Hunting wildlife for disease detection

The R-scripts to perform analysis and output results of the paper:

Atle Mysterud, Petter Hopp, Kristin Ruud Alvseike, Sylvie L. Benestad, Erlend B. Nilsen, Christer M. Rolandsen, Olav Strand, Jørn Våge and Hildegunn Viljugrein: Hunting wildlife for disease detection.

Author of R-scripts: Hildegunn Viljugrein, hildegunn.viljugrein@vetinst.no

A. The procedure to run simulations

The population estimation model (Step 1 in Paper Fig. 2) was run according to https://github.com/ErlendNilsen/CiR_usage, using model variant 2 in the r-script M1_variant_Hardangervidda_fall2018.R at the github-link. The estimated population size, sex and age class structure and demographic rates were saved to be directly read as input data. See «C. Overview of data files» below. Notice that the data set presented here is more updated than the versions used in the submitted manuscript and, as a result, slight deviations from results presented in the paper are expected. The results will be updated in a revised version of the manuscript.

The remaining steps are:

1) run **SimulatePopulationDataNf2.R**

   The main content in this r-script:

   *Read data for the area Nordfjella zone 2 (Nf2). Data are described in section B.

   *The basic functions are sourced. The basic functions are described in section C.

   *The R-script «**Pfreedom_calculations_Nf2.R**" is sourced: estimates diagnostic sensitivity, calculates probability of detecting at least one individual as positive (SSe) given the specific design prevalence and the samples tested, and updates the estimated probability of freedom of CWD infection for each year of surveillance.

   *Apply the function *SimPop18RKtot_h* or *SimPop18Ktot_h* (from **SimPopModelsPaper.R**) through the function *MakeSimPopTabRKh18* (from **SimFunctionsPaper.R**) to simulate a specific harvest strategy and summarize the output of *n* iterations as a list of arrays, some with 3 dimensions: 6 (number of sex and age classes) x P (numer of simulatied years) x nsim (number of simulations). The list of arrays is among others: Htab (number of hunted animals), Ntab (number of preharvest animals) and Xtab (number of postharvest animals).

   For each specific harvest strategy the two R-scripts **"PlotScenarioNf2.R"** and **"MakeSummaryVar.R"** are sourced. If the epidemiological parameters (e.g., relative risk and year-specific design prevalence) are changed from basic level, also **Pfreedom_calculations_Nf2.R** is sourced again.

2) run **SimulatePopulationDataHvidda.R**

   The main content in this r-script:

   *Read data for area Hardangervidda (Hvidda or HV)

   *The basic functions (R-scripts above) are sourced

   *The R-script «**Pfreedom_calculations_HV.R**" is sourced: estimates diagnostic sensitivity, calculates SSe (probability of detecting at least one individual as positive given the specific

design prevalence and the samples tested), and updates the estimated probability of freedom of CWD infection for each year of surveillance.

*Apply the function *SimPop18RKtot_h* or *SimPop18Ktot_h* (from **SimPopModelsPaper.R**) through the function *MakeSimPopTabRKh18* (from **SimFunctionsPaper.R**) to simulate a specific harvest strategy and summarize the output of *n* iterations as a list of arrays, some with 3 dimensions: 6 (number of sex and age classes) x P (numer of simulatied years) x nsim (number of simulations). The list of arrays is among others: Htab (number of hunted animals), Ntab (number of preharvest animals) and Xtab (number of postharvest animals).

For each specific harvest strategy the two R-scripts "**PlotScenarioHV.R**" and "**MakeSummaryVar.R**" are sourced. If the epidemiological parameters (e.g., relative risk and year-specific design prevalence) are changed from basic level, also **Pfreedom_calculations_HV.R** is sourced again.

3) run **SummarizeResTabNf2.R** – Summarize output as shown in Supplementary tables 2 and 3.

4) run **SummarizeResTabHV.R** – Summarize output as shown in Supplementary tables 4 and 5.

5) run **MakeFiguresPaper.R** – Produce the figures in the paper.

B. Overview of data files

Nmat_xx: the estimated (Bayesian population estimation model) pre-harvest population numbers for the two areas (xx=HV or Nf2) summarized as mean of the posterior distribution. There is one row for each year (2015-2019). N0f, N0m, N1f, N1m, Nadf, Nadm and Total denote female and male calves, female and male yearlings, adult females, adult males and total numbers.

Nmatsd_xx: is the standard deviation of the estimated population numbers for the two areas (xx=HV or Nf2), one row for each year (2015-2019).

Demorates_xx: the estimated demographic rates from the Bayesian population estimation model, summarized as mean and standard deviation of the posterior distribution for each of the two areas (xx=HV or Nf2). f: fertility rate of adult females, phi1: summer survival rate of calves, phi3: winter survival rate.

nHarvested_xx: the number of harvested reindeer for the two areas, one row for each year. H0f, H0m, H1f, H1m, Aadf, Hadm denote female and male calves, female and male yearlings, adult females and adult males.

hrates_xx: the proportion of each sex and age class being harvested, one row for each year (2016-2018). H0f, H0m, H1f, H1m, Aadf, Hadm denote female and male calves, female and male yearlings, adult females and adult males.

Stab_xx: The yearly number of samples tested for CWD for the two areas (xx=HV or Nf2), specified for yearlings, adult females (Adf) and adult males (Adm), as well as pRLN, the proportion of samples tested with RLN (lymph nodes) in addition to brain sample. There is one row for each year (2016-2019). For Nordfjella zone 2 there is an extra row representing the extraordinary harvest of adult males during the winter 2018/2019.

C. R-scripts with basic functions and procedures applied for simulations and estimations

"**SimPopModelsPaper.R**": the population simulation model are defined for harvest strategy 1 «ordinary» (*SimPop18Ktot_h*) and for harvest strategy 2 «proactive» with operational sex ratio (functions *SimPop18RKtot_h* or *SimPop18RKtot_hcalf*)

Functions_modelling_dSe1.R: the modelling of how test sensitivity is depending on time since infected (infection stage) and type of tissue tested (Viljugrein et al 2018).

"**Functions_PrDetectingDisease.R**": define functions to estimate surveillance sensitivity (SSe) based on the number and type of samples tested, the design prevalence, the relative risk of infection and the diagnostic sensitivity of the different types of samples tested. Time since infection is randomly drawn from a time line.

simSSe3L: use this function when population sizes are given as mean and standard deviation.

simSSe3LmeanDSE_i: use this function when input data (mod1) is from each of *i* simulations from the population simulation model.

simSSe3L_RRi: same as simSSe3L, bur includes stochasticity in RR as given by a betapert distribution.

simSSe3LmeanDSE_RRi: same as simSSe3LmeanDSE_i, but includes stochasticity in RR

SimFunctionsPaper.R:

defines the function *MakeSimPopTabRKh18* to run the population simulation model (from SimPopModelsPaper.R). The output of *n* iterations is saved as an object *mod1*. The output mod1 is a list of arrays: e.g., Htab (number of hunted animals), Ntab (number of preharvest animals) and Xtab (number of postharvest animals). Htab, Ntab and Xtab are 3-dimentional (6 (number of sex and age classes) x P (number of simulated years) x nsim (number of simulations).

defines the funtion *pfree.estPYR* to update the probability of freedom after each of P years, based on the Surveillance sensitivity (SSe, in the form of a matrix with one row for each of *i* simulations and one column for each year P) and the probability of introduction (vectors of length P)

defines the function *calcPFreedom10meanDSE_i* which utilizes *simSSe3LmeanDSE_i* to calculate surveillance sensitivity (frequency distribution from i iterations) for each of P years for the output of the population simulation (mod1). Similarly, in case of stochastic RR, the function *calcPFreedom10_RRi* utilizes *simSSe3LmeanDSE_RRi* to calculate surveillance sensitivity for each of P years for the output of the population simulation (mod1). The design prevalence is specified as a vector (one value for each of k years). In the next step the function *pfree.estPYR* is used to update the probability of freedom for each of P years. The output of the estimated probability of freedom is also given as a frequency distribution (one estimate for each of i iterations)

defines the function *MakeSummaryPstarScenarioVar* which utilizes *calcPFreedom10* to calculate probability of freedom for the output of the population simulation (mod1) for a range of different design prevalences (set by 2, 4, 6 or 10 infected individuals, as well as for the varying design prevalence as specified for Nordfjella zone 2). Before running the function *MakeSummaryPstarScenarioVar*, the function calcPFreedom10 has to be set equal to the function *calcPFreedom10meanDSE_i or calcPFreedom10_RRi*. The frequency distribution of the estimated probability of freedom is summarized as a table with mean estimates (one row for each design prevalence and one column for each year) and similar tables summarizing the lower (2.5%) and upper (97.5%) percentiles.

defines the plot function *MakeProbPlotNf2* (applied in PlotScenarioNf2.R)

defines the plot function *MakeProbPlot* (applied in PlotScenarioHV.R)

Pfreedom_calculations_Nf2.R and **Pfreedom_calculations_HV.R:** The R-scripts run to calculate estimated probability of freedom for specific design prevalence, relative risk of infection and number and type of samples tested (2016-2019) for the two populations.

Other R-scripts that are used for summarizing and presenting results:

PlotScenarioNf2.R / PlotScenarioHV.R :

The function *assembleDatA* are used on the output from the population simulation model (mod1) to summarize population data in a table dtab.

The function MakeSummaryPstarScenarioVar is utilized to summarize probability of freedom over years with simulated data, given as mean (probfreeS), the lower 2.5th (probfreeSLow) and upper 97.5th (probfreeSHigh) percentiles of the frequency distribution.

The function *MakeProbPlot* is applied to plot the development of estimated probability of freedom over time (probfreeS) from simulated data (2015-2028), showing a separate line for each specified design prevalence (set by 2,4, 6 or 10 infected individuals).

MakeSummaryVar.R : applies the dtab (output from PlotScenarioXX.R) to summarize population number and structure in a table «tabV».