

Coupling Spatiotemporal Disease Modeling with Diagnosis

Martin Mubangizi and Catherine Ikae Athina Spiliopoulou

Department of Computer Science
Makerere University
mmubangizi,cikae@cit.mak.ac.ug

School of Informatics
University of Edinburgh
a.spiliopoulou@sms.ed.ac.uk

John A. Quinn

Department of Computer Science
Makerere University
jquinn@cit.mak.ac.ug

Abstract

Modelling the density of an infectious disease in space and time is a task generally carried out separately from the diagnosis of that disease in individuals. These two inference problems are complementary, however: diagnosis of disease can be done more accurately if prior information from a spatial risk model is employed, and in turn a disease density model can benefit from the incorporation of rich symptomatic information rather than simple counts of presumed cases of infection. We propose a unifying framework for both of these tasks, and illustrate it with the case of malaria. To do this we first introduce a state space model of malaria spread, and secondly a computer vision based system for detecting plasmodium in microscopical blood smear images, which can be run on location-aware mobile devices. We demonstrate the tractability of combining both elements and the improvement in accuracy this brings about.

1 Introduction

An important aspect of sustainability is the management of disease, whether it be of humans, plants or animals. One common problem is to estimate the spatial density of a particular disease, given limited and noisy observations, and another is to diagnose disease in individual cases at particular locations. These two tasks are usually done in isolation. Informally, a doctor may be aware of outbreaks of human disease in particular places or seasonal variations in disease risk, and they may interpret test results accordingly. But generally the diagnosis is not formally coupled with estimates of disease risk.

The tasks of mapping disease density over space and time and of diagnosing individual cases are complementary, however. A “risk map” can be used to give a prior in diagnosis of an individual with a known location. In turn, the results of individual diagnoses can be used to update the map in a more effective way than simply making hard decisions about infection statuses and using summary count data for the update. The potential for combining maps and diagnosis in this way has come about with the possibility of performing diagnosis with networked location aware devices that can carry out the necessary calculations.

A drawback to combining symptom information during the inference of spatial density, particularly if these observations are continuous, is that probabilistic inference scales exponentially with the number of samples per time frame. In this paper we propose a unified probabilistic framework for which inference is tractable, taking the case of malaria as an example.

Our case study, malaria, is one of the most prevalent diseases in the world, with about 3.3 billion people (i.e. half of the world’s population) at risk, the poorest countries being the most affected. There is need to monitor and predict malaria incidence across infected countries to allow for early intervention initiatives. Control of the disease also depends on accurate diagnosis of individuals, in order to treat those who need it without the drug resistance caused by over-diagnosis.

In section 2, we describe how the two tasks of disease density estimation and disease diagnosis can be unified. In section 3, we present how this unification can be applied to the case of malaria. In section 4, we present results that show a) how a symptom model of individuals can be tractably used for density estimation and prediction, with superior accuracy at population level to traditional models, and b) conversely, that inclusion of the density model can improve diagnosis of diseases, taking an example of photomicroscopical diagnosis of malaria.

2 Dynamical models of infectious disease

We briefly review types of models used for estimating and predicting spatiotemporal density of disease given count data. Historically there has been a shift from “hand-crafted” models with parameters set through the judgement of experts, towards the learning of parameters and model structure from data. ARMA, ARIMA, linear and log-linear regression methods have been popular for modelling and predicting disease rates. More recently, probabilistic graphical modelling has been applied to disease count data, particularly for inference tasks such as outbreak prediction (Xia and Garrick 2006; Cooper and Dash 2004).

One interesting development has been the use of novel observation types in order to infer disease density, and in particular to gain early warning of trends that might be indicative of an outbreak. New sources of data that are being considered include sales of over-the-counter drugs (Hogan

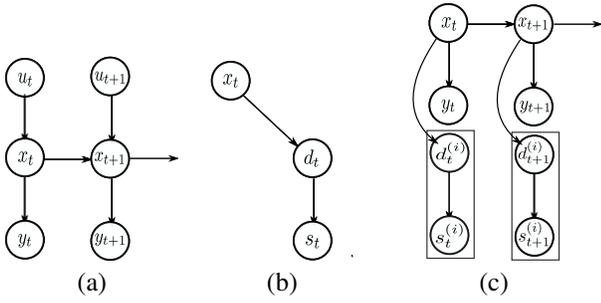


Figure 1: (a) Generalized model of disease rate dynamics, relating control inputs u_t , latent variables x_t (including underlying risk) and observations y_t (such as disease counts); (b) symptom-disease model used in diagnosis of an individual patient relating environmental risk factors x_t , disease status of an individual $d_t^{(i)}$ and symptom presented $s_t^{(i)}$; (c) model combining disease rate dynamics with individual cases.

and Wagner 2006), absenteeism from work or school (Lenert et al. 2006), chief complaint recorded in hospital visits, emergency call center records, physiologic and space-based sensors, internet search term frequency and ambulance dispatches.

In general, it is possible to express all of these approaches in the graphical model form depicted in Figure 1(a). This is the well-known state space form, in which control inputs $u_{1:T}$ (such as environmental or demographic factors) affect latent variables $x_{1:T}$ (such as vector densities, population susceptibility or immunity) which in turn affect observed quantities $y_{1:T}$ (normally infection counts, but potentially including the more exotic observation types mentioned above). Dynamics on the latent variables model the spread of the disease in space and time.

2.1 Incorporation of symptoms and diagnosis

Next we consider the diagnosis of a disease state d in an individual case given presentation of symptoms s . A generative model is a common device for using disease symptoms to diagnose disease cases. Furthermore, the underlying risk factors x may be taken into account during diagnosis, at least informally, by the agent performing that diagnosis. For example, if a certain disease is known to be very common, cases with ambiguous symptoms are more likely to be diagnosed as positive. A graphical model representing the elements of this diagnosis process for individual cases is shown in Figure 1(b). The diagnosis process can also be represented by equation 1.

$$p(d|s_{1:n}) \propto p(s_1, \dots, s_n|d) \int p(d|x) p(x) dx \quad (1)$$

where $s_{1:n}$ is a vector of symptoms.

The two models in Figure 1(a) and Figure 1(b) can be combined to get the model in Figure 1(c). This is possible

when we conceive of x_t , the underlying risk factors, as being the same in both models and $y_t \perp\!\!\!\perp d_t^{(i)}, s_t^{(i)} | x_t$, which means that the different types of observations are independent from each other given the hidden disease state. Note that in this model the input control variable, u_t , has been dropped and there is plate around the disease case, $d_t^{(i)}$, and disease symptom, $s_t^{(i)}$, pair, meaning that we expect multiple instances from the set of individuals \mathcal{S}_t sampled at time t . We define $\mathcal{O}_t \equiv \{y_t, s_t^{(i)} | i \in \mathcal{S}_t\}$, the observed data at time t . The joint probability distribution of the model in Figure 1 (c), can be written as

$$p(x_{1:T}, \{d_{1:T}^{(i)}\}, \mathcal{O}_{1:T}) = p(x_1) \prod_{t=2}^T p(x_t|x_{t-1}) \times \left[\prod_{t=1}^T p(y_t|x_t) \prod_{i \in \mathcal{S}_t} p(s_t^{(i)}|d_t^{(i)}) p(d_t^{(i)}|x_t) \right]. \quad (2)$$

We could for example parameterize this as a variation of the linear dynamical system (LDS), with the vector x_t representing the latent underlying disease risk at a set of locations. The locations might be arbitrarily small cells on a regular grid, though it is common for them to be irregular administrative regions due to the availability of data.

In order to combine the tasks of spatiotemporal modelling and diagnosis, the timing of the different inference requirements must be taken into account. Diagnosis needs to be carried out on the spot, while the updates to the density model can happen at the end of the time frame. We can think about the sequence of calculations needed in order to estimate x_t and $\{d_t^{(i)}\}$ given $\mathcal{O}_{1:t}$. Being a state space model, this can be carried out with a pair of recursive steps known as prediction and correction. Prediction for this model is the following:

$$\hat{p}(x_t|\mathcal{O}_{1:t-1}) = \int p(x_t|x_{t-1}) \hat{p}(x_{t-1}|\mathcal{O}_{1:t-1}) dx_{t-1} \quad (3)$$

$$\hat{p}(d_t^{(i)}|\mathcal{O}_{1:t-1}) \propto \int p(d_t^{(i)}|x_t) \hat{p}(x_t|\mathcal{O}_{1:t-1}) dx_t \quad (4)$$

Correction is then done with the following:

$$\hat{p}(d_t^{(i)}|\mathcal{O}_{1:t}) \propto p(s_t^{(i)}|d_t^{(i)}) \hat{p}(d_t^{(i)}|\mathcal{O}_{1:t-1}) \quad (5)$$

$$\hat{p}(x_t|\mathcal{O}_{1:t}) \propto p(\mathcal{O}_t|x_t) \hat{p}(x_t|\mathcal{O}_{1:t-1}) \quad (6)$$

where

$$p(\mathcal{O}_t|x_t) = p(y_t|x_t) \prod_{i \in \mathcal{S}_t} \sum_{d_t^{(i)}} p(s_t^{(i)}|d_t^{(i)}) p(d_t^{(i)}|x_t). \quad (7)$$

Step (3) can be done at the beginning of a new time frame, and using this result steps (4) and (5) can be done at the instant diagnosis is required for an individual case. Finally at the end of the time frame, all the symptom information can be incorporated in step (6).

3 Parameterising the model for malaria

Malaria is endemic in many regions of the world but has highly variable spatial density even at a fine scale, depending on terrain, climate, population density and a number

of other factors. A number of studies have been done to estimate and predict malaria density. Representative examples of these studies include (Loha and Lindtjorn 2010) using ARIMA with disease counts and meteorological factors (rainfall, temperature and relative humidity) to predict falciparum malaria incidence in Ethiopia and (Gomez-Eliphe et al. 2007) using ARIMA in forecasting malaria incidence using monthly disease cases, climatic factors (rainfall and temperature) and normalized difference vegetation index (NDVI) from satellite images.

Because of the high spatial variability of the disease, we can expect benefits in diagnosis by incorporating this information when the location of the person being tested is known. There are several methods of diagnosing malaria, foremost among them visual inspection of blood cells. Malaria is caused by the presence of the parasite genus *plasmodium*, and the gold standard test is microscopical analysis of a stained blood sample in order to visually identify such parasites (Murray et al. 2008). Diagnosis is important as leaving the disease untreated frequently leads to death, whereas taking the treatment based only on symptoms leads to drug resistance and possibly the failure to treat diseases with similar early symptoms (fever, joint pain) such as meningitis. However, in the geographical areas in which malaria is prevalent, there is frequently a shortage of experts. Hence there has been increasing interest in carrying out this diagnosis automatically with computer vision methods.

In vision terms this is an object detection problem, and some previous work is reviewed in (Tek, Dempster, and Kale 2009). There has also been work in comparing these methods with other forms of diagnosis (Andrade et al. 2010). (Ross et al. 2006) uses neural networks with morphological features to identify red blood cells and possible parasites present on a microscopic slide. The results obtained with this technique were 85% recall and 81% precision using a set of 350 images containing 950 objects. Color space and morphological heuristics were employed to segment red blood cells and parasites by using an optimal saturation threshold (Makkapati 2009) using a set of 55 images. Multi-class parasite identification, attempting to classify the type and life cycle stage of detected parasites has also been attempted (Tek, Dempster, and Kale 2010).

In the remainder of this section, we discuss the estimation of x_t and $d_t^{(i)}$, beginning with the simple case where only count data $y_{1:t}$ is available. In section 3.3 we describe inference of $d_t^{(i)}$ with image features.

3.1 Inference and learning with count data

In the following we take the latent risk to have linear-Gaussian state transitions,

$$p(x_t|x_{t-1}) \sim \mathcal{N}(x_t|Ax_{t-1}, Q) \quad (8)$$

in which A is the transition matrix and Q is the transition covariance. The disease counts can then be modelled by a linear-Poisson observation likelihood on each dimension of x_t and y_t ,

$$p(y_{t,j}|x_{t,j}) \sim \text{Poisson}(y_{t,j}|c_j g(x_{t,j})) \quad (9)$$

where c_j is a scaling factor and $g(\cdot)$ is a link function normally used to keep the Poisson rate positive, though in the experiments described here the hidden states always take on positive values and thus here $g(\cdot) = \cdot$. We assume that the observed disease counts at different locations are independent given the hidden rates and thus the likelihood of the data under this model is a product of the one-dimensional Poisson distributions. Possible parameterizations of the disease and symptom variables are discussed in section 3.2.

We now describe the processes of inferences and learning in this method, giving details only where they vary from the standard linear-Gaussian dynamical system. Unlike the case of Gaussian transitions and Gaussian observations, exact inference of x_t in this model given observations $y_{1:T}$ is intractable. The exact posterior is a product of Gaussian state transitions and Poisson likelihood terms, which has no closed-form representation. A simple and effective method for inference, however, is the Sequential Importance Resampling (SIR) algorithm, summarized for this model in Algorithm 1.

Parameter estimation can be carried out with expectation maximization (EM) steps as in the linear Gaussian LDS (Ghahramani and Hinton 1996), using SIR in place of the filtering E-step, to estimate A and Q . We make one alteration to the standard EM updates in this work, which is to employ shrinkage (taking a weighted average of the M-step result with a diagonal matrix). This has the effect of reducing the most extreme off-diagonal coefficients, which is known to reduce test error and improve on the positive-definite quality of the resulting covariance matrix (Wolf and Ledoit 2004). It is also possible after each M-step to zero the off-diagonal elements of A and Q corresponding to pairs of locations that are known to have no direct effect on each other. We performed several runs with different randomised initial parameters during training and chose the parameters with highest likelihood in order to mitigate the problem of local minima.

The observation process differs from the LDS, so the estimator for c_j must be derived separately. Doing so gives the M-step update

$$\tilde{c}_j = \frac{1}{T} \sum_{t=1}^T \frac{y_{t,j}}{g(\langle x_{t,j} | \mathcal{O}_{1:T} \rangle)}. \quad (10)$$

This estimation is not strictly required, however: in the following experiments we simply take $c_j = 1$ for all j .

3.2 Inference with symptom data

We now consider the case of inference with symptomatic information $\{s_t^{(i)}\}$. As an example, take the case that the disease status $d_t^{(i)} \in \{0, 1\}$ in each individual is distributed conditional on the underlying risk as follows:

$$p(d_t^{(i)} = 1|x_t) = 1 - \alpha \left(1 - \frac{x_t}{N}\right) \quad (11)$$

where α is the false alarm rate in a sample (representing the bias arising from the fact that people who present themselves for testing are more likely to be ill than those in the general population; this can be estimated by looking at historical

Algorithm 1: SIR with count and symptom data.

Input: Observations: $\mathcal{O}_{1:T}$ 1
Model parameters: $A, Q, p(x_1)$
Number of particles: P
Resampling threshold: N_{thr}
Output: $\hat{p}(x_t|\mathcal{O}_{1:t})$ for $t = 1 : T$.
Initialize particles $\hat{x}_1^{(p)} \propto p(x_1)$ for $p = 1 : P$
for $t = 1 : T$ **do**
 Sample P particles from the transition prior
 $T(\hat{x}_t^{(p)} \leftarrow \hat{x}_{t-1}^{(p)}) = \mathcal{N}(\hat{x}_t^{(p)} | A\hat{x}_{t-1}^{(p)}, Q)$
 Compute the importance weights
 $w_t^{(p)} \propto p(\mathcal{O}_t | \hat{x}_t^{(p)})$ (See eqn. 7)
 Normalize $w_t^{(p)} = \frac{w_t^{(p)}}{\sum_{p'} w_t^{(p')}}$
 Resample if $\frac{1}{\sum_p (w_t^{(p)})^2} < N_{thr}$
 return $\hat{p}(x_t|\mathcal{O}_{1:t}) = \sum_p w_t^{(p)} \hat{x}_t^{(p)}$
end

numbers of cases for testing and true positives), and N is the population size in the corresponding area. Conditional on an individual’s disease state, the sample $s_t^{(i)}$ could be drawn from a Gaussian distribution, for example.

With continuous, possibly multi-dimensional symptom observations, the likelihood term $p(\{s_t^{(i)}\} | x_t)$ is a mixture of many terms that is difficult to simplify analytically (the product in Eq. (7)). Therefore the normalisation coefficient in Eq. (6) might be very difficult to estimate in general. Particle filtering would have the same complexity, as we would need to propose particles which explore the full space of disease states in n individuals, making inference in this model intractable for more than a few tens of symptom observations per time frame. However, the complexity can be radically reduced by quantizing the sample.

If we split the range of $s_t^{(i)}$ into bins, then each symptom observation can be assigned to one of those bins. Given x_t , the likelihood of observing a symptom measurement assigned to a particular bin can be calculated by marginalizing out $d_t^{(i)}$. The likelihood of a set of observed symptoms can be represented as a set of frequency counts for each bin, and evaluated as a multinomial distribution. Evaluating the likelihood in this way is constant in the number of observed cases.

3.3 Vision-based diagnosis of malaria

When image data is to be used as a source of symptom information, there is a wide choice of features encodings which might be used to represent the raw pixel data more effectively. We could take standard color, shape and gradient features directly, though we propose here training discriminative classifiers on such features and using the outputs of these classifiers as the ‘symptoms’. Consider a case where using n discriminative classifiers we extract features for each patch i to form the symptom vector $s_{i,1}, \dots, s_{i,n}$.

From a training set, we can look at the class-conditional distribution of features $p(d_i | s_{i,1}), \dots, p(d_i | s_{i,n})$, where we take $d_i \in \{0, 1\}$ to denote the absence or presence of a parasite object in the i th patch. Making an assumption of conditional independence between these features given the patch class, classification can be carried out using

$$p(d_i | s_{i,1}, \dots, s_{i,n}) \propto p(s_{i,1} | d_i) \dots p(s_{i,n} | d_i) p(d_i) \quad (12)$$

that is, a Naive Bayes classifier. To relate this to inference in the coupled dynamical model of section 2, we can map $p(s_{i,1} | d_i) \dots p(s_{i,n} | d_i)$ to $p(s_t^{(i)} | d_t^{(i)})$, and $p(d_i)$ to $\hat{p}(d_t^{(i)} | \mathcal{O}_{1:t-1})$ in Eq. (5).

4 Results

This section presents experimental results for density estimation with count data, density estimation with symptom data, and diagnosis with image data.

4.1 Density estimation with count data

We demonstrate the accuracy of inference in this model with an example of data from six zones of Kampala, Uganda. This data consisted of weekly reported malaria cases of the period from 15th January 2007 to 1st February 2010. Data was obtained from the Health Management Information (HMIS) Department Kampala City Council. Two years of data was used to learn the parameters of the model. One year’s data was used to test the model.

Table 1 shows the mean absolute error of step-ahead predictions of disease counts using SIR inference, for single-dimensional state-space models (temporal) and six-dimensional state-space models (spatiotemporal). These were compared with log-linear regression, a method commonly used for this task, and a simple baseline, where the count at time $t - 1$ was taken to be the predicted count for time t . The order of the log-linear model was set to 3 after investigating the partial autocorrelation of the training data.

	MAE	SD
State space (temporal)	221.3025	109.9517
State space (spatio-temp, 0.1)	275.5585	110.6003
State space (spatio-temp, 0.3)	241.1393	107.7543
State space (spatio-temp, 0.5)	231.6659	108.3874
State space (spatio-temp, 0.7)	239.7405	105.7676
Log linear (temporal)	311.2650	266.7252
Log linear (spatio-temp)	302.4633	275.5987
Baseline ($\hat{y}_t = y_{t-1}$)	225.7778	97.0029

Table 1: Mean and standard deviations of absolute error rates for step-ahead predictions of malaria incidence in six connected regions. Particle filter inference in the state space model are compared to log-linear regression and predicting the next observation to be the same as the previous one. The number associated with the spatiotemporal models are shrinkage factors used in estimating A and Q .

The state-space temporal model has by a small margin the lowest error rate, suggesting that the assumption of a hidden disease rate evolving through time is good for disease

density modelling. Surprisingly, the second best predictor, is the simple baseline of predicting a repeat of the previous observation. We should note that the dataset for this experiment is fairly small and therefore models with many parameters are prone to over-fitting. The spatio-temporal state-space models perform comparably to the temporal one and the baseline, while the log-linear models perform considerably worse.

The particle filtering technique shown here is simple to implement but has the problem that many particles are required when the dimension of the latent variables are high. There are several extensions of the basic method intended to overcome this problem. For example, it is possible to approximate the distribution $p(x_t|x_{t-1})p(O_t|x_t)$ directly with the particle set without using a proposal distribution or importance sampling, a technique known as marginal particle filtering (Klaas, de Freitas, and Doucet 2005). Auxilliary particle filtering is a simpler approach which removes particles inconsistent with the next observation, often leading to a smaller particle set being necessary.

4.2 Density estimation with symptom data

To demonstrate this type of inference we first simulated a sequence of hidden states $x_{1:T}$. Then we used the symptom disease model to generate samples of individual binary disease states $\{d_{1:T}^{(i)}\}$ of patients, conditioned on the hidden rate x_t , using Eq. (11) with $N=5000$ and $\alpha = \frac{2}{3}$. The symptom data (temperature) for each individual case was sampled from $\mathcal{N}(37, 0.5)$ if $d_t^{(i)} = 0$, and from $\mathcal{N}(38.5, 1)$ if $d_t^{(i)} = 1$. We set the number of samples at each time step to $|\mathcal{S}_t| = 100$. Equivalent count information $y_{1:T}$, for comparison, was derived from the sampled symptom information.

To be able to compare the use of symptoms versus disease counts, we carried out estimation of $\hat{p}(x_t|\{s_1^{(i)}\}, \dots, \{s_t^{(i)}\})$ and $\hat{p}(x_t|y_{1:t})$ and studied the mean absolute error, $\frac{1}{T} \sum_{t=1}^T x_t - \hat{x}_t$. The predictions using symptom observations have a mean absolute error of 81.69 ± 2.77 , which as expected is significantly lower than the error when using count observations of 88.06 ± 4.73 . The intuition for this is that making a hard decision about disease status is error-prone; using the symptom information allows us to optimally take account of uncertainty in the disease status. If diagnosis from symptoms alone could be made with perfect accuracy, there would be no advantage in modelling symptoms. The more uncertain the diagnosis, however, the more benefit in using this information.

Figure 2 illustrates the operation of particle filtered inference using symptom data. The top panel shows samples of symptom data for 20 time steps, while the lower panel shows both the true underlying infection rate in the population and the positions of particles estimating it. Note the frequent resampling of particles as they move far away from x_t .

4.3 Diagnosis with image data

To train and evaluate malaria diagnosis, thick blood film images were collected from patients, manually diagnosed with and without *plasmodium falciparum* from Mulago Hospi-

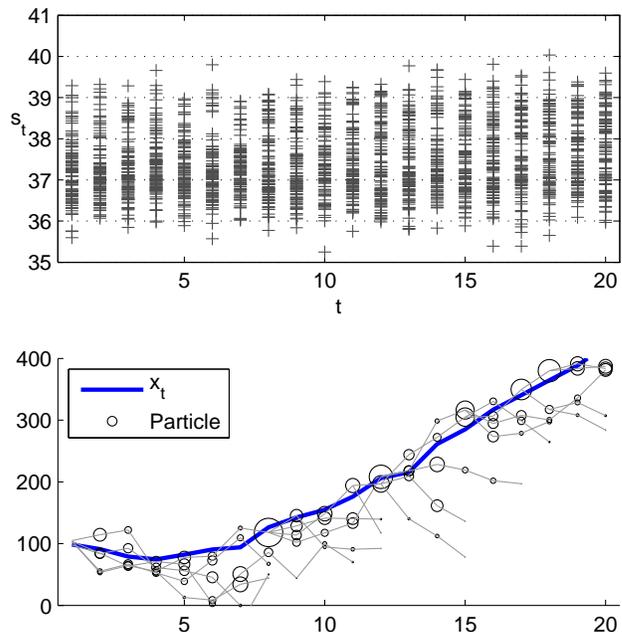


Figure 2: Particle filtering with symptom data. The upper panel shows simulated symptoms during an outbreak; at each time frame the proportion of high measurements increases. The lower panel shows the true underlying infection rate x_t , and the trajectories of point estimates $\hat{x}_t^{(p)}$. The size of the circles indicates the importance weight $w_t^{(p)}$ of the corresponding particle.

tal, Kampala, Uganda. Images were collected with a Brunel SP100 microscope and Motic MC1000 microscope camera (Figure 3, left) from blood film samples of 133 patients. After discarding images which were out of focus or otherwise poor quality, the data set contained 2703 images, 800 of these reserved for testing. These images were annotated by a team of laboratory technicians, in which the Pascal VOC (Everingham et al. 2010) software was used to draw bounding boxes around malaria parasites which were visible in the blood images (Figure 3, right). The images fell broadly into three categories: hyper-parasitaemic, in which several dozen parasites might be visible in a single image; parasitaemic, in which up to ten parasites might be visible in one image; and negative, where no objects were recorded. A total of 49,900 parasite objects were recorded in the image set¹. Figure 4 shows examples of these objects compared to clutter in the images which is easily confounded with parasites, illustrating the difficulty of the detection task.

Detection was implemented by training three discriminative classifiers. First, a boosted cascade of Haar-like features (Viola and Jones 2001) was used. Two boosted sets of decision trees were trained on sliding-window patches, one with dense SURF descriptors (Bay et al. 2008) of each patch, and another with central moments of a Canny-filtered

¹We intend to make the complete dataset of images and annotations publically available in the near future.



Figure 3: Image capture using a dedicated microscope camera (left). Object detection code running on a smartphone mounted over the microscope eyepiece (centre). Example blood smear image with ground truth, showing bounding boxes around parasites provided by an expert (right).

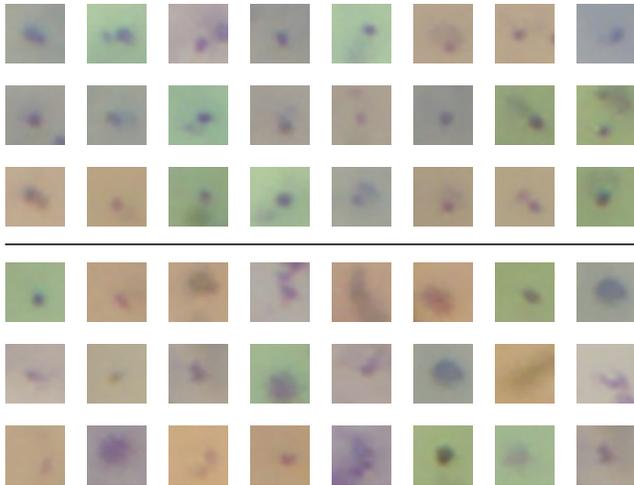


Figure 4: Image patches close to the decision boundary. The upper group are parasites, while the lower group are clutter (such as platelets, spots of dust on one of the lenses or staining solution artifact).

contour image of each patch. The output of these classifiers for each patch forms a three dimensional symptom vector $s_{i,1}, \dots, s_{i,3}$. In the case of the cascade classifier $s_{i,1} \in \{0, 1\}$, whereas for the other features $s_{i,2}, s_{i,3} \in \mathbb{R}$ taking the weighted sum of votes from each of the set of boosted decision trees. The class-conditional distributions of these features $p(d_i|s_{i,1}), \dots, p(d_i|s_{i,3})$ are computed and used with $p(d_i)$ in Eq. (12) to perform diagnosis.

System deployment We investigated two modes of deployment. First, we looked at using a dedicated microscope camera for capturing a live video feed, with object detection code running on a laptop. In order to do this we implemented a Linux USB driver for the proprietary Motic camera hardware. Because the imaging apparatus was identical

to that used in testing, the accuracy results below generalized well to live testing, and several video frames per second could be analysed. Second, we tried deploying detection code on a Huawei Ideos phone running Android 2.2 and costing around \$100. The camera of the phone was applied directly to the microscope eyepiece using an eyepiece clamp and ring magnet to hold the camera in place (Figure 3, centre). Detection was implemented for this platform using the OpenCV library. Parasites could be distinguished in images obtained this way, though lower image quality negatively impacted detection accuracy – we are still in the process of quantifying the performance of the smartphone-based system. One frame could be processed in around 10 seconds on the phone. Both deployments are intended to incorporate the spatiotemporal framework (downloading location-specific priors and uploading symptoms).

Table 2 shows results of the detection accuracy with different detection methods. Using a generative (Naive Bayes) model of the discriminative features at a threshold of 0.5 gives inferior accuracy to the cascade detector in terms of recall and F-score. However it has better precision and allows the incorporation of knowledge of the prior.

We examine the effect of such information by carrying out a separate test with the generative classifier. The set of 800 test images was split into 10 partitions with different concentrations of parasite objects. The classifier was supplied with the correct prior (the ratio of actual positive patches in that partition). With this information, we see that the threshold on the posterior can be increased giving much higher precision with usable recall. Precision is more important in this application as we can compensate for low recall at test time by scanning more images from each blood sample.

5 Discussion

We have presented a probabilistic model combining dynamic estimation of disease infection rates with diagnosis of individuals at given locations. Combining these tasks can lead to higher accuracy for both. Continuous symptom in-

		Precision	Recall	F-score
Cascade		0.678	0.866	0.761
Boost-SURF		0.346	0.802	0.484
Boost-moments		0.236	0.555	0.331
Generative	$t=0.5$	0.698	0.734	0.716
	$t=0.6$	0.755	0.296	0.426
	$t=0.7$	0.781	0.050	0.093
Generative*	$t=0.5$	0.726	0.733	0.730
	$t=0.6$	0.754	0.701	0.726
	$t=0.7$	0.795	0.629	0.702
	$t=0.8$	0.841	0.496	0.624
	$t=0.9$	0.907	0.203	0.332

Table 2: Parasite detection performance for discriminative classifiers and generative classifier. * denotes that the test set was partitioned and the correct prior probability of each patch in a partition being a positive match was supplied to the classifier. Precision and recall are calculated from the posteriors of the generative classifier by thresholding at probability t .

formation can be incorporated in constant time, though one area of future work is in finding tractable inference methods for cases in which the latent state has hundreds or thousands of dimensions.

Acknowledgements

We thank Ian Munabi for input on microscopical diagnosis, and Alfred Andama, Vincent Wadda, Steven Ikodi and Patrick Byanyima for assisting in data collection. We also thank the anonymous reviewers for their feedback. The work was partly funded by a Mobile Healthcare for Africa award from Microsoft Research. AS was supported by a University of Edinburgh Innovation Initiative Grant (GR000290).

References

Andrade, B. B.; Reis-Filho, A.; Barros, A. M.; Souza-Neto, S. M.; Nogueira, L. L.; Fukutani, K. F.; Camargo, E. P.; Camargo, L. M.; Barral, A.; Duarte, A.; and Barral-Netto, M. 2010. Towards a precise test for malaria diagnosis in the Brazilian Amazon: comparison among field microscopy, a rapid diagnostic test, nested PCR, and a computational expert system based on artificial neural networks. *Malaria Journal* 9:117.

Bay, H.; Ess, A.; Tuytelaars, T.; and Gool, L. V. 2008. SURF: Speeded Up Robust Features. *Computer Vision and Image Understanding* 110(3):346–359.

Cooper, G. F., and Dash, D. H. 2004. Bayesian Biosurveillance of Disease Outbreaks. In *Proceedings of the 20th Annual Conference on Uncertainty in Artificial Intelligence (UAI-04)*.

Everingham, M.; Van Gool, L.; Williams, C. K. I.; Winn, J.; and Zisserman, A. 2010. The PASCAL Visual Object Classes (VOC) challenge. *International Journal of Computer Vision* 88(2):303–338.

Ghahramani, Z., and Hinton, G. 1996. Parameter Estimation for Linear Dynamical Systems. Technical report, Department of Computer Science, University of Toronto.

Gomez-Elipse, A.; Otero, A.; van Herp, M.; and Aguirre-Jaime, A. 2007. Forecasting malaria incidence based on monthly case reports and environmental factors in Karuzi, Burundi, 1997-2003. *Malaria Journal* 6(1):129.

Hogan, W. R., and Wagner, M. M. 2006. Sales of over-the-counter healthcare products. In Wagner, M. M.; Moore, A. W.; and Aryel, R. M., eds., *Handbook of Biosurveillance*. Elsevier Inc, 1 edition. chapter 22, 321–331.

Klaas, M.; de Freitas, N.; and Doucet, A. 2005. Toward Practical N^2 Monte Carlo: the Marginal Particle Filter. In *Proceedings of the Twenty-First Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-05)*.

Lenert, L.; Johnson, J.; Kirsh, D.; and Aryel, R. M. 2006. Absenteeism. In Wagner, M. M.; Moore, A. W.; and Aryel, R. M., eds., *Handbook of Biosurveillance*. Elsevier Inc, 1 edition. chapter 24, 361–368.

Loha, E., and Lindtjorn, B. 2010. Model variations in predicting incidence of Plasmodium falciparum malaria using 1998-2007 morbidity and meteorological data from south Ethiopia. *Malaria Journal* 9(1):166.

Makkapati, V.V.; Rao, R. 2009. Segmentation of malaria parasites in peripheral blood smear images. *IEEE International Conference on Acoustics, Speech and Signal Processing*.

Murray, C. K.; Gasser, R. A.; Magill, A. J.; and Miller, R. S. 2008. Update on rapid diagnostic testing for malaria. *Clinical microbiology reviews* 21(1):97–110.

Ross, N.; Pritchard, C.; Rubin, D.; and Duse, A. 2006. Automated image processing method for the diagnosis and classification of malaria on thin blood smear. *Med Biol Eng Comput* 44:427–436.

Tek, F. B.; Dempster, A. G.; and Kale, I. 2009. Computer vision for microscopy diagnosis of malaria. *Malaria Journal* 8:153.

Tek, F. B.; Dempster, A. G.; and Kale, I. 2010. Parasite detection and identification for automated thin blood film malaria diagnosis. *Computer Vision and Image Understanding* 114(1):21–32.

Viola, P., and Jones, M. J. 2001. Rapid object detection using boosted cascade of simple features. *CVPR*.

Wolf, M., and Ledoit, O. 2004. Honey, I shrunk the sample covariance matrix. In *International Conference on Stochastic Finance*.

Xia, J., and Garrick, W. L. 2006. A Bayesian network for outbreak detection and prediction. In *Proceedings of the Twentieth Conference on Artificial Intelligence (AAAI-06)*.