

Appendix E. Landscape scale effects of transport infrastructure: Best practice survey protocol and data analysis

Aims

The aim of the survey is to assess the effect of linear infrastructure (road and rail) on bats on a landscape and population scale using a robust, standardised method. This is done by measuring changes in bat abundance and diversity in relation to the proximity of a road or railway.

Overview of survey protocol and analysis

- The basic survey consists of walked transects 1 km either side of and perpendicular to the road or railway, with bat activity recorded using bat detectors during 10 min stationary spot checks at 100 m intervals from the road or railway. Weather and habitat variables are also recorded at each spot check. Transects are walked both towards and away from the road or railway.
- A minimum of ten transects must be carried out at each site to detect changes in overall bat activity or activity of the most common species. More transects may be required to detect changes in abundance of less common species.
- To increase speed of analysis and remove the subjectivity of identification, bat calls are identified to species using automated software. The data are then entered into a spreadsheet and statistical analysis methods are used to detect changes in bat abundance and diversity related to the road or railway. A robust statistical analysis is needed to overcome the complications caused by habitat variability and the intrinsically high variability of bat activity data.
- The method can be applied before construction along the proposed road or rail route to provide baseline data on bat activity and diversity in the area, and then repeated during and after construction to assess the impact of the road or railway. We advise surveying over at least two seasons for baseline activity data, and for a minimum of three years post-construction. Frequency of survey will depend upon the projects goals, but effects on bats may be detected during construction and may persist for many years.

Detailed guidelines

Number of surveys and timing

A minimum of ten transects must be carried out at each site for the method to have sufficient power to detect changes in overall bat activity, or activity of the most common species (e.g. *Pipistrellus* species). This must include either 10 independent transects (with five walked towards the road or railway, and five walked away), or 5 independent transects (each walked twice, once away from and once towards the road or railway). Transects are walked in both directions to take into account the effect of time on bat activity throughout the evening. There will be natural peaks in bat activity at certain times after sunset and walking transects in different directions will ensure that this does not bias the results.

More transects may be required to detect changes in abundance of less common species. The accuracy of your model estimates and significance levels will increase if you increase transect number until the species of interest is present at >70% of spot checks. More transects may also be required if the recommended direct sampling bat detectors are not used, as the losses in sampling time that occur with other types of detector may reduce the accuracy of model estimates. Surveys are best done June-August inclusive. May and September are acceptable, but bat activity may be lower than in other months and behaviour may not be

typical of mid-summer. To compare activity levels between years, surveys should be completed during the same months of each year.

Transect selection and route planning

Where possible, an equal number of transects should be selected on each side of the road or railway and equal numbers walked towards and away from the road or railway. Transects should be located along minor roads, bridleways or public footpaths. Minor roads should be unmarked local C roads or country lanes. Transect routes must be approximately perpendicular to the road or railway with a suitable route that can be walked up to 1 km (perpendicular distance) from the road or railway (see Figures D1 and D2). To avoid pseudo-replication, transects should be located at least 500 m apart. Large areas of water and human habitation should be avoided.

Potential transect routes may be selected initially by looking at Ordnance Survey maps and Google Earth, but must always be visited on the ground during the daytime to check access and suitability, and to familiarise surveyors with the route. Consider a daylight dry run to check your ability to walk the route safely, handle the equipment and take appropriate notes. You should also check that each transect can be completed within approximately two hours (although this timeframe should be sufficient for routes that are approximately perpendicular to the road).



Figure D1. Example of transect routes (marked in red) perpendicular to an A road (green line) © Crown Copyright/database right 2013. An Ordnance Survey/EDINA supplied service.



Figure D2. Example of a transect route. Spot checks are shown as circles and labelled with the perpendicular distance from the road. © 2013 Google Earth © 2013 Getmapping plc.

Team size and safety

The surveys should be done by at least two people for reasons of safety. Individual surveyors may walk separate transects on the same night, but only if it is possible to maintain contact via mobile phone or two way radios. Someone 'at home' should know of your whereabouts and itinerary, and an emergency plan should be in place.

Although transect routes are located on public rights of way and will not require specific permission to access them, it is good practice to notify residents or landowners when the route passes through or close to private property such as farm buildings or houses. This will avoid arousing suspicion when walking the routes at night. It may also be beneficial to notify the local police.

Safety must be considered when walking transects along minor roads where there may be traffic. For example, high visibility clothing should be worn, surveyors should walk/stand off the road where possible or step off the road when traffic approaches, and care should be taken not to distract passing motorists with head torches.

Weather conditions

Surveys should only be done in 'good' weather: temperature at the start of the survey $>7^{\circ}\text{C}$, wind $<20\text{ km/h}$ ($\sim 12\text{ mph}$), no rain. Weather conditions should be recorded at each spot check using a handheld thermometer/anemometer. If weather conditions deteriorate significantly during a transect (wind, rain, temperature below 7°C), then it should be abandoned and repeated on another night when weather conditions are suitable.

Surveyor and equipment details

Creating transect routes and marking spot check locations

Once transect routes have been selected and checked, spot check locations should be measured and marked along each route, and loaded onto a handheld GPS device for navigation during the survey. Spot checks must be located every 100 m (shortest, i.e. perpendicular distance) from the road or railway between 0 and 1000 m (Figure D2). Distances can be measured and spot check locations marked using online mapping tools such as Google Earth or Digimap (www.edina.ac.uk/digimap). The transect routes and spot check locations can then be saved and converted as necessary to be loaded onto a GPS device. We provide detailed instructions for this on page 15. After loading onto your GPS device, all transect routes and spot checks should be checked on the ground during daylight to ensure that routes are correct and no errors have been made.

Equipment

For each transect, the following equipment is required:

- ***Clipboard, datasheet and pen/pencil***
Information will need to be entered onto a datasheet at each spot check. An example datasheet has been provided on page 16.
- ***Head torch***
Head torches will be required for navigation and to fill in the datasheet when light levels are low. To avoid disturbance, surveyors should only switch on lights when needed and use the lowest light levels compatible with safety, or red light setting, with the beam angled downwards.
- ***A handheld GPS device***
Most GPS units are now accurate to 3-4 m in the UK.
- ***Heterodyne bat detector (optional)***
Surveyors may wish to carry a heterodyne bat detector to hear bats during the survey if a detector that does not emit sound (e.g. some direct-sampling detectors) is being used to record bat activity.

- **Bat detector and recording device (see below)**

We advise using full spectrum direct-sampling detectors (e.g. Pettersson D500X, www.batsound.com) to maximise sound recording and sound quality, and for compatibility with the suggested automatic software (BatClassify) for reliable identification of most UK bat species. See below for appropriate settings. Other types of bat detector can be used but may not be compatible with Bat Classify and results may be less accurate. Other software may be used, but accuracy of identification should be checked.

- **Data cards for detectors**

These will vary with recorder settings, but as a guide a 16 GB of memory should be sufficient per survey for a Pettersson D500X detector. Carry spares in case of card failure during a survey.

- **A thermometer and anemometer**

A wide range of devices are available from £15. We found the Technoline EA 3010 to be reliable.

- **Two way radios or mobile phones**

To maintain contact between surveyors or someone 'at home' during the survey

- **Spare batteries for all equipment**

Bat detector settings

Bat detector settings must be kept constant for all surveys. If multiple bat detectors are used to complete the surveys at a site, they must be of the same type and be calibrated to the same sensitivity level. The sensitivity level must be set so that it is high enough to record bats flying around each spot check, but not so high that it is constantly triggered by noise (e.g. wind, surveyor movements). This may need to be tested in the field. A recording length of 1 s is ideal for the automatic extraction and identification of bat calls. The settings we suggest for the Pettersson D500x are given below. The timer on the Pettersson D500x should be set to start recording at the survey start time, and continue for the duration of the survey.

Pettersson D500X direct-sampling detector

User Profile Menu:

SAMP. FREQ = 500
PRETRIG = OFF
REC. LEN = 1
HP-FILTER = YES
AUTOREC = YES
T.SENSE = VERY HIGH

Recording Settings Menu:

INPUT GAIN = 40
TRIG LEVEL = 80
INTERVAL = 0

Recorders

Most detectors will need to be connected to recording devices (with the exception of direct-sampling detectors which have an in-built recording function). Any digital device that can record WAV, FLAC or MP3 files at 16 or 32 bit and 320 kbps (256 or 196 kbps also acceptable) is suitable e.g. Edirol R-09HR, Zoom 2, or equivalent. WAV and FLAC files are better for archiving, but are much larger files. MP3 files can be batch converted to WAV or FLAC format (using free software e.g. Foobar, www.foobar.org) for analysis with BatClassify or other software.

The Survey

Surveys should commence 30 min after sunset, and be completed within approximately two hours. Surveyors should be set up at the first spot check with the equipment in plenty of time before the start of the survey. The survey information should be filled in at the top of the datasheet. At each spot check the surveyor must record bat activity for exactly 10 min and note down the relevant information on the datasheet before walking briskly (aim for 6 km/h) to the next spot check. The bat detector should be held at approximately waist height pointing upwards and away from the surveyor at all times.

The information to be recorded is as follows:

- **Spot check start time**
To the nearest second, to ensure that all spot check recordings are comparable
- **Spot check end time**
Exactly 10 min after the spot check start time. This can be filled in straight after the spot check start time so that the surveyor knows when to stop recording and walk to the next spot check
- **Track number 'From'**
The track number on the bat detector at the start of the spot check/10 min recording
- **Track number 'To'**
The track number on the bat detector at the end of the spot check /10 min recording
- **Weather variables**
Record temperature and wind speed during the spot check. Be consistent with when you take these measurements e.g. halfway through every spot check
- **Habitat grade**
The habitat at each spot check should be recorded as a grade between 1 and 5 (see Table 1 below). If the habitat differs on each side of the path, the highest habitat grade should be recorded
- **Path type**
The path that the transect is walked on (C for minor road, B for bridleway, FP for footpath)
- **Notes**
Enter any additional information or points of interest that are not covered by the other columns

Habitat grade

Table 1. The criteria used to classify spot check habitat types

Grade	Habitat type
1	Fence or wall lining road/path & open fields beyond
2	Hedges/shrubby verges lining road/path & open fields beyond
3	Intermittent medium trees/bushes lining road/path & open fields beyond
4	Intermittent tall trees lining road/path & open fields beyond
5	Continuous tall tree cover lining road/path with woodland &/or open fields beyond

Preparation checklist

- If needed, do you have you the landowner's permission to visit the site?
- Have you found sufficient transects and loaded routes and spot checks onto a GPS?
- Have you checked transect routes on the ground
- Are surveyors familiarised with the protocols and survey sheet?
- Dry run?

Survey summary/checklist

- Have you considered safety (risk assessment) and met the necessary safety requirements e.g. high visibility clothing?
- Have you checked the weather forecast?
- Equipment – do you have it all, does it work correctly? (GPS, detectors, recorder, headtorch, datasheet, thermometer/anemometer, spare batteries/memory cards)

- Start 30 min after sunset
- 10 min recordings of bat calls at 100 m intervals along transect
- Record habitat and weather variables at each spot check
- Download, organise and back up data after each survey

An alternative to walked transects would be the placement of automated, stationary loggers at the same measurement points along the 1 km transects. This would require a greater outlay in equipment, but would allow data to be gathered simultaneously over longer time periods. Analysis methods would be the same. Similarly, other commercial bat detectors/loggers could be used but the reliability of the call extraction and identification software would need to be tested. Care would be needed to ensure all loggers are fully functional and have the same sensitivity and settings.

Post-survey data handling

As soon as possible after the survey, download the data from the bat detector. Change the name of files or save in appropriately named folders so that they can be easily identified and located. File the datasheet. Reset the detectors, format memory cards and recharge batteries as needed ready for the next survey.

To organise and enter the data into spreadsheets, set up a Microsoft Excel Workbook for your site. Create a sheet for each transect, a master sheet to collate the totals, and a final modified master sheet with the totals formatted as required for analysis (see instructions below).

Step 1. Species identification

The first step in processing the data from each transect is to run it through bat identification software.

Bat identification software

We advise that you use BatClassify (www.bitbucket.org/chrisscott/batclassify) if using Pettersson detectors. Identification software is available for some other bat detectors. Automated call extraction and identification is a field in its infancy and accuracy is still improving, sometimes from a poor and undocumented baseline. We advise against using software unless it comes with a clear statement of how well it extracts calls, and how accurate it is at identifying species.

BatClassify is free, Open Source software with a user-friendly interface that has been tested. The software extracts all but the weakest bat calls and identifies most UK bat species with high accuracy using Pettersson detectors. Preliminary assessment with files from other bat detector brands, such as EcoObs Batcorder (www.ecoobs.de), Elekon Batlogger (www.elekon.ch), Batbox Griffin (www.batbox.com) and Wildlife Acoustics SM2 (www.wildlifeacoustics.com) suggest that it will be useful with these too, but it has not been rigorously tested.

Instructions for downloading and running BatClassify are given on page 13.

Run the data from your first transect through the software. The output of BatClassify, the 'Results.csv' file, is a simple spreadsheet file that will open automatically in Excel if double-clicked. Copy the data from the output file into a worksheet in your Microsoft Excel workbook and name the worksheet with the relevant transect name, e.g. Transect A. Repeat this for each of your transects, creating and naming a new worksheet for each one, until you have run all of your data through the software.

Open the worksheet for your first transect and look at the output file. It gives the date of analysis on the top line beneath which is a line for each sound file analysed, giving the filename, date and time, and the probability of occurrence (between 0 and 1) for each species within the sound file, as follows:

Date									
Filename	Date	Time	Bbar	Malc	Mbec	Mbra Mmys	Mnat	Paur	etc.
Site1_001	29.06.14	22:48:55	0.02	0.92	0.02	0.04	0.02	0.04	
Site1_002	29.06.14	22:50:11	0.03	0.01	0.03	0.20	0.95	0.01	

An acceptable default threshold for correct identification would be > 0.9.

Step 2. Assign a species to each bat call

Add a column to the output file called 'Species' (this will be column Q). This will be used to assign a species to each bat call in the spreadsheet using the probability of occurrence.

Copy and paste the following formula into the first row of your 'Species' column and fill the formula down the remaining rows (highlight the cell you have entered the formula in, hover the mouse over the bottom right corner of the cell so that a cross symbol appears and double click).

```
=IF(E2>=0.9,"Bbar",IF(F2>=0.9,"Malc",IF(G2>=0.9,"Mbec",IF(H2>=0.9,"MbraMmys",IF(I2>=0.9,"Mdau",IF(J2>=0.9,"Mnat",IF(K2>=0.9,"NSL",IF(L2>=0.9,"Paur",IF(M2>=0.9,"Ppip",IF(N2>=0.9,"Ppyg",IF(O2>=0.9,"Rfer",IF(P2>=0.9,"Rhip","NA"))))))))))))
```

This will assign a species name to each row when the probability of occurrence is equal to or greater than 0.9. For probabilities of less than 0.9, "NA" will be displayed. These sound files are likely to contain noise only and can be discarded. To delete these rows, highlight the 'Species' column that you created and select 'Sort & Filter... Filter' on the 'Home' tab. Click on the arrow that will appear at the top of the 'Species' column and untick all of the boxes except "NA". This will display only the unwanted rows. Highlight them all, right click and select delete. Click on the 'Species' column filter again and select all of the boxes. Only your results that were assigned a species will then be displayed.

BatClassify is capable of extracting a very high percentage of calls and identifying them to species. Weak calls may be missed, but identification of these will in any case be less certain. Even though the software has successful ID rates in the 90-100% range, errors will be inevitable when processing large numbers of bat calls. Scan through your results, and if you have reason to question any of them (e.g. a rare species not recorded in the area before, or outside its known range), the recordings can be viewed as spectrograms and inspected for obvious misidentification, perhaps due to extraneous noise interfering with analysis. The sonograms should match the expected form of the species or genus. Modify the species identification if required.

Step 3. Assign a spot check distance to each bat call

Each bat call must be labelled with the spot check (distance from the road/railway) where it was recorded. Create another new column in your output file called 'Distance' (this will be column R). Refer to your survey datasheet for that transect, pair up the track numbers with those for the start and end of each spot check and label each bat call in your output file with a distance corresponding each spot check. If you recorded bat calls continuously throughout the whole of the transect (as with the Pettersson D500X), then you will need to discard the calls that were recorded when walking between spot checks (this can be done using the filter tool as described above).

Step 4. Create totals for each spot check

Create another column in your output file and call it 'Count code' (this will be column S). In the first row copy and paste the following formula, and then fill the formula down the rest of the column.

=CONCATENATE(R2,Q2)

This will give you an identifier for each row which combines the spot check distance and the species e.g. '100Ppip' for *P. pipistrellus* at a spot check located 100 m from the road.

Create a totals table within the same spreadsheet, starting at the top of column W, as follows:

W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AH
Distance (m)	Bbar	Malc	Mbec	MbraMmys	Mdau	Mnat	NSL	Paur	Ppip	Ppyg	Rfer	Rhip	Total passes	Total species
0														
100														
200														
300														
400														
500														
600														
700														
800														
900														
1000														

To calculate the totals, a unique formula must be entered into each cell of the table. This will tell Excel to count a row when it contains a count code specific to that distance and species. For example:

For Bbar at a 0 m use the formula: =COUNTIF(S:S, "0Bbarb")

For Ppip at 800 m use the formula: =COUNTIF(S:S, "800Ppip")

Enter the appropriate formula into each cell of the table for each species.

For the 'Total passes' column, enter the following formula and fill down the column. This adds together the number of bat passes for all species.

=SUM(X2:A12)

NB. To exclude *Nyctalus/Eptesicus* species (NSL) from your total number of bat passes use the formula below. This is advised as these species are likely to be unaffected by roads/rail, which may influence the ability to detect overall effects (see Discussion in the main report).

=SUM(X2:AC2,AE2:A12)

For the 'Total species' column, enter the following formula and fill down the column. This counts each species when it is present.

=COUNTIF(W2:AH2,">0")

NB. To exclude *Nyctalus/Eptesicus* species (NSL) from your total number of bat species use the formula below.

=COUNTIF(X2:AC2, ">0") + COUNTIF(AE2:A12, ">0")

Step 5. Repeat for each transect

Repeat the steps above with the data from each transect in separate worksheets in your workbook.

Step 6. Create a master spreadsheet

Create a new worksheet in your workbook and label it 'Master'. The master spreadsheet contains totals and data for each spot check per transect (see example on page 14). Enter data in the following columns, beginning in column A:

Route:	The label/identifier for each transect route (e.g. A, B, C, D...)
Direction:	The direction the route was walked, either towards or away (from the road/railway)
Date	The date the transect was walked (format cells as 'Date')
Distance (m):	The distance of the spot check from the road/railway
Sunset:	The sunset time for the day the transect was completed in hours and min (format cells as 'Time')
Sunset (min):	The sunset time above converted into minutes (sunset time x 1440). Use the following formula where E2 is the 'Sunset' column: Enter '= E2*1440' in the first row and fill the formula down the column (format cells as 'Number')
Spot start time:	The start time of each spot check in hours and min (format cells as 'Time')
Spot start time (min):	The start time of each spot check in minutes. Use the following formula where G2 is the 'Spot start time' column: Enter '=IF(G2<0.9, ((G2*1440)+1440), (G2*1440))' in the first row and fill the formula down the column (format cells as 'Number')
Time after sunset (min):	The time of each spot check expressed as the number of minutes after sunset. Use the following formula where F2 is the 'Sunset (min)' column and H2 is the 'Spot start time (min)' column Enter '=H2-F2' and fill the formula down the column (format cells as 'Number')
No. bat passes:	The total number of bat passes per spot check (with or without NSL). Copy and paste the values from the relevant worksheet for each transect
No. bat species:	The total number of bat species per spot check. Copy and paste the values from the relevant worksheet for each transect
Bbar, Malc etc:	Each column is the number of bat passes for each individual species. There will be a column for each species (our example spreadsheet has had some removed to save space). Copy and paste the values from the relevant worksheet for each transect
Temp:	The temperature recorded at each spot check
Wind:	The wind speed recorded at each spot check
Hab:	The habitat grade recorded at each spot check

Step 7. Create a spreadsheet formatted for use in R.

Create a new worksheet in your workbook and label it 'R'. Highlight and copy all of the data in your master spreadsheet, and paste the data (using 'Paste Values') into your R worksheet (this removes the formulas and copies the numbers only). Modify the columns as follows:

Route:	Keep this column
Direction:	Delete this column
Day:	Add this column if transects have been walked twice. For the first night a transect is walked enter '1', and for the second night enter '2'
Date:	Delete this column
Distance (m):	Rename as 'Dist'
Sunset:	Delete this column
Sunset (min):	Delete this column
Spot start time:	Delete this column
Spot start time (min):	Delete this column
Time after sunset (min):	Rename as 'Time'
No. bat passes:	Rename as 'Pass'
No. bat species:	Rename as 'Species'
Bbar, Malc etc:	Keep these columns
Temp:	Delete this column
Wind:	Delete this column
Hab:	Keep this column

Step 8. Save your R worksheet as a CSV file.

Open a new blank Excel workbook (with 'File...New'). Copy and paste the data from your R worksheet into the new workbook, and save as a CSV (Comma delimited) file (with 'File'...'Save As'...'Save as type: CSV). Name the CSV file with your site reference e.g. Site 1. Do not save your R worksheet directly as a CSV file from your main workbook as this will cause the other worksheets to be lost.

Data analysis

Generalised Estimating Equations (GEE)

The transect data is analysed using Generalised Estimating Equations (GEE) in the R program to test for the effect of road/rail on bat activity while taking account of auto-correlation that may be present between spot checks conducted along the same route and on the same night. We provide the full R code and detailed instructions for this in this Appendix, which uses the R CSV file created above.

This method can be repeated for individual bat species that are found to be present at more than 20% of spot checks, and we provide R code to identify these species. However, it should be noted that the results and significance levels will be approximate for species with zero counts of more than 70% (see Appendix C for more information).

This method can also be repeated for the number of bat species. The number of species detected at each spot check is counted and converted to a proportion of the total number of species detected across the whole site. GEEs are then fit to the data using the same methods as above but with a binomial link. See instructions below.

Multiple years of survey

If transect studies have been repeated over more than one year, the data from each year must be analysed separately. Comparing the magnitude of road/rail effects between two years of survey is not sufficient to draw reliable conclusions as to whether effect sizes have changed, as the effect size predicted by the model may vary due to other factors.

The data collected from a transect study may also be used as baseline activity data, and we suggest using t-tests or one way anova as a means to compare levels of bat activity over two or more years. Use of this method is untested and instructions are not provided, but both tests are simple to carry out in R, Microsoft Excel, SPSS, etc.

Presenting the results

Model output

An example and explanation of the final model output from R is given in the detailed instructions with the R code. When reporting your results the estimate, standard error and significance of the distance from the road/railway variable and any other significant variables in your final model should be displayed in a table, along with the Scale and Correlation Parameters. The Wald statistic and significance level should be reported individually for each variable in the text of your results section, along with the effect size expressed as a percentage (see below).

Model predictions

Instructions are given for producing tables of model predictions. The predicted difference in bat activity over the range of your explanatory variables can be expressed as a percentage as a way of displaying the magnitude of effects. For example the percentage change in the number of predicted bat passes between 0 and 1000 m from the road/railway can be calculated using the following formula:

$$\left(\frac{\text{predicted no. of bat passes at 1000 m} - \text{predicted no. of bat passes at 0 m}}{\text{predicted no. of bat passes at 0 m}} \right) \times 100$$

This formula can also be modified to calculate the difference in bat activity between different habitat grades, or between different times after sunset.

N.B. model predictions for the number of bat species are expressed as probabilities and these should be presented in your results directly rather than calculating a percentage change.

Plots

Plots of predictions should be displayed for the effect of distance from the road/railway for total bat activity, the number of bat species and for individual species where possible. You may also wish to provide plots of habitat and time effects if significant in your final model.

Examples for all of the above can be found in the results section of the main report.

Drawing conclusions

Habitat and time effects

The habitat and time variables may or may not be significant, and this does not matter in terms of the results. There are biological reasons why each may affect bat activity (discussed in the main report) and they must be included in the final model if significant so that we can accurately model the road/rail effects.

Distance effects

The distance from the road/railway variable is retained in all models so that the effect can be investigated regardless of statistical significance.

Care must be taken when interpreting distance effects from your final model. A negative effect of distance from the road/railway in your model means that the number of bat passes is decreasing as you move further away from the road, therefore there is no detrimental road effect. A positive effect of distance from the road in your model means that the number of bat passes is increasing as you move further away from the road, and it can be assumed that the road/railway is having a negative effect on local bat populations.



If the effect of distance from the road/railway is negative in your final model (i.e. no detrimental road effect), or the effect of distance is positive (i.e. a negative road effect) but less than 20% with a significance level of $P > 0.05$ then it can be assumed that the road or railway is not having a detrimental effect on local bat populations on a landscape scale.



If there is a positive effect of distance from the road/railway (i.e. a negative road effect) which is over 20% (regardless of statistical significance) or less than 10% with a significance level of $P < 0.05$, then it can be assumed that the road is having a detrimental effect on local bat populations.

N.B. Surveys must have been carried out to the exact guidelines described above and with at least the minimum amount of replicates advised in order to safely draw these conclusions. Any reduction in the amount of data collected may result in road/rail effects not being detected correctly by the analysis.

These criteria apply to the model results for species-specific bat activity. However, it should be noted that for less abundant species (e.g. those present at fewer than 30% of spot checks), estimates and significance levels will be approximate and may be unreliable.

The number of bat species provides another measure of the impact of the road, but as this is usually correlated with bat activity, we suggest drawing conclusions on the magnitude of road/rail effects from the models of bat activity and supplementing the results with the models for the number of bat species.

Although we give guideline figures every case should be considered carefully in relation to the local and national scarcity of the species at risk. Rare or vulnerable bat species may need special consideration.

BatClassify

Step 1. Download

Find 'BatClassify_win32.zip' at <https://bitbucket.org/chrisscott/batclassify/downloads> and right click and select 'Save link as...' to download the file. Unzip the file contents to a location of your choice.

Step 2. Run BatClassify

The BatClassify.exe is found in the 'bin' folder inside the 'BatClassify' folder previously extracted. Double-click the .exe file to bring up the BatClassify user interface.

If the recordings to be analysed were made with a time-expansion detector and are long continuous files, go to step 3. If the recordings are from a direct-sampling detector go straight to step 4.

Step 3. Split time-expansion recordings

Continuous time-expansion recordings must be broken up into smaller files prior to analysis. The 'split' process extracts non-silent regions of the recording, saving each as a separate file.

Click the 'Split' tab of the user interface. Select the folder containing the recording(s) to be split using the directory browser (click 'Browse...'). Next select the folder where the output files should be saved, as above. Click run. If a large number of files were selected for splitting, a progress bar will indicate the percentage of work complete.

Step 4. Analyse recordings

Select the Analyse tab of the user interface. Select the folder containing the recording(s) to be analysed using the directory browser (click 'Browse...'). Next select the folder where the output files should be saved, as above. Select whether spectrograms of extracted calls should be saved using the check box. Click run. If a large number of files were selected for splitting, a progress bar will indicate the percentage of work complete.

Step 5. View results

Once analysis has finished, the 'Results.csv' file will be found in the output folder selected in step 4. CSV files can be viewed in a number of programs e.g. Microsoft Excel, OpenOffice and Google Docs. Analysis results are organised with a row for each recording, and a column for each species. Classifier output for each species or species group is a number between 0 and 1, and represents an estimate of presence within that recording. For example, a recording containing calls from a lesser horseshoe only should have values near 0 for all species except Rhip, where values closer to 1 indicate increasing confidence in presence. Note that multiple species may be present in a single recording, in which case more than one value in that row may be non-zero.

Example of Master spreadsheet

N.B. Some of the species columns have been removed for display purposes

Route	Direction	Date	Distance (m)	Sunset	Sunset (mins)	Spot start time	Spot start time (mins)	Time after sunset (mins)	No. bat passes	No. bat species	Bbarb	Ppip	Ppyg	Mbec	Mdau	etc.	Temp (°C)	Wind (km/h)	Habitat
A	towards	01/07/2014	0	21:23:00	1283	23:46:00	1426	143	0	0	0	0	0	0	0	0	9.7	0	5
A	towards	01/07/2014	100	21:23:00	1283	23:34:00	1414	131	0	0	0	0	0	0	0	0	9.9	0	5
A	towards	01/07/2014	200	21:23:00	1283	23:23:00	1403	120	0	0	0	0	0	0	0	0	9.9	0	5
A	towards	01/07/2014	300	21:23:00	1283	23:12:00	1392	109	5	2	0	0	0	0	0	1	10.2	0	5
A	towards	01/07/2014	400	21:23:00	1283	23:00:00	1380	97	0	0	0	0	0	0	0	0	10.1	0	5
A	towards	01/07/2014	500	21:23:00	1283	22:49:00	1369	86	8	1	0	0	0	0	0	0	10.4	0	5
A	towards	01/07/2014	600	21:23:00	1283	22:38:00	1358	75	0	0	0	0	0	0	0	0	10.8	0	5
A	towards	01/07/2014	700	21:23:00	1283	22:26:00	1346	63	10	2	0	3	0	0	0	0	11.7	0	5
A	towards	01/07/2014	800	21:23:00	1283	22:15:00	1335	52	59	2	0	51	0	0	0	0	11.8	0	5
A	towards	01/07/2014	900	21:23:00	1283	22:04:00	1324	41	3	1	0	3	0	0	0	0	11.8	0	5
A	towards	01/07/2014	1000	21:23:00	1283	21:53:00	1313	30	15	1	0	15	0	0	0	0	11.8	0	5
A	away	03/07/2014	0	21:22:00	1282	21:52:00	1312	30	0	0	0	0	0	0	0	0	17.8	2.4	5
A	away	03/07/2014	100	21:22:00	1282	22:03:00	1323	41	25	2	0	10	0	0	0	0	17.4	0	5
A	away	03/07/2014	200	21:22:00	1282	22:14:00	1334	52	15	1	0	0	0	0	0	0	16.8	0	5
A	away	03/07/2014	300	21:22:00	1282	22:25:00	1345	63	0	0	0	0	0	0	0	0	16.5	0	5
A	away	03/07/2014	400	21:22:00	1282	22:36:00	1356	74	88	2	0	68	0	0	0	0	16.2	0	5
A	away	03/07/2014	500	21:22:00	1282	22:48:00	1368	86	78	1	0	78	0	0	0	0	15.9	0	5
A	away	03/07/2014	600	21:22:00	1282	22:59:00	1379	97	0	0	0	0	0	0	0	0	16.4	0	5
A	away	03/07/2014	700	21:22:00	1282	23:11:00	1391	109	0	0	0	0	0	0	0	0	16.2	0	5
A	away	03/07/2014	800	21:22:00	1282	23:22:00	1402	120	0	0	0	0	0	0	0	0	16.1	0	5
A	away	03/07/2014	900	21:22:00	1282	23:33:00	1413	131	0	0	0	0	0	0	0	0	16	0	5
A	away	03/07/2014	1000	21:22:00	1282	23:44:00	1424	142	0	0	0	0	0	0	0	0	16.1	0	5

Creating transect routes and marking spot check locations in Google Earth

- Open Google Earth (available to download at <http://google.com/earth>)
- Create a folder in 'My Places' for your road or rail site (select 'Add', 'Folder') and name appropriately e.g. Site 1. Make sure that this folder is selected in the 'Places' window before adding Placemarks and Paths below.
- Zoom into the first transect route with the road or railway visible.
- Reset the Tilt (the view may become tilted when zooming in, which will affect your measurements) by selecting 'View' on the tool bar, then 'Reset', 'Tilt'.
- Select the Ruler tool (either by clicking on the ruler icon at the top of the screen, or select 'Tools', 'Ruler').
- Measure the direct line distance from the road or railway to find the point on the transect route that is located 1000 m away.
- Mark this point with a Placemark (select by clicking on Placemark icon or select 'Add', 'Placemark'). Move the Placemark pin to your spot check location and label with the spot check name (e.g. 1000 m).
- Repeat for all 11 spot checks along your transect.
- Draw over the transect route with the Path tool (click on Path icon or select 'Add', 'Path') and label the transect route with a unique name (e.g. Transect A).
- Repeat for all transects for your road or rail site.
- Select your site folder (click on it in the Places window) and save as a Kml file (right click, 'Save Place As...', type in a file name and select Kml for 'Save as type', click 'Save').

The KML file must be converted to a file type that is compatible with your GPS device. GPX is a standard format compatible with most devices. Free software is available to convert file types online. We used GPS Visualizer (<http://gpsvisualizer.com>) which is a free online conversion utility. Select 'Convert to GPX', upload your KML file and select 'Convert'. Your converted file can then be downloaded, saved and imported into your GPS software to view and transfer to your GPS device.

Survey datasheet

Date:		Site:		Surveyor name (s):	
Sunset time:		Transect route:		Cloud cover (%)	Start
Start time:		Weather:		End	

Distance (m)	Spot check time		Track no.		Temp (°C)	Wind speed (km/h)	Habitat grade (1 - 5)	Path type Minor road (C) Footpath (FP) Bridleway (B)	Notes
	Start	End	From	To					
0									
100									
200									
300									
400									
500									
600									
700									
800									
900									
1000									

Analysis in R

R is an open source program capable of advanced analysis and plotting methods unavailable with most commercial software. R can be downloaded for free from <http://www.r-project.org/>. Manuals are available on the website, and help pages for specific packages can be accessed by typing '?' followed by the package name into R.

We give detailed step-by-step instructions for the analysis below with the relevant R code in grey and what it means in pink. R runs on the code you input and it is very sensitive to syntax, punctuation, spaces, spelling, case, etc. At the end of this Appendix we also provide a continuous section of text with only the R code and a brief explanation of the function of each segment. This code can be copied and pasted directly into an R script file. We advise working from a script file rather than directly into the R console, as the code may be saved, modified and re-run easily. If code is modified, take care to ensure the structure remains correct. The analysis can be done step by step or sections of code can be combined to perform multiple steps in one go. However, if errors occur it is best to go back to a single step approach to identify problems.

Generalised Estimating Equations (GEE)

Preparing your data

Read your R CSV file into R and give it a name e.g. site1 and attach your data:

```
site1<-read.csv("C:\\Transects\\Site1.csv", header=T)
attach(site1)
```

Replace highlighted text with the file location in your computer directory and the file name of your spreadsheet.

Attaching the data assigns it as a dataframe for further use in R.

Tell R the nature of the variables (e.g. numeric or factors):

```
site1$Route<-as.factor(site1$Route)
site1$Day<-as.factor(site1$Day)
site1$Dist<-as.numeric(site1$Dist)
site1$Time<-as.numeric(site1$Time)
site1$Pass<-as.numeric(site1$Pass)
site1$Species<-as.numeric(site1$Species)
site1$Bbar<-as.numeric(site1$Bbar)
site1$Malc<-as.numeric(site1$Malc)
site1$Mbec<-as.numeric(site1$Mbec)
site1$MbraMmys<-as.numeric(site1$MbraMmys)
site1$Mdau<-as.numeric(site1$Mdau)
site1$NSL<-as.numeric(site1$NSL)
site1$Paur<-as.numeric(site1$Paur)
site1$Ppip<-as.numeric(site1$Ppip)
site1$Ppyg<-as.numeric(site1$Ppyg)
site1$Rfer<-as.numeric(site1$Rfer)
site1$Rhip<-as.numeric(site1$Rhip)
site1$Hab<-as.factor(site1$Hab)
```

Create new variable (called 'RouteNight') to label observations from the same night and same transect:

```
site1$RouteNight<-factor(ifelse(site1$Day=="1",paste(site1$Route, ".1", sep=""),
paste(site1$Route, ".2", sep="")))
```

This new variable is used for the autoregressive correlation structure of the GEE, and is only needed if transects have been walked twice, otherwise the 'Route' variable is used instead.

Install the packages required for the analysis for GEE (select an appropriate CRAN mirror when prompted):

```
install.packages("geepack")
install.packages("MESS")
install.packages("xlsx")
install.packages("plotrix")
```

N.B. Packages only need to be installed once, this step can be skipped for future analyses.

Load the installed packages into R:

```
library(geepack)
library(MESS)
library(xlsx)
library(plotrix)
```

Specifying the model

Transform number of bat passes:

```
LPass<-log(Pass+1)
```

The number of bat passes is transformed to a log(count+1) to be used as the dependent variable in the GEE model.

Pass is used for total bat activity. For individual species replace with relevant species name e.g. Ppip.

Run the GEE models:

```
M1<-geeglm(LPass ~ Dist + Hab + Time, family=gaussian, data=sitel,
  id=RouteNight, corstr="ar1", std.err="fij")
M2<-geeglm(LPass ~ Dist + Hab + poly(Time,2,row=TRUE), family=gaussian,
  data=sitel, id=RouteNight, corstr="ar1", std.err="fij")
M3<-geeglm(LPass ~ Dist + Time, family=gaussian, data=sitel, id=RouteNight,
  corstr="ar1", std.err="fij")
M4<-geeglm(LPass ~ Dist + poly(Time,2,row=TRUE), family=gaussian,
  data=sitel, id=RouteNight, corstr="ar1", std.err="fij")
M5<-geeglm(LPass ~ Dist + Hab, family=gaussian, data=sitel, id=RouteNight,
  corstr="ar1", std.err="fij")
M6<-geeglm(LPass ~ Dist, family=gaussian, data=sitel, id=RouteNight,
  corstr="ar1", std.err="fij")
```

There are six models named M1 to M6. Each model has the transformed number of bat passes as the dependent variable and a different combination of explanatory variables (distance from the road/railway 'Dist', habitat grade 'Hab' and time after sunset 'Time').

Time after sunset is entered as either a linear term (for a linear relationship with the number of bat passes), or as a quadratic term (for a curved relationship with the number of bat passes).

A first order autoregressive model "ar1" is used to account for auto-correlation between spot checks conducted along the same route and on the same night. The blue highlighted text is the variable that describes this. If each transect was walked once only use the variable 'Route'. If transects were walked twice use the new variable 'RouteNight' as above.

The fully iterated jackknife estimation principle is used (std.err="fij") to avoid bias due to a small number of clusters (< 30 transects completed).

A Gaussian distribution with an identity link is used and should give the best fit to the data.

Compare the QICu values of the models and select the model with the lowest QICu value as the final model:

```
print(QIC(M1,M2,M3,M4,M5,M6),digits=5)
```

QICu (Quasilikelihood under the Independence model Criterion) values are used to compare GEE models and evaluate which predictors best explain the response variable in the model. A model with a lower QICu value has a better fit and is more reliable.

If one or more models have similar QICu values making selection difficult, the significance of individual variables can be tested by comparing two models with and without that variable:

```
anova(M5,M6)

Analysis of 'Wald statistic' Table

Model 1 LPass ~ Dist + Hab
Model 2 LPass ~ Dist
  Df      X2 P(>|Chi|)
1  4 13.905 0.007603 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

If the result for a variable is significant ($P < 0.05$, shown by one or more stars ''), then it contributes significantly to the model and should be retained.*

For example, the code above compares the models M5 and M6 (which include the distance variable with and without the habitat variable). The resulting output suggests that habitat grade contributes significantly to the fit of the model and should be retained in the final model.

To view the model output of the final model use the command below and replace the highlighted text with your final model name:

```
summary(M5)
```

An example of a model output is shown below (this will differ depending on your final model structure):

```
Call:
geeglm(formula = LPass ~ Dist + Hab, family = gaussian, data = sitel,
        id = RouteNight, corstr = "ar1", std.err = "fij")

Coefficients:
              Estimate   Std.err   Wald Pr(>|W|)
(Intercept) 1.0875014 0.3965028  7.523 0.00609 **
Dist         0.0014949 0.0003746 15.926 6.59e-05 ***
Hab2         0.5510206 0.4927199  1.251 0.26343
Hab3         1.1087058 0.4459873  6.180 0.01292 *
Hab4         0.5757255 0.5789921  0.989 0.32005
Hab5         0.7211035 0.4177602  2.979 0.08433 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Estimated Scale Parameters:
              Estimate Std.err
(Intercept)   3.317   0.355

Correlation: Structure = ar1 Link = identity

Estimated Correlation Parameters:
              Estimate Std.err
alpha       0.4382  0.1659
Number of clusters: 14 Maximum cluster size: 11
```

The estimate, standard error, test statistic (Wald) and significance (P) of each explanatory variable are given in the model output. A variable has a significant effect when $P < 0.05$. This is shown by one or more stars ''.*

A positive estimate indicates a positive effect of an explanatory variable on the independent variable (number of bat passes), and a negative estimate indicates a negative effect. If there is a quadratic term for time after sunset included in the model this will have both a positive and negative estimate which models the curved shape of the effect.

As habitat grade is a factor (categorical variable), estimates are given for each level and are in comparison with the lowest habitat grade (in this case habitat grade 1) as a reference point. The above output shows that habitat grade 3 has significantly higher bat activity than grade 1. The other habitat grades do not differ significantly from grade 1. This may be because there is no difference or there may be too little data.

The effect sizes appear small in the model output as they are on a log scale. Model predictions are used to interpret the magnitude of these effects (see below).

The Correlation Parameter is the amount of correlation (on a scale of 0 to 1) between two sequential spot checks along the same transect route on the same night. The correlation is moderate in the example above (typically < 0.3 is low, $0.3 - 0.7$ is moderate and > 0.7 is high).

The number of clusters (transects completed) and maximum cluster size (number of spot checks) as specified in the correlation structure are also given.

The final model output can be saved by either highlighting the text in the R console and copy/pasting to a word processing program such as Notepad or Word, or you can save directly to a text file using the code below (replace yellow highlighted text with your model name, and blue highlighted text with the location and desired file name for saving):

```
capture.output(summary(M5), file="C:\\Transects\\Summary.txt")
```

When reporting your results, the estimate, standard error and significance of the variables in the final model should be displayed in a table, along with the Scale and Correlation Parameters.

The Wald statistic and significance level should be reported for each variable in the text of your results section.

For variables with more than one level (e.g. habitat grade and the quadratic term for time), single test statistics and significance levels can be obtained by using the `anova()` function described above to compare the final model with and without the variable included.

View plots of residuals (replace highlighted text with your model name):

```
plot(resid(M5))  
plot(fitted(M5), resid(M5))
```

Residuals should not be clustered. Some clustering may occur for less abundant species due to high numbers of zeros, but models should be acceptable if species are present at more than 20% of spot checks.

Model predictions

We can use predictions generated by the final model to look at the effect of distance from the road/railway (and other variables) on bat activity.

Distance from the road/railway

Create a new dataset from which to make predictions:

```
New_dist<-data.frame(Dist = seq(0,1000, length = 101), Hab = c("5"), Time =  
c(60))
```

The code above creates a list of distances from the road (every 10 m) and holds habitat constant at grade 5, and time at 60 min after sunset. The highlighted values can be changed as required, or removed if the habitat or time variables are not retained in the final model.

Produce model predictions for the new dataset:

```
logPass_predict_dist<-predict(M5,newdata=New_dist,type="response")
```

This will generate the predicted number of bat passes from your final model for each distance in the dataset. Replace the highlighted text with the final model name.

NB. Ignore the warning message in R (it is produced because a dummy dataset is being used).

Combine the predictions with the new dataset as a table, and view it in R:

```
Predict_dist<-cbind(New_dist,logPass_predict_dist)  
Predict_dist
```

To save the table of predictions in an Excel file (replace blue highlighted text with desired file location and name for saving):

```
write.xlsx(Predict_dist, "C:\\Transects\\Predictions_distance.xlsx")
```

Create a plot of the predictions:

```
par(mar=c(4,4,4,4.5))  
count.overplot(Dist,LPass, cex.axis=1.2, pch=20, tol=0.2, xlab="Distance  
(m)", ylab = "log(bat passes+1)", cex.lab=1.5, font=1, las=2, xaxt="n")  
xax<-c(0,200,400,600,800,1000)  
axis(1, at=xax, cex.axis=1.2, font=1)  
original_scale<-c(0,2,5,10,20,50,100,200,500,1000)  
original_scale_position<-log(original_scale+1)  
axis(side=4,at=original_scale_position,labels=original_scale,  
cex.axis=1.2,las=2)  
mtext(side = 4, line = 3.5, "No. bat passes",cex=1.5)  
lines(New_dist$Dist, logPass_predict_dist, lwd=2, lty=1)
```

The code above will plot a line of the predicted effect of distance from the road/railway according to the final model (see Figures 3.1, 3.4 and 3.7 in the full report for examples).

The left y axis shows bat passes on a log scale as predicted by the model, and the right axis shows the number of bat passes on the original scale for reference.

The points represent the data, with numbers for replicate points.

Habitat and time effects

If time after sunset and/or habitat grade are significant variables in the final model, model predictions and plots can be produced for these variables also using the code as follows.

Habitat:

```
New_hab<-data.frame(Dist=c(500), Hab=c("1","2","3","4","5"), Time=c(60))
logPass_predict_hab<-predict(M5, newdata=New_hab, type="response")
Predict_hab<-cbind(New_hab, logPass_predict_hab)
Predict_hab
write.xlsx(Predict_hab, "C:\\Transects\\Predictions_habitat.xlsx")

par(mar=c(4,4,4,4.5))
plot(as.numeric(Hab), LPass, type="n", cex.axis=1.2, xlab="",
     ylab="log(bat passes+1)", cex.lab=1.5, font=1, yaxt="n", xaxt="n")
axis(side=2, cex.axis=1.2, las=2)
xax<-c(1,2,3,4,5)
axis(1,at=xax, labels=c("1","2","3","4","5"), cex.axis=1.5,
     tck=0.01,lwd=1.5)
mtext(side = 1, line = 2.75, "Habitat grade", cex=1.5)
original_scale<-c(0,2,5,10,20,50,100,200,500,1000)
original_scale_position<-log(original_scale+1)
axis(side=4,at=original_scale_position,labels=original_scale,
     cex.axis=1.2,las=2)
mtext(side = 4, line = 3.5, "No. bat passes",cex=1.5)
points(New_hab$Hab, logPass_predict_hab, cex=1.5, pch=18, lwd=2)
lines(New_hab$Hab, logPass_predict_hab, lty=2, lwd=1)
```

Time:

```
New_time<-data.frame(Dist=c(500), Hab=c("5"), Time=seq(30,160,length=14))
logPass_predict_time<-predict(M5, newdata=New_time, type="response")
Predict_time<-cbind(New_time,logPass_predict_time)
Predict_time
write.xlsx(Predict_time, "C:\\Transects\\Predictions_time.xlsx")

par(mar=c(4,4,4,4.5))
count.overplot(Time,LPass, cex.axis=1.2, pch=20,tol=1.5,xlab="Time (min)",
     ylab = "log(bat passes+1)",cex.lab=1.5,font=1,las=2,xaxt="n")
xax<-c(30,60,90,120,150)
axis(1,at=xax, labels=c("30","60","90","120","150"),cex.axis=1.2,font=1)
original_scale<-c(0,2,5,10,20,50,100,200,500,1000)
original_scale_position<-log(original_scale+1)
axis(side=4,at=original_scale_position,labels=original_scale,
     cex.axis=1.2,las=2)
mtext(side = 4, line = 3.5, "No. bat passes",cex=1.5)
lines(New_time$Time, logPass_predict_time,lwd=2, lty=1)
```

Species-specific models

Use the code below to determine which species are abundant enough for individual analyses (i.e. present at more than 20% of spot checks).

Calculate the percentage of zero counts per species:

```
Bbarb0<-((sum(Bbarb==0))/(nrow(sitel))*100)
Malc0<-((sum(Malc==0))/(nrow(sitel))*100)
Mbec0<-((sum(Mbec==0))/(nrow(sitel))*100)
MbraMmys0<-((sum(MbraMmys==0))/(nrow(sitel))*100)
Mdau0<-((sum(Mdau==0))/(nrow(sitel))*100)
Mnat0<-((sum(Mnat==0))/(nrow(sitel))*100)
NSL0<-((sum(NSL==0))/(nrow(sitel))*100)
Paur0<-((sum(Paur==0))/(nrow(sitel))*100)
Ppip0<-((sum(Ppip==0))/(nrow(sitel))*100)
Ppyg0<-((sum(Ppyg==0))/(nrow(sitel))*100)
Rhip0<-((sum(Rhip==0))/(nrow(sitel))*100)
Rfer0<-((sum(Rfer==0))/(nrow(sitel))*100)
```

Combine as a table with the total counts for each species:

```
percent_zero<-rbind(Bbarb0, Malc0, Mbec0, MbraMmys0, Mdau0, Mnat0, NSL0,
  Paur0, Ppip0, Ppyg0, Rhip0, Rfer0, c("Total"))
no_passes<-rbind(sum(Bbarb), sum(Malc), sum(Mbec), sum(MbraMmys),
  sum(Mdau), sum(Mnat), sum(NSL), sum(Paur), sum(Ppip), sum(Ppyg),
  sum(Rhip), sum(Rfer), sum(Pass))
Species_counts<-data.frame(percent_zero,no_passes)
Species_counts
```

Export table to an Excel file (replace highlighted text with desired file location and name for saving):

```
write.xlsx(Species_counts, "C:\\Transects\\Species_counts.xlsx")
```

For species with < 80% zero counts, GEE analysis may be carried out using methods identical to those described above, by substituting 'Pass' and 'LPass' for the relevant species in the R code e.g. 'Ppip' and 'LPpip'.

Model for the number of species

The GEE method is also used to produce a model for the number of species, with the following modifications.

Convert the 'Species' variable to proportion data by creating and combining two new variables:

```
site1$Species_fail<-7-site1$Species
site1$Sp<-cbind(site1$Species, site1$Species_fail)
```

The species counts are converted to proportion data by calculating the number of species that you failed to record out of the total recorded across the whole site (called 'Species_fail', replace highlighted number with the total number of bat species that you recorded across your site), and then creating a matrix (called 'Sp') of two columns for successes (number of species recorded) and failures (number of species not recorded).

The GEE methods above can be repeated substituting 'LPass' with 'Sp' in the model formulas, and specifying the family as binomial, as in the example below:

```
M1<-geeglm(Sp~Dist+Hab+Time, family=binomial, data=site1,id=RouteNight,
  corstr="ar1", std.err="fij")
```

The binomial family is used to model the proportion data that we have created consisting of successes and failures.

Final model selection is carried out using identical methods to those described above.

Model predictions are also carried out in the same way (replacing the highlighted text as previously) e.g. for distance effects:

```
New_dist<-data.frame(Dist = seq(0,1000, length = 101), Hab = c("5"), Time =
  c(60))
Sp_predict_dist<-predict(M5,newdata=New_dist,type="response")
Species_predict_dist<-cbind(New_dist,Sp_predict_dist)
Species_predict_dist
write.xlsx(Species_predict_dist,"C:\\Transects\\Species_Predict_distance.xlsx"
)
```

As the number of species is modelled as proportion data, the results are expressed as probabilities (i.e. the probability of recording more bat species).

Plots of model predictions differ from those for bat activity and show predicted probabilities (see code for each variable below).

Distance effects:

```
par(mar=c(4,4.1,2,2))
plot(New_dist$Dist, Sp_predict_dist, type="l", lwd=2, ylim=c(0,0.4),
      xlab="Distance (m)", ylab="Probability(species)", cex.lab=1.5,
      cex.axis=1.2, yaxt="n")
axis(side=2, cex.axis=1.2, las=2)
```

Habitat effects:

```
par(mar=c(4,4.1,2,2))
plot(as.numeric(New_hab$Hab), Sp_predict_hab, type="l", lwd=1, lty=2,
      ylim=c(0,0.4), xlab="Habitat grade", ylab="Probability(species)",
      cex.lab=1.5, cex.axis=1.2, yaxt="n")
axis(side=2, cex.axis=1.2, las=2)
points(New_hab$Hab, Sp_predict_hab, cex=1.5, pch=18)
```

Time effects:

```
par(mar=c(4,4.1,2,2))
plot(New_time$Time, Sp_predict_time, type="l", lwd=2, ylim=c(0,0.4),
      xlab="Time (min)", ylab="Probability(species)", cex.lab=1.5,
      cex.axis=1.2, yaxt="n")
axis(side=2, cex.axis=1.2, las=2)
```

At the end of your R session, detach you data and clear the console:

```
detach(site1)
rm(list = ls())
```


R CODE - copy and paste into R console

```
#Load data
sitel<-read.csv("C:\\Transects\\Sitel.csv", header=T)
attach(sitel)

#Assign variable types
sitel$Route<-as.factor(sitel$Route)
sitel$Day<-as.factor(sitel$Day)
sitel$Dist<-as.numeric(sitel$Dist)
sitel$Time<-as.numeric(sitel$Time)
sitel$Pass<-as.numeric(sitel$Pass)
sitel$Species<-as.numeric(sitel$Species)
sitel$Bbar<-as.numeric(sitel$Bbar)
sitel$Malc<-as.numeric(sitel$Malc)
sitel$Mbec<-as.numeric(sitel$Mbec)
sitel$MbraMmys<-as.numeric(sitel$MbraMmys)
sitel$Mdau<-as.numeric(sitel$Mdau)
sitel$NSL<-as.numeric(sitel$NSL)
sitel$Paur<-as.numeric(sitel$Paur)
sitel$Ppip<-as.numeric(sitel$Ppip)
sitel$Ppyg<-as.numeric(sitel$Ppyg)
sitel$Rfer<-as.numeric(sitel$Rfer)
sitel$Rhip<-as.numeric(sitel$Rhip)
sitel$Hab<-as.factor(sitel$Hab)

#Create new variable for observations from the same night and same transect route
sitel$RouteNight<-factor(iffelse(sitel$Day=="1",paste(sitel$Route,".1",sep=""),
    paste(sitel$Route, ".2",sep="")))

#Install and load packages
install.packages("geepack")
install.packages("MESS")
install.packages("xlsx")
install.packages("plotrix")
library(geepack)
library(MESS)
library(xlsx)
library(plotrix)

#Transform number of bat passes
LPass<-log(Pass+1)

#Run GEE models for total bat activity
M1<-geeglm(LPass ~ Dist + Hab + Time, family=gaussian, data=sitel, id=RouteNight,
    corstr="ar1", std.err="fij")
M2<-geeglm(LPass ~ Dist + Hab + poly(Time,2,raw=TRUE), family=gaussian,
    data=sitel, id=RouteNight, corstr="ar1", std.err="fij")
M3<-geeglm(LPass ~ Dist + Time, family=gaussian, data=sitel, id=RouteNight,
    corstr="ar1", std.err="fij")
M4<-geeglm(LPass ~ Dist + poly(Time,2,raw=TRUE), family=gaussian, data=sitel,id
    =RouteNight, corstr="ar1", std.err="fij")
M5<-geeglm(LPass ~ Dist + Hab, family=gaussian, data=sitel, id=RouteNight,
    corstr="ar1", std.err="fij")
M6<-geeglm(LPass ~ Dist, family=gaussian, data=sitel, id=RouteNight,
    corstr="ar1", std.err="fij")

#QICu values
print(QIC(M1,M2,M3,M4,M5,M6), digits=5)
#Compare models
anova(M5,M6)
#View model output
summary(M5)
#Save model output
```

```

capture.output(summary(M5), file="C:\\Transects\\Summary.txt")

#Plot residuals
plot(resid(M5))
plot(fitted(M5), resid(M5))

#Generate model predictions for distance variable
New_dist<-data.frame(Dist = seq(0,1000, length = 101), Hab = c("5"), Time =
  c(60))
logPass_predict_dist<-predict(M5,newdata=New_dist,type="response")
Predict_dist<-cbind(New_dist,logPass_predict_dist)
Predict_dist
#Save table of predictions as Excel file
write.xlsx(Predict_dist, "C:\\Transects\\Predictions_distance.xlsx")

#Create a plot of the predictions
par(mar=c(4,4,4,4.5))
count.overplot(Dist,LPass, cex.axis=1.2, pch=20, tol=0.2, xlab="Distance (m)",
  ylab = "log(bat passes+1)", cex.lab=1.5, font=1, las=2, xaxt="n")
xax<-c(0,200,400,600,800,1000)
axis(1, at=xax, cex.axis=1.2, font=1)
original_scale<-c(0,2,5,10,20,50,100,200,500,1000)
original_scale_position<-log(original_scale+1)
axis(side=4,at=original_scale_position,labels=original_scale, cex.axis=1.2,las=2)
mtext(side = 4, line = 3.5, "No. bat passes",cex=1.5)
lines(New_dist$Dist, logPass_predict_dist, lwd=2, lty=1)

#Generate table and plot of predictions for habitat variable
New_hab<-data.frame(Dist=c(500), Hab=c("1","2","3","4","5"), Time=c(60))
logPass_predict_hab<-predict(M5, newdata=New_hab, type="response")
Predict_hab<-cbind(New_hab, logPass_predict_hab)
Predict_hab
write.xlsx(Predict_hab, "C:\\Transects\\Predictions_habitat.xlsx")
par(mar=c(4,4,4,4.5))
plot(as.numeric(Hab), LPass, type="n", cex.axis=1.2, xlab="",
  ylab="log(bat passes+1)", cex.lab=1.5, font=1, yaxt="n", xaxt="n")
axis(side=2, cex.axis=1.2, las=2)
xax<-c(1,2,3,4,5)
axis(1,at=xax, labels=c("1","2","3","4","5"), cex.axis=1.5, tck=0.01,lwd=1.5)
mtext(side = 1, line = 2.75, "Habitat grade", cex=1.5)
original_scale<-c(0,2,5,10,20,50,100,200,500,1000)
original_scale_position<-log(original_scale+1)
axis(side=4,at=original_scale_position,labels=original_scale, cex.axis=1.2,las=2)
mtext(side = 4, line = 3.5, "No. bat passes",cex=1.5)
points(New_hab$Hab, logPass_predict_hab, cex=1.5, pch=18, lwd=2)
lines(New_hab$Hab, logPass_predict_hab, lty=2, lwd=1)

#Generate table and plot of predictions for time variable
New_time<-data.frame(Dist=c(500), Hab=c("5"), Time=seq(30,160,length=14))
logPass_predict_time<-predict(M5, newdata=New_time, type="response")
Predict_time<-cbind(New_time,logPass_predict_time)
Predict_time
write.xlsx(Predict_time, "C:\\Transects\\Predictions_time.xlsx")
par(mar=c(4,4,4,4.5))
count.overplot(Time,LPass, cex.axis=1.2, pch=20,tol=1.5,xlab="Time (min)",
  ylab = "log(bat passes+1)",cex.lab=1.5,font=1,las=2,xaxt="n")
xax<-c(30,60,90,120,150)
axis(1,at=xax, labels=c("30","60","90","120","150"),cex.axis=1.2,font=1)
original_scale<-c(0,2,5,10,20,50,100,200,500,1000)
original_scale_position<-log(original_scale+1)
axis(side=4,at=original_scale_position,labels=original_scale, cex.axis=1.2,las=2)
mtext(side = 4, line = 3.5, "No. bat passes",cex=1.5)
lines(New_time$Time, logPass_predict_time,lwd=2, lty=1)

```

```

#Calculate the percentage of zero counts per species
Bbarb0<-((sum(Bbarb==0))/(nrow(sitel))*100)
Malc0<-((sum(Malc==0))/(nrow(sitel))*100)
Mbec0<-((sum(Mbec==0))/(nrow(sitel))*100)
MbraMmys0<-((sum(MbraMmys==0))/(nrow(sitel))*100)
Mdau0<-((sum(Mdau==0))/(nrow(sitel))*100)
Mnat0<-((sum(Mnat==0))/(nrow(sitel))*100)
NSL0<-((sum(NSL==0))/(nrow(sitel))*100)
Paur0<-((sum(Paur==0))/(nrow(sitel))*100)
Ppip0<-((sum(Ppip==0))/(nrow(sitel))*100)
Ppyg0<-((sum(Ppyg==0))/(nrow(sitel))*100)
Rhip0<-((sum(Rhip==0))/(nrow(sitel))*100)
Rfer0<-((sum(Rfer==0))/(nrow(sitel))*100)

#Combine as a table and export to an Excel file
percent_zero<-rbind(Bbarb0, Malc0, Mbec0, MbraMmys0, Mdau0, Mnat0, NSL0, Paur0,
  Ppip0, Ppyg0, Rhip0, Rfer0, c("Total"))
no_passes<-rbind(sum(Bbarb), sum(Malc), sum(Mbec), sum(MbraMmys), sum(Mdau),
  sum(Mnat), sum(NSL), sum(Paur), sum(Ppip), sum(Ppyg), sum(Rhip), sum(Rfer),
  sum(Pass))
Species_counts<-data.frame(percent_zero,no_passes)
Species_counts
write.xlsx(Species_counts, "C:\\Transects\\Species_counts.xlsx")

#Run species-specific GEE models e.g. Ppip
LPpip<-log(Ppip+1)
Ppip1<-geeglm(LPpip ~ Dist + Hab + Time, family=gaussian, data=sitel,
  id=RouteNight, corstr="ar1", std.err="fij")
Ppip2<-geeglm(LPpip ~ Dist + Hab + poly(Time,2,raw=TRUE), family=gaussian,
  data=sitel, id=RouteNight, corstr="ar1", std.err="fij")
Ppip3<-geeglm(LPpip ~ Dist + Time, family=gaussian, data=sitel, id=RouteNight,
  corstr="ar1", std.err="fij")
Ppip4<-geeglm(LPpip ~ Dist + poly(Time,2,raw=TRUE), family=gaussian,
  data=sitel, id=RouteNight, corstr="ar1", std.err="fij")
Ppip5<-geeglm(LPpip ~ Dist + Hab, family=gaussian, data=sitel, id=RouteNight,
  corstr="ar1", std.err="fij")
Ppip6<-geeglm(LPpip ~ Dist, family=gaussian, data=sitel, id=RouteNight,
  corstr="ar1", std.err="fij")

#QICu values
print(QIC(Ppip1, Ppip2, Ppip3, Ppip4, Ppip5, Ppip6),digits=5)
#Compare models
anova(Ppip5,Ppip6)
#View model output
summary(Ppip5)
#Save model output
capture.output(summary(Ppip5), file="C:\\Transects\\Summary_Ppip.txt")
#Plot residuals
plot(resid(Ppip5))
plot(fitted(Ppip5), resid(Ppip5))

#Generate table and plot of predictions for distance variable
New_dist<-data.frame(Dist = seq(0,1000, length = 101), Hab = c("5"), Time =
  c(60))
logPpip_predict_dist<-predict(Ppip5,newdata=New_dist,type="response")
Predict_dist_Ppip<-cbind(New_dist,logPpip_predict_dist)
Predict_dist_Ppip
write.xlsx(Predict_dist_Ppip, "C:\\Transects\\Predictions_distance_Ppip.xlsx")
par(mar=c(4,4,4,4.5))
count.overplot(Dist,LPpip, cex.axis=1.2, pch=20, tol=0.2, xlab="Distance (m)",
  ylab="log(bat passes+1)", cex.lab=1.5, font=1, las=2, xaxt="n")
xax<-c(0,200,400,600,800,1000)
axis(1, at=xax, cex.axis=1.2, font=1)
original_scale<-c(0,2,5,10,20,50,100,200,500,1000)

```

```

original_scale_position<-log(original_scale+1)
axis(side=4,at=original_scale_position,labels=original_scale, cex.axis=1.2,las=2)
mtext(side = 4, line = 3.5, "No. bat passes",cex=1.5)
lines(New_dist$Dist, logPpip_predict_dist, lwd=2, lty=1)

#Generate table and plot of predictions for habitat variable
New_hab<-data.frame(Dist=c(500), Hab=c("1","2","3","4","5"), Time=c(60))
logPpip_predict_hab<-predict(Ppip5, newdata=New_hab, type="response")
Predict_hab_Ppip<-cbind(New_hab, logPpip_predict_hab)
Predict_hab_Ppip
write.xlsx(Predict_hab_Ppip, "C:\\Transects\\Predictions_habitat_Ppip.xlsx")
par(mar=c(4,4,4,4.5))
plot(as.numeric(Hab), LPpip, type="n", cex.axis=1.2, xlab="",
      ylab="log(bat passes+1)", cex.lab=1.5, font=1, yaxt="n", xaxt="n")
axis(side=2, cex.axis=1.2, las=2)
xax<-c(1,2,3,4,5)
axis(1,at=xax, labels=c("1","2","3","4","5"), cex.axis=1.5, tck=0.01,lwd=1.5)
mtext(side = 1, line = 2.75, "Habitat grade", cex=1.5)
original_scale<-c(0,2,5,10,20,50,100,200,500,1000)
original_scale_position<-log(original_scale+1)
axis(side=4,at=original_scale_position,labels=original_scale, cex.axis=1.2,las=2)
mtext(side = 4, line = 3.5, "No. bat passes",cex=1.5)
points(New_hab$Hab, logPpip_predict_hab, cex=1.5, pch=18, lwd=2)
lines(New_hab$Hab, logPpip_predict_hab, lty=2, lwd=1)

#Generate table and plot of predictions for time variable
New_time<-data.frame(Dist=c(500), Hab=c("5"), Time=seq(30,160,length=14))
logPpip_predict_time<-predict(Ppip5, newdata=New_time, type="response")
Predict_time_Ppip<-cbind(New_time,logPpip_predict_time)
Predict_time_Ppip
write.xlsx(Predict_time_Ppip, "C:\\Transects\\Predictions_time_Ppip.xlsx")
par(mar=c(4,4,4,4.5))
count.overplot(Time,LPpip, cex.axis=1.2, pch=20,tol=1.5,xlab="Time (min)",
               ylab ="log(bat passes+1)",cex.lab=1.5,font=1,las=2,xaxt="n")
xax<-c(30,60,90,120,150)
axis(1,at=xax, labels=c("30","60","90","120","150"),cex.axis=1.2,font=1)
original_scale<-c(0,2,5,10,20,50,100,200,500,1000)
original_scale_position<-log(original_scale+1)
axis(side=4,at=original_scale_position,labels=original_scale, cex.axis=1.2,las=2)
mtext(side = 4, line = 3.5, "No. bat passes",cex=1.5)
lines(New_time$Time, logPpip_predict_time,lwd=2, lty=1)

#Convert the species variable to proportion data
sitel$Species_fail<-7-sitel$Species
sitel$Sp<-cbind(sitel$Species, sitel$Species_fail)

#Run GEE models for the no. of bat species
Sp1<-geeglm(Sp ~ Dist + Hab + Time, family=binomial, data=sitel, id=RouteNight,
            corstr="ar1", std.err="fij")
Sp2<-geeglm(Sp ~ Dist + Hab + poly(Time,2,raw=TRUE), family= binomial,
            data=sitel, id=RouteNight, corstr="ar1", std.err="fij")
Sp3<-geeglm(Sp ~ Dist + Time, family= binomial, data=sitel, id=RouteNight,
            corstr="ar1", std.err="fij")
Sp4<-geeglm(Sp ~ Dist + poly(Time,2,raw=TRUE), family= binomial, data=sitel,id
            =RouteNight, corstr="ar1", std.err="fij")
Sp5<-geeglm(Sp ~ Dist + Hab, family= binomial, data=sitel, id=RouteNight,
            corstr="ar1", std.err="fij")
Sp6<-geeglm(Sp ~ Dist, family= binomial, data=sitel, id=RouteNight, corstr="ar1",
            std.err="fij")

#QICu values
print(QIC(Sp1, Sp2, Sp3, Sp4, Sp5, Sp6),digits=5)
#Compare models
anova(Sp5,Sp6)

```

```

#View model output
summary(Sp5)
#Save model output
capture.output(summary(Sp5), file="C:\\Transects\\Summary_Species.txt")
#Plot residuals
plot(resid(Sp5))
plot(fitted(Sp5), resid(Sp5))

#Generate table and plot of predictions for distance variable
New_dist<-data.frame(Dist = seq(0,1000, length = 101), Hab = c("5"), Time =
c(60))
Sp_predict_dist<-predict(Sp5,newdata=New_dist,type="response")
Predict_dist_Sp<-cbind(New_dist,Sp_predict_dist)
Predict_dist_Sp
write.xlsx(Predict_dist_Sp, "C:\\Transects\\Predictions_distance_Species.xlsx")
par(mar=c(4,4.1,2,2))
plot(New_dist$Dist, Sp_predict_dist, type="l", lwd=2, ylim=c(0,0.4),
xlab="Distance (m)", ylab="Probability(species)", cex.lab=1.5,
cex.axis=1.2, yaxt="n")
axis(side=2,cex.axis=1.2,las=2)

#Generate table and plot of predictions for habitat variable
New_hab<-data.frame(Dist=c(500), Hab=c("1","2","3","4","5"), Time=c(60))
Sp_predict_hab<-predict(Sp5, newdata=New_hab, type="response")
Predict_hab_Sp<-cbind(New_hab, Sp_predict_hab)
Predict_hab_Sp
write.xlsx(Predict_hab_Sp, "C:\\Transects\\Predictions_habitat_Species.xlsx")
par(mar=c(4,4.1,2,2))
plot(as.numeric(New_hab$Hab), Sp_predict_hab,type="l", lwd=1,lty=2,
ylim=c(0,0.4), xlab="Habitat grade", ylab="Probability(species)",
cex.lab=1.5, cex.axis=1.2, yaxt="n")
axis(side=2, cex.axis=1.2, las=2)
points(New_hab$Hab, Sp_predict_hab, cex=1.5, pch=18)

#Generate table and plot of predictions for time variable
New_time<-data.frame(Dist=c(500), Hab=c("5"), Time=seq(30,160,length=14))
Sp_predict_time<-predict(Sp5, newdata=New_time, type="response")
Predict_time_Sp<-cbind(New_time,Sp_predict_time)
Predict_time_Sp
write.xlsx(Predict_time_Sp, "C:\\Transects\\Predictions_time_Species.xlsx")
par(mar=c(4,4.1,2,2))
plot(New_time$Time, Sp_predict_time, type="l", lwd=2, ylim=c(0,0.4), xlab="Time
(min)", ylab="Probability(species)", cex.lab=1.5, cex.axis=1.2, yaxt="n")
axis(side=2, cex.axis=1.2, las=2)

#detach data and clear console
detach(site1)
rm(list = ls())

```