular month of the core wet season in a particular RBMU. The core wet season was defined as the period when rainfall associated with the northeast monsoon is typically heaviest: September and October for RBMUs in the northwestern Peninsula, October and November for other west-coast RBMUs, and November and December for east-coast RBMUs.

We manually recorded RBMU-level data on monthly flood events (floods, days flooded) from a 9-volume report that compiled and synthesized data from the 1980–2000 annual flood reports prepared by the DID's 11 state offices in the Peninsula (KTA Tenaga 2003). Data were available for a varying number of years across the states (median = 14 years; range 5–19 years). We recorded data for 35 RBMUs, which accounted for 97 % of the Peninsula's surface area. We then removed observations for time periods that occurred after the completion of known river-engineering projects. These projects could potentially confound the effects of deforestation. For example, an absence of flood events in a deforested RBMU could result from a successful flood-mitigation project instead reflecting the lack of an effect of deforestation. We used multiple sources to identify different types of river-engineering projects: dams for flood control, water supply, and irrigation (Drainage and Irrigation Department 2013); hydroelectric dams (Energy Commission Malaysia 2013); miscellaneous dams shown on online DID maps (Drainage and Irrigation Department 2010); and flood-mitigation projects (KTA Tenaga 2003).

This data-cleaning led to four RBMUs being dropped from the sample, and it reduced the number of observations for several others. The final 31 RBMUs accounted for 78 % of the Peninsula's area. They ranged in size from 560 to 29,300 km^2, with a mean area of 3,337 km^2 (standard deviation = 5,410 km^2). Twenty-three of them were larger than 1,000 km^2. The final number of observations in the regression models was 716 for floods and 636 for days flooded. These totals differ because information on duration was missing for some floods.

The Malaysian Department of Agriculture has conducted mid-decadal land-use surveys since 1966–1967 (Wong 1971). The department granted us restricted access to GIS layers from the 1984–1985, 1997–1998, and 2004–2005 surveys. This enabled us to measure the area of each RBMU in five land uses—inland forest, wetland forest, oil palm, rubber, and urban—with a sixth, residual category of other uses, mainly crops other than oil palm and rubber. Inland forest refers to tropical rainforests located mostly in hilly or mountainous interior regions. The coastal plains of the Peninsula were once covered by lowland dipterocarp rainforests, but most of this forest type had already been converted to nonforest land uses by 1984–1985 (Vincent and Hadi 1993; Vincent and Mohamed Ali 2005). Wetland forest refers to peat-swamp forests, which are found in permanently wet (but not saline) areas near the coast, and mangroves, which grow in the intertidal zone. Both inland and wetland forests include a mix of virgin and logged forests, which the land-use surveys do not distinguish. We linearly interpolated the areas of all six land uses for intervening years between the survey dates, and we expressed them as percentages of an RBMU's area.

The detail on land-use type enabled us to estimate a series of three models that revealed the marginal effect of deforestation on flood events in a progressively more detailed manner. The first model included a single, aggregate nonforest land-use variable. It provided an estimate of the marginal effect averaged across all types of forest and all land uses to which forests were converted. The second model included the four nonforest land uses as separate variables. It provided estimates of marginal effects differentiated by the type of land use to which forests were converted, but still averaged across all types of forest. The third model added the wetland forest percentage as a fifth land-use variable. This changed the definition of the reference forest category (i.e., the land use that was omitted from the model) from all forest to inland forest. The third model thus provided estimates of the marginal effects of deforestation of inland forests differentiated by the type of land use to which such forests were converted.

Comparing the results of the second and third models enabled us to investigate whether the marginal effects of deforestation were the same for all types of forest or differed between inland and wetland forests. The tropical forest hydrology literature provides reason to expect the effects to differ. In the case of inland forests, it suggests three mechanisms whereby forests can mitigate flooding, if those mechanisms are sufficiently strong and are not overwhelmed by prolonged, heavy rainfall (Bruijnzeel 2004). The first is soil retention. Rates of erosion tend to be lower for tropical forests than for perennial or annual crops. Deeper soils are able to store more water, and lower erosion typically reduces the amount of sediment in rivers and thus maintains channel depth so that rivers can receive more runoff before overflowing their banks. The second mechanism is infiltration. The land surface is typically more permeable in forests than in other land uses, both agricultural and especially urban, where much of the surface is paved. In forests, a larger proportion of rainfall thus enters the soil instead of running off quickly. The third mechanism is evapotranspiration. Compared to other vegetation, forests remove more water from the soil through transpiration, thus enhancing soil water storage by creating a larger soil moisture deficit, and they provide a larger surface area of leaves and stems from which rain can evaporate before reaching the surface.

The hydrology of tropical wetland forests is not as well understood, but research in Malaysia suggests that converting them to other uses might decrease flooding instead of increasing it. Results of a study of a peat-swamp forest in the west coast state of Selangor suggested that "instead of the peat swamp forest holding back excess water, it allows the runoff to gush out rather quickly during storms" (Zulkifli et al. 2004, p. 111). The explanation is that peat-swamp forests are located in flat coastal terrain and are fully saturated; hence, rainfall tends to flow out instead of being retained. Drainage structures associated with conversion can reduce the water level in these forests and thus reduce the rate of outflow during rainfall events. The same study reported that as a result of such "drainage alteration," "the extent and frequency of floods might have been considerably reduced" (p. 116).

The interpolated area estimates undoubtedly contained measurement error, as actual land uses surely did not change smoothly from year to year. Random measurement error tends to bias regression coefficients toward zero, and fixed effects tend to amplify this attenuation bias (Wooldridge 2002, pp. 73–76, 311–313). There is no apparent reason to expect interpolation to cause nonrandom measurement error,[8] and so it follows that the interpolation of land-use areas caused our estimates of the effects of deforestation on flood events to be conservative (i.e., underestimated).

The annual frequency of the interpolated land-use variables differed from the monthly frequency of the flood-event variables. Conventional robust standard errors can severely underestimate true standard errors when a covariate varies at a lower frequency than the dependent variable (Moulton 1986). Unlike the underestimation of parameters, this is not a conservative bias, as it exaggerates the significance of the covariate's effect on the dependent variable. This problem can be addressed by clustering standard errors at the lower frequency, which in our case meant clustering them by year for each RBMU.[9] Clustering is an asymptotic correction that requires a large number of clusters, with 40–50 clusters being the rule-of-thumb (Angrist and Pischke 2009, Ch. 8). The number of clusters was 364 in the model for floods and 350 in the model for days flooded. Clustering also corrected for serial correlation in the errors between the two core wet-season months within a given year (Zeger and Liang 1986).[10]

Although fixed effects control for many factors that could confound the marginal effect of deforestation on flood events, they obviously do not control for all of them. In particular, they do not control for factors that vary over time in different ways across RBMUs. One such factor is rainfall, which can affect not only flooding but also deforestation (e.g., by impeding logging and burning when forests are cleared). Another important factor is flood recording. Floods are more likely to be recorded in RBMUs that have more people to report them and more river-monitoring stations to detect elevated streamflow. These RBMUs are likely to be ones with less forest, both because human activities drive deforestation and because river monitoring tends to be more intensive in locations with more people and more economic activity. Omission of controls for rainfall and flood recording could thus result in biased and inconsistent estimates.

For these reasons, our models included covariates for rainfall, population density, and river monitoring. All three variables varied by RBMU, with rainfall and river monitoring varying by both year and month and population density varying by year. To construct the rainfall variable, we downloaded monthly data for 121 DID rainfall stations (Drainage and Irrigation Department 2010) and obtained data for two additional stations directly from the department. We chose stations with complete or nearly complete data series that were located upstream from or near flood-prone areas of each RBMU, based on maps in the DID synthesis report (KTA Tenaga 2003). The rainfall variable used in the model was the simple average of the readings across the stations for a given month of a given year in a given RBMU. The number of rainfall stations in most RBMUs was too small to justify using spatial interpolation (e.g., kriging) to generate more precise estimates. We estimated the annual population density in each RBMU by aggregating and interpolating district-level data from the 1980, 1991, and 2000 Malaysian Population and Housing Censuses (Department of Statistics 1981, 1992, 2001) and dividing by RBMU area. We used online information to identify river-monitoring stations operated by the DID in each RBMU (Drainage and Irrigation Department 2010).

[8] Similarly, given that the Department of Agriculture conducts the land-use surveys using the same procedures in all parts of Peninsular Malaysia, there is no apparent reason to expect measurement error in the 1984–1985, 1997–1998, and 2004–2005 surveys to be nonrandom across RBMUs or land uses.

[9] Clustering is a second reason that we did not use the AIC or BIC to guide model specification: these statistics should not be used when data have a clustered structure (Hilbe 2011, p. 69).

[10] The 10-month gap from 1 year to the next meant that we did not need to cluster the standard errors at an even lower frequency, e.g., the entire sample period for a given RBMU, because the gap interrupted the serial correlation process between years.

Table 1 Descriptive statistics for sample in Table 3 (models for days flooded). No. observations: 636

Variable	Mean	Std. Dev.	Minimum	Maximum
RBMU area (km^2)	3,343	5,234	554	28,643
Dummy: West Coast	0.50	0.50	0	1
Dummy: East Coast	0.50	0.50	0	1
No. floods per month	0.18	0.42	0	3
No. days flooded per month	1.11	3.36	0	31
Flood deaths per month	0.07	0.75	0	13
Flood evacuees per month	122	941	0	16,690
% Forest	48.2	20.9	11.2	91.1
% Inland forest	36.8	20.9	5.1	85.0
% Wetland forest	11.5	11.7	0.1	58.7
% Nonforest	51.8	20.9	8.9	88.8
% Oil palm	19.4	13.2	0.0	54.1
% Rubber	17.4	13.8	0.1	59.3
% Urban	2.2	2.4	0.0	11.4
% Other nonforest	12.7	10.2	2.2	60.5
Rainfall per month (mm)	354	268	31	1,956
Dummy: 1st month	0.52	0.50	0	1
Population per km^2	154	107	12	484
New stations	0.60	1.29	0	7

This information included station establishment dates. We set the river-monitoring variable equal to the cumulative number of new stations in each RBMU since 1983.

The fourth and final time-varying control in the regression models was a dummy variable for the first month of the core wet season. Flood events should be less common in that month, because the soil is less saturated. Because the months in the core wet season differed across RBMUs, this variable varied by RBMU and month but not by year.

Table 1 shows descriptive statistics for the 636-observation sample used in the models for days flooded.[11]

3 Results

3.1 Changes in Land Use Between 1984–1985 and 2004–2005

Figure 1 displays the 1984–1985 and 2004–2005 land-use survey data for the entire Peninsula. As can be seen, inland forests accounted for much more of total forest area than did wetland forests: 88 % in 1984–1985, and 93 % in 2004–2005. Not surprisingly then, most deforestation during the 20-year interval occurred in inland forests, 76 % of the total area deforested. The area of inland forest declined from 52 % of the Peninsula's area in 1984–1985 to 44 % in 2004–2005. The annual deforestation rate for the two forest types combined was 0.77 %/year, which was nearly four times the global rate during 1990–2000, 0.20 %/year (FAO 2010, Table 2.4).

[11] Descriptive statistics differed little for the 716-observation sample used in the models for number of floods.

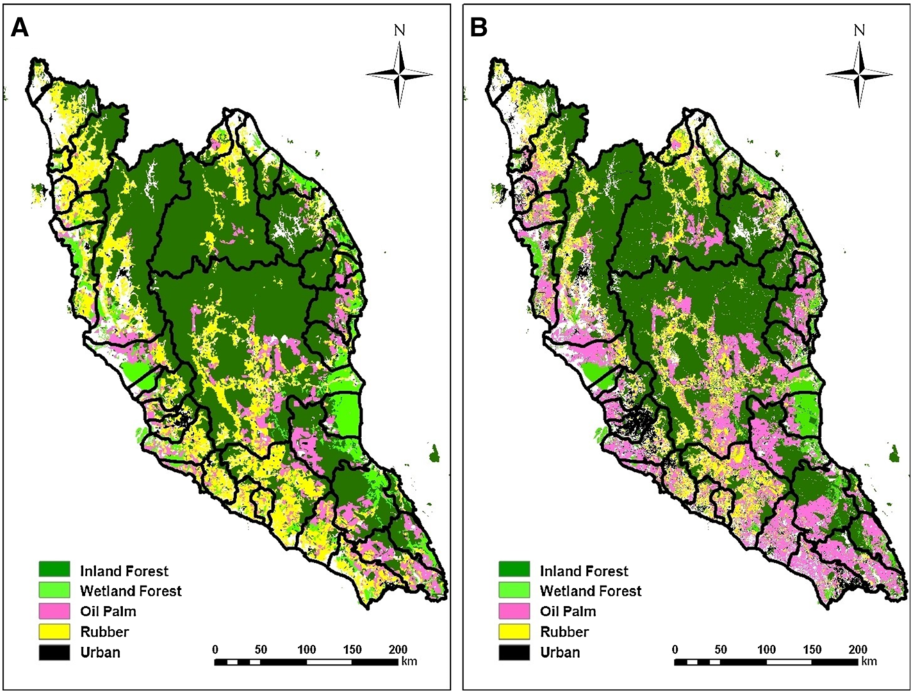

Fig. 1 Land use in Peninsular Malaysia in 1984–1985 (**a**) and 2004–2005 (**b**). *Black lines* show river basin management unit (RBMU) boundaries

Consistent with prior research on deforestation in Peninsular Malaysia (Vincent and Hadi 1993), the land-use surveys indicate that oil palm was the principal land use that replaced inland forests. Forty-five percent of the loss of inland forest between 1984–1985 and 2004–2005 was due to conversion to oil palm. Twenty-two percent was due to conversion to rubber, which had been the leading cause of deforestation earlier in the twentieth century (Vincent and Hadi 1993). The expansion of oil palm came at the expense of not only forest but also rubber, whose area decreased from 15 % of the Peninsula's area in 1984–1985 to 10 % in 2004–2005, while oil palm's area increased from 9 % to 20 %.

Forest trends in the 31 RBMUs were very similar to the trends in the overall Peninsula: for example, inland forests accounted for 88 % of forest area in 1984–1985 and 90 % in 2004–2005, and 76 % of deforestation between the two dates; and the area of inland forest declined from 54 % of the aggregate area of the 31 RBMUs in 1984–1985 to 48 % in 2004–2005. Oil palm plantations were already established in all 31 RBMUs in 1984–1985, and they covered 11 % of the aggregate area at that point. By 2004–2005, their area had increased in every RBMU, and they covered 22 % of aggregate area. Forty-two percent of the incremental oil palm area came from sites that were in rubber in 1984–1985, while 36 % came from inland forest.

3.2 Effects of Deforestation on Monthly Number of Floods

Table 2 shows Poisson regression results for the series of three models described in Sect. 2.2 applied to the monthly number of floods. The models differ only in terms of the land-

Table 2 Effect of land use on number of floods per month: Poisson regression models

Variables[a]	(1)[b]	(2)[c]	(3)[d]
% Nonforest	−0.0691		
	(0.170)		
% Oil palm		−0.0170	0.00771
		(0.786)	(0.906)
% Rubber		−0.146	−0.126
		(0.104)	(0.150)
% Urban		−0.329*	−0.262
		(0.098)	(0.212)
% Other nonforest		0.0109	0.0325
		(0.935)	(0.814)
% Wetland forest			0.0825
			(0.350)
ln(Rainfall)	1.14***	1.12***	1.13***
	(0.000)	(0.000)	(0.000)
Dummy: 1st month	0.309	1.15	0.864
	(0.477)	(0.536)	(0.626)
ln(Population density)	−1.73*	−1.37	−1.39
	(0.067)	(0.167)	(0.169)
New stations	−0.0352	−0.0189	−0.0315
	(0.730)	(0.868)	(0.787)
FE[e]: RBMUs	Y	Y	Y
FE: Years	Y	Y	Y
FE: Months	Y	Y	Y
P value, deviance GOF[f]	1.000	1.000	1.000
P value, Pearson GOF	0.998	0.994	0.996
Observations	716	716	716
Log-likelihood	−364.6	−362.4	−362.1

Asterisks indicate commonly used significance levels: *** $P < 0.01$; ** $P < 0.05$; * $P < 0.1$

[a] *P* values are shown in parentheses below parameter estimates. They refer to two-sided *z*-tests of null hypothesis that parameter estimates equaled zero and are based on robust standard errors clustered by RBMU-year (364 clusters)

[b] Aggregate nonforest land use; reference forest type = all forest

[c] Disaggregated nonforest land use; reference forest type = all forest

[d] Disaggregated nonforest land use; reference forest type = inland forest

[e] *FE* fixed effects, *Y* yes

[f] *GOF* goodness-of-fit statistic. Null hypothesis: conditional mean = conditional variance

use variables, with model (1) including a single, aggregate nonforest variable; model (2) decomposing the aggregate nonforest variable into its four components (oil palm, rubber, urban, other); and model (3) including wetland forest in addition to the four nonforest variables, which has the effect of making inland forest the reference forest type. *P* values for the deviance and Pearson overdispersion tests are given at the bottom of the table, and they indicate that the data are conditionally Poisson distributed. A close match between the observed distribution of floods in the regression sample and a Poisson distribution can also be seen in Fig. 2. We therefore did not reestimate the models using the hurdle estimator.

P values are large for nearly all land-use variables in all three models, indicating little evidence of significant effects. The exception is the urban variable in model (2), which has a *P* value less than 0.1 and a negative effect: conversion of forest to urban use was associated with a reduced number of floods. The significance of this variable fell, however, when wetland forest was added as a separate variable in model (3). This loss of significance suggests that conversion of wetland forests, but not inland forests, to urban use might have reduced the number of floods. This is consistent with the Malaysian hydrological research cited earlier (Zulkifli et al. 2004).

Of the remaining variables, the clearest evidence of a significant effect is for rainfall. Because the rainfall variable was logarithmically transformed, the parameter estimate on it can be interpreted as an elasticity. All three estimates are slightly greater than one, which indicates that flooding increased more than proportionately with rainfall.

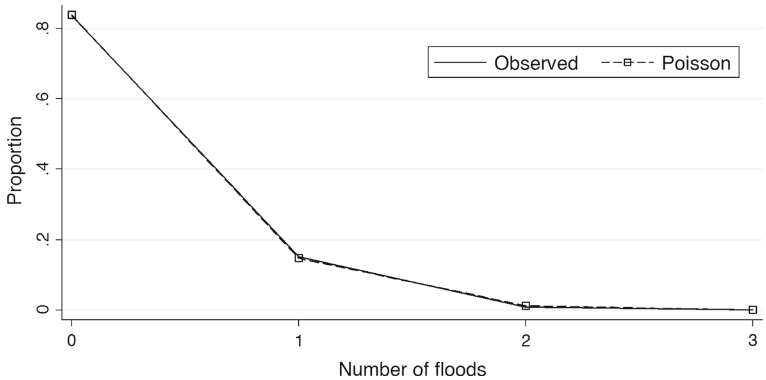

Fig. 2 Share of observed number of months in the regression sample ($n = 716$) with the indicated number of floods, compared to shares predicted by a Poisson distribution with the same mean (0.272 floods per month). Variance of number of floods in the regression sample = 0.232, which is similar to the mean. No month in the sample had more than three floods

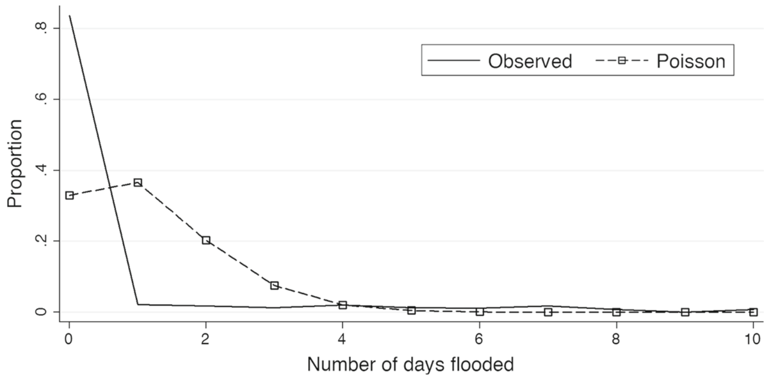

Fig. 3 Share of observed number of months in the regression sample ($n = 636$) with the indicated number of days flooded, compared to shares predicted by a Poisson distribution with the same mean (1.11 days flooded per month). Variance of number of days flooded in the regression sample = 11.3, which exceeds the mean and indicates that the data are overdispersed. Twenty-two months in the sample had more than 10 days flooded; the maximum was 31. Note that the share of zeros in the sample is much larger than the number predicted by the Poisson distribution

3.3 Effects of Deforestation on Monthly Number of Days Flooded

Figure 3 shows the distribution of days flooded per month for the regression sample. It reveals clear evidence of an unusually large number of zeros compared to the expected number for a Poisson distribution. Hence, it is not surprising that overdispersion tests in Poisson regression models for days flooded corresponding to models (1)–(3) in Table 2 led us to reject the null that the data were conditionally Poisson distributed ($P = 0.000$ for all six tests). We therefore reestimated the models using the hurdle estimator.

Table 3 shows the hurdle-model results. The first three columns show results for the logit component, which predicts whether a month had a nonzero number of days flooded. As in Table 2, P values are large for most of the land-use variables. The exception is again the urban variable in model (2), which now has a P value less than 0.05. The P value rises above

Table 3 Effect of land use on number of days flooded per month: hurdle regression models

Variables[a]	Component 1: logit model			Component 2: zero-truncated Poisson		
	(1)[b]	(2)[c]	(3)[d]	(1)	(2)	(3)
% Nonforest	−0.0580			0.0214		
	(0.592)			(0.711)		
% Oil palm		0.105	0.180		0.199*	0.268***
		(0.433)	(0.244)		(0.065)	(0.002)
% Rubber		−0.203	−0.136		0.190*	0.265***
		(0.295)	(0.515)		(0.051)	(0.002)
% Urban		−0.859**	−0.731		0.0517	0.359*
		(0.036)	(0.106)		(0.811)	(0.094)
% Other nonforest		0.00870	0.000692		−0.302*	−0.102
		(0.977)	(0.998)		(0.061)	(0.366)
% Wetland forest			0.171			0.380***
			(0.441)			(0.000)
ln(Rainfall)	3.28***	3.33***	3.34***	0.220	0.278	0.218
	(0.000)	(0.000)	(0.000)	(0.371)	(0.280)	(0.387)
Dummy: 1st month	−1.01**	−1.05**	−1.04**	−0.124	−0.0180	−0.0236
	(0.013)	(0.016)	(0.017)	(0.420)	(0.917)	(0.882)
ln(Population density)	−3.03	1.13	2.01	1.26	3.84**	4.97***
	(0.207)	(0.685)	(0.492)	(0.167)	(0.012)	(0.001)
New stations	−0.116	−0.0198	−0.0438	−0.251***	−0.136	−0.219**
	(0.545)	(0.926)	(0.842)	(0.0041)	(0.219)	(0.021)
FE[e]: RBMUs	Y	Y	Y	Y	Y	Y
FE: Years	Y	Y	Y	Y	Y	Y
FE: Months	Y	Y	Y	Y	Y	Y
Observations	636	636	636	636	636	636
Log-likelihood	−416.0	−405.3	−400.6	−416.0	−405.3	−400.6

Asterisks indicate commonly used significance levels: *** $P < 0.01$; ** $P < 0.05$; * $P < 0.1$

[a] P values are shown in parentheses below parameter estimates. They refer to two-sided z-tests of null hypothesis that parameter estimates equaled zero and are based on robust standard errors clustered by RBMU-year (350 clusters).

[b] Aggregate nonforest land use; reference forest type = all forest

[c] Disaggregated nonforest land use; reference forest type = all forest

[d] Disaggregated nonforest land use; reference forest type = inland forest

[e] *FE* fixed effects, *Y* yes

0.1 in model (3) after wetland forest is added, but it is less than the corresponding P value in Table 2. Evidence that conversion of wetland forest to urban use reduced flood events is thus somewhat stronger than in Table 2.

The similarity of the signs and significance levels for the land-use variables in the Poisson models in Table 2 and the logit component of the hurdle models in Table 3 is not surprising.[12] As Fig. 2 shows, very few months had more than one flood (nine had two floods, and one had three), so the dependent variable in the Poisson model in Table 2 has mostly values of zero or one. It thus differs little from the binary dependent variable in the logit model in Table 3. Another similarity is that rainfall has a highly significant effect in both Table 2 and the logit portion of Table 3. The only notable difference is that the number of days flooded is now significantly lower (P = 0.013–0.017) during the first month of the wet season, which is expected because the ground is less saturated at the beginning of the core wet season than at the end.

[12] The parameter estimates cannot be directly compared, as they are interpretable as semi-elasticities in the Poisson model and marginal effects on log odds ratios in the logit model.

Reduced measurement error is the likely explanation for the greater significance of the logit results compared to the Poisson results in Table 2. The sample in Table 3 excludes 80 observations from the sample in Table 2 that lacked information on flood duration. This missing information could signal that other information on flooding during those months was also not recorded accurately or completely, including information on the number of floods. This measurement error in the dependent variable would be expected to increase the standard errors in the models in Table 2 (Wooldridge 2002, pp. 71–72).

The last three columns of Table 3 show results for the zero-truncated Poisson component of the hurdle model. There is no evidence of a significant effect of deforestation on days flooded in model (1). Evidence is stronger in model (2), with P values for three of the four disaggregated land uses lying between 0.05 and 0.1. Results for model (2) also reveal that the insignificance of the aggregate nonforest variable in model (1) is a consequence of offsetting effects of the disaggregated land uses: conversion of forest to oil palm and rubber had a positive effect (days flooded increased), while conversion to other nonforest, nonurban land uses had a negative effect (days flooded decreased).

Results for model (3) further reveal that conversion of inland forests, not wetland forests, was primarily responsible for the positive effects of oil palm and rubber: adding wetland forest to the model causes these effects to increase sharply in magnitude and become highly significant. The parameter estimates indicate that converting 1 % of an RBMU from inland forest to oil palm increased the number of days flooded by 26.8 %, while converting 1 % from inland forest to rubber increased it by a statistically identical amount, 26.5 %.[13] Multiplying the estimates by the sample means for the oil palm and rubber variables (19.4 and 17.4, respectively), yields the following elasticities: a 1 % increase in oil palm area increased the number of days flooded by 5.2 %, while a 1 % increase in rubber area increased it by 4.6 %. The number of days flooded per month thus increased much more than proportionately with the area of these crops.

Wetland forest has a positive and highly significant effect in model (3), which implies that deforestation *reduced* flooding when the forests converted were wetland forests.[14] The effect was large (semi-elasticity = 38.0 %). As suggested earlier, improved drainage when wetland forests were converted could explain why the effect was the opposite of that for inland forests. Conversion of inland forests to urban use also had a large effect (semi-elasticity = 35.9 %), but it was not estimated very precisely ($P = 0.094$).

Regarding the other variables in the second component of model (3), population density had the expected positive effect, presumably due to increased flood reporting. The cumulative number of new monitoring stations had an unexpected negative effect, however, and it was significant at $P = 0.021$. A possible explanation is that rivers in economically important areas were not only monitored more closely but were also the locations of flood-mitigation projects that were not fully recorded in the sources we consulted. If so, then the expected positive effect of the monitoring-stations variable due to increased flood recording could have been outweighed by the negative effect of the variable proxying for unobserved flood-mitigation projects. Both rainfall and the first-month dummy were much less significant than in the logit component. This loss of significance suggests different roles for hydrological and socioeconomic factors in the flooding process: flooding occurred only under particular hydrological

[13] These effects refer to months when flooding occurred, not all months in the sample, as they are based on coefficients from the zero-truncated component of the hurdle model.

[14] One RBMU contained a much higher percentage of wetland forest than the others: 59 % as of 1984, with the next highest percentage being 32 %. Results changed little if we excluded that RBMU from the sample. For example, the coefficients on oil palm, rubber, and wetland forest in model (3) when estimated for that sample were 0.280 ($P = 0.001$), 0.275 ($P = 0.002$), and 0.312 ($P = 0.003$).

conditions (high rainfall, saturated soils), but if those conditions were present, then the extent of flooding (as measured by days flooded) was influenced primarily by socioeconomic factors (land use, population density, flood monitoring and mitigation efforts).[15]

3.4 Omitted Variables Bias and Alternative Estimators

Table 4 sheds light on the bias that would have occurred in the second component of the hurdle model if we had controlled less carefully for confounding factors. Models (4)–(6) are identical to model (3) in Table 3, except they exclude particular groups of controls: model (4) excludes the RBMU fixed effects; model (5) excludes the fixed effects for years and months; and model (6) excludes the time-varying covariates. Omitting either group of fixed effects severely attenuates the effects of nearly all of the land-use variables. For example, the effects of the oil palm and rubber variables are one to two orders of magnitude smaller than in model (3) in Table 3. Omitting the time-varying covariates does not bias the effects nearly as much, although the effects of oil palm and wetland forest are much smaller than in Table 3 while the urban effect is much larger.[16] Collectively, these results confirm that identifying the effect of deforestation on days flooded requires controlling carefully for confounding factors.

Model (7) is likewise identical to model (3) in Table 3, except the regression sample now includes observations from RBMUs with known river-engineering projects. Compared to the sample in Table 3, this sample includes more RBMUs (35 instead of 31) and more observations (759 instead of 636). The estimated effects of oil palm and rubber are very similar to those for model (3) in Table 3, but the estimated effects of the other land uses change substantially, with some increasing and some decreasing. Failing to control for river-engineering projects thus can bias the estimated effects of deforestation on days flooded in divergent ways that depend on the land use to which forests are converted.

Table 5 shows the sensitivity of the results to the estimator applied to the data. Results are shown for three alternative estimators: a logit-negative binomial hurdle model (model (8)), a zero-inflated Poisson model (model (9)), and a regular Poisson model (model (10)). With one exception, the models are otherwise identical to version (3) of the logit-Poisson hurdle model in Table 3. The exception is the zero-inflated Poisson model, which requires selecting variables to explain the excess zeros. We selected the variables with $P < 0.05$ in the logit component of model (3) in Table 3: rainfall, the first-month dummy, and the monthly fixed effects.

[15] We also considered the effect of using samples that defined wet months in ways other than the core wet season. In one variant, we expanded the sample by extending the wet season to include the preceding and following months. One would expect this to reduce the absolute value of the regression coefficients on the land-use variables, given that the coefficients reflect mean monthly effects on days flooded and that extending the sample causes those effects to be averaged over a period that includes months when there is less rainfall, less saturated soils, and thus less flooding. (In the extreme case of no flooding at all, the effects of land use on days flooded during those months would necessarily be zero.) This is what we found; for example, the coefficients on oil palm and rubber were reduced to 0.147 and 0.155, respectively ($P < 0.05$ for both). In a second variant, we ignored seasonality and included months with rainfall in the 70th percentile. This yielded a sample with 670 observations, similar to the samples in Table 3. Coefficients on the land-use variables retained their signs but became more significant and smaller: the positive ones became less positive, and the negative one became more negative. The coefficients on oil palm and rubber were 0.192 and 0.212, respectively ($P < 0.001$ for both).

[16] Flooding could be affected by time-varying factors other than the three included in model (3). One example is regional income; another is forest management, which under the Malaysian constitution is the responsibility of state governments. To investigate the influence of such factors, we added annual trends for the nine states in the sample to model (3). With the exception of the "other" land-use variable, which remained negative and insignificant, coefficients on all the other land-use variables increased. For example, the coefficients on oil palm and rubber were 0.385 and 0.602, respectively ($P < 0.01$ for both).

Table 4 Investigating omitted variables bias in hurdle regression models for number of days flooded per month

Variable[b]	Component 2: zero-truncated Poisson[a]			
	(4)[c]	(5)[d]	(6)[e]	(7)[f]
% Oil palm	−0.00288	−0.0149	0.127*	0.311***
	(0.749)	(0.738)	(0.070)	(0.001)
% Rubber	0.0320***	0.00348	0.268***	0.281***
	(0.000)	(0.955)	(0.000)	(0.000)
% Urban	−0.0797*	0.0389	0.642***	0.0532
	(0.095)	(0.822)	(0.001)	(0.564)
% Other nonforest	0.00129	−0.140	−0.189	−0.413**
	(0.824)	(0.257)	(0.205)	(0.013)
% Wetland forest	−0.0198	0.255**	0.201*	0.0705
	(0.112)	(0.014)	(0.099)	(0.647)
ln(Rainfall)	0.292	0.421**		0.555**
	(0.152)	(0.030)		(0.012)
Dummy: 1st month	0.422*	−0.237*		−0.0915
	(0.088)	(0.096)		(0.630)
ln(Population density)	−0.216	2.69**		4.61***
	(0.221)	(0.031)		(0.001)
New stations	−0.110*	−0.235**		0.142
	(0.081)	(0.012)		(0.274)
FE[g]: RBMUs	N	Y	Y	Y
FE: Years	Y	N	Y	Y
FE: Months	Y	N	Y	Y
RBMUs with known river-engineering projects	N	N	N	Y
Observations	636	636	636	759
Log-likelihood	−518.4	−443.5	−458.7	−597.6

[a] Results are not shown for the logit component of the hurdle model, which included the same variables as the zero-truncated Poisson component

[b] *P* values are shown in parentheses below parameter estimates. They refer to two-sided *z*-tests of null hypothesis that parameter estimates equaled zero and are based on robust standard errors clustered by RBMU-year (416 clusters in model (7), 350 clusters in the others)

Asterisks indicate commonly used significance levels: *** $P < 0.01$; ** $P < 0.05$; * $P < 0.1$

[c] Model (4) differs from model (3) in Table 3 by omitting fixed effects for RBMUs

[d] Model (5) differs from model (3) in Table 3 by omitting fixed effects for years and months

[e] Model (6) differs from model (3) in Table 3 by omitting time-varying controls

[f] Model (7) differs from model (3) in Table 3 by including observations from RBMUs with known river-engineering projects

[g] *FE* fixed effects, *Y* yes, *N* no

The results for the logit-negative binomial hurdle model in Table 5 differ little from those for the corresponding logit-Poisson hurdle model in Table 3. Results differ more for the zero-inflated Poisson model, which is not surprising given that this model imposes more restrictive assumptions to deal with the excess zeros problem. Nevertheless, the effects of the land-use variables in this model show the same patterns of signs, significance levels, and relative magnitudes as in Table 3. Results differ the most for the regular Poisson model, with the effect of rubber decreasing sharply, the effect of the other nonforest land use variable increasing sharply, and the effect of urban land use reversing sign. These results illustrate the severity of the bias that can occur if a Poisson model is used when the number of zeros exceeds the number expected for a conditional Poisson distribution. They also indicate, however, that our findings on the effect of deforestation on days flooded are not especially sensitive to the particular estimator chosen to address that problem: a logit-

Table 5 Alternative estimators for modeling the effect of land use on number of days flooded per month

Variables[a]	(8)[b]	(9)[c]	(10)[d]
% Oil palm	0.269***	0.322***	0.335**
	(0.002)	(0.007)	(0.015)
% Rubber	0.270***	0.302***	0.0767
	(0.002)	(0.003)	(0.504)
% Urban	0.375*	0.431	−0.421
	(0.076)	(0.112)	(0.138)
% Other nonforest	−0.0897	−0.0733	−0.358
	(0.432)	(0.557)	(0.170)
% Wetland forest	0.403***	0.504***	0.244
	(0.000)	(0.000)	(0.204)
ln(Rainfall)	0.235	0.192	2.33***
	(0.366)	(0.513)	(0.000)
Dummy: 1st month	−0.0260	−7.39***	−9.17***
	(0.871)	(0.001)	(0.006)
ln(Population density)	5.00***	6.09***	5.41**
	(0.001)	(0.001)	(0.023)
New stations	−0.222**	−0.254**	−0.168
	(0.015)	(0.017)	(0.420)
FE[e]: RBMUs	Y	Y	Y
FE: Years	Y	Y	Y
FE: Months	Y	Y	Y
Observations	636	636	636
Log-likelihood	−398.9	−439.7	−792.5

Asterisks indicate commonly used significance levels: *** $P < 0.01$; ** $P < 0.05$; * $P < 0.1$

[a] P values are shown in parentheses below parameter estimates. They refer to two-sided z-tests of null hypothesis that parameter estimates equaled zero and are based on robust standard errors clustered by RBMU-year (350 clusters)

[b] Differs from model (3) in Table 3 by using a logit-negative binomial hurdle model instead of a logit-Poisson hurdle model. Results shown are for the second component of the model (i.e., the zero-truncated Poisson)

[c] Differs from model (3) in Table 3 by using a zero-inflated Poisson model instead of a logit-Poisson hurdle model. Results shown are for the count part of the model, not the part that explains zero inflation

[d] Differs from model (3) in Table 3 by using a regular Poisson model instead of a logit-Poisson hurdle model

[e] *FE* fixed effects, *Y* yes

Poisson hurdle model, a logit-negative binomial hurdle model, or a zero-inflated Poisson model.

3.5 Effects of Deforestation on Flood-Related Deaths and Evacuations

Data from the DID synthesis report indicate that the mean numbers of deaths and evacuees per day flooded in the regression sample were 0.065 and 110, respectively. This information can be combined with the regression parameters to estimate the effects of deforestation on flood-related deaths and evacuations. We focused on the effects of converting inland forests to oil palm and rubber, given these crops' dominance as causes of deforestation.

As stated earlier, the regression parameter for oil palm in model (3) in Table 3 implies that converting 1 % of an RBMU from inland forest to oil palm resulted in a 26.8 % increase in the number of days flooded per month. Combining this with the sample means for days flooded per month (1.11) and RBMU area (3,343 km^2) and recalling that the core wet season included 2 months, we estimate that converting 1 km^2 of inland forest to oil palm increased the mean number of days flooded during the core wet season by 0.018.[17] For the means given in the

[17] $2 \times (0.268 \times 1.11)/(0.01 \times 3343) = 0.018.$

previous paragraph, this implies that the additional km^2 of oil palm led to 0.0012 additional deaths and 2.0 additional evacuees per year. Parallel calculations for the conversion of 1 km^2 of inland forest to rubber yield 0.018 additional days flooded, 0.0011 additional deaths, and 1.9 additional evacuees per year.

4 Discussion and Conclusions

Brauman et al. (2007) called for more specificity in research on hydrological services provided by forests. Our findings respond to this call in three ways. First, by focusing on a particular country that plays a large role in forest loss to oil palm and rubber, we were able to create a dataset that accounted for flood events much more completely than the DFO Global Archive used by recent cross-country econometric studies. Unlike those studies, we found robust evidence that deforestation can increase the number of days flooded per month in large catchments during heavy-rainfall periods of the year. Second, whether deforestation increases days flooded depends on the land use to which forests are converted, with our analysis providing stronger evidence of an increase in the case of conversion to oil palm and rubber than in the case of conversion to urban or other nonforest uses. Third, deforestation's effects also depend on the type of forest that is converted, with inland forest, where most deforestation occurred in Peninsular Malaysia, and wetland forest having opposite effects. Our results thus demonstrate the importance of using disaggregated data on nonforest land use and forest type when studying the effects of deforestation on hydrological services.

We suspect that the lack of evidence of a significant effect of deforestation on the number of floods per month was due to an inherent problem in defining that variable, a problem which does not affect the number of days flooded per month. This problem has not been noted in prior studies that have estimated the effects of deforestation on flood frequency (e.g., Bradshaw et al. 2007; Van Dijk et al. 2009; Ferreira and Ghimire 2012; Ferreira et al. 2013), and it is most easily explained by a simple example. Suppose that three floods occur in a given RBMU in a given month, with each flood lasting 1 day and each flood separated from the prior one by one day, for a total of 3 days flooded during the 5-day interval. Now imagine the same situation with more conversion of inland forest to oil palm and rubber, such that flooding also occurs on the two intervening days. The number of days flooded would increase from 3 to 5, but the number of floods would decrease from three to one, as the 3 day-long floods in the initial situation merge into a single 5-day flood. This simple example implies that the number of days flooded is a more reliable flood measure than the number of floods, which in turn implies that the results in Table 3 provide a more reliable indication of the effects of deforestation on flooding than do the results in Table 2.

Peninsular Malaysia had the world's largest area of both oil palm and rubber at the end of the twentieth century (Vincent and Hadi 1993), but these crops are expanding rapidly in other parts of Asia and beyond (Li et al. 2007; Ziegler et al. 2009; Koh et al. 2011; Gutiérrez-Vélez and DeFries 2013). Their expansion into tropical forests has attracted particular concern due to biodiversity losses and carbon emissions (Li et al. 2007; Koh and Ghazoul 2010; Koh et al. 2011; Carlson et al. 2012). Our results add increased flooding to the list of the potential negative environmental consequences of converting tropical forests to oil palm and rubber.

Determining whether these consequences are sufficiently large to justify protecting forests against conversion to these crops would require benefit-cost analysis, which is beyond the scope of this paper. We can, however, identify two key features that such an analysis should have. First, it should account for spatial variability in both the costs and benefits of deforestation. It is possible that flood-related costs could exceed the benefits generated by oil palm and rubber in some locations but not others. Second, it should employ a definition of costs that extends beyond flood-related deaths and evacuations (Pattanayak and Wendland 2007). To the extent possible, it should also include other flood-related costs (e.g., property damage, injuries, nonfatal illnesses), losses of other hydrological services (e.g., soil retention, water purification), and the lost value of other forest ecosystem services, especially those related to biodiversity and carbon.

Acknowledgments This study was funded by the Global Environment Facility through the United Nations Development Programme (MAL/04/G31). Additional support was provided by the Government of Malaysia, through the Ministry of Natural Resources and Environment and the Forest Research Institute Malaysia, and by the Center for International Forestry Research. The cooperation of the Departments of Agriculture, Drainage and Irrigation, and Statistics of the Government of Malaysia is gratefully acknowledged as are helpful comments by M Jeuland, D Richter, and E Sills.

References

Alila Y et al (2009) Forests and floods: A new paradigm sheds light on age-old controversies. Water Resour Res 45: doi:10.1029/2008WR007207

Angrist JD, Pischke JS (2009) Mostly harmless econometrics: an empiricist's companion. Princeton University Press, Princeton

Bennett G et al (2013) Charting new waters. Forest Trends, Washington

Bradshaw CJA et al (2007) Global evidence that deforestation amplifies flood risk and severity in the developing world. Glob Change Biol 13:2379–2395

Brauman KA et al (2007) The nature and value of ecosystem services. Annu Rev Environ Resour 32:67–98

Bruijnzeel LA (2004) Hydrological functions of tropical forests. Agric Ecosyst Environ 104:185–228

Cameron CA, Trivedi PK (2013) Regression analysis of count data, 2nd edn. Cambridge University Press, Cambridge

Carlson KM et al (2012) Committed carbon emissions deforestation and community land conversion from oil palm plantation expansion in West Kalimantan Indonesia. Proc Natl Acad Sci USA 109:7559–7564

Department of Statistics (1981) General report of the population and housing census 1980. Kuala Lumpur, Malaysia

Department of Statistics (1992) General report of the population and housing census 1991. Kuala Lumpur, Malaysia

Department of Statistics (2001) General report of the population and housing census 2000. Kuala Lumpur, Malaysia

Drainage and Irrigation Department (2010) National hydrological networks. http://h2o.water.gov.my/v2/fail/Invstations/Inventory.html

Drainage and Irrigation Department (2013) Dam information. http://www.water.gov.my/our-services-mainmenu-252/dams/dam-informations-mainmenu-834?lang=en

Energy Commission Malaysia (2013) Electricity supply industry in Malaysia: performance and statistical information 2011. Putra Jaya, Malaysia

FAO (2010) Global forest resources assessment 2010: main report. Food and Agriculture Organization of the United Nations, Rome

FAO and CIFOR (2005) Forests and floods. Food and Agriculture Organization of the United Nations, Rome and Center for International Forestry Research Bogor, Indonesia

Ferraro PJ et al (2012) Forest figures. Rev Environ Econ Policy 6:20–44

Ferreira S, Ghimire R (2012) Forest cover socioeconomics and reported flood frequency in developing countries. Water Resour Res 48: doi:10.1029/2011WR011701

Ferreira S et al (2013) Does development reduce fatalities from natural disasters? New evidence for floods. Environ Dev Econ 18:649–679

Gutiérrez-Vélez RH, DeFries R (2013) Annual multi-resolution detection of land cover conversion to oil palm in the Peruvian Amazon. Remote Sens Environ 129:154–167

Hilbe JM (2011) Negative binomial regression, 2nd edn. Cambridge University Press, Cambridge

Huber PJ (1967) The behavior of maximum likelihood estimates under nonstandard conditions. In: Neyman LM (ed) Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, vol 1. University of California Press, Berkeley, pp 221–233

Jonkman SN (2005) Global perspectives on loss of human life caused by floods. Nat Hazards 34:151–175

Kahn ME (2005) The death toll from natural disasters: the role of income, geography, and institutions. Rev Econ Stat 87:271–284
Kennedy P (2008) A guide to econometrics, 3rd edn. Blackwell, Malden, Massachusetts
Kiersch B (2000) Land use impacts on water resources: a literature review. In: Calder IR (ed) Land-water linkages in rural watersheds. Food and Agriculture Organization of the United Nations, Rome
Kinzig AP et al (2011) Paying for ecosystem services–promise and peril. Science 334:603–604
Koh LP, Ghazoul J (2010) Spatially explicit scenario analysis for reconciling agricultural expansion forest protection and carbon conservation in Indonesia. Proc Natl Acad Sci USA 107:11140–11144
Koh LP et al (2011) Remotely sensed evidence of tropical peatland conversion to oil palm. Proc Natl Acad Sci USA 108:5127–5132
Kramer RA et al (1997) Ecological and economic analysis of watershed protection in Eastern Madagascar. J Environ Manage 49:277–295
KTA Tenaga (2003) Updating of condition of flooding in Malaysia–main report. Drainage and Irrigation Department, Kuala Lumpur
Lambert D (1992) Zero-inflated Poisson regression with an application to defects in manufacturing. Technometrics 34:1–14
Li H et al (2007) Demand for rubber is causing the loss of high diversity rain forest in SW China. Biodivers Conserv 16:1731–1745
Millennium Ecosystem Assessment (2005) Ecosystems and human well-being: synthesis. Island Press, Washington
Moulton BR (1986) Random group effects and the precision of regression estimates. J Econom 32:385–397
Mullahy J (1986) Specification and testing of some modified count data models. J Econom 33:341–365
Ogden FL et al (2013) Effect of land cover and use on dry season river runoff runoff efficiency and peak storm runoff in the seasonal tropics of Central Panama. Water Resour Res 49: doi:10.1002/2013WR013956
Pattanayak SK, Kramer RA (2001) Worth of watersheds. Environ Dev Econ 6:123–146
Pattanayak SK, Wendland KJ (2007) Nature's care: diarrhea, watershed protection and biodiversity conservation in Flores, Indonesia. Biodivers Conserv 16:2801–2819
Pattanayak SK et al (2010) Show me the money. Rev Environ Econ Policy 4:254–274
Ritschl A (2009) Spurious significance of treatment effects in overfitted fixed effect models. London School of Economics
Van Dijk AJM et al (2009) Forest-flood relation still tenuous. Glob Change Biol 15:110–115
Vincent JR, Hadi Y (1993) Malaysia. In: National Research Council (ed) Sustainable agriculture and the environment in the humid tropics. National Academy Press, Washington
Vincent JR, Mohamed Ali R (2005) Managing natural wealth: environment and development in Malaysia. Resources for the Future, Washington
White HL (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48:817–838
Wong IFT (1971) The present land use of West Malaysia 1966. Ministry of Agriculture and Lands, Kuala Lumpur
Wooldridge JM (2002) Econometric analysis of cross section and panel data. MIT Press, Cambridge
Zeger SL, Liang KY (1986) Longitudinal data analysis for discrete and continuous outcomes. Biometrics 42:121–130
Ziegler AD, Fox JM, Xu J (2009) The rubber juggernaut. Science 324:1024–1025
Zulkifli Y et al (2004) Runoff estimate from the North Selangor Peat Swamp Forest. In: Abdul Rahim N (ed) Water: forestry and land use perspectives, Technical Documents in Hydrology No. 70, International Hydrological Programme. UNESCO, Paris