Modeling diapause termination of Rhipicephalus appendiculatus using statistical tools to detect sudden behavioral changes and time dependencies.

N. Speybroeck[†*]
P.J. Lindsey[‡]
M. Billiouw[†]
M. Madder[†]
J.K. Lindsey[§]
D.L. Berkvens[†]

October 4, 2002

Abstract: This article presents statistical methodology, which can be used to analyze longitudinal binary responses for which a sudden change in the response occurs at a point in time. This includes exploratory tools such as cumulative probability plots, transition matrices, and change-point models and more advanced statistical techniques such as generalized auto-regression models and hidden Markov chains. This methodology is applied to analyze a study on the activity of Rhipicephalus appendiculatus, the major vector of Theileria parva.
The model shows that activity and non-activity act in an absorbing way meaning that once a tick becomes active it shows a tendency to remain active. Activity suddenly changes around the longest day of the year. The change-point model indicates that this happens on December 10. The reaction of ticks on acceleration and changes in rainfall and temperature indicates that ticks can sense climatic changes. The study revealed the underlying states during diapause development of the adult tick of R. appendiculatus. These not visually observable states form the basis of the tick's behavior and activity pattern. The three different states could be related to phases during the dynamic event of diapause development and post-diapause activity in R. appendiculatus. The first hidden state could be defined as a non-responsive dormant phase. Ticks in this state progressively terminate diapause as their physiological age increases. The second hidden state is a responsive dormant phase, a reversible state during which a tick responds to unfavorable and favorable conditions by becoming active or returning to a dormant condition. The third hidden state is a non-dormant phase in which ticks react to microclimatic conditions. The methods presented are especially valuable for studying behavioral diapause, although they can be used for a morphogenetic diapause too.

Keywords: Behavior, ecology, diapause, generalized auto-regression models, hidden Markov chains, Rhipicephalus appendiculatus, ticks.

[†]Institute for Tropical Medicine, Nationalestraat 155, 2000 Antwerp, Belgium.
[*]Author for corresponding — Telephone: + 32 3 247 62 73 and Email: nspeybroeck@itg.be.
[‡]EURANDOM, P.O. Box 513, 5600 MB Eindhoven, The Netherlands.
[§]Center for Statistics, Limburgs Universitair Centrum, Universitaire Campus, 3590 Diepenbeek, Belgium.

1 Introduction

In ecology, an organism often remains in a resting phase for a certain period for instance to synchronize its life cycle with fitting climatic conditions. An example of this phenomenon is the diapausing behavior of *Rhipicephalus appendiculatus*, vector of *Theileria parva*, the causative agent of the bovine disease East Coast fever. Under some circumstances and at a certain point in time, the organism becomes active in order to continue its life cycle. By not taking into account this sudden change of activity in time in an analysis of the activity patterns, the influence of other explanatory variables, which are measured, might be misinterpreted.

Short and Norval (1981) discussed the principal factors governing the seasonal occurrence of *R. appendiculatus* in eastern and southern Africa. They concluded that adult *R. appendiculatus* are active under specific sets of temperature, rainfall/humidity, and day length. In Zambia adult tick activity is generally confined to the well-marked wet season from November to April, probably to ensure that the most vulnerable stages of the lifecycle, namely eggs and larvae, are exposed to conditions of low vapor pressure deficit. In the case of *R. appendiculatus*, numbers of adults on the hosts increase after the onset of the rains (MacLeod, 1970; Pegram *et al.*, 1986; Pegram and Banda, 1990; Berkvens *et al.*, 1995; Berkvens *et al.*, 1998). The main factor responsible for this phenology is thought to be day length where a long photoperiod terminates the state of diapause and induces host seeking in the wet months (Short *et al.*, 1989; Pegram and Banda, 1990). Pegram *et al.* (1984) suggested that the lifecycle of *R. appendiculatus* shows a pattern of seasonal activity controlled mainly by rainfall, including the amount of precipitation, length of the rainy season, number of rainy and cloudy days, and atmospheric humidity. The diapausing behavior of *R. appendiculatus* in Eastern Province of Zambia was reported in Berkvens *et al.* (1995). Under the prevailing conditions at the beginning of the dry season, unfed adults entered a behavioral diapause. The latter was terminated after the onset of the next rains, apparently not after a long day signal, but due to weakening photoperiodic maintenance of the diapause because of increased physiological age of the ticks (Berkvens *et al.*, 1995). This could also be demonstrated in laboratory conditions (Madder *et al.*, 2002).

The biological aim of this paper is to study tick activity under quasi-natural conditions during the 1986-87 rainy seasons taking into account the precipitation, temperature, relative humidity, day length, and origin of the ticks. The interest lies in presenting and applying methodology for detecting a sudden change in the behavior of an organism and in the investigation of the remaining importance of explanatory variables after having accounted for the change in behavior and for the dependency of measurements on the same subject in time.

Guttorp (1995) and MacDonald and Zuchini (1997) provide mathematical background on the topic of hidden Markov chains. In particular the work of Guttorp (1995) can be used as a reference for this methodology. In Mac Donald and Zuchini (1997), the main example consists of a single binary time series recording the behavior of locusts. The theory and applications of generalized[1] auto-regression models have been thoroughly discussed by Lindsey (1999, pp. 62-68). Lindsey (2001, pp. 57-60) fitted a change point model for Poisson distributed responses and also indicated that hidden Markov chains could have been used to detect the change point.

Section 2 concentrates on describing the tick activity study. This also includes some easy-to-use exploratory tools to inspect visually a sudden change of binary data in time. These tools are presented in Subsection 2.1 and consist of cumulative probability plots, transition matrices, and change-point models. Section 3 provides the information and references necessary to understand auto-regression models and hidden Markov chains used in the analysis. Section 4 presents the two main analyses, carried out on the activity patterns of female *R. appendiculatus*, and their respective results. These results are then compared and discussed in Section 5. The data used and the computer code to reproduce the analyses and figures presented in this paper are available on the second[2] and fourth[3] authors' web sites or on request from the first author.

2 Design of the tick activity study

The experiment was carried out at Chipata (13°39'S, 32°30'E) in the Eastern Province of Zambia. On the 11th of July 1986, newly moulted adults of four local *R. appendiculatus* strains were released into circular gauze columns (Nylon mesh PES243, 243µm pores, Swiss Bolting Company) of 5cm diameter

[1] Note that the term 'generalised' is used in this context to indicate the possibility of choosing distributions other than the Gaussian one to model the response.

[2] http://euridice.tue.nl/~plindsey/

[3] http://www.luc.ac.be/~jlindsey/

and 1m tall. The bottom end of each column was glued to a polyvinilchloride open cylinder, which was secured in the soil to a depth of about 10 cm. This allowed the ticks access to the soil. The top of the column was tied to a metal wire frame. The columns were placed under a cover of hessian sacks. No water was sprinkled on the shed during the dry season. Three columns per strain and per sex each contained five female or male ticks were established as just described. Colored numbered tags used by beekeepers to identify their queens were applied to the ticks, allowing individual identification of each tick.

The following stocks of ticks originating from the Eastern provinces were used in this trial:

- *R. appendiculatus* (Michembo, Eastern province, Zambia 1986), replete P_1-female collected from a local Zebu ox at Michembo (altitude 1040 m) on 23 January 1986.
- *R. appendiculatus* (Genda, Eastern province, Zambia 1985), replete P_2-female collected from a local Zebu ox at Genda (altitude 1150 m) on 7 February 1985, P_1-adult fed on rabbit, replete P_1-female collected on 10 January 1986.
- *R. appendiculatus* (Nkolowondo, Eastern province, Zambia 1986), replete P_1-female collected from a local Zebu ox at Nkolowondo (altitude 1080 m) on 23 January 1986.
- *R. appendiculatus* (Lundazi, Eastern province, Zambia 1986), replete P_1-female collected from a local Zebu ox at Lundazi (altitude 1250 m) on 15 January 1986.

Note that there are major differences among these stocks: Ticks from the Michembo have only recently appeared there and are still settling, whereas ticks from the Genda and Nkolowondo locations have been there for a long time. On the other hand, Lundazi is located closer to the Equator and has a longer rainy season and a smaller difference in day length. The activity behavior of the latter ticks could therefore be different from that of ticks from any of the three other locations.

Male ticks showed little and irregular activity and were excluded from the study. Because of mortality and loss of tags, relevant data were collected on the following numbers of females: 7 from Michembo, 8 from Genda, 9 from Nkolowondo, and 9 from Lundazi. Each tick was observed daily around 10 in the morning between October 1, 1986 and March 31, 1987.

They were recorded as either absent from the column, present but inactive, or present and active. A tick was considered to be active when it reacted to the presence of the recorder by actually moving in the column, usually upwards. Waving of the first pair of legs was not deemed enough to constitute full activity and was not recorded as such. When ticks were unobserved for more than 7 consecutive days, they were considered to be dead.

Besides the tick's origin (four localities), several other explanatory variables are measured over time, day of measurement (from day 1 to day 182). Ambient temperature (average temperature in °C) and relative humidity (in %) were recorded by means of a thermohygrograph. This was calibrated weekly against a wet-and-dry-bulb thermometer. The rainfall (in mm), minimum and maximum temperature of the day (in °C) were recorded daily at the study site, and day length (in hours) was calculated using an algorithm described in Duffett-Smith (1985).

Some of these were also used to construct additional explanatory variables such as the average daily vapor pressure deficits (in mm Hg), some change-point indicators, and various cumulative and lagged variables. The daily average vapor pressure deficits (*vpd*) was calculated according to Rosenberg *et al.* (1983) as

$$vpd = 4.58\left(1-\frac{rh}{100}\right)e^{\frac{17.269t}{t+237.3}}$$

where *rh* is the daily relative humidity and *t* is the daily temperature. Two indicators were also created: a rain and a day length change-point variable respectively indicating by zeroes the period before and by ones the period after the highest rainfall peak (which occurred on December 11 or day 69) and the longest day (which occurred on December 21 or day 79). Hence, a change-point indicator variable is a dummy variable being zero before and one after a certain event in time.

Lagged variables are also used in the models presented in Section 4 in order to take into account dependence on previous observed values. A lagged variable is obtained by carrying forward in time the observations recorded. Hence, the following is an example of an explanatory variable and its corresponding lag(1) variable.

`100001100011110000·`	An arbitrary explanatory variable.
`·100001100011110000`	The corresponding lag(1) variable.

Notice that by lagging the explanatory variable once, the first observation of the series is lost.

For the response (see state dependence in Section 3), lagged variables up to three previous values were constructed, whereas for explanatory meteorological variables, only lagged variables up to two previous values were created for the modeling process described in Section 4. From a biological point of view, it is sensible to consider such lagged variables because the short term history of each particular individual most probably influences its current and further behavior.

Finally, only the inactive and active responses collected are considered in this paper although the response collected is actually trichotomous. This decision was made due to the few observations (mainly at the end of an individuals observation series) actually collected in the absent or unobserved category and due to the heterogeneity of the observations in this category. Indeed, towards the end of the study, an individual observed in this category could be hiding or dead. Thus, when a tick was unobserved, the observation was considered to be missing (at random).

2.1 Visualization

The binary response is coded as zero if the tick was inactive and as one if the tick was active. Unfortunately, no pattern can be seen from individual profiles of longitudinal binary responses, because these profiles always overlap exactly at the value one and zero. Hence, the cumulative probabilities over time for the ticks are shown in Figure 1a. Note that cumulative probability plots give an idea of the general trends occurring in the entire sample but cannot represent any individual fluctuations over time. The probability of a tick being observed inactive is represented by the height of the continuous line in Figure 1a and the probability being observed active by the height above that line. From this plot, it can be seen that the probability of a tick being inactive is very high at the start of the study, so that the probability of observing an active tick at the start of the study is almost null. The probability of inactivity decreases at some point in time (which will be investigated) and remains lower until the end, so that the probability of activity increases but fluctuates over time. The vertical dotted and dashed-dotted lines respectively represent the maximum rainfall peak (day 69) and the longest day (day 82).

Figure 1b shows the recorded day lengths. They are smoothly increasing until December 21 (day 82 represented by the vertical dashed-dotted line) and then smoothly decreasing until the end of the study. Figure 1c shows the recorded rainfall during the study. The rainfall suddenly increases and obtains its maximum peak on day 69 (represented by a vertical dashed line). From then on the precipitation level is higher than at the start of the study. Figure 1d to Figure 1f respectively show the recorded relative humidity level, the vapor pressure deficit, and the temperature during the study. These are all highly related to the amount of rainfall: a change in their patterns occurs around the maximum rainfall peak day.

2.1.1 Transition matrices

An estimate of a Markov chain can be obtained using lagged response variables by creating a two-way contingency table. Thus, a first-order two-state Markov chain can be summarized as

$$\begin{pmatrix} \pi_{0|0} & \pi_{1|0} \\ \pi_{0|1} & \pi_{1|1} \end{pmatrix}$$

where for example $\pi_{0|1}$ is the conditional probability of no event following an event. This is called a transition matrix because it describes the changes between time points. One may also consider a second-order Markov chain, which will yield a three-way contingency table.

For the data at hand, the transition matrix is

$$\begin{pmatrix} 0.94 & 0.06 \\ 0.35 & 0.65 \end{pmatrix}$$

Thus, the conditional (transition) probability of no activity given that there was also no activity the previous day is 0.94. This is much higher than when there was activity the previous day (0.35). The conditional probability of activity given that there was activity during the previous day is 0.65. This is also much higher than when there was no activity the previous day (0.06). The marginal probability of being active regardless of the activity pattern of the previous day is 0.14. Hence, a female tick also active on the previous day is 4.56 times more likely to be currently active than this probability. On the other hand, a female tick not active on the previous day is 0.43 times less likely to be currently active.

The data for a second-order Markov chain are given in Table 1. Note that none of the numbers of observations reported in this three-way contingency table include those recorded on the two first days of the study (due to the lagged variables).

Hence, the corresponding transition matrix is

$$\begin{pmatrix} \pi_{0|00} & \pi_{1|00} \\ \pi_{0|01} & \pi_{1|01} \\ \pi_{0|10} & \pi_{1|10} \\ \pi_{0|11} & \pi_{1|11} \end{pmatrix} = \begin{pmatrix} 0.95 & 0.05 \\ 0.75 & 0.25 \\ 0.48 & 0.52 \\ 0.28 & 0.72 \end{pmatrix}$$

Thus, the probability of activity is increasing respectively if there was no activity during the last two days, there was activity only two days ago, there was activity only on the previous day, and there was activity during both previous days. On the other hand, the probability of being inactive decreases respectively to the amount and distance in time of previous active days.

2.1.2 Change-point models

A change-point can be included in most models. This is done by adding to the regression model an indicator variable, which is zero before a certain point in time and then becomes one. The point in time can then be estimated by selecting the model yielding the best fit (usually judge by the smallest minus log-likelihood).

Change-point models will be discussed for nominal responses, the binary case being a special case (obtained by simplification). If **Y** is a random r-vector, it is said to have a multinomial distribution with r categories, n draws, and $\mathbf{p}=(p_1,\ldots p_r)$ success probabilities, so that

$$\Pr(\mathbf{Y}=\mathbf{y}) = \frac{n!}{y_1!\ldots y_r!} p_1^{y_1} \ldots p_r^{y_r}$$

where $y!$ denotes the factorial of y.

In general having observed the outcome y, the aim is to estimate the parameters **p**, using maximum likelihood. The likelihood L(**p**) (given the outcome y) is the probability of the observed value as a function of the unknown parameters **p**. The maximum likelihood estimator of **p** is the value $\hat{\mathbf{p}}$, which maximizes the likelihood. Minimizing minus the logarithm of the likelihood can facilitate computations. In a simple change point model, two values of **p** are estimated simultaneously respectively one before (p_a) and another after a point in time (p_b). More complex models can easily be fitted by e.g. making p_a and/or p_b depend on observed explanatory variables. The example model shown with only one change point can also be adapted to include more change points if necessary.

Hence, fitting this model to the binomial response data (tick active=1, tick non-active=0) will allow a change in the average response at some unknown point in time. Figure 2a shows the corresponding negative normed likelihood for the binomial response. Day 69, which corresponds to December 7, was estimated as the optimal change-point. Figure 2b shows the fit of this model. Once the tick is active, a cyclic pattern of activity is visible.

3 Statistical background

This section gives a detailed description of the models used throughout Section 4 in order to provide a better understanding of the different analyses performed on the tick activity dataset.

Models can be classified into two categories, namely individual and population models. The difference between these depends on the regression constraints applied: they may be applied respectively to each of the individual observed series and to the population average. Only individual models are considered in this paper.

Several types of dependencies among observations on a subject in time can be considered. In this paper, we shall be using three different ones. State dependence adjusts the model by taking the actual previous observation into account. In serial dependence, the difference between the model prediction at the previous time point and the actual previous observation is used to take the individuals history into account in the model. Finally, hidden Markov chains allow the undergoing biological process to switch over time among several hidden states (or behaviors). Such dependencies are called spells.

Hence, state dependence, serial dependence, and spells are three different types of history dependencies; they all allow the subject's history to be taken into account by the model but in different ways. Sometimes, it is useful to consider several of these different types of dependencies together in a model.

3.1 Generalized auto-regression models

For observations collected at equally spaced times, a Markov process can be of order M, denoted by AR(M), and can be written as

$$\mu_t = \alpha_0 + \sum_{h=1}^{M} \rho_h y_{t-h} \tag{1}$$

where μ_t is the location parameter of the distribution, α_0 characterizes the intercept, ρ is the Markov process parameter (constrained between 0 and 1) that quantifies the dependence strength on the previous response, and y_{t-h} is the h^{th} previous response with respect to time t (Lindsey, 1999, p. 50).

Equation (1) can be generalized further for instance by using an appropriate link function g(μ_t) to describe the location of the distribution (Lindsey, 1999, pp. 35-37). This yields a generalized auto-regression model because the present location parameter depends directly on the previous state of the subject (given by the previous responses). Models with this type of dependence will from now on be referred to as state dependence models.

On the other hand, models where the location of the distribution does not depend directly on the previous response but on the difference between the previous response and its predictor

$$\mu_t = \alpha_0 + \sum_{h=1}^{M} \rho_h \left(y_{t-h} - \mu_{t-h} \right)$$

will from this point be referred to as serial dependence models. The implementation of these latter dependencies for non-Gaussian distributions is more complex than the implementation of a state dependency. This is mainly due to the estimation method, which must be carried out iteratively because each fitted value depends on that at the previous point in time. First and second order serial dependence models are considered in this paper.

A change point could also be combined with an auto-regression model, but this will not be performed in this paper. Indeed, we preferred to consider only change-point variables, which could biologically be interpreted such as the longest day or the start of the rainy season.

Finally, it is also important to assess whether the observations collected are following a stationary or non-stationary process. So far, the biological process producing the observations collected has been assumed to be in equilibrium (or a stable state) suggesting that a stationary Markov process should be considered to model the observations. But this might not be the case; indeed the biological process may not yet have reached equilibrium and is therefore still evolving suggesting that a non-stationary dependence process should be considered. In this latter case, the dependence structure will have two parameters (φ and ρ). These parameters do not measure the dependence strength on just the previous residual but capture a recursively fading dependence on all previously collected observations. Note that when the first parameter (φ) tends to zero the dependence structures indicates that a stationary Markov process, with parameter (ρ) measuring the dependence strength on the previous residual, is more suitable model for the observations considered.

3.2 Hidden Markov chains

The dependence structure in the previously discussed models was constructed conditionally by taking each subject's previous observed history into account. Dependence among each subject's observations will now be induced differently for the case where the responses are generated in one of several different unknown states. The subjects are not restricted to be influenced by the same set of external factors (explanatory variables) in each of these states, so that a (linear or non-linear) regression is required for each one. The model can sometimes be simplified by allowing common parameters among these different regression equations. Now, the dependence is induced conditional on the subject's previous hidden state history rather than its previous observed history. This type of dependence can be used to model spells.

All possible changes of state over time must be taken into account by the probability model. Each subject's possible "path" through the states corresponds to a product of conditional probabilities. Because the dependence is conditioning on each subject's previous state history, it is calculated by multiplying the probability of a particular subject being in each state at each given time point by a transition probability. This ensures (because they are probabilities) that summing all these products of conditional probabilities together produces the joint probability over time and all possible states for a particular subject. The sample probability is then obtained by multiplying these sums of products of conditional probabilities together over all individuals, yielding a likelihood.

One last problem remains concerning the initial conditions. At the first time point, no previous information is available to estimate the probability of being in a particular state. This can be solved either by introducing $s-1$ additional initial marginal probability parameters for s states or by using the stationary marginal transition probabilities. This latter case assumes that stationarity of the hidden process has been reached. This solution is chosen. Then $s(s-1)$ transition probabilities must be estimated along with the regression parameters.

Consider a simple case with two states and three time points. As above for ordinary Markov chains, the transition probabilities are $\pi_{1|1}$, $\pi_{1|2}$, $\pi_{2|1}$, and $\pi_{2|2}$ and the marginal probabilities are $\delta = \frac{\pi_{1|2}}{\pi_{2|1}+\pi_{1|2}}$ and $1-\delta$ respectively for state 1 and state 2 (MacDonald and Zucchini, 1997, pp. 67–68). The joint probability over time and all possible states for subject i is then

$$\begin{aligned}
&\Pr\left(Y_{i1}=k_1,\ldots,Y_{im}=k_m\right)= \\
&\quad\delta\ \Big\{\Pr\left(Y_{i1}=k_1|S_{11}\right)\ \ \pi_{1|1}\left[\Pr\left(Y_{i2}=k_2|S_{21}\right)\ \pi_{1|1}\Pr\left(Y_{i3}=k_3|S_{31}\right)+\pi_{2|1}\Pr\left(Y_{i3}=k_3|S_{32}\right)\right] \\
&\qquad\qquad+\pi_{2|1}\left[\Pr\left(Y_{i2}=k_2|S_{22}\right)\pi_{1|2}\Pr\left(Y_{i3}=k_3|S_{31}\right)+\pi_{2|2}\Pr\left(\mathrm{Y}_{i3}=k_3|\mathrm{S}_{32}\right)\right]\Big\} \\
&\quad+\left(1-\delta\right)\Big\{\Pr\left(Y_{i1}=k_1|S_{12}\right)+\pi_{1|2}\left[\Pr\left(Y_{i2}=k_2|S_{21}\right)\ \pi_{1|1}\Pr\left(Y_{i3}=k_3|S_{31}\right)+\pi_{2|1}\Pr\left(Y_{i3}=k_3|S_{32}\right)\right] \\
&\qquad\qquad+\pi_{2|2}\left[\Pr\left(Y_{i2}=k_2|S_{22}\right)\ \pi_{1|2}\Pr\left(Y_{i3}=k_3|S_{31}\right)+\pi_{2|2}\Pr\left(Y_{i3}=k_3|S_{32}\right)\right]\Big\}
\end{aligned}$$

where $\Pr(Y_{ij}=k_j|S_{jh})$ is the probability of the response being observed in category k at time j for subject i given it is in state h at time j. Notice that this collapses to the product of independent probabilities if there is only one state.

Unfortunately, the number of terms expands exponentially as the number of time points increases and quadratically for the number of states considered. This expression is therefore not computationally feasible as it stands for any reasonable number of states and/or time points. However, it can be rearranged in a recursive form over time (MacDonald and Zucchini, 1997, pp. 78–79; Lindsey, 1999, p.73). The joint probability over time and all possible states for subject i can then be written as

$$\Pr\left(Y_{i1}=k_1,\ldots,Y_{im}=k_m\right)=\boldsymbol{\delta}^T\prod_{j=1}^{m}\left(\boldsymbol{\pi}D_{ijk_j}\right)\mathbf{J}^T \tag{2}$$

where $\boldsymbol{\delta}$ is a row vector containing the marginal probabilities, $\boldsymbol{\pi}$ is the transition matrix, D_{ijk_j} is an $s\times s$ matrix containing on the diagonal the probabilities of the response being observed in category k at time j for subject i given the various possible states, and $\mathbf{J}$ is a row vector of ones.

The likelihood can then be obtained by multiplying Equation (2) over all subjects. Note that at the first time point the Equation (2) reduces to

$$\Pr\left(Y_{i1}=k_1\right)=k_1\boldsymbol{\delta}^T D_{ijk_j}\mathbf{J}^T$$

because the vector of marginal probabilities $\boldsymbol{\delta}$ is obtained by solving $\boldsymbol{\delta}^T(\mathbf{I}-\boldsymbol{\pi})=0$.

Because the responses are binary, the probabilities at time j contained on the diagonal of D_j will be obtained using the binary logistic parameterization.

A hidden Markov chain can be illustrated for instance by considering ticks being active or not (a binary time series). The tick can now be in one of two unobservable states, a state where its metabolism is low and another state where it is high. There are no ways to measure this directly but a tick can behave differently (and shows a different activity outcome) depending on which state it is in. Each of the two possible events might be generated by one of the two Bernoulli distributions. The process switches from the one to the other according to the state of the hidden Markov chain, in this way generating dependence over time. Analogous models can be constructed for binary, nominal, or ordinal responses (but only the binary models are appropriate for analyzing the data at hand).

3.2.1 Filtered and recursive probabilities

Two types of recursive probabilities can now be extracted from this model.

These are obtained from the intermediate values

$$\zeta_{ijr} = \sum_{o=1}^{s} \zeta_{i,j-1,o}\pi_{or} \Pr(Y_{ij} = k_j \mid S_{jr})$$

calculated while constructing the joint probability over time and all possible states for subject i are required, where

$$\zeta_{i1r} = \delta_r \Pr(Y_{i1} = k_1 \mid S_{1r})$$

is obtained for the first time point.

A filtered probability is the probability that a specific subject is in a particular hidden state given this subject's previous state history. Hence, the probability of subject i being in state r at time j is

$$\xi_{ijr} = \frac{\zeta_{ijr}}{\sum_{o=1}^{s} \zeta_{ijo}}$$

which is obtained by standardizing the ζ_{ijr}.

The probabilities of the response being observed in category k at time j for subject i can then be calculated by

$$\varphi_{ij} = \sum_{o=1}^{s} \xi_{ijo} \Pr(Y_{ij} = k_j \mid S_{jo})$$

which are the recursive probabilities for subject i.

4 Analyses of tick activity

4.1 Modeling strategy

Because the modeling process is exploratory, the inference criterion used for comparing the models under consideration is their ability to fit the observed data that is how probable they make the data. In other words, models are compared directly through their minimized minus log likelihood. When the numbers of parameters in models differ, they are penalized by adding the number of estimated parameters, a form of the Akaike information criterion (AIC, see Akaike, 1973; Lindsey and Jones, 1998). Smaller values indicate more preferable models. This criterion allows direct comparisons among models that are not required to be nested.

AICs are only comparable if they are calculated by fitting models based on the same (data and) number of observations. Hence, care must be taken when working with lagged variables. As lag(1), lag(2), and lag(3) response variables (previous, second previous, and third previous day activity status) are considered during the analysis presented in this section, all the exploratory methods and analyses included in this paper are based on the tick dataset with all observations recorded during the three first days removed. This allows all desired comparisons among results presented in this paper.

The analysis is started by fitting an intercept or null model to provide a reference point for comparison with further fitted models. Models, each containing only one of the different explanatory variables, are then fitted and sorted in ascending order of their AIC value. The explanatory variable model with the lowest AIC is selected as the starting point for the model building process. Additional explanatory variables are then added to this model according to their AIC value and are only kept in the model if the AIC reduces. If the final model contains change-point variables, interactions between change-point variables and the explanatory variables present in the model are considered. This step will provide additional information on differences in behavior before and after the change-point. The last modeling step is then to check whether any explanatory variables are no longer significant and could therefore be deleted from the model. Care must be taken during this last step, because the model must remain hierarchically valid. In other words, a non-significant main effect will not be removed from the model if this explanatory variable is still present in an interaction term.

Finally, for all models presented, a logit link was used as a one-to-one transformation of the expected probability of a tick being active (which ranges from zero to one) to the right hand side of the regression equation (which can range between minus and plus infinity). Hence, when the right hand side of the regression equation evaluates to zero, the tick is then as probable to be active as inactive (the expected probability of the tick being active is one half).

4.2 Statistical computations

The analyses presented in this paper are performed using special packages in R (Ihaka and Gentleman, 1996). R is a fast S-Plus clone freely available (http://cran.r-project.org) under the GNU license, which we thank Robert Gentleman, Ross Ihaka, and the R core group for developing.

The fourth author has written a collection of R packages for repeated measurements and non-linear regression, which can be downloaded from his web page (http://www.luc.ac.be/~jlindsey/rcode.html). The generalized auto-regression models and the hidden Markov chains can be fitted respectively using the *gar* and *hidden* functions provided by the package *repeated*[4].

The functions to fit change point models, the code used to perform the analyses, and the tick dataset can be obtained on the second[5] and fourth[6] authors' web sites or on request from the first author.

4.3 Generalized auto-regression models

To begin the analysis, a binomial regression is fitted that only contains the intercept parameter. This one parameter model has an AIC of 2146. The AIC can then be lowered to 1371.3 by adding a non-stationary dependence. This model now contains three parameters.

Adding state dependence on the previous three days activity status improves the model by lowering the AIC to 1327.4 (six-parameter model). The cumulative rainfall, longest day indicator, maximum temperature, change in maximum temperature and in rainfall, acceleration in maximum temperature and in rainfall, vapor pressure deficit, Lundazi location, and interaction between the longest day indicator and both the previous day activity status and the vapor pressure deficit are the explanatory variables to enter the model. This seventeen-parameter model has now an AIC of 1290.1 and regression equation

$$
\begin{aligned}
\text{logit}(\mu) = \textit{0.436} &+ 2.754 \times \text{previous day activity status} + 0.474 \times \text{second previous day activity status} \\
&+ 0.310 \times \text{third previous day activity status} + 0.001 \times \text{cumulative rainfall} \\
&+ \textit{0.321} \times \text{longest day indicator} - 0.154 \times \text{maximum temperature} \\
&+ 0.091 \times \text{change in maximum temperature} - 0.055 \times \text{acceleration in maximum temperature} \\
&+ 0.022 \times \text{change in rainfall} - 0.012 \times \text{acceleration in rainfall} \\
&- \textit{0.0001} \times \text{vapour pressure deficit} - 0.669 \times \text{Lundazi location} \\
&- 0.975 \times \text{longest day indicator} \times \text{previous day observation} \\
&+ 0.140 \times \text{longest day indicator} \times \text{vapour pressure deficit}
\end{aligned}
$$

where μ is the expected (or average) probability of being active and the three italized coefficients are not significantly different from zero. The dependence parameters for the non-stationary process are 0.916 (φ) and 0.702 (ρ). The first parameter (φ) clearly indicates that the biological process is still evolving over time and that the dependence structure can therefore not be simplified to a first-order Markov process.

The effect of a variable can be judged by its regression coefficient along with its actual observed range. Thus, the maximum temperature is the most important factor in the model obtained (because the daily maximum temperature varies between 19 and 38˚C and is on average 23˚C). Because the maximum temperature coefficient is negative, an increase in temperature corresponds to a decrease in the probability of a tick being active. Indeed, even if a tick has been active during the three previous consecutive days, this tick's current probability of being active will be below one half for (constant or increasing) temperatures above 24.3°C.

The tick's activity history is also quite important. Indeed, the probability of a tick being active increases according to the number of previous days it has already been active. This indicates that once a tick becomes active it will have a greater probability of remaining active. The change in amount of rainfall respectively increases and decreases the probability of a tick being active each time a rainy period starts and ends. This factor will have the most impact on the probability of a tick being active when the rainy season starts and hence can be interpreted as one of the signals indicating to the tick to start becoming active.

Another one of the signals that could be indicating to the tick to start becoming active is the longest day but in the model obtained the coefficient of the main effect is not significantly different from zero.

---

[4] Note that this package also requires the *rmutil* package to be installed.

[5] http://euridice.tue.nl/~plindsey/

[6] http://www.luc.ac.be/~jlindsey/

Note that the AIC rises from 1290.1 to 1296.1 when the rainy season indicator replaces the longest day indicator. On the other hand after the longest day (December 21 or day 79), the effect of the previous day activity status decreases and the vapor pressure deficit proportionally affects the probability of a tick being active (due to the two interaction terms with the longest day indicator).

On the other hand, the cumulative rainfall can be interpreted as a non-linear aging effect, which slowly obliges the probability of the tick being active to increase over time regardless of any other factors.

The acceleration in rainfall is the fourth most important factor influencing the probability of a tick being active, closely followed by the change and acceleration in maximum temperature. Figure 3 represents the change in probability due to the acceleration in rainfall or in maximum temperature, as accelerations are not straightforward to visualize[7].

The model presented also distinguishes ticks from the Lundazi location, because the probability of these ticks being active remains lower throughout the entire study. From Figure 4 six out of the nine ticks from Lundazi are only active for a very short period of time (including a tick that is never observed as being active) compared to ticks from other locations.

The effect of these explanatory variables along with the ones of the state and serial dependence can also be seen from Figure 4. The underlying population curve indicated by the continuous line keeps slightly increasing (due to the cumulative rainfall) and fluctuating over time (due to the meteorological explanatory variables). The effect of adding a state and then a serial dependence can also clearly be seen. These individual (or recursive) curves are respectively represented by the dashed and dotted lines. As expected by such dependencies, a tick's probability of being active is pulled down or pushed up according to its history and is therefore closer to its observed activity status represented by the filled circles.

4.4 Hidden Markov chains

As in the previous section, the analysis is started by fitting the intercept model in order to provide a reference point for comparison with further fitted models. This intercept model is identical to the one fitted for the generalized auto-regression model with an AIC of 2146.8.

The AIC is then lowered to 1326.3 by introducing three hidden states. When s states are introduced in the model, each one of these states is allowed to have a separate regression and the transition matrix has $s\times(s-1)$ parameters minus any of these fixed at zero. Hence, the last model fitted contains eight parameters because each state has a separate regression containing only an intercept parameter and the transition matrix has six parameters minus one fixed at zero.

The final model resulting from this modeling process has an AIC of 1265.1 and contains thirty-seven parameters. Thus, it fits considerably better than the final generalized auto-regression model obtained in Section 4.3 but it is also much more complex. The regression equations for the three states are

$$\begin{aligned}\text{logit}(\mu_1) = &-593.332 + 0.543\times\text{cumulative rainfall} + 1.246\times\text{rainfall} - 0.301\times\text{change in rainfall}\\ &+ 0.312\times\text{acceleration in humidity} + 44.658\times\text{Genda location}\\ &+ 31.278\times\text{Nkolowondo location}\end{aligned}$$

$$\begin{aligned}\text{logit}(\mu_2) = &\ 5.528 + 1.944\times\text{previous day activity status} + 0.877\times\text{second previous day activity status}\\ &+ 3.054\times\text{longest day indicator} + 3.004\times\text{rainy season indicator}\\ &- 0.261\times\text{maximum temperature} + 0.409\times\text{change in maximum temperature}\\ &- 0.294\times\text{acceleration in maximum temperature} - 0.029\times\text{rainfall}\\ &+ 0.026\times\text{change in rainfall} - 0.074\times\text{humidity} + 0.104\times\text{change in humidity}\\ &- 0.068\times\text{acceleration in humidity} - 0.634\times\text{Lundazi location}\end{aligned}$$

$$\begin{aligned}\text{logit}(\mu_3) = &\ 6.793 + 1.668\times\text{previous day activity status} + 0.213\times\text{second previous day activity status}\\ &+ 0.522\times\text{third previous day activity status} - 0.230\times\text{maximum temperature}\\ &+ 0.012\times\text{rainfall} + 0.032\times\text{change in rainfall} - 0.021\times\text{acceleration in rainfall}\\ &- 0.032\times\text{humidity} - 0.021\times\text{change in humidity} - 0.392\times\text{Genda location}\\ &+ 1.040\times\text{Nkolowondo location}\end{aligned}$$

where all explanatory variables are significant.

[7] The acceleration is obtained by the following formula:
$[(\text{today} - \text{previous day}) - (\text{previous day} - \text{second previous day})] = \text{today} - 2\times\text{previous day} + \text{second previous day}$

The transition matrix is

$$\begin{pmatrix} 0.980 & 0.020 & 0.000 \\ 0.022 & 0.970 & 0.008 \\ 0.000 & 0.013 & 0.987 \end{pmatrix}$$

and the stationary distribution is (0.409, 0.368, 0.224).

The transition matrix shows that a tick cannot change directly from behavior described by the first (or third) hidden state to the behavior described by third (or first). For such a change in behavior to occur the tick must always go through the transitional behavior described by the second hidden state. From this transition matrix, it can also be seen that the probability of remaining in a particular state is quite high and that the probability of changing from a state to the next one and from this state back to the previous are almost identical.

The three states can be interpreted as follows. The first hidden state corresponds to a behavior where the ticks are in diapause but takes into account an aging effect. Indeed, the probability of a tick being active (when following this type of behavior) will only tend to reach one half towards the end of the study once the cumulative rainfall becomes large enough. At that point in time, the only meteorological factors that will affect a tick's behavior are the rainfall, changes in rainfall, and acceleration in humidity. Ticks from the Genda and Nkolowondo locations have a constant higher probability of being active than ticks from the other two locations when following this type of behavior. This indicates that ticks from these two locations age faster and will become active earlier if they remained throughout the study in this type of behavior.

The second hidden state corresponds to a behavior where the tick is waiting for indications as to when to become active and then, once this has occurred, corresponds to an out-of-diapause behavior. The probability of a tick being active (when following this type of behavior) will only be greater than one half once the rainy season starts due to the presence of the maximum temperature in this part of the model. If this indicator is not sufficient to trigger certain ticks out-of-diapause, then this will most likely occur once the longest day of the year has been reached. After the longest day is reached (on December 21 or day 79), this hidden state describes the influences of external factors on a possible behavior followed by ticks out-of-diapause. Ticks from the Lundazi location have a slightly lower probability of being active than ticks from one of the other three locations when following this type of behavior.

The third hidden state only describes an out-of-diapause behavior, which is influenced by external factors. The probability of a tick being active, when following this type of behavior, is slightly smaller and more dependent on meteorological factors than the out-of-diapause (or later) behavior described for the second hidden state. The rainy season and longest day indicators are not present in this part of the model and hence do not provide the major increase in the probability of a tick being active once these events have occurred (as was the case in the second hidden state). Ticks from the Genda and Nkolowondo locations respectively have a slightly lower and higher probability of being active than ticks from the other two locations when following this type of behavior.

The differences in behavior of colonizers (ticks from the Michembo location), settled down ticks (from Genda and Nkolowondo locations), and ticks collected closer to the Equator (Lundazi location) are also captured by this model and described by linear shifts of the probability of being active. Hence, ticks just have a lower or higher probability of being active from the start and the external factors triggering and influencing activity have the same effect on all of them.

This model also points out three different subgroups of ticks present at all four locations. These subgroups respectively characterize "early", "optimal", and "late" reactors. An "optimal" reactor would be a tick that starts in the first hidden state and therefore would just be aging. It then would go to the second state where it would be waiting for a signal indicating that it is time to start being active. The next step would be to go to the third hidden state where meteorological conditions would drive the amount of activity. Finally, it might go back to the second and then to third hidden state indicating a last intense period of activity, which would be less influenced by external factors. Such ticks would be individuals three and four on Figure 5, individuals nine to twelve on Figure 6, individuals sixteen and eighteen on Figure 7, and individuals twenty-six, twenty-eight, and twenty-nine on Figure 8.

An "early" reactor would be a tick that starts directly in the third hidden state and therefore is already very active at the beginning of the study. Such a tick does not need to age before being active nor wait for a signal indicating the beginning of favorable conditions for being active. Hence, its behavior is straight away influenced by meteorological conditions. Such ticks would be individuals one and two on Figure 5, individual eight on Figure 6, and individual twenty-five on Figure 8.

A "late" reactor would be a tick that starts in the first hidden state and remains aging for quite a long time. It switches to the second hidden state after the signals indicating the beginning of favorable conditions for being active have occurred. Because it has waited too long for optimal conditions to be active, it must be very active during the remaining time of the study. Thus, such ticks remain mostly in the second hidden state, although they can on an occasion briefly go to the third hidden state. Towards the very end of the study, aging might catch up with such ticks and switch them back to the behavior described by the first hidden state. Such ticks would be individuals five, six and seven on Figure 5, individuals thirteen, fourteen, and fifteen on Figure 6, individuals seventeen and nineteen to twenty-four on Figure 7, and individuals twenty-six and thirty to thirty-three on Figure 8. Note that individual thirty-three is very peculiar, because it never becomes active. Hence, it remains throughout the study in the first hidden state, which indicates that eventually the aging process will force the tick to become active. Nevertheless, forced activity due to the tick's aging process was not observed during this study for this female tick.

The fit of this model can be assessed from Figure 9. Indeed, it can clearly be seen that the underlying population curve indicated by the continuous line increases slightly and fluctuates accordingly to the meteorological explanatory variables from the start of the rainy season or from the longest day until the end of the study. The dashed and dotted lines represent the individual (or recursive) curves. The dashed curve shows the effect of adding state dependence. Finally, the probability of an individual being active given the optimal path through the hidden states for its activity history is represented by the dotted line.

5 Discussion

This article had two purposes: (1) propose new techniques that could be used in a general ecological context where a sudden change of the behavior of an organism occurs, (2) the application on itself is important because new information on the behavior of adult *R. appendiculatus* could be assessed.

Three different types of dependencies among observations on a subject in time were considered. State dependence, serial dependence, and spells (hidden Markov models); they all allow the subject's history to be taken into account by the model but in different ways. State dependence adjusts the model by taking the actual previous observation into consideration. In serial dependence, the difference between the model prediction at the previous time point and the actual previous observation is used to take the individuals history into account in the model. Finally, hidden Markov chains allow the undergoing biological process to switch over time among several hidden states (or behaviors), called spells. Disregarding the serious dependence that exists in the data at hand would have resulted in erroneous conclusions. Hidden Markov chains are especially appropriate to study post-dormancy behavior, portrayed in this study by the activation of *R. appendiculatus* after behavioral diapause. Behavioral diapause involves a temporary interruption in a hierarchical sequence of behavioral patterns (Belozerov, 1982). Belozerov (1982) mentioned a weakening of the photoperiodic diapause maintenance with increasing age of the tick. In contrast, morphogenetic diapause comprises all categories of diapause whereby a development process is temporarily suspended or interrupted. An example of the latter form of diapause is the delay in oviposition of engorged *Amblyomma variegatum* females (Pegram et al., 1988). Change point models probably suffice for analyzing morphogenetic diapause as the state is always observable but the unobservable levels that occur in behavioral diapause can probably only be analyzed by using hidden states through hidden Markov chain models. As pointed out by Hodek (1996), the end of diapause can be indicated by a covert potential to be active or by an overt activity. In the former situation (adverse) climatic conditions prevent post-diapause ticks being active, forcing them into quiescence until better climatic conditions occur. In the latter, activity has been observed but sometimes long after the end of diapause. Because it is extremely difficult to distinguish between dormancy levels in the first situation, hidden Markov chain models could be useful in clarifying the hidden states.

R. appendiculatus can be in three different states. These states could be related to phases during the dynamic event of diapause development and post-diapause activity in *R. appendiculatus*. The first hidden state could be defined as a non-responsive dormant phase known as a horotelic (gradual) completion of diapause (Hodek, 1996). Ticks in this state progressively terminate diapause as their physiological age increases (Madder, 2002). The second hidden state is a responsive dormant phase most comparable with quiescence, a reversible state during which a tick responds to unfavorable and favorable conditions by becoming active or returning to a dormant condition. The third hidden state is a non-dormant phase in which ticks react to microclimatic conditions. It can be remarked that Eastern Province in Zambia is a transition zone, where uni- and bi-voltine populations are observed (Berkvens *et al.*, 1998). This necessitates flexibility in the behavior of ticks, as indicated in this study.

All the models indicate that temperature plays a major role in the activity-pattern of *R. appendiculatus,* higher temperatures leading to lower activity. In this data set the dependency is very strong. The model shows that activity and non-activity act in an absorbing way meaning that once a tick becomes active it shows a tendency to remain active. The auto-regression models further indicate that activity suddenly changes around the longest day, also indicated by the change-point model. The reaction of ticks on acceleration in rainfall, temperature indicates that they sense climatic changes. The change-point model can only detect an irreversible change of the behavior in time. The hidden Markov chains allow subjects to move back to the state where it came from.

Computationally all the above methods are very demanding and this is probably the reason that the methodology in this paper is relatively unused. With increasing computing power, the here presented methods may be used more frequently.

References

[1] Akaike, H. (1973) Information Theory and an Extension of the Maximum Likelihood Principle. in Petrov, B.N. and Csàki, F., eds. *Second International Symposium on Information Theory*, Budapest: Akadémiai Kiadó, pp. 267-281.

[2] Belozerov, V.N. 1982. Diapause and biological rhythms in ticks. In: Obenchain, F.D. and Galun, R. (Eds). Physiology of ticks. Pergamon Press, Oxford.

[3] Berkvens, D.L., Pegram, R.G., and Brandt, J. (1995) A study of the diapausing behaviour of *Rhipicephalus appendiculatus* and *Rhipicephalus zambeziensis* under quasi-natural conditions in Zambia. *Medical and Veterinary Entomology,* **9**, 307-15.

[4] Berkvens, D.L, Geysen DM, Chaka G, Madder M, and Brandt JR. (1998) A survey of the ixodid ticks parasitising cattle in the Eastern province of Zambia. *Medical and Veterinary Entomology,* **12**, 234-40.

[5] Duffett-Smith, P. (1985). Astronomy with your personal computer. Cambridge University Press, Cambridge.

[6] Guttorp, P. (1995) *Stochastic Modelling of Scientific Data.* London: Chapman and Hall.

[7] Hodek, I. (1996). Diapause development, diapause termination and the end of diapause. European Journal of Entomology, **93**, 475-487.

[8] Ihaka, R. and Gentleman, R. (1996) R: a language for data analysis and graphics. *Journal of Computational Graphics and Statistics,* **5**, 299-314.

[9] Lindsey, J.K. (1999, 2nd edn.) *Models for Repeated Measurements.* Oxford: Oxford University Press.

[10] Lindsey, J.K. (2001) *Nonlinear Models for Medical Statistics.* Oxford: Oxford University Press.

[11] Lindsey, J.K. and Jones, B. (1998) Choosing among generalized linear models applied to medical data. *Statistics in Medicine,* **16**, 59-68.

[12] MacDonald, I.L. and Zuchini, W. (1997). *Hidden Markov and other Models for Discrete-Valued Time Series.* London: Chapman and Hall.

[13] MacLeod, J. (1970). Tick infestation patterns in the southern province of Zambia. *Bulletin of Entomological Research,* **60**, 253-274.

[14] Madder, M., Speybroeck, N., Brandt, J. Tirry, L., Hodek, I., and Berkvens, D. (2002). Geographic variation in diapause response in the African brown ear tick Rhipicephalus appendiculatus (Acari: Ixodidae). *Experimental and Applied Acarology*, in press.

[15] Pegram, R.G., Perry, B.D., and Shells, H.F. (1984) Seasonal activity of the parasitic and non-parasitic stages of cattle tics in Zambia. in Griffiths, D.A. and Bowmann, C.E., eds. *Acarology VI, Vol. 2,* Chichester: Ellios Horwood, pp. 1183-1188.

[16] Pegram, R.G., Perry, B.D., Musisi, F.L. And Mwanaumo, B. (1986). Ecology and phenology of ticks in Zambia: seasonal dynamics on cattle. *Experimental and Applied Acarology*, **2**, 25-45.

[17] Pegram, R.G., Mwase, E.T., Zivkovic, D., and Jongejan, F. (1988). Morphogenetic diapause in *Amblyomma variegatum* (Acari: Ixodidæ). *Medical and Veterinary Entomology*, **2**, 301-307.

[18] Pegram, R.G. and Banda, D.S. (1990) Ecology and phenology of cattle ticks in Zambia: Development and survival of free-living stages. *Experimental and Applied Acarology,* **8**, 291-301.

[19] Rosenberg, N.J., Blad, B.L., and Verma, S.B. (1983, 2nd edn.) *Microclimate: The biological environment.* New York: John Wiley and Sons.

[20] Short, N.J. and Norval, R.A.I. (1981) Regulation of seasonal occurrence in the tick *Rhipicephalus appendiculatus* Neumann, 1901. *Tropical Animal health and Production,* **13**, 19-26.

[21] Short, N.J., Floyd, R.B., Norval, R.A.I., and Sutherst, R.W. (1989). Survival and behaviour of unfed stages of the ticks *Rhipicephalus appendiculatus*, *Boophilus decoloratus* and *B. microplus* under field conditions in Zimbabwe. *Experimental and Applied Acarology,* **6**, 215-236.

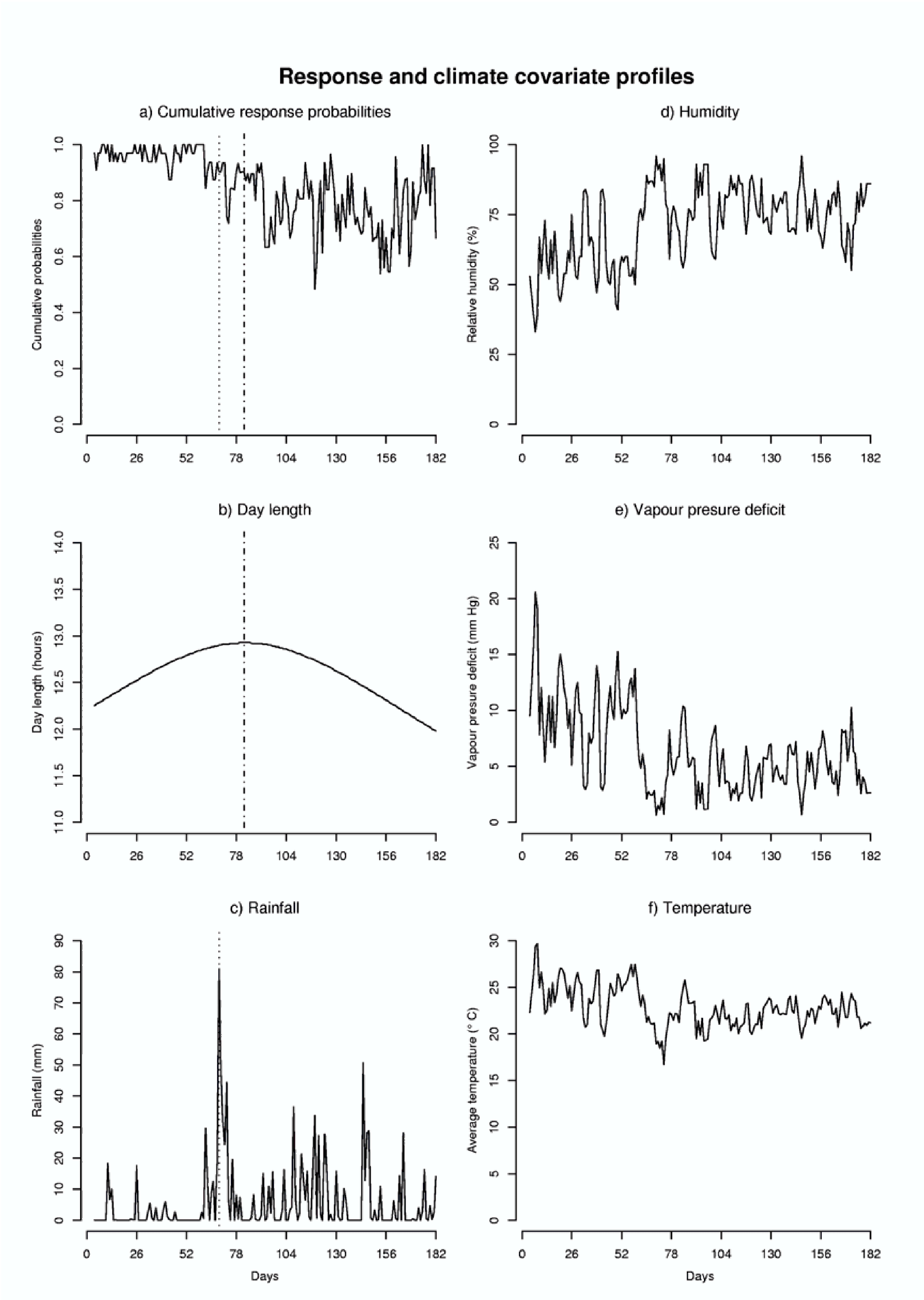

Figure 1: a) Plot of the cumulative probabilities over time for the observed responses. The height of the continuous line represents the overall probability curve of being inactive. On the other hand, the height above this continuous line represents the overall probability curve of being active. The vertical dotted and dash-dotted lines respectively represent the rainiest and the longest days. b) Plot of the day length (in hours) over time. c) Plot of the rainfall (in mm) over time. d) Plot of the relative humidity (in percentage) over time. e) Plot of the vapor pressure deficit (in mm Hg) over time. f) Plot of the average temperature (in °C) over time.

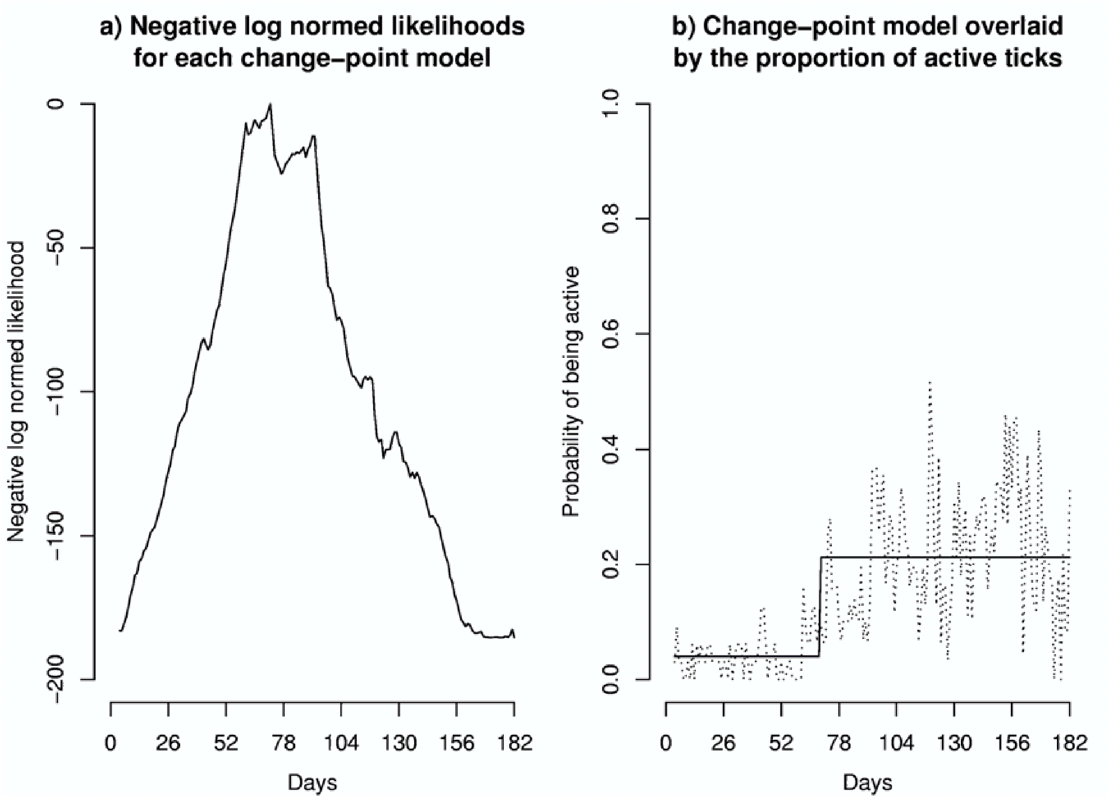

Figure 2: a) Negative log normed likelihoods for the change-point model with a change in activity occurring respectively on each corresponding day. b) Fit of the change-point model for the data (solid line) overlaid by the daily average probability of tick activity (dotted line).

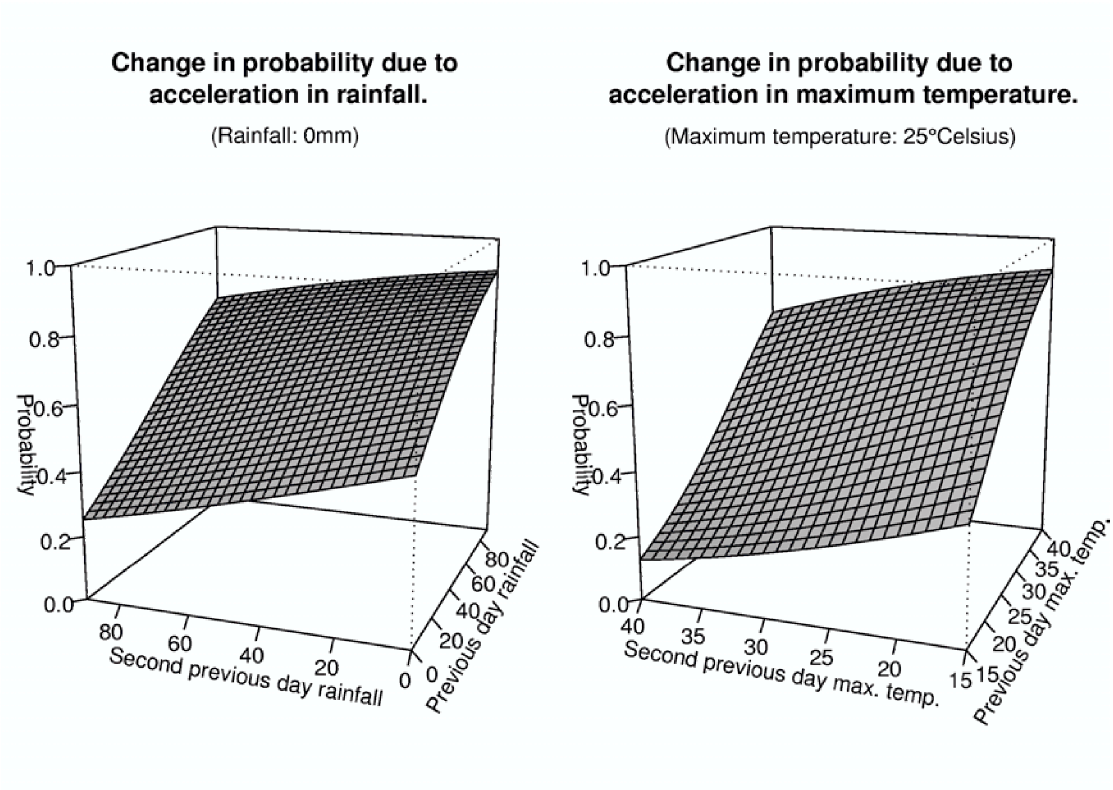

Figure 3: Plots of the change in probability due to the acceleration in rainfall (left panel) and in maximum temperature (right panel) for the generalized auto-regression model. Note that there is no change when the probability is equal to one half. In other words, when the previous and second previous observation of rainfall (or maximum temperature) are both (respectively) equal to the current rainfall (or maximum temperature).

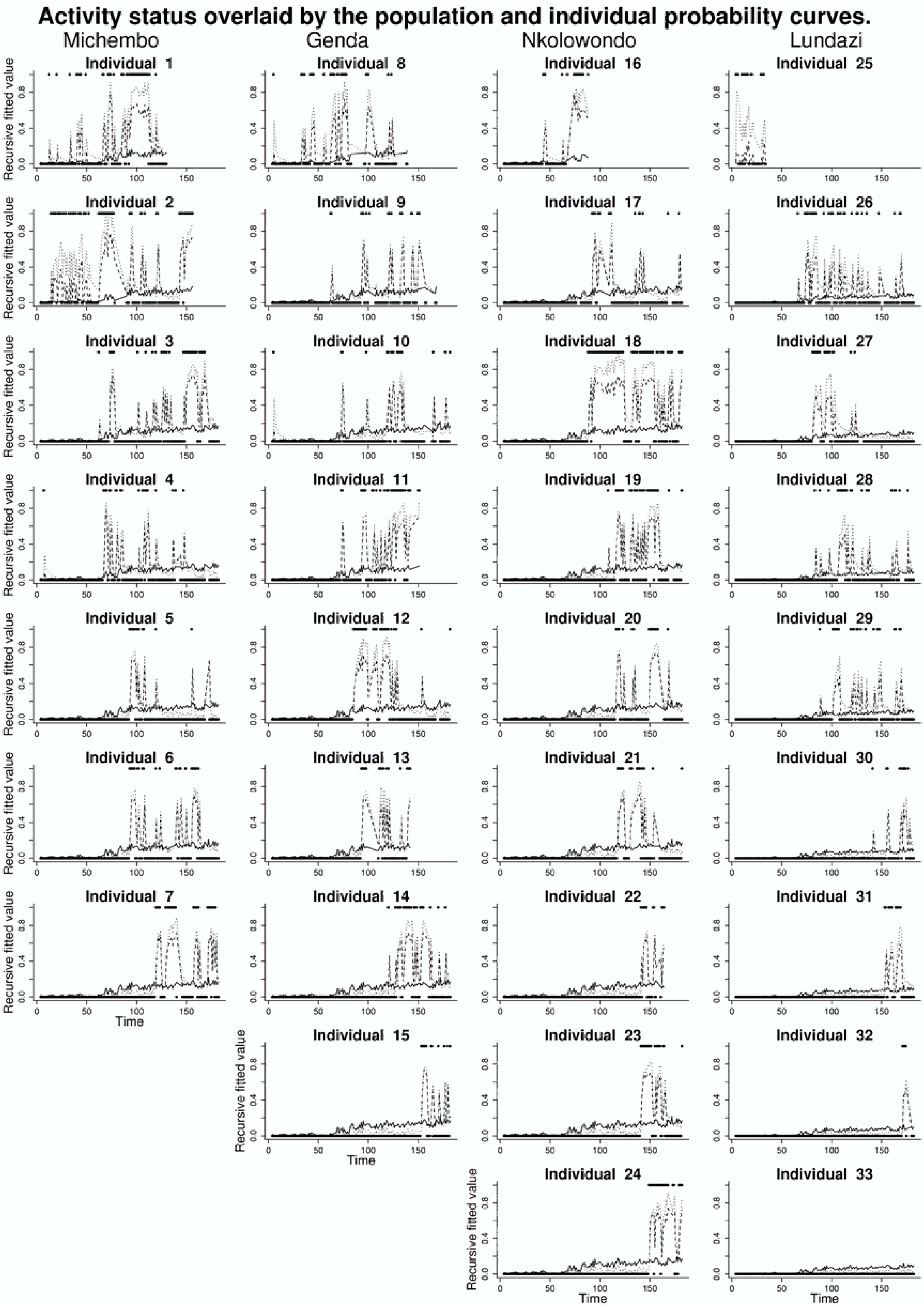

Figure 4: Plots of the activity status (filled circles where 0 and 1 respectively indicates inactive and active) for the thrity-three ticks included in the study. The height of the solid line indicates the underlying population probability curve of being active for the generalized auto-regression model. The dashed and dotted lines represent individual (or recursive) curves obtained for this same model. The height of the dashed curve represents a tick's probability of being active taking only the state dependence into account, whereas the dotted curve is obtained by taking also the serial dependence into account.

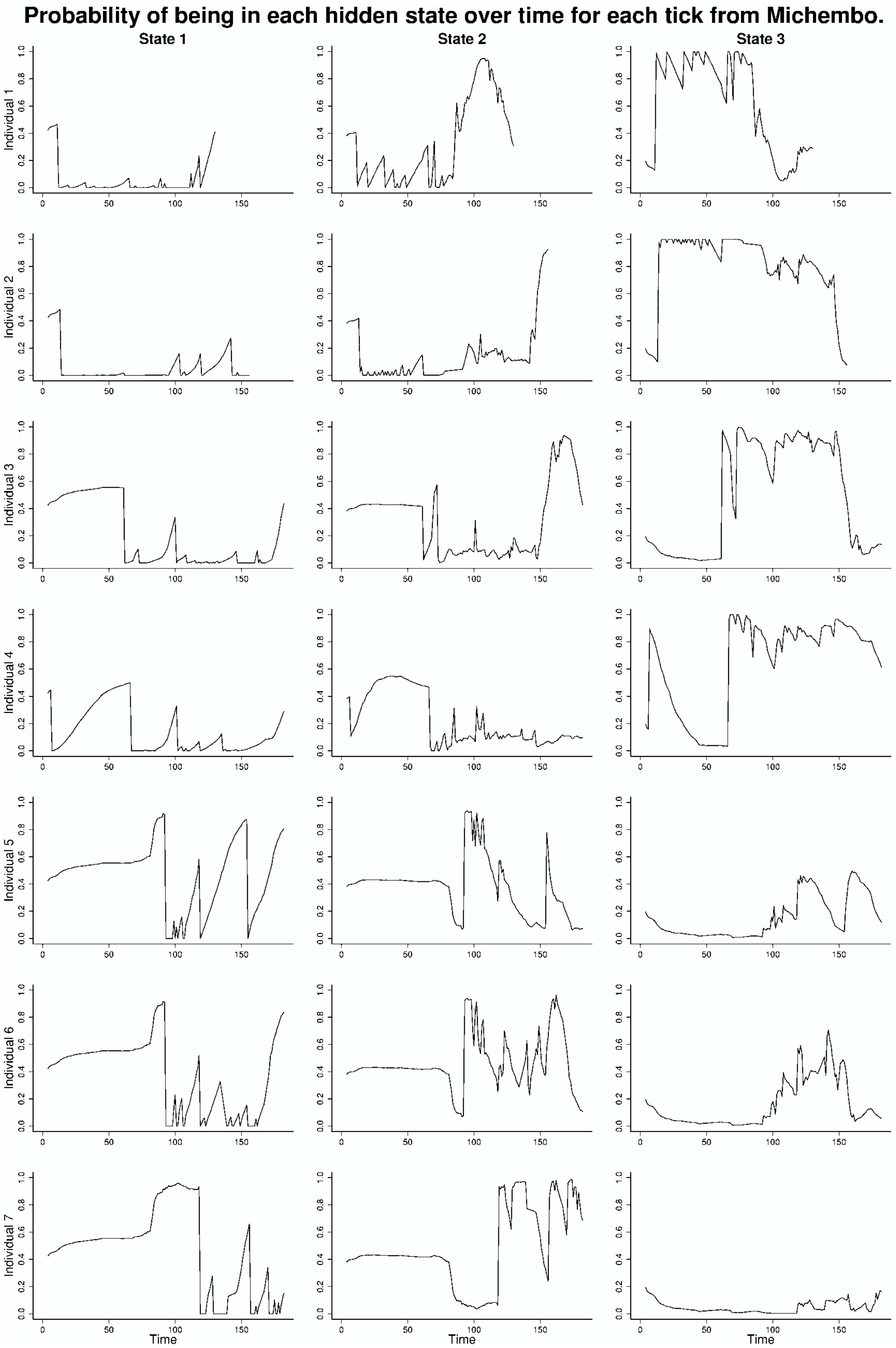

Figure 5: Plots of the probability of being in a particular hidden state over time for each tick from the Michembo location for the three hidden states.

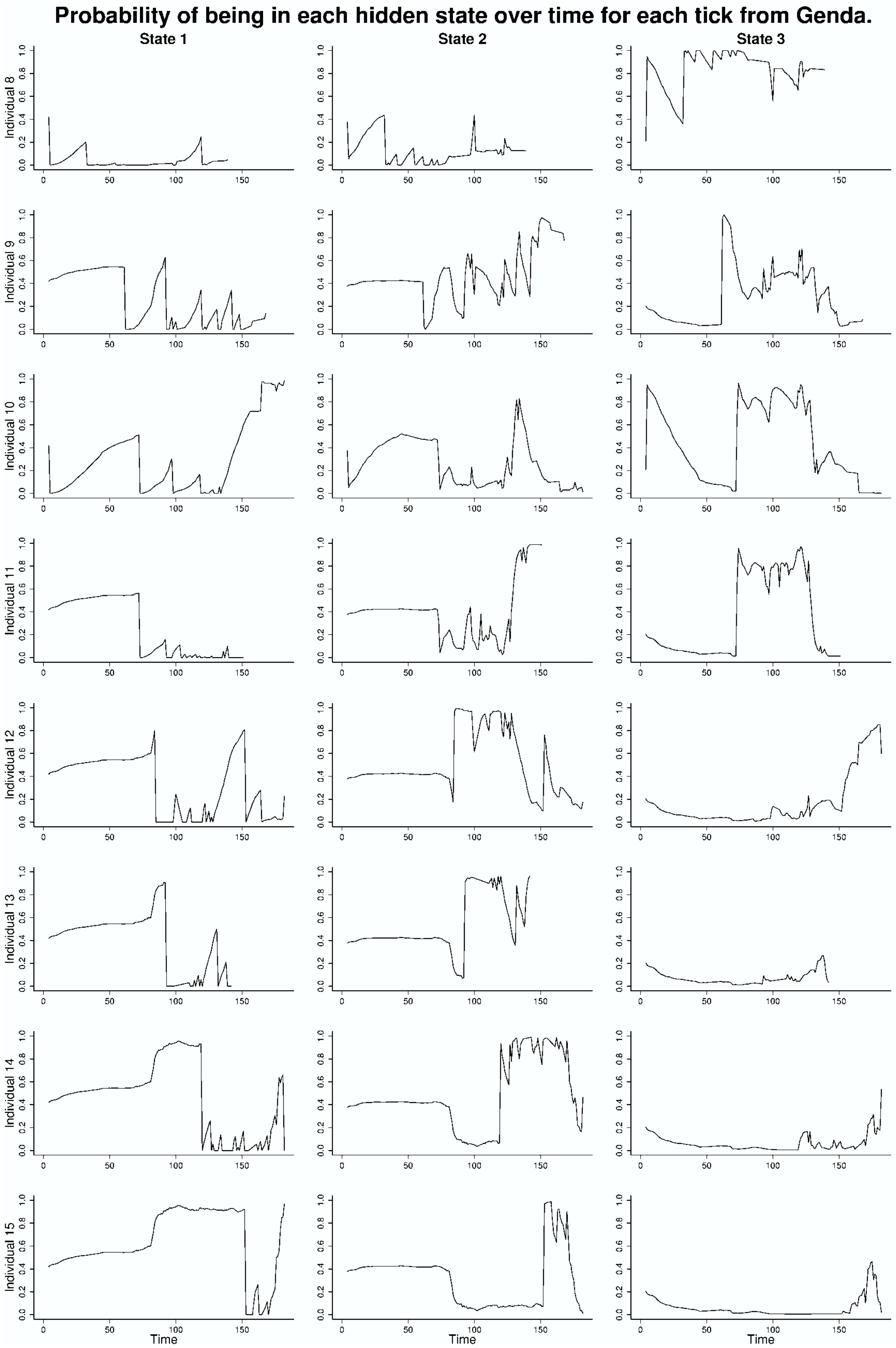

Figure 6: Plots of the probability of being in a particular hidden state over time for each tick from the Genda location for the three hidden states.

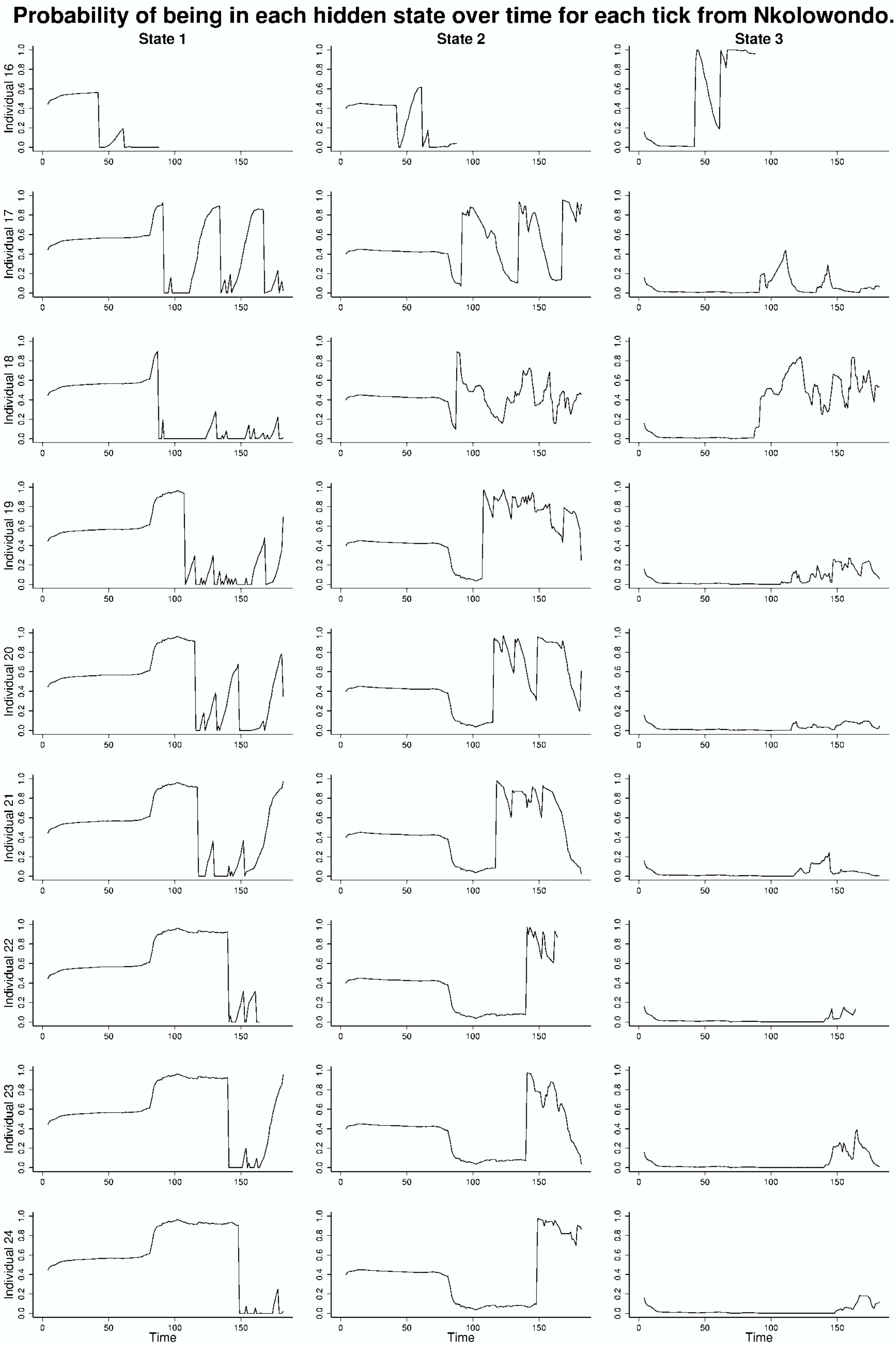

Figure 7: Plots of the probability of being in a particular hidden state over time for each tick from the Nkolowondo location for the three hidden states.

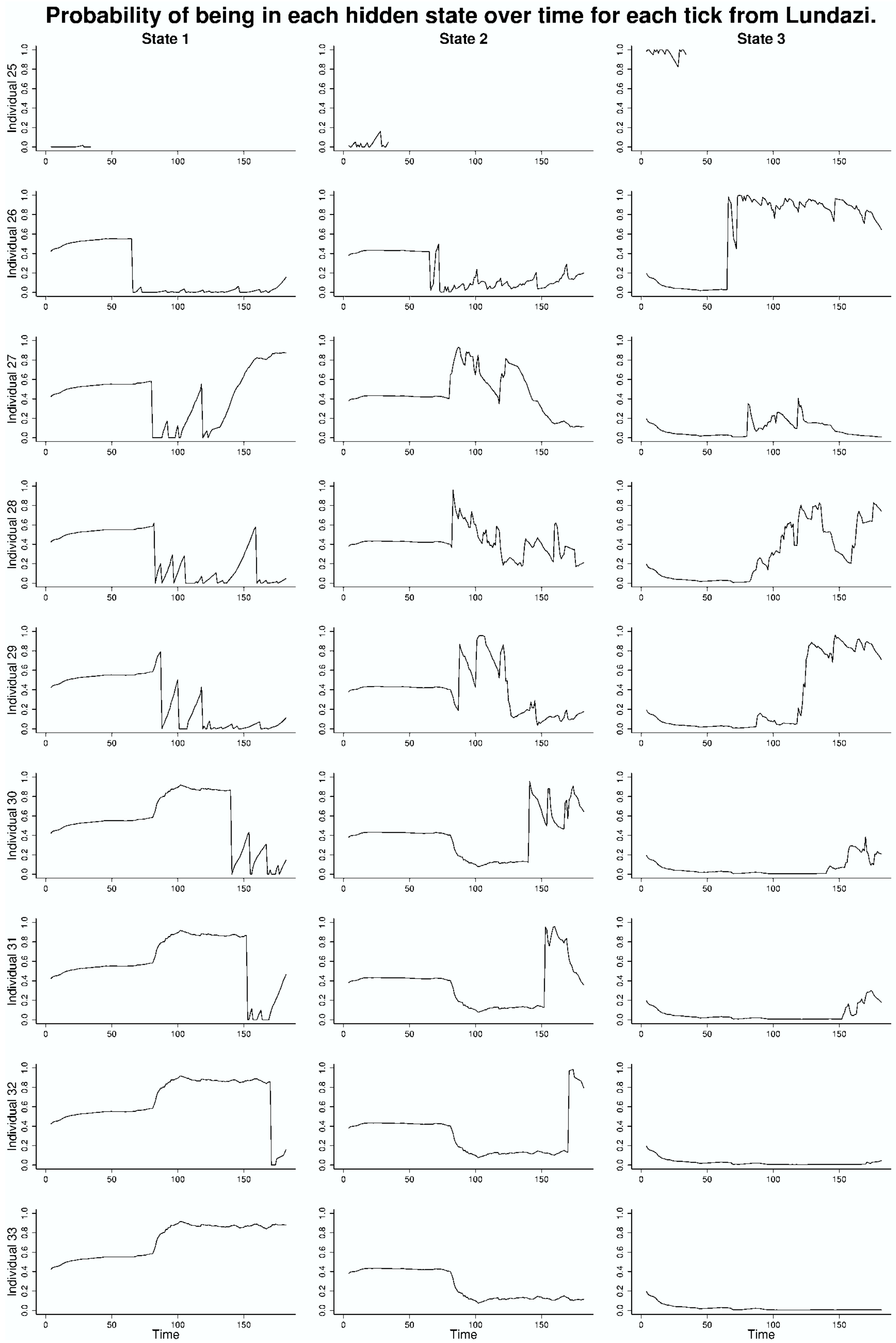

Figure 8: Plots of the probability of being in a particular hidden state over time for each tick from the Lundazi location for the three hidden states.

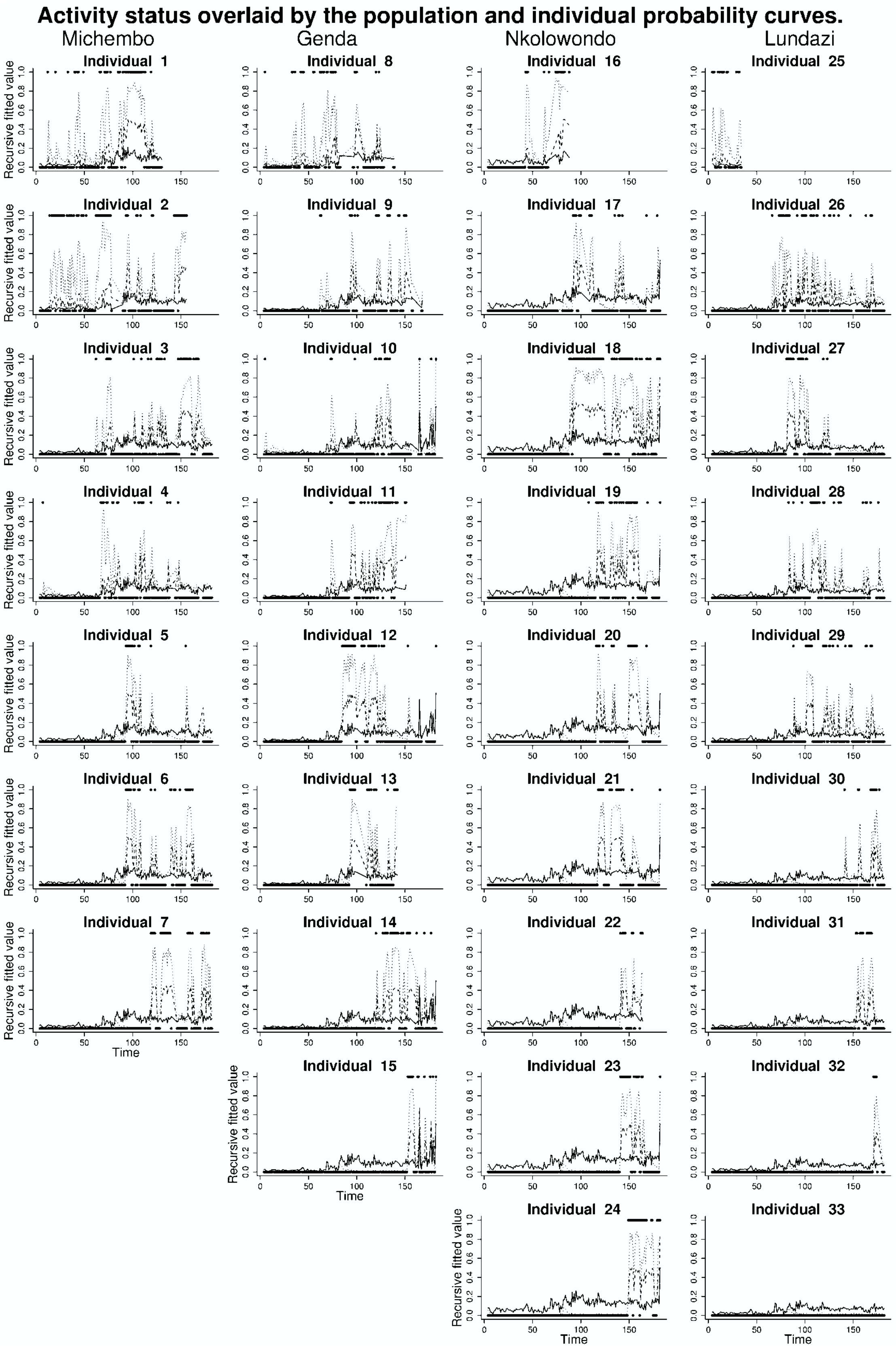

Figure 9: Plots of the activity status (filled circles where 0 and 1 respectively indicates inactive and active) for the 33 ticks. The height of the solid line indicates the underlying population probability curve of being active for the hidden Markov chain model. The dashed and dotted lines represent individual (or recursive) curves obtained for this same model. The height of the dashed curve represents a tick's probability of being active taking only the state dependence into account, whereas the dotted curve is obtained by taking also a tick's optimal path through the hidden states into account.

Table 1: Three-way contingency table of the observed tick activity status.

	Lag(1) response Inactive		Lag(1) response Active		
	Lag(2) response				
Response	Inactive	Active	Inactive	Active	Total
Inactive	4055	192	128	128	4503
Active	212	64	137	333	746
Total	4267	256	265	461	5249