Prediction of bat presence close to wind sites

Written by Hugo BERNARD
Supervisors: Paul DOUKHAN – Université de Cergy-Pontoise, Yves LE BRAS – BIOTOPE and Lionel TRUQUET - ENSAI

1 – Context of the study

Every year, many bat fatalities occur due to collisions with wind turbines. Being able to predict when bat are active in wind farms is a key to mitigate this threat on their populations.

This internship aims at building a model to predict the presence of bats in wind farms. The predicted probabilities can be used in a constrained dynamic programing algorithm to optimize curtailment strategies. The probability will especially be used in the following equation:

$$E[J_k]=(1-q_k)[(1-F_k(\delta_a))(w_k+\alpha_{i+1,k+1})+F_k(\delta_a)\alpha_{i,k+1}]+q_k(E[w_k]+\alpha_{i,k+1})$$

The idea is to calculate the expectation of future electrical production given curtailement decision at time k given using:

- Presence probability of bat $1-q_k$
- Reward vectors (expected productions revenues) w_k
- Distribution function of the reward vector $F_k(\delta_a)$
- Expected electrical production at step k for a fixed state of the constraint $\alpha_{i,k+1}$ or $\alpha_{i+1,k+1}$

The previous algorithm requires the presence probability of bats in the wind turbines vicinity.

2 – Model

We decided to use a generalized additive model to estimate the bat presence probabilities close to wind sites. The model equation is defined by:

$$logit(P_i)=\beta_0+s(site)+s_1(X_{1,i})+s_2(X_{2,i})+s_3(X_{1,i},X_{2,i})$$

Where:

- P_i presence probability of bat
- s a random effect for each site
- s_1 and s_2 smooth functions
- $X_{1,i}$ day of the year
- $X_{2,i}$ hour of the day
- s_3 an interaction term between hours and days

We chose this model because it does not require the hypothesis of linearity unlike generalized linear model. It is also able to catch cyclic events present in explicative variables. The time variables also partially capture meteorological trends which impact presence probability of bats. To apply this non-parametric method each numerical discrete variable has been divided by the number of distinct values that it has.

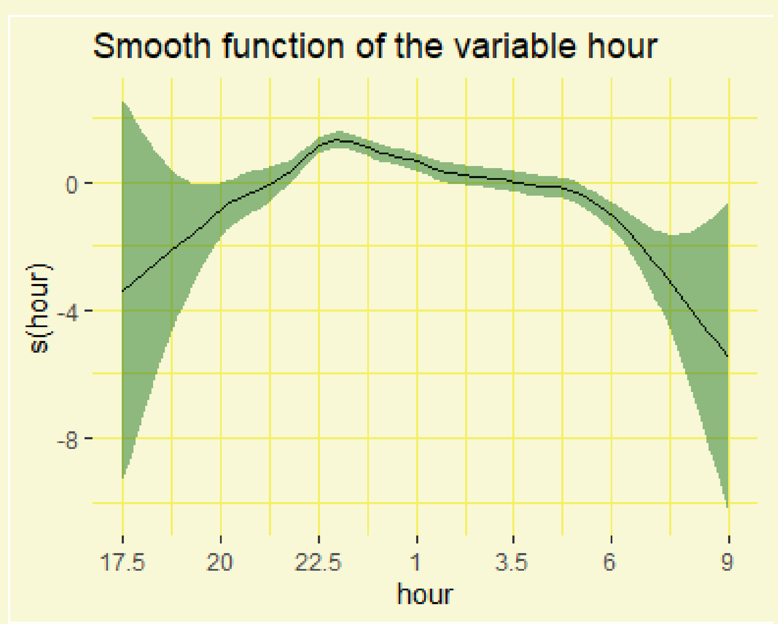
<u>Model assumptions:</u>

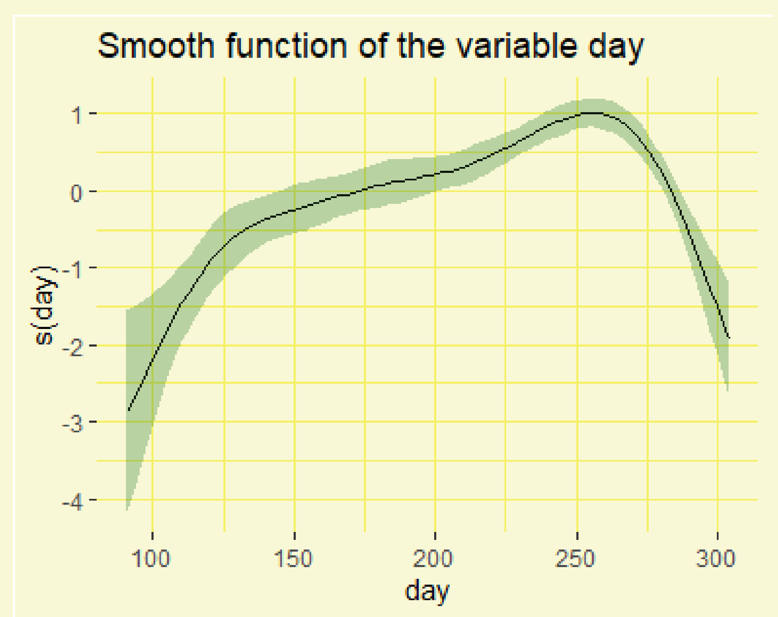
- The explained variable Y_i must be independent
- The explained variable follows a binomial distribution
- Smooth functions must be additive
- Smooth functions must respect spline properties

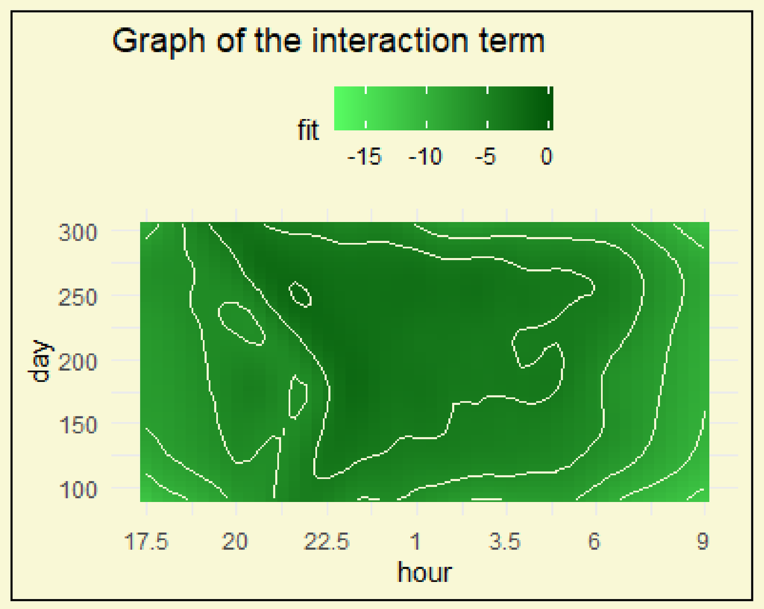
<u>Identifiability constraint:</u>

- For smooth functions: $\sum s(x_j)=0$
- For interaction term: $\begin{cases}\sum\int s(x_1,x_2)dx_1=0\\ \sum\int s(x_1,x_2)dx_2=0\end{cases}$

3 – Results







Smooth functions

The curves of the smooth functions are represented on the left frames. All these functions are used to model presence probability of bats.

We can point out that:

- Splines used are cubic splines
- Functions are not wiggly due to the likelihood penalty.
- Confidence intervals are larger outside of the periods of bat activity

Tiny bit of interpretation : bat activity is maximal in summer, at the beginning of nights. There is an interaction between hour and day : at a given hour the presence probability is different in October and in August. This is due to the day becoming shorter and to differences in meteorological conditions.

Model selection

We performed a model selection by using the Akaike Information Criteria (AIC) and the ROC (Receiver Operator Characteristic). This selection lead us to choose the full model (model 6).

The ROC curve of our model also measures the performance of the model.The curve shape indicates that the model is efficient. The area under the curve of this model is **0.96.**

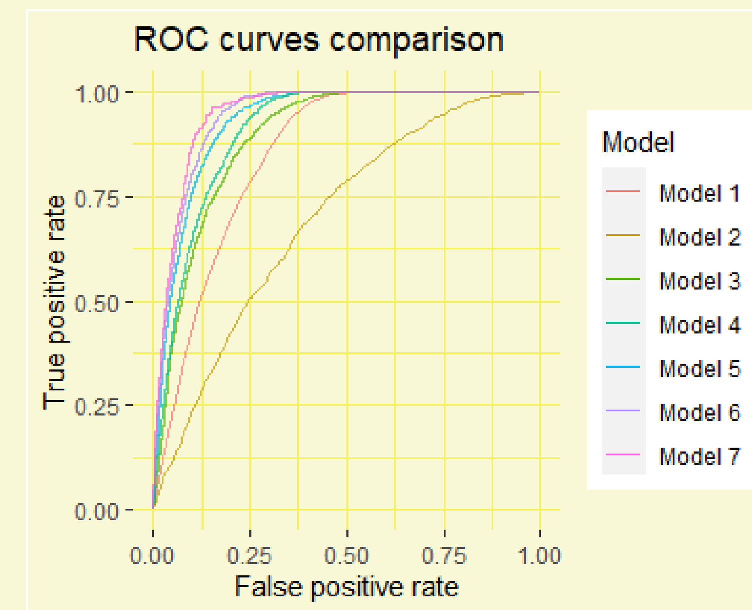
Model adequation

<u>Basis dimension:</u>
We checked that the basis dimension was large enough to build each smooth function.

<u>Study of residuals:</u>
We gathered individual residuals and we plotted the grouped residuals and fitted values. We did not notice any pattern.

<u>Concurvity:</u>
Concurvity happens when two explanatory variables are in relation. The study of this issue showed that there was no concurvity between our explanatory variables.

<u>Serial correlation :</u>
Study of serial correlation revealed an issue which lead us to add lagged dependent variables to the model (model 7).



4 - Conclusion

Generalized additive models allowed us to model bat presence/absence in the wind turbines vicinity. We proceeded to several checks to guarantee the model validity and the model efficiency.
The fitted smooth functions highlight non linear patterns. Despite being very simple, the model achieves predictive performances of 0,96 for the area under the curve (AUC).
Some adjustments might be needed depending on the data available for the wind sites studied.