A method for identifying households at high risk for mosquito borne illnesses.

Dominic D. LaRoche, Melanie L. Bell, and Kacey C. Ernst

February 29, 2016

1 Abstract

Mosquito borne illnesses are a significant threat to public health in countries where these diseases are either endemic or epidemic. Concerted efforts have been made in the past decade to reduce and in some cases eliminate mosquito borne diseases with the use of prophylactic interventions. The World Health Organization recommends preferential administration of interventions to those with the highest disease burden. However, previous research has also identified the benefit of additionally targeting interventions at those with the highest risk of infections. We develop a methodology for identifying the highest risk households on landscape by combining both health risk and infection risk. We use census information to determine health risk and we use a topographic wetness index to determine infection risk. We implement the method to evaluate the administration of interventions at two sites in Kenya. We find preferential administration of interventions at the high-elevation epidemic-prone site but not at the low-elevation endemic site. Our results have important implications for assessing the administration of interventions in the battle against mosquito borne illnesses.

2 Introduction

Malaria and other mosquito borne illnesses are considered a significant threat to public health and a socio-economic burden in countries where these diseases are either endemic or epidemic [8]. Concerted efforts have been made in the past decade to reduce and in some cases eliminate malaria specifically. Many national strategic plans to reduce or eliminate malaria are in their third generation. Spatial targeting of high risk areas is a strategy that has been recommended but few studies have assessed if government programs are achieving differential coverage in high risk areas [15].

The government of Kenya developed the National Malaria Strategy 2009-2017 in response to the ongoing threat of malaria [11]. This strategy outlined 6 objectives, the first of which is to have at least 80% of people living in malaria risk areas using appropriate malaria preventive interventions. The two primary non-pharmaceutical interventions identified in the plan are Indoor Residual Spraying (IRS) and Long Lasting Insecticidal Nets (LLINs). The strategy outlined for achieving the intervention objective included the initial mass distribution of LLINs where malaria is either endemic (western lowlands) or epidemic-prone (western highlands); followed by routine distribution of LLINs to pregnant women and children under 1 year of age and a subsidized sale of LLINs. The strategy also outlined the use of widespread IRS followed by focal treatments in epidemic-prone areas.

The World Health Organization recommends prioritizing the administration of interventions to pregnant women and young children followed by progressively achieving intervention coverage of all community members. The preferential administration of interventions to pregnant women and young children reflects the disproportionate disease burden borne by this group [4]. However, previous research has identified the benefit of additionally targeting interventions at those with the highest risk of infections [15]. Identifying individuals

which are both vulnerable to infection and likely to be exposed to an infected mosquito is therefore a priority. However, existing distribution campaigns do not typically account for both infection risk and disease burden simultaneously.

The Topographic Wetness Index (TWI) [1] derived from freely available remotely-sensed topographic data has been previously investigated as a tool for assessing risk of malaria infection [6, 5]. TWI can potentially identify areas where water is likely to pool and Anopheles densities are likely to be higher. This method, therefore, has the potential to both inform new distribution campaigns and evaluate the efficacy of existing campaigns. Our primary objective was to use topographic data, combined with a household census of demographics and intervention use, to develop a method for identifying high risk households. Because of the wide variety of TWI algorithms used in practice, and the relative complexity of these methods, we further wanted to investigate the sensitivity of the results to a more easily implemented TWI algorithm. We applied the method to two sites in Kenya where malaria is either endemic or epidemic-prone. Since policies for intervention administration, as well as topographic features, differed between the epidemic-prone and endemic regions we also sought to compare the sensitivity of the method to choice of TWI algorithm between these two regions.

3 Statistical Methodology

Every household on a landscape will vary with respect to both the risk of exposure to mosquitoes and the number of at risk individuals in the household. We develop a method to combine these two risk factors into an overall risk score for each household within a management area in order to identify the households which will yield the greatest benefit from the application of limited prophylactic resources.

3.1 Topographical Wetness Index

The Topographical Wetness index was originally introduced by Beven and Kirby in 1979 ([1]). TWI combines a measure of the amount of upstream drainage area with the local slope to determine the amount of wetness likely to accumulate at a point and is defined by Beven and Kirby as:

$$ln\frac{a}{tanb},$$

where a is the local up-slope area (local area draining to the point) and $tanb$ is the local slope in radians. The TWI is designed to predict the amount of water that is likely to flow into an area, based on surface topology, and the rate at which this water will flow out of an area. Areas with high in-flow and low out-flow are likely provide habitat for breeding mosquitoes.

Since TWI was first defined numerous methods have been developed to calculate it and several review papers have been published [13, 17] as well as alternative modeling strategies developed [9]. The primary goal of using TWI in this application is to identify areas where mosquitoes are likely to breed. However, it is unclear what effect differing TWI algorithms would have on our results. To determine the sensitivity of our method to choice of TWI algorithm, we implement 2 different algorithms which differ in their calculation of the up-slope area and local slope: 1) the SAGA Wetness Index [2] (hereafter "general TWI"), and 2) a simplified terrain wetness index which simply identifies depressions without regard to up-slope area (hereafter "restricted TWI"). We carried out all analyses using the statistical programming language R version 3.2.3 [14], with the exception of the SAGA TWI which was calculated with the SAGA open-source GIS software [3].

Details for calculation of the general TWI are available in the software documentation available at

In contrast to the relatively complicated calculation of a general TWI, we suggest a simple model based identifying depressions with low outflow and disregarding the up-slope area. We first identify depressions

and valleys by identifying pixels which are lower than the average of their neighbors. We do this by first calculating the average elevation of the surrounding pixels for each pixel in the landscape at three resolutions of increasing size:

$$\mu_{i,j} = \frac{\sum_{p \in j} e_p}{|p \in j|},$$

where $\mu_{i,j}$ is the average elevation of the pixels in the window of size j surrounding pixel i. The size of the windows is somewhat arbitrary but the idea is to identify small depressions from large ones so we suggest $90m \times 90m$, $210m \times 210m$, and $330m \times 330m$. These sizes may be adjusted or tuned to a particular landscape if information is available to do so. We then subtract the mean elevation from the actual elevation at each pixel from the three calculated averages and sum these differences.

$$\delta e_{i,j} = e_i - \mu_{i,j}$$
$$\Delta e_i = \sum_{j \in 1,2,3} \delta e_{i,j}$$

We then identify areas with low outflow by calculating the variance of the aspects of neighboring pixels. Since aspect is measured on a circular scale from 1 to 360 the variance of two aspects which are in fact quite close to each other but on opposite sides of the circular reset would be artificially high. For example, the aspect of two points facing close to due north could be 359 and 1. The variance of these two points would be $var(359, 1) = 64082$. However, the variance of two aspects of equal distance but not on opposite sides of the circular reset would be $var(1, 3) = 2$. In order to mitigate this difficulty we first translate each aspect to a cardinal direction denoted 1, 2, 3, or 4. We then calculate the variance of the aspects of the cells in each of 3 moving windows the same size as above ($90m \times 90m$, $210m \times 210m$, and $330m \times 330m$) and sum the resulting variances for each pixel:

$$\sigma_i = \sum_{j \in 1,2,3} var(a_{i \in j}),$$

where $a_{i \in j}$ are the set of aspects in the moving window j around pixel i. The final wetness index is then calculated as,

$$w_i = -1 \times \Delta e_i \times \sigma_i.$$

Relatively large positive values of w_i are expected to have higher wetness than small positive numbers. Negative numbers indicate a ridge or peak and are therefore expected to be dry.

We assign each household a risk for exposure to mosquitoes (exposure risk hereafter) by deriving a continuous risk surface over the study area from each TWI algorithm. We assume the mosquito exposure risk of a household is inversely related to the distance to one or more of these high-wetness areas. Therefore, we apply a Gaussian filter with $\sigma = 10$ to create a weighted average of mosquito risk for each cell in the study area. The Gaussian filter is an isometric 2-dimensional smoothing function with a Gaussian kernel. The value of σ determines the degree of smoothing and should be scaled appropriately to match the resolution of the data. The value of σ can be tuned for a specific application if data is available and appropriate to do so. Each house is then assigned the risk score of the cell it is in or, for high resolution topographical data such as LiDAR (laser radar), the average of the cells a property occupies.

3.2 Individual Health Risk

Each household varied with respect to both the risk of exposure to mosquitoes and the number of at-risk individuals in the household. The household risk formulas will vary depending on the disease under study and the most vulnerable population(s) for that disease. We recommend a simple additive risk score, like

the score formula provided below developed for malaria, based on expert opinion or relevant literature. The sole purpose of the health risk score is to differentiate households with high risk from those with low risk. More complicated formulas can be constructed but we believe a simple formula will be easier to interpret and adequate for most applications.

3.3 Overall Risk

To be at risk for a poor outcome a person must 1) come in contact with a disease harboring mosquito, and 2) be inherently vulnerable to infection (e.g. very young or very old). We create two risk scores representing each of these types of risk for every household on the landscape. Since these risks will be calculated on different scales we center at 0 and standardize risk scores so that they are scale-independent. These two scores can then be combined with a weighted sum to create an overall risk score for the household. Weights in the sum are determined by expert opinion, in our example below we use equal weights. This method lends itself well to the addition of additional risk scores, such as the distance to a health care provider, or even risk attenuating factors such as the presence of window screens, which can all be summed together with appropriate weights.

3.4 Example Analysis

Prior to the initiation of a community-based research program, two study sites in western Kenya were mapped and a census was taken. These two sites represent the western highland (hereafter epidemic-prone, N = 3380) and lowland (hereafter endemic, N = 604) populations. Our objective was to determine if the highest risk households were given preference in the administration of interventions in the form of both bed nets and indoor residual spraying.

We collected demographic information for each occupant including age and sex. Both sites have had partial treatment with both LLINs and IRS and household heads provided initial information about LLIN ownership and government administration of household IRS in the previous six months. Additional information for each participant was also collected such as age, sex, and relation to the head of the household.

We assigned an individual-based health risk score (health risk hereafter) to each household with the following formula:

$$Health\ Risk\ Score = (2 \times\ Children\ \leq 1) + (1 <\ Children\ \leq 5) + (2 \times\ Pregnant\ Women)$$

We assigned twice the weight to children under 1 and pregnant women since they have been previously identified as high risk [10, 16, 12]. This formula is tailored for malaria exposure but can be adjusted based on *a priori* information about the individual risks for a particular mosquito borne disease. For example, for the Zika virus children may have a relatively low health risk whereas pregnant women and sexually active women may have much higher risk.

We added the standardized household health risk with the standardized household exposure risk to create a combined risk. We then determined if high risk households are more likely to have received either a bed-net or aerial spraying with a logistic model;

$$log(\frac{p}{1-p}) = \beta_0 + \beta_1 \times\ Combined\ Household\ Risk,$$

where p = Probability of a house having a treatment. If e^{β_1} is > 1 and statistically significant ($\alpha = 0.05$) then high-risk households are more likely to receive treatment. We used a restricted cubic spline function to determine if there was a linear relationship between the log odds of treatment and combined risk. We modeled

the high and low sites separately because these sites differed substantially in topography and intervention administration protocols. We also modeled the two interventions separately since they are administered under different protocols. Finally, in order to determine the sensitivity of this method to the choice of TWI algorithm, we repeated the analysis using both the general and restricted TWI.

3.5 Results

The odds of receiving either a bed net or aerial spraying were higher for households with higher combined risk, but only at the high site (1). For each 1 standard deviation increase in combined risk at the high site the probability of receiving a bed net increased 27% (OR: 1.27, 95% CI: 1.18, 1.35) and the probability of aerial spraying increased 15% (OR: 1.15, 95% CI: 1.03, 1.29). At the low site, we found no preferential administration of either treatment to high combined risk households. We found some evidence, from the fitting of a restricted cubic spline, of a non-linear relationship between the log-odds of net use and combined risk at the low site. However, accounting for non-linearity in the model did not change the results so we only report the linear model here.

The probability of bed net use at the high site was more strongly associated with health risk, whereas the probability of aerial spraying at the high site was more strongly associated with exposure risk (2). However, We did not find the same pattern at the low site where we found households with high exposure risk were actually significantly less likely to receive aerial spraying (OR: 0.35, 95% CI: 0.14, 0.83).

Site	Treatment	OR	Lower 95% CI	Upper 95% CI
High	Net	1.35	1.26	1.44
	Spray	1.13	1.00	1.26
Low	Net	1.09	0.86	1.39
	Spray	0.9	0.61	1.34

Table 1: Odds of receiving a treatment as a function of combined risk.

		Health Risk			Exposure Risk		
Site	Treatment	OR	Lower 95% CI	Upper 95% CI	OR	Lower 95% CI	Upper 95% CI
High	Net	1.36	1.27	1.45	1.01	0.93	1.10
	Spray	1.08	0.96	1.22	1.32	1.14	1.53
Low	Net	1.21	0.94	1.55	0.58	0.31	1.10
	Spray	1.08	0.67	1.74	0.34	0.15	0.79

Table 2: Odds of treatment from risk of either mosquito exposure or malaria risk.

3.6 Sensitivity Analysis

The use of the restricted TWI algorithm identified fewer regions as high risk than the general TWI at both sites and was less likely to assign exposure risk to areas near a channel (fig. 1). The use of the restricted TWI based risk surface in the combined risk score increased the odds of a high-risk labelled household to have received a treatment for both the high and low sites, although the increase in OR for the low site remained non-significant (Table 3).

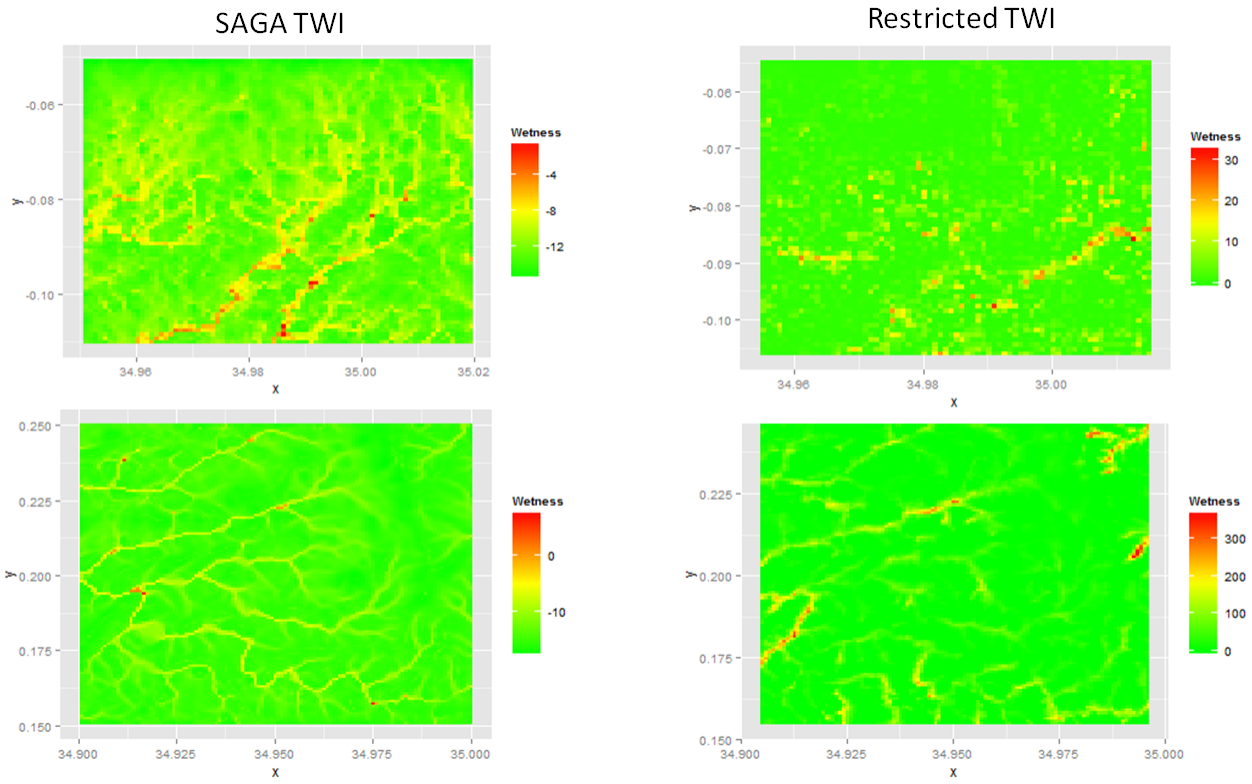

Figure 1: Comparison of the general TWI (left) and the restricted TWI (right) showing the lower extent of channels identified as moderate to high risk by the restricted TWI. The restricted TWI also identified several areas missed by the general TWI.

		General TWI			Restricted TWI		
Site	Treatment	OR	Lower 95% CI	Upper 95% CI	OR	Lower 95% CI	Upper 95% CI
High	Net	1.35	1.26	1.44	1.45	1.33	1.59
	Spray	1.13	1.00	1.26	1.22	1.05	1.42
Low	Net	1.09	0.86	1.39	1.33	0.93	1.90
	Spray	0.90	0.61	1.34	1.10	0.59	2.03

Table 3: Comparison of results when constructing combined risk from the restricted TWI or general TWI.

4 Discussion

Our methodology is motivated by the need to improve the efficiency of intervention administration protocols and evaluate existing protocols. Our example analysis evaluates existing protocols in which the administration of bed nets target pregnant women. Therefore, we would expect that households with young children would be more likely to have LLINs. We found that health risk was associated with an increased probability of ILLN use at the high site but not the low site. Similarly, under current protocols, aerial spraying is intended to target households at high risk for mosquito exposure so we would expect that households with high exposure risk would be associated with aerial spraying. Again, this is the pattern we observed at the high site but not the low site where we found the opposite association.

The general TWI algorithm identified channels at both sites as moderate to high risk. However, while these channels are likely to have higher wetness, the water may drain too quickly to provide quality breeding habitat for mosquitoes. Use of the restricted TWI algorithm identified fewer channels as having moderate to high exposure risk while identifying some areas not identified by the general TWI algorithm (fig. 1). Therefore, the restricted TWI algorithm may out-perform the general TWI in identifying mosquito breeding sites. Our results comparing both TWI algorithms indicates a stronger association between the restricted TWI and intervention use indicating better alignment between the restricted TWI and mosquito risk assessments on the ground. This was particularly noticeable when evaluating the OR of a household receiving indoor residual spraying at the low site where the association was at least in the preferable direction, albeit not significantly. We believe our method can be used to quickly evaluate the overall risk for a given household and improve the efficiency of distribution protocols.

Our method does not depend on accurately identifying the absolute risk of any given household, but rather the relative risk of every household on the landscape. We believe the restricted TWI described in appendix A provides a simple and easily implemented algorithm for identifying high risk areas which may perform better than more complicated TWI algorithms based on the calculation of up-slope areas. However, the sensitivity of our results to choice of TWI algorithm suggests that the TWI should be validated with additional information such as infection data. This has been done previously (Cohen2008, Cohen2010), but only with a single algorithm at a single site. We recommend further evaluation of TWI algorithms with actual mosquito borne infection information at multiple sites, preferably with different topographies.

References

[1] K. J. BEVEN and M. J. KIRKBY. "A physically based, variable contributing area model of basin hydrology / Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant". en. In: *Hydrological Sciences Bulletin* 24.1 (1979), pp. 43–69. ISSN: 0303-6936. DOI: 10.1080/02626667909491834. URL: http://www.tandfonline.com/doi/abs/10.1080/02626667909491834.

[2] J Böhner. *Soil Regionalisation by Means of Terrain Analysis and Process Parameterisation*. Ispra, 2002.

[3] J Böhner and T Selige. "Spatial prediction of soil attributes using terrain analysis and climate regionalisation". In: *Gottinger Geographische Abhandlungen* (2006). URL: http://www.pe.wzw.tum.de/publikationen/pdf/sd663.pdf.

[4] Teun Bousema et al. "Hitting hotspots: spatial targeting of malaria for control and elimination." In: *PLoS medicine* 9.1 (2012), e1001165. ISSN: 1549-1676. DOI: 10.1371/journal.pmed.1001165. URL: http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001165.

[5] Justin M Cohen et al. "Local topographic wetness indices predict household malaria risk better than land-use and land-cover in the western Kenya highlands." In: *Malaria journal* 9 (2010), p. 328. ISSN: 1475-2875. DOI: 10.1186/1475-2875-9-328. URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2993734{\&}tool=pmcentrez{\&}rendertype=abstract.

[6] Justin M Cohen et al. "Topography-derived wetness indices are associated with household-level malaria risk in two communities in the western Kenyan highlands." En. In: *Malaria journal* 7.1 (2008), p. 40. ISSN: 1475-2875. DOI: 10.1186/1475-2875-7-40. URL: http://malariajournal.biomedcentral.com/articles/10.1186/1475-2875-7-40.

[7] O. Conrad et al. "System for Automated Geoscientific Analyses (SAGA) v. 2.1.4". English. In: *Geoscientific Model Development* 8.7 (2015), pp. 1991–2007. ISSN: 1991-9603. DOI: 10.5194/gmd-8-1991-2015. URL: http://www.geosci-model-dev.net/8/1991/2015/gmd-8-1991-2015.html.

[8] Margaret Crouch. *The Second National Health Sector Strategic Plan of Kenya: NHSSP II 20052010*. Tech. rep. Republic of Kenya Ministry of Health. URL: http://www.healthpolicyinitiative.com/Publications/Documents/796{_}1{_}796{_}1{_}NHSSP{_}II.pdf.

[9] T. Grabs et al. "Modeling spatial patterns of saturated areas: A comparison of the topographic wetness index and a dynamic distributed model". In: *Journal of Hydrology* 373.1-2 (2009), pp. 15–23. ISSN: 00221694. DOI: 10.1016/j.jhydrol.2009.03.031. URL: http://www.sciencedirect.com/science/article/pii/S002216940900211X.

[10] S Gupta et al. "Immunity to non-cerebral severe malaria is acquired after one or two infections." In: *Nature medicine* 5.3 (1999), pp. 340–3. ISSN: 1078-8956. DOI: 10.1038/6560. URL: http://www.ncbi.nlm.nih.gov/pubmed/10086393.

[11] Ministry of public health and sanitation. *National Malaria Strategy 2009-2017*. Tech. rep. July. Ministry of public health and sanitation, 2009, p. 116. URL: https://www.google.com/url?sa=t{\&}rct=j{\&}q={\&}esrc=s{\&}source=web{\&}cd=1{\&}ved=0CB4QFjAAahUKEwiXj76j-IzHAhWGVT4KHZD7AFA{\&}url=http://www.c-hubonline.org/sites/default/files/resources/main/Kenya{_}National{_}Malaria{_}Strategy{_}2009-2017.pdf{\&}ei=9GG{_}VZfFKIar-QGQ94OABQ{\&}usg=AFQj.

[12] C Menendez, A F Fleming, and P L Alonso. "Malaria-related anaemia." In: *Parasitology today (Personal ed.)* 16.11 (2000), pp. 469–76. ISSN: 0169-4758. URL: http://www.ncbi.nlm.nih.gov/pubmed/11063857.

[13] P. F. Quinn, K. J. Beven, and R. Lamb. "The in(a/tan/β) index: How to calculate it and how to use it within the topmodel framework". In: *Hydrological Processes* 9.2 (1995), pp. 161–182. ISSN: 08856087. DOI: 10.1002/hyp.3360090204. URL: http://doi.wiley.com/10.1002/hyp.3360090204.

[14] R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2015. URL: https://www.r-project.org/.

[15] Julianna Schantz-Dunn and Nawal M Nour. "Malaria and pregnancy: a global health perspective." In: *Reviews in obstetrics & gynecology* 2.3 (2009), pp. 186–92. ISSN: 1941-2797. URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2760896{\&}tool=pmcentrez{\&}rendertype=abstract.

[16] R W Snow et al. “Estimating mortality, morbidity and disability due to malaria among Africa’s non-pregnant population.” In: *Bulletin of the World Health Organization* 77.8 (1999), pp. 624–40. ISSN: 0042-9686. URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2557714{\&}tool=pmcentrez{\&}rendertype=abstract.

[17] R. Sørensen, U. Zinko, and J. Seibert. “On the calculation of the topographic wetness index: evaluation of different methods based on field observations”. English. In: *Hydrology and Earth System Sciences* 10.1 (2006), pp. 101–112. ISSN: 1607-7938. DOI: 10.5194/hess-10-101-2006. URL: http://www.hydrol-earth-syst-sci.net/10/101/2006/hess-10-101-2006.html.