

Modelling the distributions of *Culicoides* bluetongue virus vectors in Sicily in relation to satellite-derived climate variables

B. V. PURSE^{*}, A. J. TATEM[†], S. CARACAPPA[‡], D. J. ROGERS[†],
P. S. MELLOR^{*}, M. BAYLIS^{*} and A. TORINA[‡]

^{*}Institute for Animal Health, Pirbright, U.K., [†]TALA Research Group, Department of Zoology, University of Oxford, U.K. and [‡]Istituto Zooprofilattico Sperimentale della Sicilia, Palermo, Sicily

Abstract. Surveillance data from 268 sites in Sicily are used to develop climatic models for prediction of the distribution of the main European bluetongue virus (BTV) vector *Culicoides imicola* Kieffer (Diptera: Ceratopogonidae) and of potential novel vectors, *Culicoides pulicaris* Linnaeus, *Culicoides obsoletus* group Meigen and *Culicoides newsteadi* Austen. The models containing the 'best' climatic predictors of distribution for each species, were selected from combinations of 40 temporally Fourier-processed remotely sensed variables and altitude at a 1 km spatial resolution using discriminant analysis. Kappa values of around 0.6 for all species models indicated substantial levels of agreement between model predictions and observed data. Whilst the distributions of *C. obsoletus* group and *C. newsteadi* were predicted by temperature variables, those of *C. pulicaris* and *C. imicola* were determined mainly by normalized difference vegetation index (NDVI), a variable correlated with soil moisture and vegetation biomass and productivity. These models were used to predict species presence in unsampled pixels across Italy and for *C. imicola* across Europe and North Africa. The predicted continuous presence of *C. pulicaris* along the appenine mountains, from north to south Italy, suggests BTV transmission may be possible in a large proportion of this region and that seasonal transhumance (seasonal movement of livestock between upland and lowland pastures) even in *C. imicola*-free areas should not generally be considered safe. The predicted distribution of *C. imicola* shows substantial agreement with observed surveillance data from Greece and Iberia (including the Balearics) and parts of mainland Italy (Lazio, Tuscany and areas of the Ionian coast) but is generally much more restricted than the observed distribution (in Sardinia, Corsica and Morocco). The low number of presence sites for *C. imicola* in Sicily meant that only a restricted range of potential *C. imicola* habitats were included in the training set and that predictions could only be made within this range. Future modelling exercises will use abundance data collected according to a standardized protocol across the Mediterranean and, for Sicily in particular, should include non-climatic environmental variables that may influence breeding site suitability such as soil type.

Key words. *Culicoides obsoletus*, *Culicoides pulicaris*, *Culicoides imicola*, biting midge, bluetongue, climate, discriminant analysis, distribution, risk map, Sicily.

Introduction

Between 1998 and 2002, an unprecedented epizootic of the OIE List A disease, bluetongue (BT) has occurred in the Mediterranean Basin. The disease has affected numerous countries, including several unaffected by previous outbreaks and, in some, has persisted for up to 4 years. Several hundred thousand sheep have been killed to date (Baylis, 2002; Mellor & Wittmann, 2002). BT has expanded into areas where *Culicoides imicola* Kieffer, the main European vector, is absent and has spread further north than ever before, suggesting the involvement of novel vector species such as members of the *C. obsoletus* Meigen and/or *C. pulicaris* groups Linnaeus. In response, vector and virus surveillance systems are being set up across the Mediterranean Basin with the aim of mapping vector distribution and producing detailed predictive risk maps of *Culicoides*-borne disease (Calistri *et al.*, 2003; Capela *et al.*, 2003).

For the production of risk maps, it is pertinent to determine the relationship between particular vectors and the probability of virus transmission in sampled regions, as attempted in a previous study for Sicily (Torina *et al.*, 2004). In addition, because the distribution and abundance of particular vectors will rarely be known across all regions of interest, it is essential to model the relationship between environmental factors and the distribution of different vectors (Baylis *et al.*, 1998, 1999) so that risk-maps can be formulated for unsampled regions on the basis of their environment.

Previous such models have found a range of climatic determinants for *C. imicola*, the main Mediterranean vector of bluetongue virus (BTV). Of those models where climatic predictor variables were derived from weather station data (interpolated in the latter study), Rawlings *et al.* (1998) found high summer temperatures and dry summer conditions (relative humidity <60%) to be an important determinant of distribution in Iberia, whereas Conte *et al.* (2003) found annual mean daily minimum temperatures, relative humidities and mean altitude to be significant predictors in Italy. Models using satellite-derived climate predictors in several different regions have all found normalized difference vegetation index (NDVI), a proxy for soil moisture and thus for suitable *C. imicola* breeding sites, to be an important determinant of the distribution of this species in Morocco (Baylis & Rawlings, 1998), Portugal (Tatem *et al.*, 2003) and South Africa (Baylis *et al.*, 1999), and in the latter, annual values of minimum land surface temperature were also important.

However, until now, no attempts have been made to model the relationship between environmental factors and the distribution of other potential *Culicoides* vectors. Although all *Culicoides* species share the same basic habitat requirements, i.e. presence of host for bloodmeals and breeding sites for egg and larval development (Mellor *et al.*, 2000), they differ in their life-history characteristics, and in turn, the extent to which their distribution and abundance is affected by environmental factors. For example, in Israel, *C. imicola* and *C. circumscriptus* breed

in areas that are rich in organic matter but *C. imicola* prefers those without surface water (Braverman *et al.*, 1974), a preference that should be reflected in the relationship between their distributions and water availability.

The first report in this series (Torina *et al.*, 2004) presented *Culicoides* surveillance data from 911 catches in 269 sites distributed evenly across Sicily collected between 2000 and 2002. Here the distribution of different vectors in Sicily is modelled in relation to satellite-derived climate variables and the climatic determinants of distribution are compared between vectors. These relationships are extrapolated to predict the distribution of these vectors across Italy and these predictions are compared with results from other surveillance programmes in an attempt to validate the models.

Materials and methods

Culicoides catch data

The maximum abundances of *C. obsoletus*, *C. pulicaris* and *C. newsteadi* from May to December (269 sites) and for *C. imicola* from July to December (248 sites) were obtained as described in Torina *et al.* (2004). Briefly, maximum abundances were calculated per site across trap catches (1–3 nights) taken at the time of the peak annual abundance of *Culicoides* (cf. Baylis *et al.*, 1997). Because *C. imicola* reaches peak numbers later than *C. pulicaris*, *C. obsoletus*, *C. newsteadi* and *C. circumscriptus* (Ortega *et al.* 1998), maximum abundance for this species was calculated only for sites sampled between July and December. Abundance data were divided into zero and non-zero catches for each species for presence-absence modelling. Although the presence of hosts are obviously a prerequisite for the survival of blood-feeding *Culicoides*, *Culicoides* abundance was not found to be highly dependent on the number and type of livestock in previous studies (Baylis *et al.*, 1997). However, to minimize any effect of livestock factors, all traps were placed very close to livestock and usually next to cattle. The geographical co-ordinates (latitudes and longitudes) of all trap sites were determined using a Garmin GPS 12 receiver.

Satellite-derived climate data

The altitude of each trapping site was derived from the 1 × 1 km spatial resolution global topography (GTOPO30) digital elevation model obtained from the US Geological Survey (2002). Forty other environmental variables were derived from the Pathfinder Advanced Very High Resolution Radiometer (AVHRR) dataset (Rogers *et al.*, 1997), as described by Tatem *et al.* (2003), from 1 × 1 km spatial resolution imagery, from April 1992 to April 1996, downloaded from the 'Land Processes Distributed Active Archive Center' website (US Geological Survey, 2002).

The remotely sensed imagery was processed to produce monthly time series of four variables of environmental

significance normalized difference vegetation index (NDVI), middle infrared reflectance (MIR), land surface temperature (LST) and air temperature (TAIR). For example, NDVI specifically measures chlorophyll abundance and light absorption, but is correlated with soil moisture, rainfall and vegetation biomass, coverage and productivity (Campbell, 1996). MIR is correlated with the water content, surface temperature and structure of vegetation canopies (Boyd & Curran, 1998). LST is a general index of the apparent environmental surface temperature (whether soil or vegetation) and TAIR is an estimate of the air temperature a few metres above the land surface (Goetz *et al.*, 2000).

The raw monthly time series were then subjected to temporal Fourier analysis (Rogers *et al.*, 1996), a technique that extracts information about the seasonal cycles of these indices in terms of their annual, biannual and triannual cycles, each described by their phase and amplitude. This analysis removes noise from the original satellite signal, leaving a smoothed picture of seasonal change and achieves data reduction and ordination, the latter in a way that has an obvious biological interpretation in terms of seasonal cycles (Rogers, 2000). This process produced 10 individual variables for each of the environmental and meteorological indices. These were the amplitude and phase of the annual, biannual and triannual cycles (six variables), the mean, maximum and minimum of the combined (annual + biannual + triannual) Fourier description of the original signal (three variables) and the variance of the original signal (one variable).

Modelling climatic determinants of different Culicoides vectors in Sicily

For all species, discriminant analysis was used to identify the combination of the 40 satellite-derived variables and altitude that most effectively allocated the 268 sites (sampled between May and December) to the observed presence-absence classes (Rogers, 2000). For *C. imicola*, the analysis was repeated using the July to December sites ($n = 142$). Equal prior probabilities for class membership and separate within-group covariance matrices were assumed. Variable selection was carried out using a forward step-wise procedure. The criterion for variable inclusion was that the addition of the chosen variable caused the greatest between-group increase in the total Mahalanobis distance. This measure is the covariance-adjusted distance between two multivariate distribution centroids, or from a sample point to a centroid (Rogers, 2000). Since the entire dataset was used for training as well as assessment, the models were initially assessed using internal validation by four methods (see later). The kappa coefficient estimates the agreement between two variables, accounting for the degree of overlap expected by chance. It ranges from 0 to 1 and values of less than 0.2 indicate poor agreement and values greater than 0.75 are considered to indicate excellent model fit (Landis & Koch, 1977; Robinson, 2000). This coefficient has been widely used in epidemiology and public health to

assess classification accuracy (Rogers *et al.*, 1996, 2002; Tatem *et al.*, 2003). Sensitivity and specificity are the proportion of correctly predicted positive observations (or presence) and negative observations (or absence), respectively. In contrast, consumer accuracy measures the proportion of predictions that are supported by the observations. When carrying out such multivariate analyses of vector distributions, there is the danger of under- or over-fitting of the training set data, either of which makes the results less generalizable to predict vector distributions in unsampled areas. More traditional regression analyses of vector distributions tend to result in the selection of one or two significant determinants of distribution (underfitting), whereas in stepwise discriminant analyses, the accuracy of extrapolation to unsampled areas has been found to be maximized when fewer predictor variables are used than were required to maximize the predictions of the training set itself (overfitting). As a compromise between under- and over-fitting of the training set data, in this study 10 variables were added to each model because, in previous models, this is the average number of steps required for the increment in the kappa value per step to level-off. This increment in kappa value was monitored at each step to determine whether significant improvement in prediction resulted from the addition of particular variables. A probability of species presence greater than 0.5 was counted as 'presence', and anything below this was put in the 'absence' class. Median values of the top four variables added to the model were compared between observed presence and absence sites using Mann-Whitney *W*-tests. Given the high degree of multicollinearity between climatic variables, these tests should be interpreted with caution but probably indicate the magnitude and direction of the effect of each predictor variable on species distribution. Only significant tests are reported here.

The coefficients from the discriminant models were then used to predict the presence-absence of each species across the rest of Italy, i.e. each 1×1 km pixel was assigned a probability between 0 (species absence) and 1 (species presence) depending on the similarity of the environmental conditions of each pixel to those of the training pixels found to contain a particular species. No prediction was made when the Mahalanobis distance between a pixel and its assigned probability or class was two or more times greater than the maximum distance observed between any one of the sites in Sicily and the classes to which they belonged. Thus predictions were limited to areas of mainland Italy where the environment and climate were broadly similar to those observed in the Sicily trap sites.

Results

Climatic determinants of different Culicoides vectors in Sicily

The model variables that best allocated the trap sites to the observed *Culicoides* species presence and absence classes are shown in Table 1. Considering the May to December

Table 1. The 10 variables, ranked in order of importance, which best allocated the Sicilian trap sites to the observed *Culicoides* species presence-absence classes. Key: K, kappa value for the model at the step when the variable was added; LST, land surface temperature; NDVI, normalized difference vegetation index; MIR, middle infrared; TAIR, air temperature a few metres above ground; DEM, altitude derived from digital elevation model; amp., amplitude; ann., annual; bi-ann., bi-annual; tri-ann., tri-annual. Where differences in variables between presence and absence sites were significant by Mann-Whitney *W*-tests, the direction of the effect is indicated as positive by ⁺ and as negative by ⁻.

Rank	July-Dec model		May-Dec models							
	<i>C. imicola</i>		<i>C. imicola</i>		<i>C. obsoletus</i>		<i>C. pulicaris</i>		<i>C. newsteadi</i>	
	Variable	K	Variable	K	Variable	K	Variable	K	Variable	K
1	LST tri-ann. phase ⁻	0.04	LST ann. amp.	0.11	LST mean ⁻	0.24	NDVI mean ⁺	0.42	MIR mean ⁺	0.26
2	MIR tri-ann. phase	0.26	NDVI tri-ann. amp.	0.15	LST variance ⁺	0.32	NDVI tri-ann. amp.	0.47	LST ann. amp.	0.30
3	NDVI ann. phase ⁻	0.45	NDVI bi-ann. phase ⁻	0.24	MIR bi-ann. amp.	0.38	NDVI variance ⁻	0.50	TAIR min. ⁺	0.37
4	NDVI mean	0.49	LST variance	0.33	TAIR min. ⁻	0.41	LST variance	0.48	LST min. ⁺	0.39
5	NDVI variance	0.55	LST min. ⁺	0.37	LST ann. amp. ⁻	0.42	LST tri-ann. amp.	0.52	TAIR tri-ann. amp. ⁺	0.43
6	MIR ann. phase	0.56	TAIR ann. phase	0.40	MIR ann. phase	0.45	TAIR variance ⁻	0.55	MIR tri-ann. amp. ⁻	0.45
7	MIR ann. amp. ⁺	0.61	MIR mean	0.43	TAIR ann. amp. ⁻	0.45	NDVI min. ⁺	0.54	NDVI min. ⁻	0.50
8	MIR mean	0.64	NDVI mean	0.54	TAIR mean ⁻	0.49	NDVI bi-ann. phase	0.55	DEM ⁻	0.52
9	LST bi-ann. amp.	0.67	LST ann. phase	0.59	NDVI max. ⁺	0.51	LST max. ⁻	0.59	LST tri-ann. amp. ⁺	0.53
10	MIR min.	0.68	LST mean ⁺	0.63	LST min. ⁻	0.54	NDVI bi-ann. amp. ⁺	0.62	NDVI bi-ann. phase	0.55

models, the most important variables in determining the presence of *C. obsoletus* were the mean and variance of LST followed by the bi-annual amplitude of MIR. The median LST was slightly higher in absence than presence sites (medians, *absent* = 43.1°C, *n* = 108, *present* = 42.9°C, *n* = 161; Mann-Whitney *W* = 20120.5, *P* = 0.01), whereas the variance in LST was slightly lower in absence than presence sites (*absent* = 9.2°C, *present* = 9.6°C; *W* = 13096.5, *P* = 0.01). Of the 10 variables added to the model for *C. obsoletus*, seven were temperature variables (LST or TAIR) and only three were related to vegetation or moisture (i.e. MIR or NDVI). The minimum TAIR was higher in absence vs. presence sites (*absent* = 16.0°C, *present* = 14.3°C; *W* = 19744.0, *P* = 0.001). The kappa value of 0.54 indicated moderate agreement between observations and model predictions. For *C. newsteadi*, mean MIR was the most important predictor of presence but this species was similarly determined mainly by temperature variables, with five such variables included in the 10 added to the model. The mean MIR (*absent* = 42.7, *n* = 93, *present* = 43.4, *n* = 176; *W* = 9957.0, *P* < 0.001), the minimum TAIR sites (*absent* = 13.3°C, *present* = 15.5°C; *W* = 9911.0, *P* < 0.001) and the minimum LST (*absent* = 40.3, *present* = 40.7; *W* = 10105.0, *P* < 0.001) were all higher in presence vs. absence sites. Altitude was also included.

Again, a kappa value of 0.55 indicated moderate agreement between observations and predictions.

In contrast, the presence of *C. pulicaris* was determined mainly by moisture and vegetation-related variables, with six NDVI variables added to the model. Mean, tri-annual amplitude of NDVI and variance of NDVI were most important. The mean NDVI was much higher (*absent* = 0.34, *n* = 166, *present* = 0.43, *n* = 103; *W* = 18241, *P* < 0.001) and the variance in NDVI was lower (*absent* = 0.009, *present* = 0.006, *W* = 11374, *P* < 0.001) in presence vs. absence sites. The kappa value for this model was 0.62 indicating substantial agreement between observations and predictions.

Although presence between July to December is likely to be a much better indicator of site suitability, the variables added to the May to December model for the presence of *C. imicola* are also shown here for comparison with the other species. The annual amplitude of LST was most important in determining the presence of *C. imicola* in these sites but four moisture and vegetation-related variables were added to the model (including tri-annual phase of NDVI that was prominent in the model for *C. pulicaris*), suggesting that this species is more moisture dependent than *C. obsoletus* and *C. newsteadi* but less so than *C. pulicaris*.

In the July to December model for the presence of *C. imicola*, the tri-annual phase of LST was selected as the most important predictor variable but in fact resulted in an extremely low kappa value until it was combined with further variables. Of the 10 variables added to the model, only two were temperature variables, five were MIR and three were NDVI. The tri-annual peak of LST was earlier in presence vs. absence sites. The kappa value of 0.68 was higher than for any of the other species models and indicated substantial agreement between observations and model predictions.

As shown above, the models for each species produced similar overall percentages of sites correctly assigned to presence-absence classes, ranging from 77 to 87%. Table 2 shows these overall percentages along with the sensitivities, specificities and consumer's accuracies for both the presence and absence classes for each species model. Sensitivities and specificities of over 70% were obtained for both classes in all species models. The high consumer accuracy of 96% for the absence class for *C. imicola* and 84% for *C. pulicaris* indicates that these species were observed to be absent in a high proportion of sites where they are predicted to be so. For *C. newsteadi* and *C. obsoletus*, the consumer accuracies for the presence class were higher than that for the absence class, indicating that these species were observed to be present in a high proportion of sites where they are predicted to be so. Consumer accuracies for all classes and all species were above 60%, suggesting that the presence-absence predictions made by all models have some value. Figure 1 illustrates the pattern of classification of traps sites in Sicily for the presence and absence of each species. Sites were correctly classified by models for all species across a wide range of territories and there was no obvious geographical clustering of misclassified sites for any species. However, a cluster of sites in Messina province is classified as false absences for *C. newsteadi*, although most sites in the province are classified as true absences. Conversely, for *C. pulicaris* a cluster of sites in Messina province is classified as false presences for *C. pulicaris*, although most sites in the province are classified as true presences. For *C. imicola* many false presences occur around the cluster of true presences in west Palermo.

Prediction maps

Culicoides obsoletus was predicted to be widespread across Sicily itself in the northern provinces of Trapani, Palermo and Messina (Fig. 2a) and in the southern province of Siracusa. It was predicted to be absent from most of the south coast of Agrigento and Caltanissetta, and from mountainous regions in Catania near Etna and in Messina (along the Nebrodi mountains – Fig. 3 shows locations of provinces, mountains and towns in Italy). It was predicted to be widespread across Sardinia, except near mount Marmora, but restricted to the coast in Corsica. Similarly, on the mainland, this species is predicted to be widespread across most provinces but absent from highly mountainous appennine regions – in Tuscany and Emilia-Romagna with the Appenino Tosco-Emilliana, in Umbria and Marche with the Appenino Umbro Marchigiano, in Abruzzo with the Appenino Abruzzese, in Basilicata with Appenino Lucano and in Calabria with Botte Bonato (Fig. 2). It was predicted to be absent from areas as high as 1056 m (Calvo mountain on Gargano point in Puglia).

Culicoides newsteadi was predicted to be widespread across all Sicilian provinces except Messina, to be widespread in western rather than eastern provinces of Sardinia but to be sporadically present in Corsica (Fig. 2b). Although this species was predicted to be widespread across the mainland, its distribution was predicted to be more patchy than that of *C. obsoletus*. It was predicted to be absent in broader regions around the mountains in Umbria, Marche, Abruzzo but present even in these regions in small patches except in Calabria, Campania and Basilicata where it was predicted to be almost wholly absent.

Culicoides pulicaris was predicted to be present in the north of Sicily, along the Nebrodi mountains in Messina and Palermo and to be absent from wide areas in the southern provinces (Fig. 2c). It was predicted to be widespread in both Corsica and Sardinia. On the mainland it was predicted to be present more or less continuously along the Appenines that run from the north to south, with only a small patch of absence corresponding to the very highest peak of Appenino Abruzzese. It was predicted to be absent

Table 2. Measures of the accuracy of the *Culicoides* species models for the presence and absence classes. *Percentage of positive observations predicted correctly, equivalent to the producer's accuracy. †Percentage of negative observations predicted correctly. For a presence-absence model, sensitivity of the absence class = specificity of the presence class and specificity of the absence class = sensitivity of the presence class. ‡Percentage of predictions observed to be correct. Key to levels of agreement indicated by kappa values from (Landis & Koch, 1977): K = 0.0–0.20, slight; K = 0.21–0.40, fair; K = 0.41–0.60, moderate; K = 0.61–0.80, substantial; K = 0.81–1.00, near perfect.

Class	May–Dec models								July–Dec model	
	<i>C. obsoletus</i>		<i>C. newsteadi</i>		<i>C. pulicaris</i>		<i>C. imicola</i>		<i>C. imicola</i>	
	Absence	Presence	Absence	Presence	Absence	Presence	Absence	Presence	Absence	Presence
Sensitivity*	78.7	76.3	72.0	77.7	79.5	78.4	88.9	75.8	81.1	91.7
Specificity†	76.3	78.7	77.7	72.0	78.4	79.5	75.8	88.9	91.7	81.1
Consumer accuracy‡	69.1	84.1	63.2	84.0	85.7	70.2	96.3	49.0	96.6	62.3
Percentage of all sites correctly predicted	77.2		74.3		79.1		87.3		83.0	
Kappa	0.54		0.55		0.62		0.63		0.68	

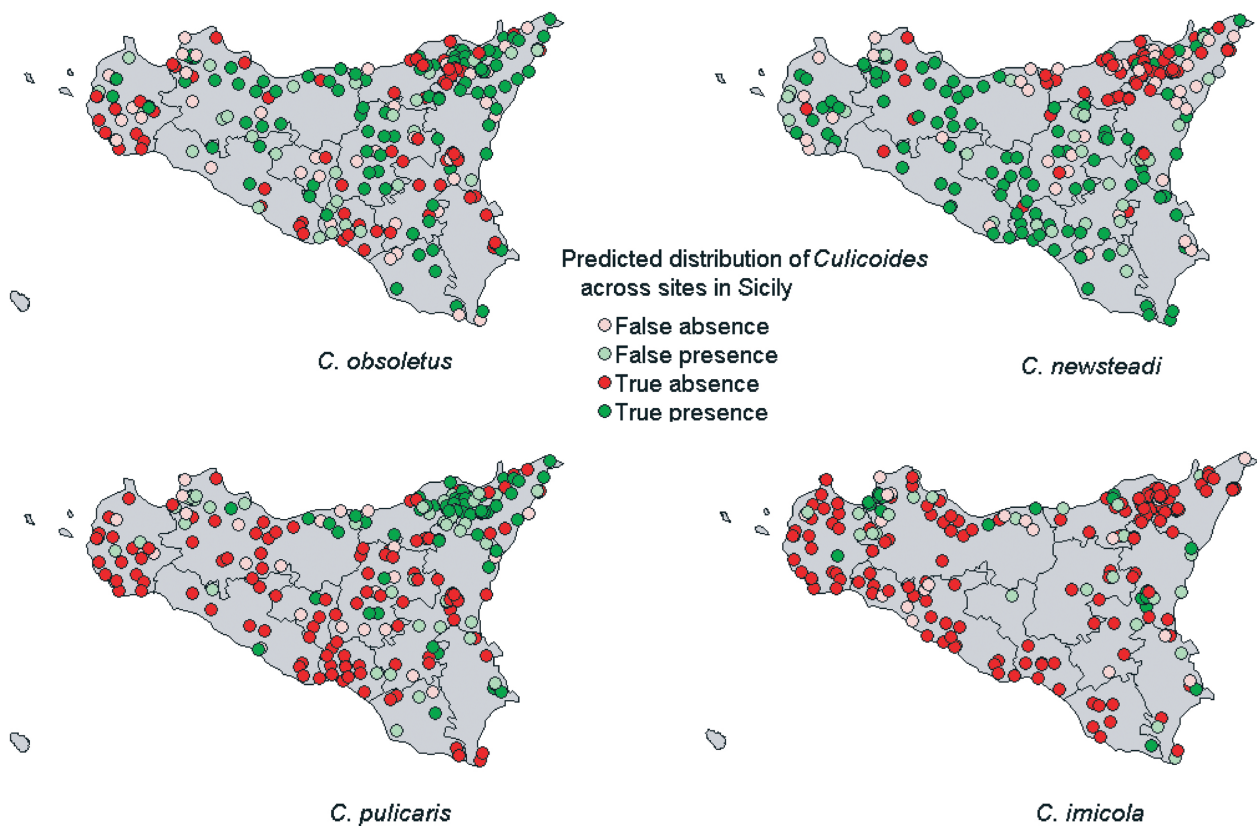


Fig. 1. Presence-absence of *Culicoides* species across trap sites in Sicily predicted by a model derived from the observed presence-absence data for each species. For *C. obsoletus*, *C. newsteadi* and *C. pulicaris*, data was taken from 269 sites, sampled between May and October. For *C. imicola*, data was taken from 248 sites, sampled between July and October.

from flatter regions of Lazio and Tuscany all along the Adriatic coast except on the Gargano point near Calvo mountain.

Culicoides imicola was predicted to be present in Sicily along the southern coast, in Trapani and in an extensive area stretching from Paterno in Catania down into Siracusa (though it was not predicted to be present in mountainous regions in this area) (Fig. 4a). Its distribution was predicted to be widespread but extremely patchy in Sardinia and restricted to only a few small patches in Corsica. On the mainland, *C. imicola* was predicted to be present along the Adriatic and Ionian coasts of Puglia and in Calabria and Basilicata along the Ionian coast. It was predicted to be patchily distributed along the Tyrrhenian coast in Tuscany, Lazio and in two patches in Campania and in fewer patches along the Adriatic coast.

The predicted distribution of *C. imicola* across the Mediterranean is consistent with recent discoveries of this species in Greece [Lesvos, Thessaly, Evia, Attica, northern Peloponnese (Corinth) – M. Patakakis, (unpublished data)] and Turkey (north-west corner of European Turkey and south coast of anatolian Turkey – Aykut Ozkul, unpublished data).

In Iberia, catch data (Capela *et al.*, 2003) shows that *C. imicola* has penetrated further north in the eastern

portion of Portugal than predicted by the model, whereas the predicted distribution in central and south-western Spain corresponds to that presented by (Rawlings *et al.*, 1997). The prediction of the presence of *C. imicola* on Mallorca and Menorca is borne out by a recent survey (Miranda *et al.*, 2003). In contrast to model predictions the species was recorded as absent on Ibiza, though only one site has been trapped. *Culicoides imicola* is predicted to have a widespread though patchy distribution along the eastern Spanish coast, but has so far been found in only a few sites such as the one in Catalonia (Sarto i Monteys & Saiz-Ardanaz, 2003). *Culicoides imicola* was not predicted to be present in Morocco where it has been observed to be widespread (Baylis *et al.*, 1997).

Discussion

Climatic determinants of different Culicoides vectors in Sicily

The differing rank and order of climatic variables added to the models of presence and absence across species indicates that the climatic determinants of distribution differ between *Culicoides* species in Sicily. Thus, predictive risk maps for BT derived entirely from distribution data for the main European vector, *Culicoides imicola*, will omit

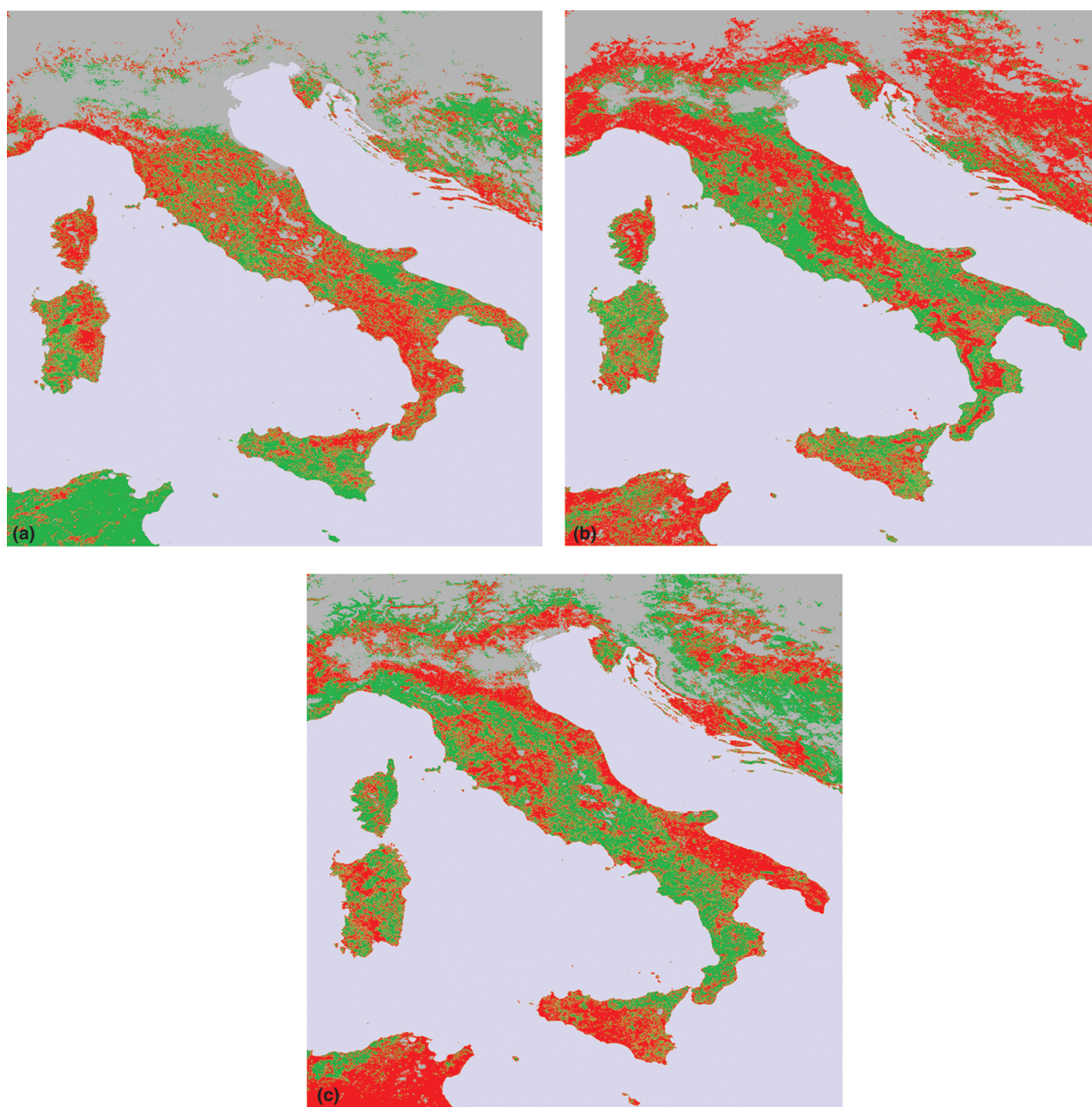


Fig. 2. Presence-absence of *Culicoides* species in Sicily and Italy predicted by a model derived from the observed presence-absence data for each species across 269 sites in Sicily. (a) *Culicoides obsoletus*. (b) *Culicoides newsteadi*. (c) *Culicoides pulicaris*. Key: green = model prediction of species presence; red = model prediction of species absence; grey = no prediction as Mahalanobis distance between a pixel and its assigned probability or class was two or more times greater than the maximum distance observed between any one of the sites in Sicily and the classes to which they belonged.

extensive regions at risk of transmission via novel vector species, whether *C. pulicaris* or *C. obsoletus*, from which BTV has been isolated recently (Caracappa *et al.*, 2003; Savini *et al.*, 2003).

The distributions of *C. obsoletus* group and *C. newsteadi* were primarily related to remotely sensed temperature variables in Sicily. Warmer, less variable temperature regimes

were associated with the absence of *C. obsoletus*. This may reflect the fact that this northern palaeoarctic species (i.e. primarily found in Europe, North Asia, North Africa) is towards the southern extreme of its range in Sicily and is probably cold-adapted, requiring relatively low temperatures for optimal development and survival. The presence of *C. newsteadi* was associated with high values of Middle

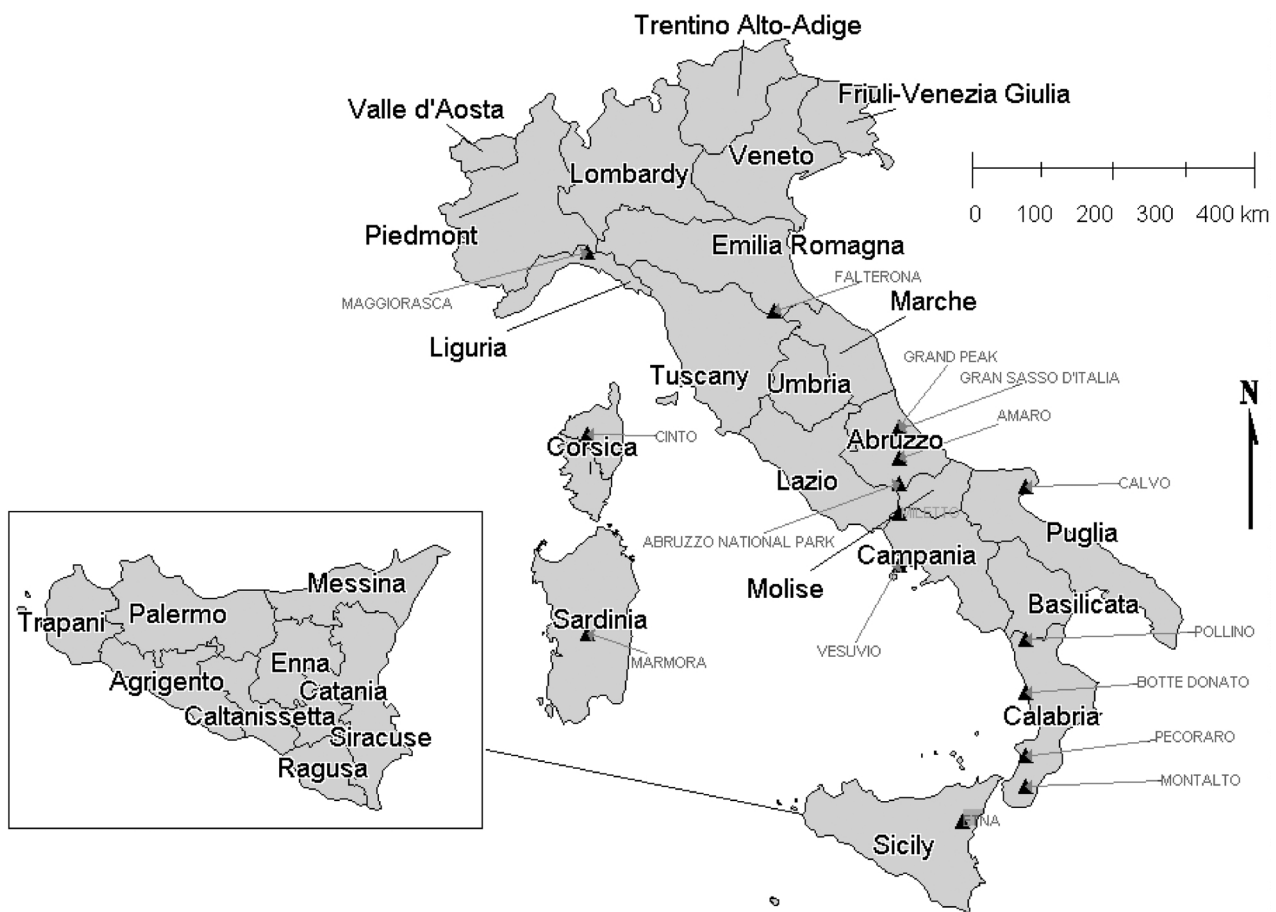


Fig. 3. Map of Italy showing provinces and mountains (closed triangles).

Infrared and high minimum temperatures perhaps reflecting high thermal requirements for development in this southern palaeoarctic species.

Normalized Difference Vegetation Index variables, themselves correlated with soil moisture, rainfall and vegetation biomass (Campbell, 1996), were the most important determinants of the distribution of *C. pulicaris*. This species was found in areas with higher mean NDVI with lower variance, suggesting that it requires microclimates with high, stable, levels of moisture for optimal survival and development. This is consistent with its breeding site requirements which are listed as wet soil and bogs (Blanton & Wirth, 1979). Although presence sites for *C. pulicaris* were at higher altitudes on average than absence sites (Torina *et al.*, 2004), NDVI variables rather than altitude itself were added to the presence-absence models. This suggests that it is the microclimatic conditions (moist, stable) frequent at high altitude rather than altitude per se that support the presence of this species.

Although a Land-Surface temperature variable was the most important predictor of *C. imicola* distribution, most other variables added to the model were NDVI or MIR variables, suggesting again that moisture levels, as well as temperature are important for the survival of this species.

The close positive relationship between NDVI and *C. imicola* distribution is consistent with previous models (Baylis *et al.*, 1998; Baylis *et al.*, 1999; Baylis *et al.*, 2001; Tatem *et al.*, 2003). This may arise because microclimate conditions that favour growth of vegetation (producing a high value of NDVI) are also those that favour breeding of *C. imicola*, since this species prefers water-saturated soils, that are high in organic matter (Braverman *et al.*, 1974).

The consumer accuracies (for both classes), sensitivities and specificities in most species models were higher than 70% and the kappa values indicated substantial levels of agreement between observations and model predictions. The values of these statistics, used for internal validation, were lower than those resulting from previous satellite-driven models for *C. imicola*, indicating that the resultant prediction maps may be less accurate. However, they remain useful for the direction of control measures. Consumer accuracies for the absence class were very high, particularly for *C. imicola* (94% of sites correctly assigned to the absence class) and *C. pulicaris* (84%) indicating that these species were found to be absent in a high proportion of sites where they were predicted to be so. Thus, areas where these two species are predicted to be absent in Sicily are recommended to be given lower priority for the application of control

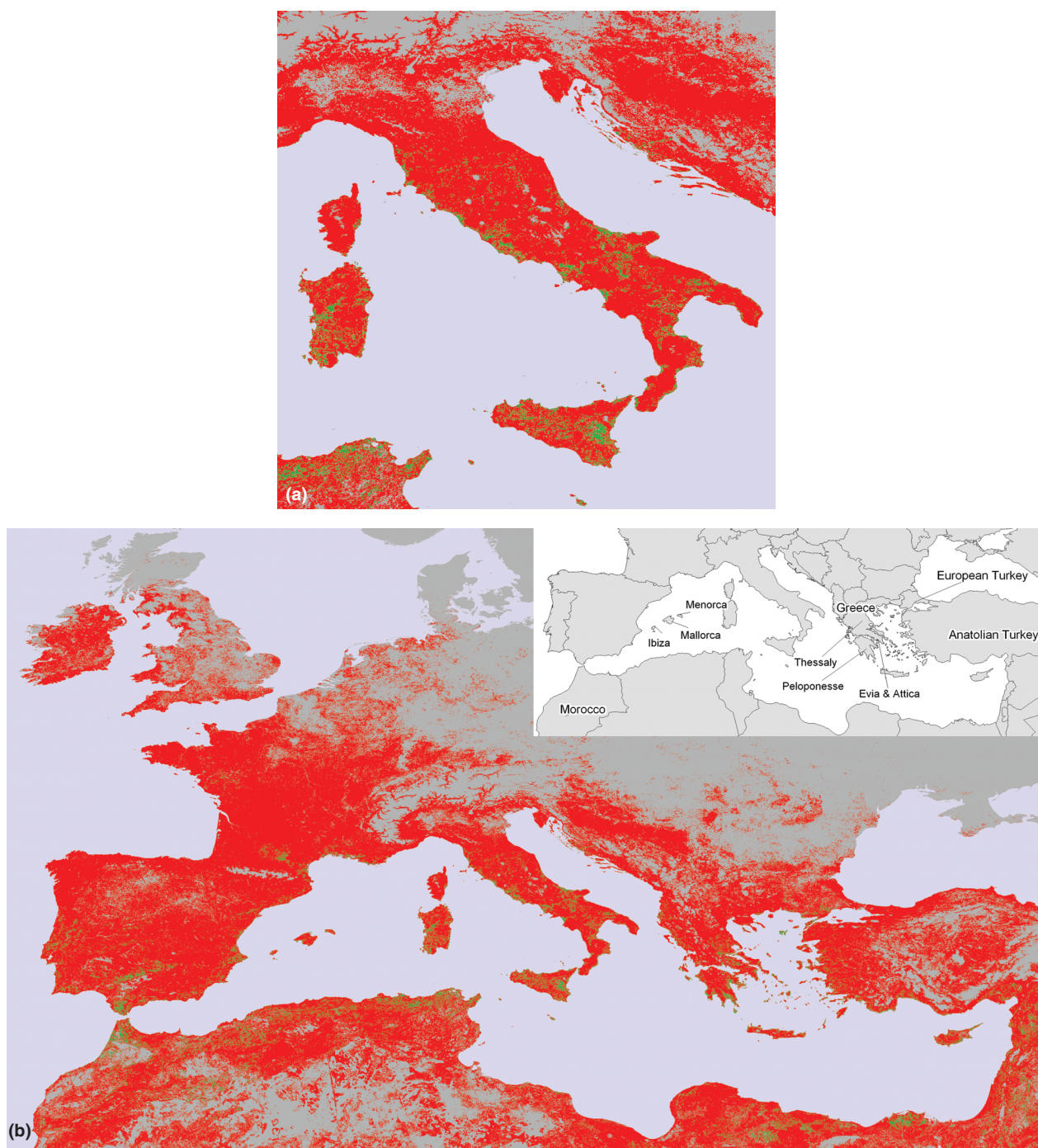


Fig. 4. Presence-absence of *Culicoides imicola* in (a) Italy and (b) the Mediterranean Basin (insert shows the places mentioned in the text) predicted by a model derived from the observed presence-absence data for *C. imicola* across 248 sites in Sicily.

measures. Such a recommendation cannot be made with the same degree of confidence with regards to the model predictions for *C. obsoletus* and *C. newsteadi* since the consumer accuracies for the absence class were slightly lower (between 60 and 70%). Patterns of classification of sites by all species models showed no particular geographical clustering and, in

the absence of high resolution layers of soil and land use information, misclassification can not be attributed to particular environmental factors at this stage. However, sites seemed more likely to be misclassified as presence sites in areas with high concentrations of true presence sites (Fig. 1). Sites seemed more likely to be classified as absence sites in

areas with high concentrations of true negative sites. This may occur because the inclusion in the training set of high concentrations of sites with similar environmental conditions produces bias in the discriminant analysis model. Where sampling is carried out primarily for surveillance during epidemics and is concentrated in outbreak sites, trap sites are likely to be more similar in environmental conditions, being suitable for transmission, than when random, grid-based sampling is carried out (Capela *et al.*, 2003).

Prediction maps for different Culicoides vectors in Sicily

Even allowing for some inaccuracy of model predictions, the large differences predicted in the ranges of the different vectors have important implications for disease surveillance and control in Sicily and mainland Italy. Most notably, *C. pulicaris* was predicted to be present more or less continuously across the Appenines that run from north to south Italy. Conte *et al.* (2003) asserted that the identification of possible, mountainous *C. imicola* free areas in central Italy could facilitate safer transhumance. Livestock are moved annually at the beginning of summer, from lowland winter pastures to highland areas until late autumn. However, such movement of infected livestock between coastal areas, at risk of BTV transmission by *C. imicola*, to mountainous areas, at the time of peak *C. pulicaris* abundance, could provide a mechanism by which the virus could be handed over between the main and the novel vector (Mellor & Boorman, 1995). The virus could then spread easily along the Appenines via passive wind dispersal of midges and cycle through *C. pulicaris* and host populations. Thus there may be few mountainous areas in which such transhumance would be safe without regular serum sampling before and after such movements were carried out. It is necessary to formulate climate models of *C. pulicaris* abundance (rather than presence) to aid determination of which of the appenines are particularly at risk of BTV transmission by this species. *Culicoides obsoletus* is predicted to be widespread across Sicily and mainland Italy, including areas where *C. pulicaris* is predicted to be absent or rare, i.e. along the Adriatic coast and in Tuscany. If this species is also widely implicated in BTV transmission in Italy and our model predictions of its range are accurate, this further increases the area of Italy that is at risk of BTV. To date, the models for these three species have only been validated internally but ideally should be validated externally. The relationship between climate and abundance may change across an insect species range since its habitat requirements may differ in core vs. marginal populations (Thomas *et al.*, 1999). However *C. obsoletus*, *C. pulicaris* and *C. newsteadi* are all palaeoartic species. The area over which the model is extrapolated is very small in relation to the total area occupied by their distributions and is reasonably distant from their southern range edges. Thus one would expect the distribution of these species in Northern Italy to be predictable from a model describing their habitat requirements in Sicily. However, detailed pixel predictions

will be made available to any workers who have surveillance data on these species for this region and would be interested in testing our models.

There was considerable regional variation in the correspondence between model predictions and observed *C. imicola* distribution, this correspondence being close in areas of Greece and Iberia (Rawlings *et al.*, 1997; Capela *et al.*, 2003; Miranda *et al.*, 2003; Sarto i Monteys & Saiz-Ardanaz, 2003) but poor in Corsica (Roger, 2002). In Morocco, *C. imicola* was observed to be widespread (Baylis *et al.*, 1997) but is predicted to be present only in the north in an area close to Tangiers. For Italy in particular, the model can be partly validated by comparison with the distribution map presented by Conte *et al.* (2003) though surveillance data is absent for extensive areas along the Adriatic coast. Our model correctly predicts the presence of *C. imicola* in coastal regions of Lazio and Tuscany, along the Ionian coast of Calabria, in Basilicata and Puglia, but in Sardinia the predicted distribution is much patchier than that observed. This could be due in part to the disparity in resolution between the two maps (the prediction map has 1 km by 1 km pixels vs. 10 km by 10 km pixels in that of Conte *et al.* (2003). Predicted areas of presence are generally much less extensive than those derived from previous satellite-driven models of *C. imicola* distribution (Baylis *et al.*, 2001; Tatem *et al.*, 2003).

Because *C. imicola* has a very restricted distribution in Sicily, only a small sample size of *C. imicola* positive sites could be included in the training set, so it is likely that a narrow range of environmental conditions that can be potentially occupied by *C. imicola* was represented. Thus only a portion of its potential distribution could be predicted from such a training set. Models where the training set included a large numbers of sites from a core area of the species distribution perform better when extrapolated to unsampled areas (Baylis & Rawlings, 1998). In the future the distribution and abundance of these species will be re-modelled using vector surveillance data from a current EU project collected according to a standardized protocol across a large area of the distribution of all species.

Due to the dependence of adult *Culicoides* activity rates on nightly weather conditions, presence or absence of a species at a particular site should be inferred from its prevalence across as many trap nights as possible. In both this model and that of Conte *et al.* (2003), it is inferred from only one or two trap nights and it is important to check whether the same relationship is obtained between climatic factors and species distribution as further presence-absence data and indeed abundance data become available.

Detailed predictive risk maps will require investigation of the precise relationship not only between remotely sensed variables and vector distribution and abundance but between these variables and the biological processes that determine population performance (Randolph, 2000). It is probable that the presence of suitable habitat for larval development and survival is the most important environmental determinant of population performance. Thus, despite the difficulties inherent in field-sampling of

Culicoides larvae in a quantitative manner (Hribar, 1990), it is essential to obtain further information on ground-based factors determining the suitability of larval habitat for different species of *Culicoides* (Kettle & Lawson, 1952; Braverman *et al.*, 1974), factors that can then be correlated with environmental variables.

In particular, it is notable that predictions both from satellite-driven climate models and models based on interpolated weather station data predict *C. imicola* to be much more widely distributed across Sicily than it is observed to be. One suggestion is that *C. imicola* having colonized the eastern Mediterranean Basin relatively recently, may not yet have had time to fill suitable habitat that is isolated by geographical position and topography from source populations. The fact that the Sicilian *C. imicola* population is a mixture of east Mediterranean and west Mediterranean haplotypes (D. Nolan, personal communication) suggests that this species is able to disperse freely to Sicily and therefore within Sicily itself. Thus it is more likely that, in Sicily, the relative absence of *C. imicola* is caused not by climatic factors but by other environmental factors that may influence its breeding sites. For example, a negative relationship was found between *C. imicola* abundance and soil sandiness in South Africa (Baylis *et al.*, 1999). Similarly, Calistri *et al.* (2003) suggest that the porous, freely draining volcanic soils with a poor moisture content that predominate in Sicily are unsuitable as *C. imicola* breeding sites. Future risk maps for *C. imicola* in Sicily should be based on the analysis of layers of both soil and climate information.

Acknowledgements

The authors would like to thank Gaspare Lo Bue, Rosa Filippi and Nicola Galati for technical assistance. Bethan Purse was supported by EU project contract no. QLK-CT-2000-00611 awarded to Drs Matthew Baylis and Philip Mellor.

References

- Baylis, M. (2002) The re-emergence of Bluetongue. *Veterinary Journal*, **163**, 1–2.
- Baylis, M. & Rawlings, P. (1998) Modelling the distribution and abundance of *Culicoides imicola*. Morocco and Iberia using climatic data and satellite imagery. *Archives of Virology* (Suppl.), **14**, 137–153.
- Baylis, M., El Hasnaoui, H., Bouayoune, H., Touti, J. & Mellor, P.S. (1997) The spatial and seasonal distribution of African horse sickness and its potential *Culicoides* vectors in Morocco. *Medical and Veterinary Entomology*, **11**, 203–212.
- Baylis, M., Bouayoune, H., Touti, J. & El Hasnaoui, H. (1998) Use of climatic data and satellite imagery to model the abundance of *Culicoides imicola*, the vector of African Horse Sickness virus, in Morocco. *Medical and Veterinary Entomology*, **12**, 255–266.
- Baylis, M., Meiswinkel, R. & Venter, G.J. (1999) A preliminary attempt to use climate data and satellite imagery to model the abundance and distribution of *Culicoides imicola* (Diptera: Ceratopogonidae) in southern Africa. *Journal of the South African Veterinary Association*, **70**, 80–89.
- Baylis, M., Mellor, P.S., Wittmann, E.J. & Rogers, D.J. (2001) Prediction of areas around the Mediterranean at risk of bluetongue by modelling the distribution of its vector using satellite imaging. *Veterinary Record*, **149**, 639–643.
- Blanton, F.S. & Wirth, W.W. (1979) The sandflies (*Culicoides*) of Florida (Diptera: Ceratopogonidae). *Arthropods of Florida and Neighboring Land Areas*, **10**, 1–204.
- Boyd, D.S. & Curran, P.J. (1998) Using remote sensing to reduce uncertainties in the global carbon budget: the potential radiation acquired in the middle infrared wavelengths. *Remote Sensing Reviews*, **16**, 293–327.
- Braverman, Y., Galun, R. & Ziv, M. (1974) Breeding sites of some *Culicoides* species (Diptera, Ceratopogonidae) in Israel. *Mosquito News*, **34**, 303–308.
- Calistri, P., Goffredo, M., Caporale, V. & Meiswinkel, R. (2003) The distribution of *Culicoides imicola* in Italy: Application and evaluation of current Mediterranean models based on climate. *Journal of Veterinary Medicine*, **50**, 132–138.
- Campbell, J.B. (1996) *Introduction to Remote Sensing*. Taylor & Francis, London.
- Capela, R., Purse, B.V., Pena, I. *et al.* (2003a) Spatial distribution of *Culicoides* species in Portugal in relation to the transmission of African horse sickness and bluetongue viruses. *Medical and Veterinary Entomology*, **17**, 165–177.
- Caracappa, S., Torina, A., Guercio, A. *et al.* (2003) Identification of a novel bluetongue virus vector species of *Culicoides*. Sicily. *Veterinary Record*, **153**, 71–74.
- Conte, A., Giovannini, A., Savini, L., Goffredo, M., Calistri, P. & Meiswinkel, R. (2003) The effect of climate on the presence of *Culicoides imicola* in Italy. *Journal of Veterinary Medicine*, **50**, 139–147.
- Goetz, S.J., Prince, S.D. & Small, J. (2000) Advances in satellite remote sensing of environmental variables for epidemiological applications. *Remote Sensing and Geographical Information Systems in Epidemiology* (ed. by S. I. Hay, S. E. Randolph and D. J. Rogers), Vol. 47, pp. 217–243. Academic Press, San Diego.
- Hribar, L.J. (1990) A review of methods for recovering biting midge larvae (Diptera: Ceratopogonidae) from substrate samples. *Journal of Agricultural Entomology*, **7**, 71–77.
- Kettle, D.S. & Lawson, J.W.H. (1952) The early stages of British biting midges *Culicoides* Latreille (Diptera: Ceratopogonidae) and allied genera. *Bulletin of Entomological Research*, **43**, 421–467.
- Landis, J.R. & Koch, G.G. (1977) The measurement of observer agreement for categorical data. *Biometrics*, **33**, 159–174.
- Mellor, P.S. & Boorman, J. (1995) The transmission and geographical spread of African horse sickness and bluetongue viruses. *Annals of Tropical Medicine and Parasitology*, **89**, 1–15.
- Mellor, P.S. & Wittmann, E.J. (2002) Bluetongue virus in the Mediterranean basin, 1998–2001. *Veterinary Journal*, **164**, 20–37.
- Mellor, P.S., Boorman, J. & Baylis, M. (2000) *Culicoides* biting midges: Their role as Arbovirus Vectors. *Annual Review of Entomology*, **45**, 307–340.
- Miranda, M.A., Borrás, D., Rincon, C. & Alemany, A. (2003) Presence in the Balearic Islands (Spain) of the midges *Culicoides imicola* and *Culicoides obsoletus* group. *Medical and Veterinary Entomology*, **17**, 52–54.
- Ortega, M.D., Mellor, R.S., Rawlings, P. & Pro, M.J. (1998) The seasonal and geographical distribution of *Culicoides imicola*, *C. pulicaris* group and *C. obsoletus* biting midges in central and southern Spain. *Archives of Virology*, **14** (Suppl.), 85–91.

- Randolph, S.E. (2000) Ticks and tick-borne disease systems in space and from space. *Remote Sensing and Geographical Information Systems in Epidemiology* (ed. by S. I. Hay, S. E. Randolph and D. J. Rogers), Vol. 47, pp. 217–243. Academic Press, San Diego.
- Rawlings, P., Pro, M.J., Pena, I., Ortega, M.D. & Capela, R. (1997) Spatial and seasonal distribution of *Culicoides imicola*. Iberia in relation to the transmission of African Horse sickness virus. *Medical and Veterinary Entomology*, **11**, 49–57.
- Rawlings, P., Capela, R., Pro, M.J. *et al.* (1998) The relationship between climate and the distribution of *Culicoides imicola* in Iberia. *Archives of Virology* (Suppl.), **14**, 93–102.
- Robinson, T.P. (2000) Spatial statistic and geographical information systems in epidemiology and public health. *Remote Sensing and Geographical Information Systems in Epidemiology* (ed. by S. I. Hay, S. E. Randolph and D. J. Rogers), Vol. 47, pp. 82–129. Academic Press, London.
- Roger, F.L. (2002) Emergence of bluetongue disease in the Mediterranean basin: modelling locations at risk for potential vectors (*Culicoides* spp.) using satellite imagery. MSc, University of London.
- Rogers, D.J. (2000) Satellites, space, time and the African trypanosomiasis. *Remote Sensing and Geographical Information Systems in Epidemiology* (ed. by S. I. Hay, S. E. Randolph and D. J. Rogers), Vol. 47, pp. 130–173. Academic Press, San Diego.
- Rogers, D.J., Hay, S.I. & Packer, M.J. (1996) Predicting the distribution of tsetse flies in West Africa using temporal Fourier processed meteorological satellite data. *Annals of Tropical Medicine and Parasitology*, **90**, 225–241.
- Rogers, D.J., Hay, S.I., Packer, M.J. & Wint, G.R.W. (1997) Mapping landcover over large areas using multispectral data derived from NOAA-AVHRR: a case study of Nigeria. *International Journal of Remote Sensing*, **18**, 3297–3303.
- Rogers, D.J., Randolph, S.E., Snow, R.W. & Hay, S.I. (2002) Satellite imagery in the study and forecast of malaria. *Nature*, **415**, 710–715.
- Sarto i Monteys, V. & Saiz-Ardanaz, M. (2003) *Culicoides* midges in Catalonia (Spain), with special reference to likely bluetongue virus vectors. *Medical and Veterinary Entomology*, **17**, 288–293.
- Savini, G., Goffredo, M., Monaco, F., de Santis, P. & Meiswinkel, R. (2003) Transmission of bluetongue virus in Italy. *Veterinary Record*, **152**, 119.
- Tatem, A.J., Capela, R., Pena, I., Mellor, P.S., Baylis, M., Purse, B.V. & Rogers, D.J. (2003) Prediction of bluetongue vector distribution in Europe and North Africa using satellite imagery. *Veterinary Microbiology*, **97**, 13–29.
- Thomas, J.A., Rose, R.J., Clarke, R.T., Thomas, C.D. & Webb, N.R. (1999) Intraspecific variation in habitat availability among ectothermic animals near their climatic limits and their centres of range. *Functional Ecology*, **13**, 55–64.
- Torina, A., Caracappa, S., Mellor, P.S., Baylis, M. & Purse, B.V. (2004) Spatial distribution of bluetongue and its vectors in Sicily. *Medical and Veterinary Entomology*, **18**, 81–89.
- US Geological Survey (2002a) NASA land processes. <http://edcdaac.usgs.gov/gtopo30/gtopo30.html>
- US Geological Survey (2002b) NASA land processes. <http://edcdaac.usgs.gov/IKM/cmp10d.html>

Accepted 19 February 2004