

# Simulating tick distributions over sub-Saharan Africa: the use of observed and simulated climate surfaces

J. M. Olwoch<sup>1</sup>, C. J. de W. Rautenbach<sup>2</sup>, B. F. N. Erasmus<sup>1</sup>, F. A. Engelbrecht<sup>2</sup> and A. S. van Jaarsveld<sup>1,3\*</sup> <sup>1</sup>Conservation Planning Unit, Department of Zoology and Entomology, University of Pretoria, Pretoria, <sup>2</sup>LRAM, Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Pretoria, and <sup>3</sup>Department of Zoology, Stellenbosch University, Stellenbosch, South Africa

## Abstract

**Aim** A broad suite of climate data sets is becoming available for use in predictive species modelling. We compare the efficacy of using interpolated climate surfaces [Center for Resource and Environmental Studies (CRES) and Climate Research Unit (CRU)] or high-resolution model-derived climate data [Division of Atmospheric Research limited-area model (DARLAM)] for predictive species modelling, using tick distributions from sub-Saharan Africa.

**Location** The analysis is restricted to sub-Saharan Africa. The study area was subdivided into 3000 grids cells with a resolution of 60 × 60 km.

**Methods** Species distributions were predicted using an established multivariate climate envelope modelling approach and three very different climate data sets. The recorded variance in the climate data sets was quantified by employing omnidirectional variograms. To further compare the interpolated tick distributions that flowed from using three climate data sets, we calculated true positive (TP) predictions, false negative (FN) predictions as well as the proportional overlaps between observed and modelled tick distributions. In addition, the effect of tick data set size on the performance of the climate data sets was evaluated by performing random draws of known tick distribution records without replacement.

**Results** The predicted distributions were consistently wider ranging than the known records when using any of the three climate data sets. However, the proportional overlap between predicted and known distributions varied as follows: for *Rhipicephalus appendiculatus* Neumann (Acari: Ixodidae), these were 60%, 60% and 70%; for *Rhipicephalus longus* Neumann (Acari: Ixodidae) 60%, 57% and 75%; for *Rhipicephalus zambeziensis* Walker, Norval & Corwin (Acari: Ixodidae) 57%, 51% and 62%, and for *Rhipicephalus capensis* Koch (Acari: Ixodidae) 70%, 60% and 60% using the CRES, CRU and DARLAM climate data sets, respectively. All data sets were sensitive to data size but DARLAM performed better when using smaller species data sets. At a 20% data subsample level, DARLAM was able to capture more than 50% of the known records and captured more than 60% of known records at higher subsample levels.

**Main conclusions** The use of data derived from high-resolution nested climate models (e.g. DARLAM) provided equal or even better species distribution modelling performance. As the model is dynamic and process based, the output data are available at the modelled resolution, and are not hamstrung by the sampling intensity of observed climate data sets (c. one sample per 30,000 km<sup>2</sup> for Africa). In addition, when exploring the biodiversity consequences of climate change, these modelled outputs form a more useful basis for comparison with modelled future climate scenarios.

\*Correspondence: A. S. van Jaarsveld, Faculty of Science, Private Bag X1, Stellenbosch 7602, South Africa. E-mail: asvanj@sun.ac.za

## Keywords

Climate data sets, ticks, Africa, distribution modelling.

## INTRODUCTION

Predictive species modelling is rapidly becoming a routine analytical procedure (for review see Guisan & Zimmerman, 2000), and with rapid advances in climate data interpolation and modelling techniques, it is now possible to predict the distributions of an array of insects, pests and pathogens using their major climate drivers (Messenger, 1959; Sutherst & Maywald, 1985; Meats, 1989; Perry *et al.*, 1990; Sutherst *et al.*, 1995; Randolph & Rogers, 1997; Coakley *et al.*, 1999; Cumming, 2000; Erasmus *et al.*, 2000, 2002; Rogers & Randolph, 2000; Randolph, 2001, 2002; Harvell *et al.*, 2002). Many of these approaches are static and probabilistic in nature but remain the primary approach available for studying the possible consequences of changing environments on species distributions.

The principal sources of climate data used for predictive distribution modelling are climate surfaces, generated by interpolating observed climate data that were sampled at varying intensities from across a region. Consequently, any differences between these climate surfaces can usually be attributed to the spatial and temporal evenness of the observed data used for interpolation. Most modern interpolation techniques are pattern-based and statistically incorporate horizontal as well as vertical (altitudinal) adjustments (see Hutchinson, 1989, 1991; Hutchinson & Gessler, 1994).

General circulation models (GCMs) have now become the primary method of simulating climates. These are coupled ocean–atmosphere models that provide three-dimensional simulations of the atmosphere. To date, GCMs have produced climate data at a horizontal resolution that is too coarse for use in predictive species modelling ( $> 100 \times 100$  km grid point resolution), especially for species that are habitat specialists or that are influenced by fine scale environmental gradients. Computational requirements usually prevent GCMs from being run at mesoscale grid resolutions (10–100 km). A potential alternative source of high-resolution climate data for use in predictive species modelling is nested limited-area models. Nested climate modelling involves the nesting of a high-resolution limited-area model within a GCM over the area of interest (for a review see McGregor, 1997). The GCM supplies the limited-area model with initial and boundary conditions. With a grid resolution of 10–100 km, the limited-area model is able to simulate some of the mesoscale properties of the circulation. This technique provides a viable alternative to the use of observed/interpolated climate surfaces for fine-scale climatic data. The resolution attained by this dynamic modelling process is essentially limited by the computing power available to the modellers.

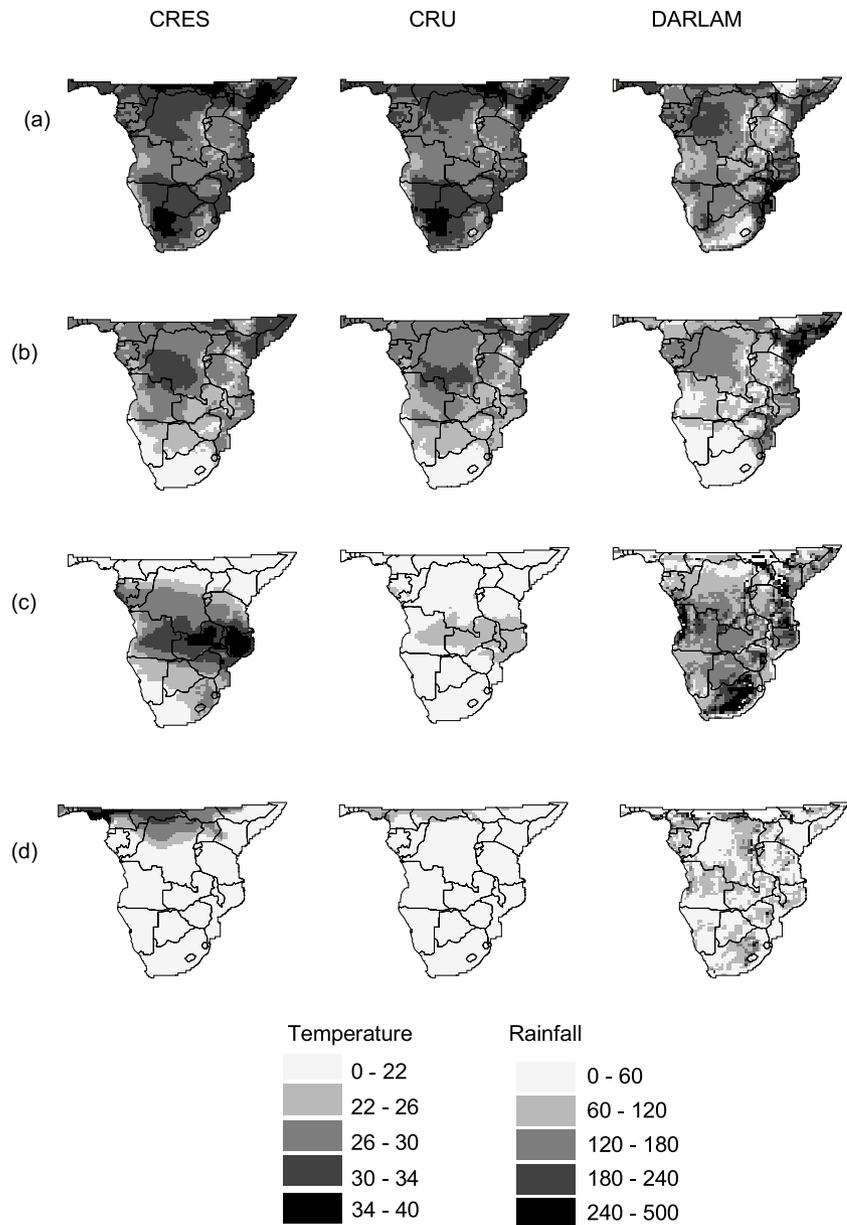
The present study used one species distribution modelling procedure (Erasmus *et al.*, 2000) and three climate data sets

(two interpolated and one modelled) to evaluate the relative performance of the climate data sets in simulating the present distributions of ticks in sub-Saharan Africa. We employed distribution data from four tick species from the genus *Rhipicephalus*. These tick species are taxonomically well known, are poorly sampled and their environmental dependencies not adequately documented (Tukahirwa, 1976; Rechav, 1981; Short & Norval, 1981; Norval *et al.*, 1982; Horak *et al.*, 1983, 1995; Perry *et al.*, 1990; Rogers & Randolph, 1993; Cumming, 1998, 1999a; Walker *et al.*, 2000). As principal vectors of numerous cattle diseases these tick species are also of significant economic importance to the continent.

## MATERIALS AND METHODS

### Center for Resource and Environmental Studies climate data

The first climate data set used forms part of a topographical and climate data base for Africa developed by the Center for Resource and Environmental Studies (CRES) (Hutchinson *et al.*, 1996) at the Australian National University. These data contain grid box values of elevation (Hutchinson, 1989), monthly mean climate of rainfall as well as daily values for maximum and minimum temperatures at a spatial resolution of  $0.05 \times 0.05^\circ$ . Climate grid box values were calculated by fitting topographically dependent climate surfaces to point climate data using procedures described by Hutchinson (1991) and Hutchinson & Gessler (1994), thereafter the surfaces were interrogated using elevations from the topographical data (McMahon *et al.*, 1995). Climate data at a sufficient spatial density to support reliable spatial interpolation were compiled. In addition to data already obtained by CRES, monthly climate data were acquired from many other research agencies. Data were collected over all available years to maximize spatial coverage, subject to the condition that rainfall averages were calculated from records of at least 5 years. The data set comprises data collected between 1920 and 1980 from *c.* 1500 temperature and 6000 rainfall stations. The error of grid values depends mainly on the accuracy of the underlying climate surfaces. The standard errors of the temperatures are *c.*  $0.5^\circ\text{C}$ , while standard errors for rainfall range between 5% and 15% (depending on data density and the spatial variability of observed mean monthly rainfall values). The mean monthly rainfall values as well as maximum temperatures for January and July are depicted in Fig. 1. The climate surfaces are relatively smooth because of their dependency on low-resolution point observations. These data were re-sampled to the finest common resolution between climate data sets, i.e.  $60 \times 60$  km Division of



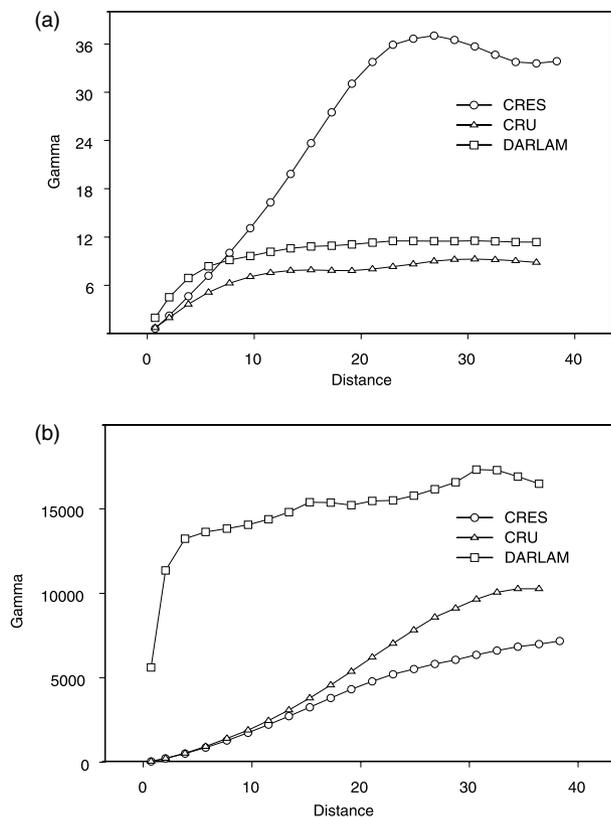
**Figure 1** Maximum temperature ( $^{\circ}\text{C}$ ) and rainfall ( $\text{mm day}^{-1}$ ) for January and July as represented by three climate data sets (CRES, CRU and DARLAM). (a) mean maximum temperature for January, (b) mean maximum temperature for July, (c) mean rainfall for January and (d) mean rainfall for July.

Atmospheric Research limited-area model (DARLAM) grid cell resolution. This resulted in data for 3000 grid cells which was within the limitations imposed by the available hardware and software. The mean values of CRES data cells contained within each  $60 \times 60$  km grid cells was used as the new re-sampled CRES data value for that particular grid cell. Although some variation was lost in this procedure, between-cell variation before re-sampling was small due to interpolation between remote data points, and therefore the predictive ability of this data was not affected. After re-sampling, 92% of the grid cells had standard deviations smaller than 5% of the mean value for that particular grid cell. The variograms of re-sampled data in Fig. 2 confirms this fact, with neighbouring cells remaining highly

autocorrelated at small distances (*c.* 60 km). We regard this small change in variation as reasonable for our purposes.

#### Climate Research Unit climate data

The 1961–90 climate data set from the Climate Research Unit (CRU) was used as a second data set (New *et al.*, 1999). The data base consists of mean monthly climate fields with a  $0.5 \times 0.5^{\circ}$  grid resolution across the global terrestrial areas, excluding Antarctica. The mean climate surfaces were constructed from a 30-year (1961–90) station observation field. The station data were interpolated as a function of latitude, longitude and elevation using thin-plate splines. The accuracy of the interpolations are assessed using cross-validation and



**Figure 2** (a) Omnidirectional variogram of January maximum temperature across sub-Saharan Africa for the CRES, CRU and DARLAM data sets; (b) Omnidirectional variogram of January rainfall across sub-Saharan Africa for the CRES, CRU and DARLAM data sets.

by comparison with other climate surfaces (New *et al.*, 1999). Considerable spatial and temporal variance of climate surface variables over southern Africa can be observed in the CRU data (Engelbrecht *et al.*, 2002). Despite this, the climate surfaces remain smooth because of their dependence on low-resolution point data. The mean monthly rainfall values as well as maximum and minimum temperatures are depicted in Fig. 1. These data were also re-sampled to  $60 \times 60$  km grid cells in the same manner as the CRES data.

#### DARLAM climate data

The Commonwealth Scientific and Industrial Research Organization (CSIRO) in Melbourne, Australia, developed the high-resolution limited-area model DARLAM for use in both short-term mesoscale atmospheric studies and long-term climate simulation experiments (Walsh & McGregor, 1995). The model is a two-time-level, semi-implicit, hydrostatic primitive equations model and follows the methodologies outlined by Mesinger & Arakawa (1976), McGregor (1993) and Phillips (1957). DARLAM employs a wide range of physical parameterization schemes to represent atmospheric processes such as cumulus convection that exist at

subgrid scale. A one-way nesting technique (Davies, 1976) is used with lateral boundary conditions supplied by the CSIRO-9 Mk 2 GCM with a R21 spectral resolution.

In the present study, nine separate 30-day simulations were performed for both January (representing mid-summer conditions) and July (representing mid-winter conditions). The simulations were performed at a horizontal grid resolution of  $60 \times 60$  km using a domain of  $100 \times 100$  grid points that covers sub-Saharan Africa. The monthly average of the nine simulations constitutes the model climatology for the month. The climatologies of rainfall and minimum and maximum temperature obtained in this manner are shown in Fig. 1. Engelbrecht *et al.* (2002) illustrated that DARLAM is capable of simulating the regional characteristics of atmospheric variables like near-surface temperature, low-level wind patterns and rainfall over sub-Saharan Africa with considerable detail. The model does, however, tend to overestimate rainfall totals over regions with a steep topography.

The DARLAM simulations were performed at the Laboratory for Research in Atmospheric Modelling (LRAM) at the University of Pretoria on a Pentium III workstation with two 550 MHz processors. It took *c.* 11-min (CPU time) to simulate a model day at the specified model resolutions.

#### Predictive species modelling

We used a single predictive species modelling approach, originally developed by Jeffree & Jeffree (1994, 1996), for predicting species distribution patterns in the present study as a standardized base for evaluating the relative performance of the different climate data sets. This model was recently modified to accept multivariate inputs to yield probability of occurrence maps for species (Erasmus *et al.*, 2000). The original model was adapted to incorporate a variety of climatic predictor variables. Not only does this improve the original model, but it also allows the production of a probability surface of suitability for each species (Erasmus *et al.*, 2000), rather than merely a presence-absence distribution model. The multivariate predictive distribution modelling procedure developed by Erasmus *et al.* (2000) was employed throughout the current study. In short, this multivariate modelling approach uses the values of selected climate variables for each cell where a species had been recorded. These are plotted on an  $n$ -dimensional scatter plot (for the  $n$ -climate variables), and mean climate values subtracted to centre values around the origin of the multidimensional scatter plot. An  $n \times n$  covariance matrix is calculated and this matrix is used as an input to calculate eigenvalues and eigenvectors for the covariance matrix. These eigenvectors form the orthogonal principle axes of an  $n$ -dimensional hyperspace with the origin representing the theoretical core of the species fundamental niche, as defined by the predictor climate variables (also see Robertson *et al.*, 2001).

The climate variable values of all grid cells are then transformed into this eigenvector space. These transformed climate variables are subsequently divided by the eigenvalues and in the resulting  $n$ -dimensional hyperspace, the distance

of any particular unsampled grid cell from the origin represents a measure of the suitability of that locality for the specific species (Robertson *et al.*, 2001). According to Austin & Meyers (1996), the fundamental niche of an organism follows a broad Gaussian distribution. Such a Gaussian distribution is best approximated by a normal distribution but given that the distance from the origin of the hyperspace is calculated by the sum of the squared eigenvector axis scores, and that a squared normal distribution is equal to a chi-square distribution, the probability of any grid cell to be suitable for the selected species, can be read off a chi-square probability table at the appropriate degrees of freedom ( $m$ , the number of climatic variables) (Robertson *et al.*, 2001). The input data comprises 3000 grid cells covering sub-Saharan Africa populated with climate variables. Grid cells, in which particular tick species were recorded, are referred to as *known records* following Erasmus *et al.* (2002). Thus, on a scatter plot of climate variables (multivariate climate space), the values of climate variables from localities where species have been recorded are used to construct a confidence region where there is a high probability that the records reflect the core range of the species. Points falling within this confidence region are then mapped back to geographical space to represent an *interpolated distribution* (ID), represented as a probability of climate suitability (see Erasmus *et al.*, 2000 for a detailed model description). Differences in model performance when employing a suit of different climate data sets were evaluated by comparing the interpolated species distributions.

### Tick data

Point localities of recorded tick observations were obtained from Cumming (1999b). These data were compiled from various tick collections (see Cumming (1999b) for a detailed list of sources). Combining data sets from different sources invariably compounds identification and distribution errors and for this reason data congruence with Walker *et al.* (2000) were used to assess data quality. This later reference provides well-illustrated distributions of *Rhipicephalus* species. Synonyms provided in this reference also solve the common problem of referring to one species using two different names. The tick species selected for this study belong to the phylum Arthropoda, Class Arachnida and Family Ixodidae. Four species (*R. appendiculatus*, *R. capensis*, *R. longus* and *R. zambeziensis*) were selected from the seventy-four *Rhipicephalus* species found in Africa because their life-history parameters and host preferences are relatively well known (Walker *et al.*, 2000). Tick point localities were assigned to particular 60 × 60 km grid cells by means of a spatial intersect. A conservative estimate of the accuracy of these point localities is 0.2° (G.S. Cumming, pers. comm.) and therefore this approximation is considered reasonable.

### Climate variables

In the present study, the climate variables employed were mean maximum temperature of January and July, mean

minimum of January and July, and mean rainfall of January and July.

### Comparison of interpolated species distributions across climate data sets

Interpolated tick distributions emanating from the three climate data sets were compared by first calculating the true positive (TP) predictions, that is the number of grid cells where the probability of climate suitability was ≥ 50% and where a tick had previously been recorded. The number of false negative (FN) predictions were also recorded, that is when the probability of climate suitability was < 50% but where a tick had previously been recorded. The comparison among climate data sets was conducted using the proportional overlap method (Prendergast *et al.*, 1993; Lombard, 1995; Reyers *et al.*, 2000). Proportional overlap was calculated as TP/TP + FN where TP is the number of true positive predictions and TP + FN represents the number of positive plus the number of negative predictions, or, the maximum number of overlapping grid cells between observed and modelled tick distributions. This method was the most appropriate vehicle of comparison as absence data for tick distributions were not available. Although receiver operator characteristic (ROC) analyses has been shown to be robust to the assumption of absence data in unsampled cells (Cumming, 2000a), the low prevalence of presence data (varying between 0.027 and 0.179) in this study means that *c.* 2500 absence records would be added if this assumption was made. We are not confident that such a large change in the number of records can still yield robust model outputs. Therefore, in this case, the absence of absence data precludes the use of ROC analyses (Erasmus *et al.*, 2002).

### Modelling random subsamples of known records

To evaluate the role played by tick data set size on the performance of the climate data sets, we performed single random draws of varying percentages of the known tick distribution records without replacement (20%, 40%, 60%, 80%, 100% of known records). By varying the input data set size and summing the proportional overlaps between predicted occurrences and known records across tick species, we were able to evaluate climate data performance for different levels of species data availability. A Kruskal–Wallis rank test between the proportional overlap values of the three climate data sets at every level of data set reduction was performed (20%, 40%, 60%, 80% and 100%).

## RESULTS

### Comparing climate data sets

The observed variance in the CRES temperature data only becomes distance invariant after *c.* 22°, which is almost three times greater than that found for the DARLAM and CRU data sets (Fig. 2a). When the variogram is interpreted at the distance of reliability (Kaluzny *et al.*, 1998), which is half the

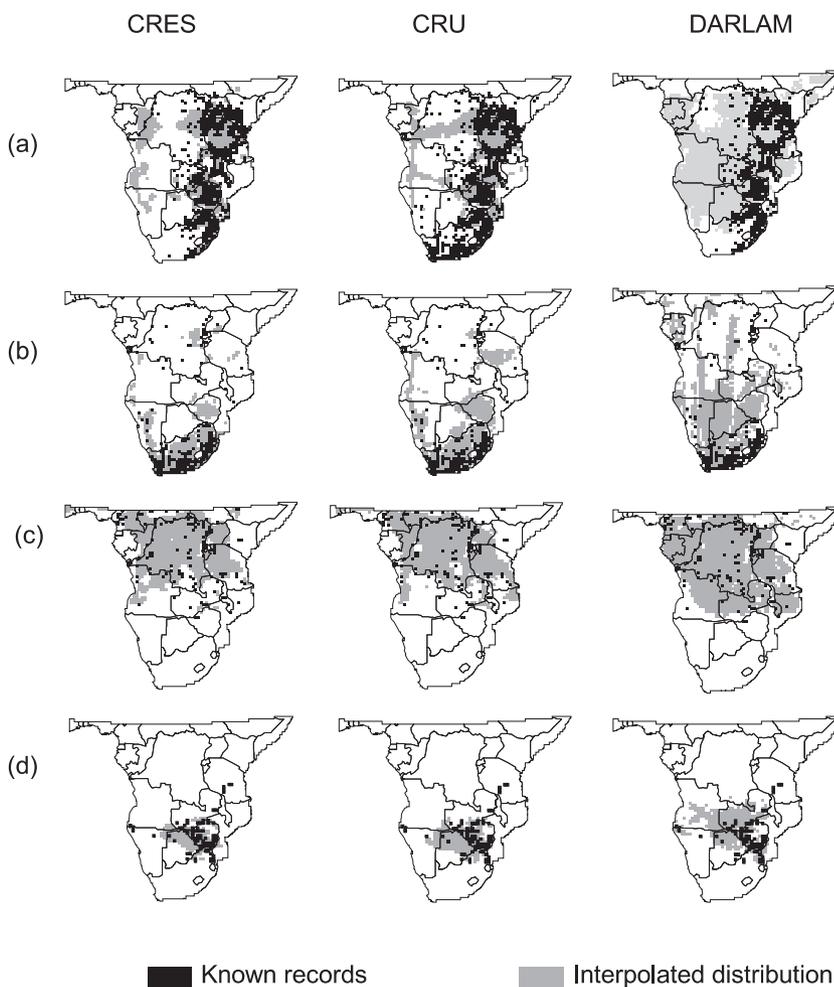
maximum distance over the field of data for an experimental variogram, CRES data are still influenced by more remote data points, whereas the autocorrelation of CRU and DARLAM data have become distance invariant. This means that any particular value in the CRES data is more strongly influenced by data up to  $15^\circ$  away than either the CRU or DARLAM data sets where the sphere of influence is much more localized (*c.*  $7^\circ$ ). The relatively small number of observations used as inputs to create the CRES climate surface, forced interpolations over larger distances than for CRU, where more observational data were available for creating the climate surface. The slope of the curve for DARLAM temperature data appears to be steeper than the other curves over short distances (Fig. 2a). This illustrates how DARLAM, a process model, is very sensitive to topography and topographical features which are usually expressed over short distances, i.e. *c.*  $2^\circ$ . DARLAM's precipitation variogram (Fig. 2b) also displays a non-monotonic behaviour.

### Interpolated tick distributions

The multivariate climate envelope model used in the present study provides probability of climate suitability values

ranging from 0 to 1 across all grid cells. It was therefore necessary to select an appropriate probability level to use for comparison between the different species. In this study, a 50% probability level of occurrence was used. This means that a tick was assumed present where the model predicted a probability of occurrence of  $\geq 50\%$  and the reverse. This follows a frequently employed convention in the field of predictive species modelling (Beard *et al.*, 1999; Brito *et al.*, 1999; Erasmus *et al.*, 2002; Van Staden *et al.*, in press; also see Walker, 1990 for even lower thresholds). When the proportion of the correctly predicted cases (TPs) were evaluated, it was found that in *R. appendiculatus*, a well-studied tick species, a TP coverage of 70% (using DARLAM data) was obtained at a probability of occurrence level of  $\geq 50\%$ . This percentage of known records captured in the ID is considered good, bearing in mind that the observed known records are not only determined by climate but also by the presence of suitable host species and other species specific micro-ecological or habitat selection factors which are not incorporated in this model.

The ID for *R. appendiculatus* (Fig. 3), when using any of the three data sets, was both visually and statistically broader than the extent of the known records, but the TP



**Figure 3** Recorded known records (KR) and interpolated distributions (ID) of ticks obtained using species predictive modelling based on each of the three climate data sets (CRES, CRU and DARLAM): (a) *R. appendiculatus*, (b) *R. capensis*, (c) *R. lungus* and (d) *R. zambeziensis*.

**Table 1** The number of true and false predictions generated for four different tick species by a predictive species distribution model (Erasmus *et al.*, 2002) using different climate data sets (CRES, CRU, DARLAM) and complete species data sets. KR, known records; TP, true positive predictions; FN, false negative predictions

	CRES	CRU	DARLAM
(a) <i>Rhipicephalus appendiculatus</i>			
KR	538	538	538
TP	315	320	372
FN	223	217	166
Proportional overlap (TP/TP + FN) (%)	60	60	70
(b) <i>R. capensis</i>			
KR	189	189	189
TP	128	119	108
FN	60	70	81
Proportional overlap (%)	60	60	57
(c) <i>R. longus</i>			
KR	129	129	129
TP	78	74	97
FN	51	55	32
Proportional overlap (%)	60	56	75
(d) <i>R. zambeziensis</i>			
KR	82	82	82
TP	47	42	51
FN	35	40	31
Proportional overlap (%)	60	50	65

predictions corresponded satisfactory with the known records (60%, 60% and 70% overlap using the CRES, CRU and DARLAM data sets, respectively) (Table 1a, Fig. 4). There was a noteworthy positive congruence between the IDs and the known records in most areas of east, central and southern Africa: Uganda, Tanzania, Zambia, Zimbabwe and South Africa.

Unlike *R. appendiculatus* that is well studied and therefore presents a more comprehensive recorded distribution, the other tick species are poorly sampled and the actual distributions are still contested in the literature. While Walker *et al.* (2000) maintain that *R. capensis* is strictly South African, Cumming (1999b) reported it mainly in South Africa with some scattered records in Rwanda, DRC, Angola and Zambia (Fig. 3). The proportional overlap correlation between ID and known records for *R. capensis* were 70%, 60% and 60% for the CRES, CRU and DARLAM climate data sets, respectively (Table 1b, Fig. 4). The highest overlap occurs in South Africa and Namibia with more than 90% of all TP records found in RSA for the three climate data sets. This resultant ID compares favourably with the Walker *et al.* (2000) distribution pattern and improves the suggested Cumming (1999b) distribution by eliminating a few scattered records in the north of South Africa and Namibia. The IDs of the other two tick species (Fig. 3) were also broader than the known records, with a 60%, 57% and 75% proportional overlap using CRES, CRU and DARLAM data for *R. longus* (Table 1c, Fig. 4) and 57%, 51% and 62% for *R. zambeziensis* (Table 1d, Fig. 4), respectively. One of the interesting and consistent results that emerged

from these analyses is the ability of DARLAM to produce a higher proportion of true predictions across all species examined (Figs 4 & 5).

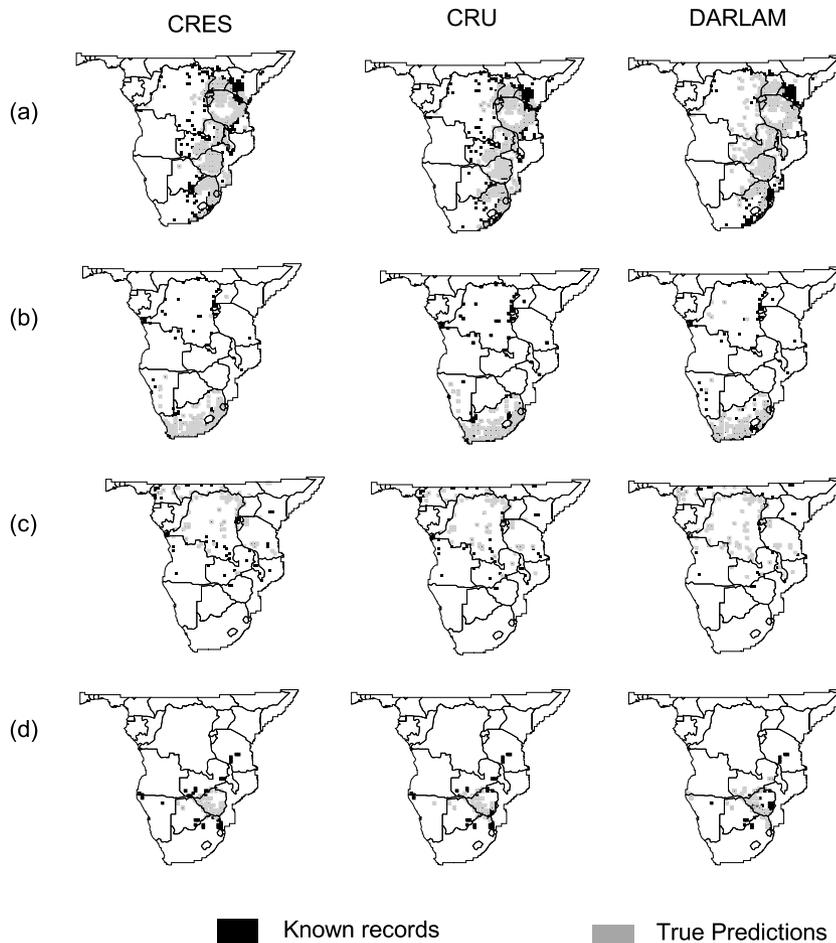
### Modelling random subsamples of known records

All three examined climate data sets seemed to be sensitive to the tick data set size in terms of their ability to correctly predict tick distributions. However, DARLAM consistently performed better with higher degree of overlap between the IDs and known records (Figs 5 & 6). Even with a 20% subsample of known records across the tick species, DARLAM was able to capture more than 50% of the known records and was consistently above 60% at higher levels of subsampling. However, none of these differences in proportional overlap reached statistical significance ( $P < 0.05$ ) suggesting that DARLAM performs at least as well as any of the observed data sets.

## DISCUSSION

### Comparing spatial data sets

When interpreting the output of climate envelope-based predictive species distribution models, it is necessary to be aware of differences in climate data that can be used as inputs. The data sets used in the present study represent three different types of climate data, namely: (1) observed data with variable coverage in space and time together with elevation influenced smoothing (CRES), (2) better coverage and 30-year averaged observed data with elevation influenced smoothing (CRU), and (3) purely process derived modelled climate data (DARLAM). Although some differences among the climate data sets are obvious, even to visual inspection (see Fig. 1), a more quantitative measure of the difference in underlying climate parameters was obtained by comparing omnidirectional variograms (Rossi *et al.*, 1992; Kaluzny *et al.*, 1998; Nielsen & Wendroth, 2001) for these data sets. In a variogram the variance of the property being measured, i.e. rainfall or temperature, is statistically expressed as a function of the distance between observations within a given spatial domain, i.e. sub-Saharan Africa (Nielsen & Wendroth, 2001). The difference between pairs of observations separated by a specific distance is plotted and the slope, position and shape of this graph gives an indication of the scale at which the underlying processes might operate. These variograms can therefore be interpreted as a measure of the degree of difference between the various data sets. For example, the emergence of non-monotonic behaviour in the DARLAM data set (Fig. 2b) is indicative of spatial variation at scales other than the sampling units (Nielsen & Wendroth, 2001). Given the fact that precipitation is the result of complex atmospheric interactions at different scales, such a non-monotonic variogram can be expected from a process-based model such as DARLAM. In the analyses conducted here, the CRES data set varied over larger spatial scales than either the CRU or DARLAM data sets. Thus, the three data sets used in this study do not only appear different to a



**Figure 4** Known records (KR) and true positive (TP) predictions for all tick species generated for each climate data set: (a) *R. appendiculatus*, (b) *R. capensis*, (c) *R. lungus* and (d) *R. zambeziensis*.

casual visual inspection (Fig. 1), but are also different with respect to the spatial dependencies of data base characteristics which influence adjacent climate values (Fig. 2).

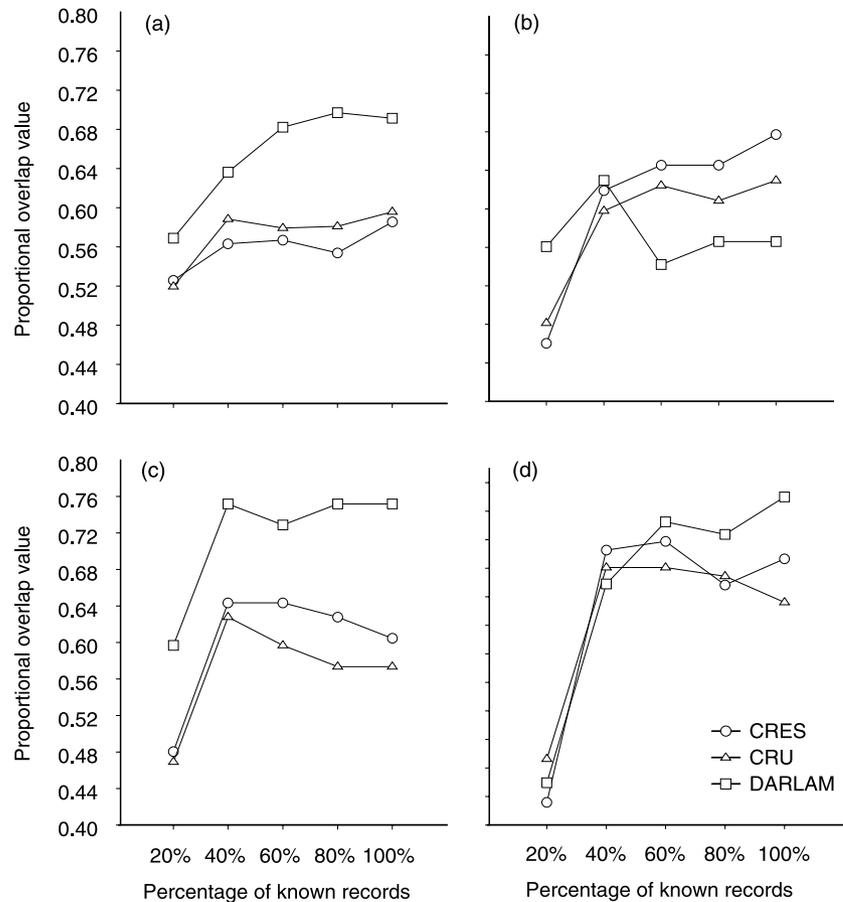
### Model evaluation

Usually the validity of a predictive species distribution model is assessed by its ability to predict both negative and positive cases (Fielding & Bell, 1997). However, the opportunistic manner in which the tick distribution data were collected precludes rigorous model evaluation. This is as true absence data are not available for these species, mainly because some areas were simply poorly sampled (Fig. 3). This same climate envelope model was, however, previously subjected to rigorous evaluation using presence–absence data resulting from a coordinated and systematic survey effort. Erasmus *et al.* (2002) used the distribution records of thirty-four bird species and tested model performance using ROC analyses (Fielding & Bell, 1997). The model performed significantly better than a random model with no discriminatory ability. The model also accurately predicted the complete known distributions for twenty-four of the thirty-four bird species, using a 20% subsample of the known records (Erasmus

*et al.*, 2002). In sum, the model performed satisfactorily and is therefore considered adequate for the purposes of the present study.

### Predicting tick distributions

At the outset it is important to acknowledge that, although we use a climate envelope approach in this study, we are not suggesting that this is an adequate basis for predicting the fine-scale distributions of tick species across the African continent. Tick distributions and abundances are more likely to be impacted by factors other than climate, such as host abundance, host resistance to ticks, acaricide use and grazing management (see Cumming, 1998, 1999a, 2002). Most of these changes interact with temperature and rainfall regimes and in the absence of experimental data. It is impossible to separate the effects of these interrelated factors. This, compounded by the fact that the broad-scale ecology of only a few tick species have been studied in detail (Cumming, 2002), means that employing climate variables as the main limiting factors for tick distributions remains the best option available. Minimum temperature, maximum temperature and rainfall have similar predictive abilities to one another;



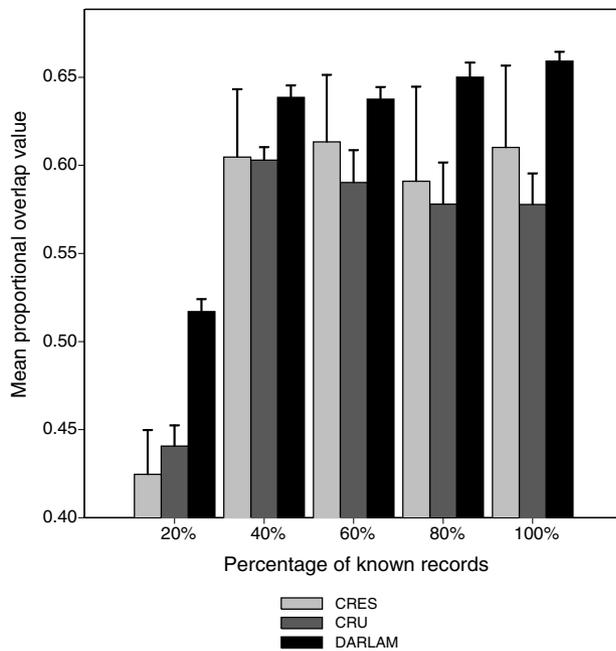
**Figure 5** Degree of proportional overlap against random subsamples of known records for all the tick species for each climate data set. These proportional overlap values were calculated at different levels of random subsampling of known records: (a) *R. appendiculatus*, (b) *R. capensis*, (c) *R. lungus* and (d) *R. zambeziensis*.

when considered together; their predictive ability increases substantially (Cumming, 2002). This is probably a consequence of their being correlated with one another. The intimate relationship between climate and tick distributions has also been reported elsewhere (Walker, 1974; Rechav, 1981; Minshull & Norval, 1982; Walker *et al.*, 2000). Taken in the context of current knowledge, we agree with Cumming (1999a) that the direct effects of climate typically determine the distributions of African ticks, but that biotic variables such as vegetation type and host distributions, may be important in creating heterogeneity in tick distributions at finer scales (Minshull & Norval, 1982) and play a subordinate role in limiting the species ranges of ticks at broad spatial scales (Cumming, 2002).

Against this background we proceeded to employ distribution data from a number of tick species that are considered sound taxa and that are widely collected. This was to evaluate the use of three different climate data sets for carrying out predictive species distribution modelling. These data sets varied in terms of the manner in which they were generated. Two data sets (CRES and CRU) represented the traditionally used data sets (see Guisan & Zimmerman, 2000; Robertson *et al.*, 2001; Erasmus *et al.*, 2002) for this type of predictive modelling activity. These are based on observed climate data and were subsequently interpolated to

generate climate surfaces for the continent (Hutchinson, 1991; Hutchinson & Gessler, 1994; McMahon *et al.*, 1995; Hutchinson *et al.*, 1996; New *et al.*, 1999). The third data set was derived from a nested limited-area model, is entirely based on a simulation of climate processes and does not use any observed meteorological data to generate climate surfaces (DARLAM) (Engelbrecht *et al.*, 2002). In all instances explored here, cross-species and degrees of data subsampling, the DARLAM data set was equal to or marginally outperformed the interpolated climate surfaces in predicting the distributions of species. This suggests that process-based modelled data sets, such as DARLAM, should in the future be considered viable or even preferred data sets for modelling the predicted distributions of species, especially when considering their ability to capture climate at fine scales relative to range sizes of species.

A further advantage of employing process-based modelled climate surfaces for predicting species distributions is that this provides improved compatibility when using these predictions as a basis for exploring the consequences of climate change for future species distributions. This improved compatibility flows from the fact that future climate surfaces can easily be generated from nested limited-area models such as DARLAM. Thus, we can easily model the future using similar process-based models as



**Figure 6** Mean value of proportional overlap (mean  $\pm$  SD) for all tick species at different levels of random subsampling of known records. This analysis is repeated for each climate data set.

used here in DARLAM but, in contrast, we cannot observe the future in order to generate climate data sets for direct comparison with the observed data derived climate data sets (CRES and CRU).

In conclusion, the recent availability of high-resolution nested limited-area models (DARLAM; Engelbrecht *et al.*, 2002) now means that there are alternative climate surfaces available for modelling species distributions at broad scales. The results obtained here suggest that the use of data derived from a nested limited-area or process model for predicting species distributions offer a viable and even a preferred alternative to using interpolated climate surfaces derived from observed climate data. This is especially true when the spatial and temporal coverage of observed data sets are poor, resulting in climate data surfaces that are strongly spatially autocorrelated over large distances.

## ACKNOWLEDGMENTS

We would like to thank the CSIRO Atmospheric Research in Australia for making DARLAM available to LRAM at the University of Pretoria. This enabled us to perform model simulations for the SADC region. The study benefited from financial support provided in part by the University of Pretoria, the Water Research Commission, the National Research Foundation, University of Stellenbosch and the global change System for Analysis, Research and Training (START; <http://www.start.org>).

## REFERENCES

- Austin, M.P. & Meyers, J.A. (1996) Current approaches to modeling the environmental niche of eucalypts: implications for management of forest biodiversity. *Forest Ecological Management*, **85**, 95–106.
- Beard, K.H., Hengartner, N. & Skelly, D.K. (1999). Effectiveness of predicting breeding bird distributions using probabilistic models. *Conservation Biology*, **13**, 1108–1116.
- Brito, J.C., Crespo, E.G. & Paulo, O.S. (1999) Modelling wildlife distributions: logistic multiple regression vs. overlap analysis. *Ecography*, **22**, 251–260.
- Coakley, S.M., Scherm, H. & Chakraborty, S. (1999) Climate change and plant disease management. *Annual Review of Phytopathology*, **37**, 399–426.
- Cumming, G.S. (1998) Host preference in African ticks (Acari: Ixodidae): a quantitative data set. *Bulletin of Entomological Research*, **88**, 379–406.
- Cumming, G.S. (1999a) Host distributions do not limit the species ranges of most African ticks (Acari: Ixodida). *Bulletin of Entomological Research*, **89**, 303–327.
- Cumming, G.S. (1999b) *The evolutionary ecology of African ticks*. Unpublished DPhil Thesis, University of Oxford, Oxford, UK.
- Cumming, G.S. (2000) Using between-model comparisons to fine-tune linear models of species ranges. *Journal of Biogeography*, **27**, 441–455.
- Cumming, G.S. (2002) Comparing climate and vegetation as limiting factors for species ranges of African ticks. *Ecology*, **83**, 255–268.
- Davies, H.C. (1976) A lateral boundary formulation for multilevel prediction models. *Quarterly Journal of the Royal Meteorology Society*, **102**, 405–418.
- Engelbrecht, F.A., Rautenbach, C.J. de W., McGregor, J.L. & Katzfey, J.J. (2002) January and July climate simulations over the SADC region using the limited-area model DARLAM. *Water SA*, **28**, 361–373.
- Erasmus, B.F.N., Kshatriya, M., Mansell, M.W., Chown, S.L. & van Jaarsveld, A.S. (2000) A modelling approach to antlions (Neuroptera: Myrmeleontidae) distribution patterns. *African Entomology*, **8**, 157–168.
- Erasmus, B.F.N., van Jaarsveld, A.S., Chown, S.L., Kshatriya, M. & Wessels, K.J. (2002). Vulnerability of South African animal taxa to climate change. *Global Change Biology*, **8**, 679–693.
- Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38–49.
- Guisan, A. & Zimmerman, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.
- Harvell, C.D., Mitchell, C.E., Ward, J.R., Altizer, S. Dobson, A.P., Ostfeld, R.S. & Samuel, M.D. (2002) Climate warming and disease risk for terrestrial and marine biota. *Science*, **296**, 2158–2162.
- Horak, I.G. Potgieter, F.T., Walker, J.B., De Vos, V. & Boomker, J. (1983) The Ixodid tick burdens of various large ruminant species in South African nature reserves. *Onderstepoort Journal of Veterinary Research*, **50**, 221–228.

- Horak, I.G., Boomker, J. & Flamand, J.R.B. (1995) Parasites of domestic and Wild animals in South Africa. XXXIV. Arthropod parasites of nyalas in northeastern Kwazulu-Natal. *Onderstepoort Journal of Veterinary Research*, **62**, 171–179.
- Hutchinson, M.F. (1989) A new procedure for gridding elevation and streamline data with automatic removal of spurious pits. *Journal of Hydrology*, **106**, 211–232.
- Hutchinson, M.F. (1991) The application of thin plate splines to continent-wide data assimilation. *Data assimilation systems* (ed. by J.D. Jasper), pp. 104–113. BMRC Research Report No. 27, Bureau of Meteorology, Melbourne, Australia.
- Hutchinson, M.F. & Gessler, P.E. (1994) Splines – more than just a smooth interpolator. *Geoderma*, **62**, 45–67.
- Hutchinson, M.F., Nix, H.A., MacMahon, J.P. & Ord, K.D. (1996) *A topographic and climatic database for Africa*, Version 1.1. Australian National University, Canberra.
- Jeffree, E.P. & Jeffree, C.E. (1994) Temperature and the biogeographical distribution of species. *Functional Ecology*, **8**, 640–650.
- Jeffree, C.E. & Jeffree, E.P. (1996) Redistribution of the potential geographical ranges of Mistletoe and Colorado Beetle in Europe in response to the temperature component of climate. *Functional Ecology*, **10**, 562–577.
- Kaluzny, S.P., Vega, S.C., Cardoso, T.P. & Shelly, A.A. (1998) *S+ Spatial Stats – user's manual for Windows® and Unix®*. Springer-Verlag, New York.
- Lombard, A.T. (1995). The problems with multi-species conservation: do hotspots, ideal reserves and existing reserves coincide? *South African Journal of Zoology*, **30**, 145–163.
- McGregor, J.L. (1993) Economical determination of departure points for semi-Lagrangian models. *Monthly Weather Review*, **121**, 221–230.
- McGregor, J.L. (1997) Regional climate modeling. *Meteorology and Atmospheric Physics* **63**, 105–117.
- McMahon, J.P., Hutchinson, M.F., Nix, H.A. & Ord, K.D. (1995) *ANUCLIM user's guide*. Draft Report, Centre for Resource and Environmental Studies, Australian National University, Canberra, Australia.
- Meats, A. (1989) Bioclimatic potential. *Fruit flies: their biology, natural enemies and control* (ed. by A.S. Robinson and G. Hooper), Vol. 3B, pp. 241–252. Elsevier, Amsterdam.
- Mesinger, F. & Arakawa, A. (1976) *Numerical methods used in atmospheric models*, Vol. 1, 66 pp. GARP Publications Series 17. WMO, Geneva, Switzerland.
- Messenger, P.S. (1959) Bioclimatic studies with insects. *Annual Review of Entomology*, **4**, 183–206.
- Minshull, J.I. & Norval, R.A.I. (1982) Factors influencing the spatial distribution of *Rhipicephalus appendiculatus* in Kyle recreational park, Zimbabwe. *South African Journal of Wildlife Research*, **12**, 118–123.
- New, M., Hulme, M. & Jones, P. (1999) Representing twentieth-century space-time climate variability. Part I: Development of a 1961–90 mean monthly terrestrial climatology. *Journal of Climate*, **12**, 829–856.
- Nielsen, D & Wendroth, O. (2001) *Spatial and temporal statistics – sampling field soils and their vegetation*. Catena-Verlag, New York.
- Norval, R.A.I., Walker Jane, B. & Colborne, J. (1982) The ecology of *Rhipicephalus zambeziensis* and *Rhipicephalus appendiculatus* (Acarina, Ixodidae) with particular reference to Zimbabwe. *Onderstepoort Journal of Veterinary Research*, **49**, 181–190.
- Perry, B.D., Lessard, R.A.I., Norval, R.A.I., Kundert, K. & Kruska, R. (1990) Climate, vegetation and the distribution of *Rhipicephalus appendiculatus* in Africa. *Parasitology Today*, **6**, 100–104.
- Phillips, N.A. (1957) A coordinate system having some special advantages for numerical forecasting. *Journal of Meteorology*, **14**, 184–185.
- Prendergast, J.R., Quinn, R.M., Lawton, J.H., Eversham, B.C. & Gibbons, D.W. (1993) Rare species, the coincidence of diversity hotspots and conservation strategies. *Nature*, **365**, 335–337.
- Randolph (2001) The shifting landscape of tick-borne zoonoses tick-borne encephalitis and Lyme borreliosis in Europe. *Philosophical Transactions of the Royal Society, London B*, **356**, 1045–1056.
- Randolph (2002) Ticks and tick-borne disease systems in space and from space. *Advances in Parasitology*, **47**, 217–243.
- Randolph, S.E. & Rogers, D. J. (1997) A generic population model for the African tick *Rhipicephalus appendiculatus*. *Parasitology*, **115**, 265–279.
- Rechav, Y. (1981) Ecological factors affecting the seasonal activity of the brown ear tick *Rhipicephalus appendiculatus*. *Tick biology and control* (ed. by G.B. Whitehead and J.D. Gibson), pp. 187–191. Tick Research Unit, Rhodes University, South Africa.
- Reyers, B., van Jaarsveld, A.S. & Krüger, M. (2000) Complementarity as a biodiversity indicator strategy. *Proceedings of the Royal Society, London B*, **267**, 505–513.
- Robertson, M.P., Caithness, N. & Villet, M.H. (2001) A PCA-based modelling technique for predicting environmental suitability for organisms from presence records. *Diversity and Distributions*, **7**, 15–27.
- Rogers, D.J. & Randolph, S.E. (1993) Distribution of tsetse and ticks in Africa: past, present and future. *Parasitology Today*, **9**, 266–271.
- Rogers & Randolph (2000) The global spread of malaria in a future warmer world. *Science*, **289**, 1763–1766.
- Rossi, R.E., Mulla, D.J., Journel, J.G., Frantz, E.H. (1992) Geostatistical tools for modeling and interpreting ecological spatial dependence. *Ecological Monographs*, **62**, 277–314.
- Short, N.J. & Norval, R.A.I. (1981) The seasonal activity of *Rhipicephalus appendiculatus* Neumann 1901 (Acari: Ixodidae) in the high veld of Zimbabwe Rhodesia. *Journal of Parasitology*, **67**, 77–84.
- Sutherst & Maywald (1985) A computerized system for matching climates in ecology. *Agriculture, Ecosystem and Environment*, **13**, 281–299.
- Sutherst, R.W., Mayward, G.F. & Skarratt, D.B. (1995) Predicting insect distribution in a changed climate. *Insects in a changing environment* (ed. by R. Harrington and N.E. Stork), pp. 59–91, Academic Press, London.
- Tukahirwa, M.E. (1976) The effects of temperature and relative humidity on the development of *R. appendiculatus* Neumann (Acari, Ixodidae). *Bulletin of Entomological Research*, **66**, 301–312.
- Van Staden, V., Erasmus, B.F.N., Roux, J., Wingfield, M.J. & van Jaarsveld, A.S. (in press) Modelling the spatial distribution

- of two important South African plantation forestry pathogens. *Forest Ecology and Management*.
- Walker, J.B. (1974) *The Ixodid ticks of Kenya. A review of present knowledge of their hosts*. Commonwealth Institute of Entomology, London.
- Walker, P.A. (1990). Modelling wildlife distributions using a geographic information system: kangaroos in relation to climate. *Journal of Biogeography*, 17, 279–289.
- Walker, J.B, Keirens, E.J. & Horak, I.G. (2000) *The genus Rhipicephalus (Acari, Ixodidae): a guide to the brown ticks of the world*. Cambridge University Press, United Kingdom.
- Walsh, K. & McGregor, J.L. (1995) January and July climate simulations over the Australian region using a limited-area model. *Journal of Climate*, 8, 2387–2403.

## BIOSKETCHES

**Jane Olwoch** is a PhD student at the University of Pretoria on a START studentship. Her main interests are in the area of climate change and its consequences for vectors and vector-borne diseases in Africa.

**Hannes Rautenbach** is Professor of Meteorology at the University of Pretoria and focuses on numerical atmospheric modelling at meso and broader scales. Research interests include numerical seasonal forecasting, climate change scenario simulations as well as short-range numerical weather prediction.

**Barend Erasmus** is a PhD student from the University of Pretoria, and works on quantitative approaches to investigating the implications of climate change for regional biodiversity resources.

**Francois Engelbrecht** is a lecturer and PhD student at the University of Pretoria. He specializes in numerical atmospheric modelling.

**Albert van Jaarsveld** is Dean of the Faculty of Science and leads a conservation biology research group from the Department of Zoology at the University of Stellenbosch. Current research interests include conservation planning, biodiversity assessment and climate change.