

Operat losowania powierzchni próbnych dla inwentaryzacji łośia w RDLP Lublin w lutym 2015

Wskazówki techniczne do realizacji prac

Przemysław Chylarecki

Muzeum i Instytut Zoologii PAN; Fundacja Greenmind

Operat losowania i losowanie

Należy zdefiniować (najlepiej w oparciu o siatkę kwadratów 1x1 km narzuconą na obszar RDLP Lublin, bo to ułatwi dalsze analizy) 3 warstwy (rozłączne obszary w granicach RDLP) uzgodnione w protokole ze spotkania 04.02.2015. Warstwy te tworzyły:

- A. nadleśnictwa (12), w których w 2012 i 2013 wykonano pędzenia łośia;
- B. nadleśnictwa, w których wtedy szacowano liczebność łośia w oparciu o całoroczne obserwacje; w tej warstwie do operatu losowania włączane mają być tylko kwadraty 1x1 km w całości zajęte przez lasy (państwowe lub prywatne);
- C. nadleśnictwa "za Sanem".

Dla każdej warstwy (z osobna) należy wykonać poniższe czynności.

1. Należy sporządzić pełną listę bloków sąsiadujących oddziałów/wydziałów, stanowiących potencjalne powierzchnie próbne. Lista ta nazywa się operatem losowania (lub ramką losowania). Dla potrzeb dalszych objaśnień przyjmujemy, że liczy ona N wierszy (=potencjalnych powierzchni próbnych).
2. Bloki sąsiadujących oddziałów będące potencjalnymi powierzchniami próbnymi powinny mieć wielkość zbliżoną do 100 hektarów
3. Dla każdej potencjalnej powierzchni próbnej należy ustalić faktyczną jej wielkość oraz dalsze atrybuty - np. procent powierzchni leśnej, strukturę gatunków panujących oraz strukturę klas wieku. Te atrybuty będą niezbędne na etapie analitycznym.
4. Jako półprodukt potrzebny w dalszych pracach należy wyliczyć faktyczną średnią wielkość powierzchni próbnych (m) w operacie losowania.
5. Należy ustalić docelową liczbę powierzchni koniecznych do wykonania liczeń w danej warstwie. W tym celu należy zsumować powierzchnie wszystkich wydziałów stanowiących operat losowania (A), ustalić 10% tej wielkości ($0.1 * A$), a następnie uzyskać docelową liczbę powierzchni próbnych (K) poprzez:

$$K = 0,1 * (A/m)$$

Dla warstw B i C należy przyjąć 5% łącznej wielkości operatu losowania $K = 0,05 * (A/m)$.

6. Należy wylosować 110% wartości K (czyli: $K1 = 1,1 * K$), co pozwoli na ewentualne usunięcie z próby powierzchni, które są niemożliwe do pędzenia w terenie.
7. Losowanie należy zrealizować w układzie, w którym poszczególne potencjalne powierzchnie próbne stanowią wiersze tabeli, a atrybuty tej powierzchni (jej identyfikator, faktyczna wielkość, atrybuty wieku i składu gatunkowego) będą kolumnami tejże tabeli lub będą powiązane relacyjnie z jej identyfikatorem.
8. Losowanie w tabeli można przeprowadzić w środowisku Excela, w środowisku GIS lub w środowisku R. Sprowadza się ono do losowego wyboru $K1$ wierszy spośród wszystkich N wierszy.
 - w środowisku R użyteczna jest funkcja `sample()`,
 - w Excelu z użyciem funkcji `LOS()` lub `RANDBETWEEN()`,
 - w GIS służą do tego predefiniowane panele w menu. Np. w QGIS jest to "losowy wybór" (=losowanie proste) oraz "losowy wybór w podzbiorach" (=losowanie proste warstwowe).
9. Po wylosowaniu $K1$ wierszy należy przejrzeć ręcznie (lub automatycznie w GIS) wylosowane powierzchnie i oflagować powierzchnie położone bliżej siebie niż 1 km.
10. W parach powierzchni zlokalizowanych bliżej niż 1 km należy odrzucić powierzchnię położoną bardziej na północ; przy jednakowych wartościach położenia na północ odrzucić tę bardziej na wschód (reguły odrzucania przykładowe).
11. Należy odrzucić powierzchnie ewidentnie niemożliwe do pędzenia - przecięte dużą rzeką, drogą wojewódzką, z dużą liczbą upraw grodzonych itd. Liczba odrzuconych powierzchni nie powinna być zbyt duża, tak by zachować do pędzeń zakładane K powierzchni. Jeśli będzie ich nieznacznie mniej lub nieznacznie więcej niż K - nic się nie stanie.

Uwaga. Powyższy schemat zakłada, że należy zdefiniować w GIS wszystkie potencjalne powierzchnie próbne, zaznaczając przyległe oddziały lub wydzielenia dbając przy tym, by ich powierzchnia wynosiła ok. 100 ha. Jest to bardzo duża praca, niemożliwa do zautomatyzowania. Można ją znacznie uprościć stosując trick opisany poniżej. Na całą warstwę (lub lepiej na całą RDLP Lublin) należy narzucić siatkę kwadratów 1x1 km (=100 ha) służącą efektywnie jako maska operatu losowania. Kwadraty należy oznaczyć jednoznacznym identyfikatorem (najlepiej unikatowo powiązany z nadleśnictwem i numerem oddziału pokrywającym się z kwadratem, np. nr-em oddziału którego lewy dolny narożnik znajduje się (a) wewnątrz kwadratu, (b) najbliższej lewego dolnego rogu kwadratu maski). Atrybuty z punktu 3 wyliczyć dla kwadratów 1x1 km. Losowanie przeprowadzić na tabeli kwadratów 1x1 km. Dopiero po losowaniu i wyłącznie dla wylosowanych kwadratów przeprowadzić operację odwzorowania każdego wylosowanego kwadratu na wzorzec realnie istniejących oddziałów i wydzieleń maksymalnie (choć nie perfekcyjnie) pokrywających się z danym kwadratem. W ten sposób wyznaczyć realne powierzchnie próbne (mioty) do pędzeń. Ustalić atrybuty z punktu 3 dla każdego faktycznie zdefiniowanego wydzielenia (powierzchni próbnej). Przeprowadzić pędzenia na tak wyznaczonych powierzchniach próbnych.

Pędzenia na powierzchniach próbnych

Kluczowe znaczenie ma uzgodnienie z wykonawcami liczenia wszelkich sytuacji spornych zaraz po wykonaniu pędzenia. Odpowiednie poprawki są wprowadzane do protokołów. Najlepiej by protokoły liczeń były osobne dla każdego miotu i zbierane były przez kierownika prac po każdym miocie.

Analiza danych

Wyniki liczeń w miotach muszą być wpisywane do tabeli, która służyła od losowania i w której są zapisane atrybuty powierzchni próbnej.

Liczba łosi w danej warstwie nie powinna być obliczana jako średnie zagęszczenie na powierzchnię próbną razy liczba powierzchni próbnych w warstwie. Taki wynik będzie bardzo nieprecyzyjny i będzie ignorował dostępną informację pozwalającą znacząco uściślić oceny całkowite. W szczególności - informację o tym, jakie atrybuty powierzchni próbnej decydują o tym że łosia nie stwierdzono *versus* toś został stwierdzony w pędzeniu. W dalszej kolejności można też wykorzystać atrybuty powierzchni do ustalenia czynników wpływających na liczbę łosi w miotach, gdzie były one stwierdzone. Takie modelowanie liczebności łosi jako funkcji atrybutów powierzchni musi jednak być przeprowadzone z użyciem modeli uwzględniających szczególne właściwości uzyskanych danych. Liczebności uzyskane dla poszczególnych powierzchni nie będą miały rozkładu normalnego (Gausa). Jak wszystkie wyniki zliczeń (tzw. *count data*) powinny mieć rozkład zbliżony do rozkładu Poissona. jednak biorąc pod uwagę niewielkie zagęszczenia łosia, jego skupiskowość oraz relatywnie nieduże powierzchnie próbne, można się też spodziewać, że dane będą charakteryzować się nadmiarową zmiennością w stosunku do danych zgodnych z rozkładem Poissona. W szczególności można oczekiwać, że dane będą zawierały dużo ("za dużo" jak na rozkład Poissona) wartości zerowych. Do takich danych stosuje się specjalne modele, które osobno szacują: (a) prawdopodobieństwo stwierdzenia zera łosi na powierzchni próbnej (jako zmienną dwumianową) oraz (b) liczebność łosi na powierzchniach, na których go stwierdzono (jako zmienną o rozkładzie Poissona, NB lub podobnym). Takie podejście stosował ongiś prof. A. Tomek do wyliczeń zagęszczeń jeleniowatych (nazywając te dwa składniki modelu "funkcją zewnętrzną" i "funkcją wewnętrzną"). Obecnie tego typu modele są rutynowo stosowane w zaawansowanej statystyce ekologicznej jako modele typu "zero-inflated" lub "zero-added"

W sumie zatem należy spróbować dopasować do danych szereg różnych klas modeli różniących się zakładaną strukturą błędu, np.:

- Poisson (PO)
- quasi Poisson
- negative binomial (NB)
- Poisson-inverse gaussian (PIG)
- zero-inflated negative binomial (ZINB)
- zero-inflated Poisson (ZIPO)
- zero-added negative binomial (ZANB)
- zero-added Poisson (ZAPO)
- zero-inflated Poisson inverse gaussian (ZIPIG)

Po ustaleniu najlepszej struktury błędu należy dopasować predyktory (atrybuty powierzchni) dające najlepsze dopasowanie modelu z wybranej wcześniej klasy.

Modele powinny obok predyktorów zawierać również - zawsze - faktyczną wielkość powierzchni próbnej jako tzw. offset.

Przykładowy model - zdefiniowany w środowisku R - może wyglądać tak:

```
modell <- glm (liczba.losi ~ powierzchnia.sosny.3klasa +  
powierzchnia.wszystkich.1klasa +  
offset(log(wielkosc.powierzchni)), family=poisson,  
data=tabela.dane.zliczen)
```

gdzie:

`powierzchnia.sosny.3klasa`, `powierzchnia.wszystkich.1klasa` to przykładowe atrybuty powierzchni użyte jako predyktory w modelu,

Wybór modelu najlepiej dopasowanego do danych (chodzi zarówno o klasę modelu = strukturę błędu, jak i wybór predyktorów) powinien nastąpić z użyciem kryterium AIC (lub AICc).

Modele te powinny być dopasowane w środowisku R z użyciem bibliotek `pscl` i `gamlss`

Po wyborze najlepszego modelu należy użyć go do przewidywania wartości liczebności łosi w każdym kwadracie operatu losowania, z użyciem generycznej funkcji `predict()`. Następnie należy podsumować wartości przewidywane modelem dla każdego kwadratu w operacie losowania. W ten sposób uzyskamy oszacowanie łącznej liczebności łosi w całym operacie losowania danej warstwy (tj. w N kwadratach). Bedzie to oszacowanie uwzględniające zróżnicowane wartości predyktorów w konkretnych kwadratach. Takie oszacowania uzyskane dla każdego kwadratu w operacie mogą być wykorzystane do sumowania w mniejszych wydzieleniach - **np. w skali obwodów łowieckich, nadleśnictw czy OHZ**. Kluczowe znaczenie ma tu wcześniejsze znalezienie wśród atrybutów powierzchni próbnych dobrych predyktorów, które pozwolą z dużym sukcesem (=precyzyjnie) przewidywać liczebność łosi - zarówno w skali całej warstwy, jak i w skali pojedynczego obwodu. Jeśli nie uda się znaleźć dobrych predyktorów, to wyniki będą mało precyzyjne.

Logistyka

- Wykonanie operatu losowania i przeprowadzenie samego losowanie wymaga udziału fachowca z dobrą znajomością programów GIS.
- Przeprowadzenie samego losowania (w GIS lub w Excel lub w środowisku R) powinno odbyć się w obecności Zespołu. Należy również doprowadzić do tego, by lokalizacja wylosowanych powierzchni została pokazana na mapie.
- Wykonanie analizy uzyskanych danych wymaga udziału fachowca z dobrą znajomością środowiska R i bibliotek `plsc` oraz `gamlss` w tymże środowisku.
- Oferuję nieodpłatną pomoc w analizie danych. Przy dobrze przygotowanych danych wejściowych (tabela z wynikami liczeń i atrybutami powierzchni) mogę ją wykonać w kilka godzin w Lublinie w obecności Zespołu (ekran komputera rzutowany na ścianę). Ewentualne

wcześniejsze udostępnienie danych pozwoli mi napisać skrypt analizy zawczasu i skrócić sesję analizy i prezentacji wyników.

Literatura (do analizy danych)

Hilbe J.M. 2014. Modeling Count Data. Cambridge Univ. Press.

Zeileis A., Kleiber C, Jackman S. 2007. Regression Models for Count Data in R. Research Report Series / Department of Statistics and Mathematics, 53. Department of Statistics and Mathematics, WU Vienna University of Economics and Business, Vienna.

Zuur A.F., Hilbe J.M., Ieno E.N. 2013. A Beginner's Guide to GLM and GLMM with R. Highland Statistics, Newburgh.

Zuur A.F., Ieno E.N., Walker N.J., Saveliev A.A., Smith G.M. 2009. Mixed Effects Models and Extensions in Ecology with R. Springer, New York.