An operational model predicting autumn bird migration intensities for flight safety
van Belle, J.; Shamoun-Baranes, J.; van Loon, E.E.; Bouten, W.

Published in:
Journal of Applied Ecology

DOI:
10.1111/j.1365-2664.2007.01322.x

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
An operational model predicting autumn bird migration intensities for flight safety

J. V. A N B E L L E , † J. S H A M O U N - B A R A N E S , * E. V A N L O O N * a n d W. B O U T E N *

*Computational Biogeography and Physical Geography, Institute for Biodiversity and Ecosystem Dynamics, Universiteit van Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, the Netherlands; and †Nature Bureau, Support Operations Branch, Directorate of Operations, Royal Netherlands Air Force, PO Box 8762, 4820 BB Breda, the Netherlands

Summary

1. Forecasting migration intensity can improve flight safety and reduce the operational costs of collisions between aircraft and migrating birds. This is particularly true for military training flights, which can be rescheduled if necessary and often take place at low altitudes and during the night. Migration intensity depends strongly on weather conditions but reported effects of weather differ among studies. It is therefore unclear to what extent existing predictive models can be extrapolated to new situations.

2. We used radar measurements of bird densities in the Netherlands to analyse the relationship between weather and nocturnal migration. Using our data, we tested the performance of three regression models that have been developed for other locations in Europe. We developed and validated new models for different combinations of years to test whether regression models can be used to predict migration intensity in independent years. Model performance was assessed by comparing model predictions against benchmark predictions based on measured migration intensity of the previous night and predictions based on a 6-year average trend. We also investigated the effect of the size of the calibration data set on model robustness.

3. All models performed better than the benchmarks, but the mismatch between measurements and predictions was large for existing models. Model performance was best for newly developed regression models. The performance of all models was best at intermediate migration intensities. The performance of our models clearly increased with sample size, up to about 90 nocturnal migration measurements. Significant input variables included seasonal migration trend, wind profit, 24-h trend in barometric pressure and rain.

4. Synthesis and applications. Migration intensities can be forecast with a regression model based on meteorological data. This and other existing models are only valid locally and cannot be extrapolated to new locations. Model development for new locations requires data sets with representative inter- and intraseasonal variability so that cross-validation can be applied effectively. The Royal Netherlands Air Force currently uses the regression model developed in this study to predict migration intensities 3 days ahead. This improves the reliability of migration intensity warnings and allows rescheduling of training flights if needed.

Key-words: bird migration, bird strikes, cross-validation, prediction, radar ornithology, robustness, weather influence

doi: 10.1111/j.1365-2664.2007.01322.x
Predicting bird migration for flight safety

Introduction

Collisions with birds pose serious safety risks for aviation and by 1999 had resulted in the loss of at least 52 civil and 283 military aircraft (Allan 2002). Although such accidents are rare and 65% of the bird strikes do not result in damage at all, the global cost of bird strikes, including the effects of cancelled flights and delays, are estimated at 1–1.5 billion USD year$^{-1}$ (Allan 2002). Most bird strikes take place during take-off and landing, but chances of bird strikes en route increase dramatically during peak periods of bird migration (Blokpoe1 1976; Sodhi 2002; Dekker, van Gasteren & Shamoun-Baranes 2003). For military training flights, which often occur at low altitudes and during nocturnal hours, these increased risks are particularly important; in contrast with civil aviation, rescheduling of military training flights is often possible. Therefore, in order to improve flight safety, several air forces issue bird migration intensity warnings, which result in flight restrictions and occasionally flight cancellations (for example see https://www.notams.jcs.mil/common/birdtam.html, accessed 24 April 2007). Migration intensity warnings for the Royal Netherlands Air Force (RNLAF) are based on radar measurements of bird densities (bird echoes km$^{-2}$) from the north of the Netherlands (Buurma 1995). Spatial extrapolation over the rest of the Netherlands and adjacent countries is based on expert judgement of observed bird flight directions and knowledge of seasonal patterns. Multiday forecasts of migration intensities were not available before this study. The great impact of migration warnings on flight operations, their spatial uncertainties and abrupt temporal changes have prompted the RNLAF to improve the reliability of the warnings and investigate the potential of migration forecasts.

Several migratory flyways pass the Netherlands. In autumn, the most important flyway consists of passerines migrating from Scandinavia towards southern Europe and Africa, crossing the east of the Netherlands in a south-westerly direction (Berthold 2001). Another very important flyway consists of birds from north-east Europe wintering in the British Isles. These birds cross the north of the Netherlands in a west-south-westerly direction and usually follow the western coast of the Netherlands to cross the North Sea further south (Lack 1963). A few times each autumn, massive numbers of migrants cross the Netherlands in a southerly direction. According to their arrival times in the second half of the night, these birds apparently depart from Norway and cross the North Sea in a direct flight, landing in the northern part of the Netherlands (Buurma 1987).

Since the late 1950s, radars have been used to quantify bird migration (Sutter 1957; Buurma 1995; Bruderer 1997; Gauthreaux & Belser 1998) and to model the relationships between bird migration and weather with multivariate statistics (Nisbet & Drury 1968; Alerstam & Ulfstrand 1972; Able 1973). In several countries regression models have been developed to prevent collisions of aircraft with birds (Blokpoe1 1969; Geil et al. 1974) and to assess the impact of wind turbines on migratory bird populations (Gauthreaux & Belser 2003; Barrios & Rodrigues 2004). Wind, rain, sea-level pressure, temperature, cloud cover and derivatives of all of these are the common predictors of migration (Richardson 1990; Zehnder et al. 2001; Erni et al. 2002; Schaub, Liechti & Jenni 2004). However, the relative contributions of these meteorological variables to predictions of migration densities differ strongly between studies. This raises the question: can weather-based models be extrapolated beyond the calibration domain to predict bird migration intensities in new situations?

In order to improve migration warnings for the RNLAF, our first objective was to determine how well existing regression models from other locations on the south-westerly migratory flyway in Europe predict migration intensities over the north of the Netherlands. We tested this by applying three existing models that describe the relationships between weather and migration intensities to Dutch weather data and measured migration densities. Each model consists of input variables and their coefficients. Our second objective was to develop a regression model that would provide more robust and precise nocturnal migratory intensity predictions for the Netherlands than the existing models. In ecological studies focusing on the influence of meteorological conditions on the spatial and temporal distribution of organisms, the variability in the data and the size of the data set are known to have a strong influence on model robustness (Beck 2002). Therefore our third objective was to determine the minimum calibration sample size for creating robust models.

Materials and methods

MIGRATION INTENSITY DATA

The S-band (10-cm wavelength) 20-MW medium-power radar (Thomson ARES, Thomson ICF, Bagneux, France) used in this study is situated in the north-east of the Netherlands and detects birds for up to 150-km distances. The measurement window is located 50–60 km from the radar, 20–60 km from the northern coast of the Netherlands, with its centre at 52°54′N and 6°11′E, and has a total surface of 850 km$^2$. The sampled altitudes in the measurement window range from 100 to 1100 m.

Nocturnal autumn migration intensities were measured from 1 August until 15 November of 2001, 2002 and 2003. During 10 min of each hour, bird echoes were extracted from the radar image with the automated tracking algorithm of ROBIN (radar observations of bird intensities and numbers, developed by TNO Defense, Security and Safety, The Hague, the Netherlands). Migration intensity was defined as the number of bird echoes per km$^2$, automatically corrected for the area with unwanted echoes from land or rain (‘clutter’) within the image (Buurma 1995). The altitude range of the radar beam is generally greater than the upper limit of...
migration, but within the measurement window the altitude distribution cannot be determined. Therefore we expressed migration intensity as echo density per unit area (km$^2$) in the 100–1100 altitude band, rather than per unit volume, as is common in radar ornithology.

Gaps in the data set were generally because of radar downtime, particularly in September 2001, and intense rain clutter. During intense rain, resulting in 100% rain clutter detection, migration intensity was set to 0. In several of the hourly measurements (<0.01% of all measurements), bird echoes were wrongly classified as rain because of the high aerial bird densities (exceeding 50 echoes km$^{-2}$). Migration intensities were then corrected manually from the clutter numbers. Another phenomenon that affected data quality was ‘anomalous propagation’, which occurred when the radar beam was trapped below a low-level temperature inversion layer. As a result, the air was sampled at lower altitudes than usual, leading to larger land clutter areas and wrongly sampled bird densities. Measurements with anomalous propagation were excluded from the analysis.

Before analysis, the radar data were pre-processed. Nights with more than 50% missing data were omitted from further analysis. Nights with complete hourly measurements were used to establish a standard hourly nocturnal trend. Missing values in the data set were then interpolated using the measured hourly intensities of that night and the standard nocturnal trend, adjusted for the length of the night. Finally, the total nocturnal migration intensity used in all subsequent analyses was the sum of hourly nocturnal migration intensities between sunset and sunrise. This resulted in 214 measurements of total nocturnal migration intensity $I_N$ (echoes km$^{-2}$ night$^{-1}$; Fig. 1; yearly summaries Table 1). To meet normality requirements for the regression analyses, migration intensities were transformed to their natural logarithm (Zar 1984).

Mean nocturnal migration intensities, measured during 1989–95 with the same radar, were used to establish a seasonal migration intensity trend. This trend served as a baseline for our models and was used as an independent input variable ($I_b$) in addition to meteorological variables that were expected to explain deviations from the seasonal trend.

### Table 1. Summary of bird measurement data: mean number of nights, means and standard deviations of migration intensities

<table>
<thead>
<tr>
<th>Year</th>
<th>$n$</th>
<th>Mean</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>58</td>
<td>17.71</td>
<td>27.48</td>
</tr>
<tr>
<td>2002</td>
<td>79</td>
<td>19.52</td>
<td>27.08</td>
</tr>
<tr>
<td>2003</td>
<td>77</td>
<td>28.35</td>
<td>46.15</td>
</tr>
</tbody>
</table>

METEOROLOGICAL DATA

Meteorological surface measurements of barometric pressure ($P$), temperature ($T$), wind speed ($S_w$), wind directions ($\alpha_w$) and precipitation ($R$) were taken from the meteorological station Eelde (53°10′N and 6°35′E), 25 km north-east of the centre of the migration measurement window (source Netherlands Meteorological Institute KNMI, De Bilt, the Netherlands). In addition,
Predicting bird migration for flight safety

wind speeds and directions at 1000, 925 and 850 hPa pressure levels (approximately 0, 750 and 1500 m altitude, respectively) were taken from sounding data measured at Emden (53°21’N and 7°13’E), 65 km north-east of the centre of the migration measurement window (source NOAA’s National Oceanic & Atmospheric Administration, Washington, US) radiosonde database; http://raob.fsl.noaa.gov/, accessed 24 April 2007).

All sounding and surface data were collected at 23:00 UTC (Universal Time Code). Additionally, hourly precipitation amounts from Eelde were integrated over the period between sunset and sunrise. Twenty-eight weather variables were tested. These included several derivatives of surface and altitudinal wind speeds and directions, barometric pressure, temperature and precipitation. A description of all input variables that were tested and how they were derived is included in Table 2.

TESTS AND BENCHMARKS

To quantify the prediction performance of models, we used the root mean squared error (RMSE) and compared the 95 percentile range of predictions with the 95 percentile range of measurements. We compared the 95 percentile range because of the technical limitations influencing measurements at extremely low and high migration intensities (see migration intensity data for more details). The RMSE of each model was also compared with two benchmarks: (i) prediction of bird migration by the measured intensity of the previous night (RMSE = 1.41) and (ii) prediction of bird migration intensity by the baseline seasonal trend ($I_n$, RMSE = 1.23).

EXISTING MODELS

To test whether existing models can be extrapolated to new locations, we applied three log-linear regression models from southern Sweden (Zehnder et al. 2001), Denmark (Geil et al. 1974) and southern Germany (Erni et al. 2002) to the Dutch bird migration and meteorology data from 2001 to 2003. These studies explained 0.44 – 0.70 of the variance in local migration intensity or traffic rate.

The model for southern Sweden, $M_S$, is based on nightly means of migration traffic rates (birds km$^{-1}$ h$^{-1}$), obtained with a passive infra red camera in fixed up-right position, recorded from 7 August to 28 October 1998 (Zehnder et al. 2001). This model relates the natural logarithm of nocturnal migration traffic rates, $I_p$, to barometric pressure, $P$ (hPa), 24-h change in barometric pressure, $\Delta P$ (hPa), wind speed, $S_v$ (m s$^{-1}$) and tail wind speed, $W_{\perp 225}$ (m s$^{-1}$):

$$I_p = 2698 - 011S_v + 006W_{\perp 225} - 002P + 003\Delta P$$  

Eqn 1

The model for Denmark, $M_D$, is based on 4 years (1968–71) of radar data from an airfield surveillance radar with a wavelength of 20 cm (Geil et al. 1974). We applied their September model to the Dutch data from September. $M_p$ relates the natural logarithm of migration intensity, $I_p$ (bird echoes km$^{-1}$), in the first half of the night to $P$ (hPa), $\Delta P$ (hPa), temperature $T$ (°C) and lateral wind drift $W_{\perp 200}$ (m s$^{-1}$), defined as the wind speed component perpendicular to a mean flight direction of 200°:

$$I_p = -5012 + 0054P + 0105T + 00291\Delta P + 0469W_{\perp 200}$$  

Eqn 2

The model for Germany, $M_G$, is based on conical scans with a pencil beam of a small-scale 3-cm wavelength tracking radar during one autumn season, 1 August to 29 October 1987 (Erni et al. 2002). For this model, four migration density measurements (bird echoes km$^{-2}$) per night were summed to calculate nocturnal migration intensity. The natural logarithm of migration intensity ($I_p$) is related to unfavourable wind ($W_u$), nightly proportion of hours with rain ($R_{pd}$), the effect of accumulated nights with rain (Acc), and two short-term trend variables that account for the increasing ($I_{\text{inc}}$) and decreasing ($I_{\text{dec}}$) migration intensities at the start and the end of the season:

$$I_p = 5.41 - 0.04I_{\text{inc}} - 0.15I_{\text{dec}} + 0.12W_u - 0.60R_{pd} + 0.13Acc$$  

Eqn 3

Quantifying bird migration intensity with different measurement techniques and for different periods may lead to different absolute numbers. Therefore we corrected each model by subtracting the mean differences between predicted and observed values. Theoretically, these correction factors are multiplicative for non-logarithmic measurements and therefore additive for the logarithmic models. Consequently, the correction does not affect the coefficients that reflect the relative importance of the explanatory variables in the models.

INPUT VARIABLES FOR NEW MODELS

We developed new regression models for the Dutch bird density and weather data to acquire more robust and precise predictions of migration intensity over the Netherlands than existing models yielded. The baseline migration intensity ($I_b$, echoes km$^{-2}$ night$^{-1}$) was included as an input variable to represent the seasonal migration intensity trend (see the Migration intensity data). We analysed 28 weather variables, including those from the existing models. Here we present only the variables included in our final models. Explanatory variables included the total nocturnal rainfall (mm) ($R_c$), the proportion of hours with rain intensity > 0.01 mm h$^{-1}$ between sunset and sunrise ($R_{pd}$), $P$, $T$ and their 24-h increments ($\Delta P$ and $\Delta T$, respectively). We quantified the influence of wind direction and speed on the bird’s flight direction and ground speed by calculating the wind profit, which is the distance the wind would carry
Table 2. Description of input variables tested in predictive models

<table>
<thead>
<tr>
<th>Variable</th>
<th>Name</th>
<th>Units</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T$</td>
<td>Temperature</td>
<td>°C</td>
<td>Surface temperature</td>
</tr>
<tr>
<td>$\Delta T$</td>
<td>24-h temperature increment</td>
<td>°C day$^{-1}$</td>
<td>24-h surface temperature increment</td>
</tr>
<tr>
<td>$\Delta T_{2}$</td>
<td>24-h temperature increment for the second half of the season</td>
<td>°C day$^{-1}$</td>
<td>Until 15 September, value was set to 0</td>
</tr>
<tr>
<td>$T_d$</td>
<td>Deviation from average temperature</td>
<td>°C day$^{-1}$</td>
<td>Deviation from average surface temperature for 1991-2000 at Eelde</td>
</tr>
<tr>
<td>$P$</td>
<td>Barometric pressure</td>
<td>hPa</td>
<td>Sea-level barometric pressure</td>
</tr>
<tr>
<td>$\Delta P$</td>
<td>24-h pressure increment</td>
<td>hPa day$^{-1}$</td>
<td>24-h increment in sea-level barometric pressure</td>
</tr>
<tr>
<td>$S_a$</td>
<td>Surface wind speed</td>
<td>m s$^{-1}$</td>
<td>Surface wind speed</td>
</tr>
<tr>
<td>$A_w$</td>
<td>Surface wind direction</td>
<td>°</td>
<td>Surface wind direction</td>
</tr>
<tr>
<td>$W_{225}$</td>
<td>Tail wind speed (225°)</td>
<td>m s$^{-1}$</td>
<td>Surface wind speed component towards a mean flight direction of 225° (Zehnder et al. 2001)</td>
</tr>
<tr>
<td>$W_{200}$</td>
<td>Lateral wind drift (200°)</td>
<td>m s$^{-1}$</td>
<td>Surface wind speed component perpendicular to a mean flight direction of 200° (Geil et al. 1974)</td>
</tr>
<tr>
<td>$W_{180}$</td>
<td>Lateral wind drift (180°)</td>
<td>m s$^{-1}$</td>
<td>Surface wind speed component perpendicular to a mean flight direction of 180° (Geil et al. 1974)</td>
</tr>
<tr>
<td>$W_u$</td>
<td>Unfavourable wind</td>
<td>–</td>
<td>Categorical variable (0,1) for unfavourable wind, based on wind profit. 1 for wind profits $W_p &lt; \text{threshold}$; the threshold was optimized manually in steps of 0·1 m s$^{-1}$</td>
</tr>
<tr>
<td>$W_u$</td>
<td>Unfavourable wind</td>
<td>–</td>
<td>Categorical variable for unfavourable wind, based on wind profit $W_p$. (Erni et al. 2002). 0 for $W_p = -7$ m s$^{-1}$, 1 for $W_p &lt; -7$ m s$^{-1}$</td>
</tr>
<tr>
<td>$W_p$</td>
<td>Mean wind profit</td>
<td>m s$^{-1}$</td>
<td>Mean of wind profits (equation 4) at 1000, 925 and 850 hPa pressure levels, calculated for a mean bird speed of 12 m s$^{-1}$ and a mean flight direction of 223°</td>
</tr>
<tr>
<td>$W_{p,m}$</td>
<td>Maximum wind profit</td>
<td>m s$^{-1}$</td>
<td>Maximum of wind profits (equation 4) at 1000, 925 and 850 hPa pressure levels, calculated for a mean bird speed of 12 m s$^{-1}$ and a mean flight direction of 223°</td>
</tr>
<tr>
<td>$W_d$</td>
<td>Mean lateral wind drift</td>
<td>m s$^{-1}$</td>
<td>Mean of wind speed components perpendicular to a mean flight direction of 223° and bird speed of 12 m s$^{-1}$ measured at 1000, 925 and 850 hPa pressure levels</td>
</tr>
<tr>
<td>$R_p$</td>
<td>Proportion of hours with rain</td>
<td>–</td>
<td>Proportion of hours with rain intensity &gt; 0·01 mm h$^{-1}$ between sunset and sunrise</td>
</tr>
<tr>
<td>$R_{a0}$</td>
<td>Proportion of hours with rain for $M_a$</td>
<td>–</td>
<td>Proportion of hours with rain defined as 0 to 3 thirds (Erni et al. 2002)</td>
</tr>
<tr>
<td>$R_t$</td>
<td>Total nocturnal rainfall</td>
<td>mm</td>
<td>Sum of hourly nocturnal rainfall between sunset and sunrise</td>
</tr>
<tr>
<td>$R_{m}$</td>
<td>Natural logarithm of total nocturnal rainfall</td>
<td>mm</td>
<td>Natural logarithm of sum of hourly nocturnal rainfall between sunset and sunrise</td>
</tr>
<tr>
<td>$R_{331}$</td>
<td>Precipitation at 23:00 UTC</td>
<td>mm</td>
<td>Precipitation from 22:00 to 23:00 UTC</td>
</tr>
<tr>
<td>$ACC_r$</td>
<td>Accumulation effect of rain</td>
<td>–</td>
<td>Accumulation effect (0–1) as a result of rain (Erni et al. 2002 and equation 5)</td>
</tr>
<tr>
<td>$ACC_w$</td>
<td>Accumulation effect of wind profit</td>
<td>–</td>
<td>Accumulation effect (0–1) as a result of number of nights with unfavourable wind (equation 5)</td>
</tr>
<tr>
<td>$I_b$</td>
<td>Natural logarithm of baseline migration intensity</td>
<td>Bird echoes km$^{-2}$ night$^{-1}$</td>
<td>Natural logarithm of seasonal migration intensity trend for the Netherlands based on measurements from 1989 to 1995</td>
</tr>
<tr>
<td>$I_{Gt}$</td>
<td>Increasing migration trend</td>
<td>–</td>
<td>Increasing migration trend in Germany (Erni et al. 2002)</td>
</tr>
<tr>
<td>$I_{Gd}$</td>
<td>Decreasing migration trend</td>
<td>–</td>
<td>Decreasing migration trend in Germany (Erni et al. 2002)</td>
</tr>
<tr>
<td>$D_{365}$</td>
<td>Day of year</td>
<td>–</td>
<td>Day of year (1–365)</td>
</tr>
<tr>
<td>$I_{b,365}$</td>
<td>Baseline migration intensity</td>
<td>Bird echoes km$^{-2}$ night$^{-1}$</td>
<td>Seasonal migration intensity trend for the Netherlands based on measurements from 1989 to 1995</td>
</tr>
</tbody>
</table>
Predicting bird migration for flight safety

The mean flight direction, \( \alpha_w \), and speed, \( S_w \), were set to 223° and 12 m s\(^{-1} \), respectively. The mean flight direction represents the mean of two predominant migratory directions over the Netherlands mainland based on expert knowledge and radar measurements. Using sounding measurements of wind, mean wind profit, \( W_p \) (m s\(^{-1} \)), was calculated as the mean of all wind profits up to pressure level 850 hPa. Note that \( W_p \) can attain negative values in headwind conditions, with strong lateral wind or with very strong tailwinds. The categorical variable for negative wind profits, \( W_{<0} \), was set to 1 for wind profits lower than a negative threshold; this threshold was optimized manually in steps of 0·1 m s\(^{-1} \). Furthermore, a variable representing the potential accumulation of migrants because of previous nights with unfavourable weather was adopted from Erni et al. (2002), so that:

\[
W_p = S_w - \sqrt{S_w^2 + S_{w,0}^2 - 2 \cdot S_w \cdot S_{w,0} \cdot \cos(\alpha_w - \alpha_{w,0})} 
\]

The mean flight direction, \( \alpha_w \), and speed, \( S_w \), were set to 223° and 12 m s\(^{-1} \), respectively. The mean flight direction represents the mean of two predominant migratory directions over the Netherlands mainland based on expert knowledge and radar measurements. Using sounding measurements of wind, mean wind profit, \( W_p \) (m s\(^{-1} \)), was calculated as the mean of all wind profits up to pressure level 850 hPa. Note that \( W_p \) can attain negative values in headwind conditions, with strong lateral wind or with very strong tailwinds. The categorical variable for negative wind profits, \( W_{<0} \), was set to 1 for wind profits lower than a negative threshold; this threshold was optimized manually in steps of 0·1 m s\(^{-1} \). Furthermore, a variable representing the potential accumulation of migrants because of previous nights with unfavourable weather was adopted from Erni et al. (2002), so that:

\[
Acc_t = \frac{1}{3} \cdot Acc_{t-1} + \frac{2}{3} \cdot U_{t-1} 
\]

where \( Acc \) is the accumulation effect, \( U \) is the variable for unfavourable weather, \( t \) relates to the current night and \( t - 1 \) relates to the previous night. \( U_{t-1} \) was set to 1 if unfavourable weather occurred during the previous night, thus increasing the potential accumulation of migrants (\( Acc_t \)). Independent accumulation variables were derived for the effect of rain (\( Acc_r \)) and the effect of unfavourable wind profit (\( Acc_w \)), based on \( R > 0 \) or \( W_p < W_{<0} \) respectively. All variables were normalized using the z-scores to enable comparison of the relative effects of the variables within each tested model.

MODEL CALIBRATION AND TESTING

The mean and variance of migration intensities (Table 1) did not differ significantly between years (ANOVA, \( P < 0.05 \)). Therefore we treated the data set as homogeneous. Multivariate log-linear regression models were identified and tested by applying cross-validation to different splits of the 3-year data set, following the guidelines for model building as put forward by Hastie, Tibshirani & Friedman (2001), Beck (2002) and Burnham & Anderson (2002). The steps in the cross-validation cycle are outlined as follows. First, the data set was split into a part for model calibration and a part for model validation. Next, all significant models up to a maximum of four explanatory variables were fitted using the calibration data. The four variables were combinations of the 28 available variables listed in Table 2 and 28 randomly permuted variables. Finally, the performance of each significant model was evaluated for the validation data, using the RMSE. The most adequate model was found by selecting those models with the lowest RMSE. In total, six data splits were made using the data in 2001, 2002, 2003, 2001 + 2002, 2001 + 2003 and 2002 + 2003 for calibration, while using the remaining data for validation. The models fitted on these different data splits were named \( M_1, M_2, M_3, M_4, M_5, M_6 \) (etc.) when using the data for 2001, 2002, 2001 + 2002 (etc.) for calibration. By including a randomly permuted dummy variable for each explanatory variable, the above framework automatically checks for random significance. Permutated variables should never be selected as explanatory by this procedure. If any permuted variables are selected, the data set or model framework is invalidated (Good 2000). According to Stone (1977), Shao (1997) and Hastie, Tibshirani & Friedman (2001), leave-one-out cross-validation is equivalent to using an Akaike’s Information Criterion (AIC) criterion in the calibration phase only, and leave-n-out cross-validation is equivalent to a Bayesian information criterion (BIC) criterion. By following the leave-n-out cross-validation procedure in our study as described above, model over-parameterization is prevented and the results are equivalent to the use of the BIC in model calibration.

To investigate the importance of calibration sample size on robustness of the input variables, we used a different approach for splitting data into calibration and validation sets. In this case the models were fitted on randomly selected calibration sets of different sizes: 30, 60, 90 and 120 data points. For each calibration set size, the data set was randomly split 2000 times into calibration and validation sets. Validation sets always contained 80 data points. Regression models were calculated and validated with each randomized set (8000 in total).

RESULTS

MIGRATION INTENSITY PREDICTION BY EXISTING MODELS

Predictions of nocturnal migration intensity by existing models were better than the benchmarks (Table 3) after correction for the different methods of measuring migration intensity. The RMSE for each model equalled about 25% of the range of the measurements. The range of predicted values was much lower than that of the measurements; for example, only 32% of the measured range was predicted by \( M_0 \) (Fig. 2).

MIGRATION INTENSITY PREDICTION BY NEW MODELS

The new regression models contained two to four significant input variables (Table 4): baseline migration
intensity $I_b$, wind profit $W_p$, 24-h trend in barometric pressure $\Delta P$ and nocturnal proportion of hours with rain $R_p$. $I_b$ and $W_p$ were the most important input variables in all the new models; together, they explained at least 40% of the measured variance. $\Delta P$ was a significant variable for all subsets of years except for 2001. $R_p$ was significant in all models based on 2 and 3 years and in $M_2$ but its contribution to the total explained variance was small compared with the contributions of other input variables. The coefficients of the input variables were of similar orders of magnitude between models. Comparison of the normalized coefficients showed that $W_p$ had the largest effect on migration intensity: the contribution of $W_p$ to the predicted values was on average 2·4 times as high as the contribution of $\Delta P$, 3·5 times as high as $R_p$ and 1·6 times as high as $I_b$.

Prediction performances of all new models were much better than the benchmarks and existing models (Table 5 and Fig. 3). Validation RMSEs were generally lower and predicted ranges were much closer to measured ranges (80–84% of the measured range; Fig. 4). Validation RMSEs of models calibrated with 2 years (RMSE 0·81–0·89) were lower, and differences between models were smaller, than for models based on single years (RMSE 0·88–0·96). Following validation of each model with independent years, it was clear that, on average, $M_2$ predictions, were too low for 2001 and 2003.

Effect of sample size on model robustness

Model structure and validation performances were more robust for larger calibration sets (Table 6). Calibrations...
with smaller data sets yielded fewer significant variables and a larger variation in variable significance. Subsamples of the data set, as small as 30 nights only, occasionally even yielded models without $I_1$ or $W_p$, significant variables in all models with larger sample sizes. RMSEs of the validations were lower and more consistent with increasing calibration sample size: the mean RMSE decreased from 1.05 ± 0.19 to 0.89 ± 0.07.

**OPERATIONAL MODEL**

For operational use in the RNLAF, we calibrated our model with variables that were significant for all three models $M_{12}$, $M_{13}$ and $M_{23}$, using the data of all 3 years (204 data points) in order to include as much information as possible. Normalized coefficients were roughly equal to the means of coefficients in $M_{12}$, $M_{13}$ and $M_{23}$. As the aim of this model was to predict actual migration intensity, the model was converted to non-normalized input variables:

$$M_{13} = 124 + 0.84I_1 + 0.154W_p + 0.043\Delta P - 0.79R_p$$  \(\text{eqn 6}\)

This yielded a modelled range of 0.05–4.31, larger than that of any other model developed in this study, and a (calibration) RMSE of only 0.81.
Table 6. Effect of calibration sample size on model structure and performance. Fraction of occurrence of input variables, median number of significant input variables and the means and standard deviations of validation RMSE for sample sizes of 30, 60, 90 and 120 data points. For each sample size, regression models were calibrated with 2000 randomized samples. Validation data sets always consisted of 80 randomly selected data points. Only variables that occurred in more than 10% of the models are shown.

<table>
<thead>
<tr>
<th>Calibration sample size</th>
<th>n = 30</th>
<th>n = 60</th>
<th>n = 90</th>
<th>n = 120</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W_p$</td>
<td>0.88</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$I_b$</td>
<td>0.59</td>
<td>0.94</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\Delta P$</td>
<td>0.18</td>
<td>0.44</td>
<td>0.63</td>
<td>0.80</td>
</tr>
<tr>
<td>$R_p$</td>
<td>0.11</td>
<td>0.19</td>
<td>0.30</td>
<td>0.47</td>
</tr>
<tr>
<td>$R_v$</td>
<td>0.13</td>
<td>0.25</td>
<td>0.25</td>
<td>0.21</td>
</tr>
<tr>
<td>$P$</td>
<td>0.14</td>
<td>0.19</td>
<td>0.24</td>
<td>0.19</td>
</tr>
<tr>
<td>$\Delta T$</td>
<td>0.10</td>
<td>0.18</td>
<td>0.19</td>
<td>0.17</td>
</tr>
<tr>
<td>Accw</td>
<td>0.09</td>
<td>0.13</td>
<td>0.12</td>
<td>0.13</td>
</tr>
<tr>
<td>Number of variables</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>RMSE</td>
<td>1.05 ± 0.19</td>
<td>0.93 ± 0.083</td>
<td>0.90 ± 0.072</td>
<td>0.89 ± 0.069</td>
</tr>
</tbody>
</table>

Discussion

Existing regression models from other locations on the south-westerly migratory flyway predict migration intensities over the north of the Netherlands better than predictions based on migration intensity of the previous night or on the seasonal trend alone. However, new models are more robust and out-perform existing models. This study emphasizes the influence of the variation and size of calibration data sets for the predictive power and robustness of regression models.

Model performance was evaluated by comparing predictions to measured migration intensities. Therefore model results and data quality need to be addressed. In specific conditions, such as high rain clutter, anomalous propagation and very high bird densities, migration intensities were corrected, partly on the basis of expert judgement. Consequently, the quality of the migration intensity measurements is higher in the middle of the range than at the extremes. This is one of the possible causes for the larger discrepancies between observations and predictions at very high and very low migration intensities (Figs 3 and 4).

Prediction performance of new models

As found in other multivariate ecological studies (Whittingham et al. 2006), it is clear that a single year of data is insufficient to identify the best and most robust model. The best newly developed models are based on subsets of 2 years. Model structures and prediction performance of newly developed models based on 2 years are very similar, but one specific year may show lower or higher migration intensities than other years. During validation of 2001, the independent model $M_{23}$ performed even better than the calibrated model $M_t$, indicating that 2001 alone contained too little variation to develop a valuable regression model. On the other hand, $M_t$ fits the calibration data better than $M_{23}$, yet performance of $M_t$ in independent years is relatively poor. These examples nicely demonstrate the trade-off between model fit and general applicability of the models that is inherent to model selection.

The increased variation in model structure and performance with smaller data subsets shows that the calibration sample size has a pronounced effect on robustness of the models. The amount of variation within the calibration set also determines model robustness. For the number of variables and the variation contained in the data for this study, the model performance strongly improves with increasing sample size up to 90 data points and variation in model structure decreases with increasing sample size throughout the test.

Influence of weather

In this study, wind profit had the strongest influence on migration intensity. The large impact of wind on migration intensity is commonly found in the literature (Alerstam 1979; Gauthreaux & Belser 1998; Liechti & Bruderer 1998; Åkesson & Hedenström 2000; Erni et al. 2002). The effect of barometric pressure trends found in this study corresponds with the expected importance of synoptic weather for migration intensity (Richardson 1990). However, in contrast with most theories and studies on the relationships between rain and migration intensity (Richardson 1990; Erni et al. 2002), the effect of tail winds in 34% of all nights. The model for Denmark ($M_d$) only uses the lateral wind drift, whereas the model for Sweden ($M_s$) uses two variables, wind speed and the tail wind component. A second cause for the differences in performance is the use of variables that are not significant for the Dutch data set. For example, $P$ in $M_s$ and $T$ in $M_d$ are not significant after recalibration of the models with the Dutch data sets. Each existing model appears to be site and perhaps even time specific, therefore over-fitting of the models is probably a major cause of the differences in model performance on new data. This is particularly likely for $M_t$ and $M_d$, as these two models are based on single years of measurement and are not validated with independent years.
Predicting bird migration for flight safety

873

The overestimation of low intensities requires further attention, as this may lead to false flight restrictions, which can be very costly.

In addition to flight safety, there are many other sectors that can benefit from migration intensity predictions. For example, migration predictions are of importance for the study and reduction of the impact of wind turbines on bird migration (Drewitt & Langston 2006; Hüppop et al. 2006), outbreaks and the spread of zoonotic pathogens such as West Nile virus and avian influenza (Reed et al. 2003; FAO 2005), and local efforts to reduce the obstruction to migration posed by tall buildings lit at night (e.g. Lights out NY, http://www.nycaudubon.org/ NYCASBirdWatch/safelightupdates, accessed 24 April 2007). The study of migration stopover ecology and particularly the arrival at and departure from stopover sites can greatly benefit from improved understanding of the relationship between weather and migratory activity (Schaub, Liechi & Jenni 2004; Bulyuk & Tsvey 2006). Furthermore, improved knowledge of environmental factors on bird migration is necessary to understand the evolutionary forces behind migration (Pulido, Berthold & van Noordwijk 1996) and the impact of climate change (Sparks 1999).

Extrapolation of regression models to new locations should be discouraged. Clearly, many existing models are only valid locally. Future research will be directed towards radar measurements from adjacent countries to get more insight into the spatial distribution of migrants, the relationships between migration intensities at different locations and the development of spatially explicit dynamic models of migration intensity. Cross-validation should be used to avoid overparameterization in developing models. To make effective use of cross-validation in model development, data sets should represent inter- and intraseasonal variability of the study site well.

Currently, the regression model developed in this study is in operational use in the Royal Netherlands Air Force. Every morning, new predictions of bird migration intensities for 3 days ahead are calculated from the weather conditions forecasted by NOAA. The model thus improves the reliability of the information needed for warnings for high bird migration intensities and, if necessary, rescheduling military flight training in the Netherlands.

Acknowledgements

We would like to thank L. Buurma of the Royal Netherlands Air Force for initiating this research, and H. Van Gasteren, also of RNLAF, for pre-processing all the radar data. Meteorological data were obtained from KNMI and NOAA. Furthermore we would like to thank B. Erni and the Vogelwarte Sempach for their ideas and the use of their original data for comparison, as well as P. Stephens, and two anonymous referees for constructive comments on a previous version of this manuscript.
This work was carried out in the context of the Virtual Laboratory for e-Science project (www.vl-e.nl). This project is supported by a BSIK grant from the Dutch Ministry of Education, Culture and Science (OC and W) and is part of the ICT innovation program of the Ministry of Economic Affairs (EZ).

References


Received 19 July 2006; final copy received 7 March 2007

Editor: Philip Stephens