### JOURNAL PUBLICATION CITATION: Sara Stoudt, Perry de Valpine, William Fithian. 2023. Non-parametric identifiability in species distribution and abundance models: why it matters and how to diagnose a lack of it using simulation. *Journal of Statistical Theory and Practice.*

### Data S1

### Identifiability simulation study setup, analysis, and resulting figures for presence-only v. presence absence data for estimating prevalence, single-visit v. double-visit data for estimating occurrence, and single-visit v. double-visit data for estimating abundance

### Sara Stoudt [of the material provided in DataS1.zip]

Sara Stoudt  
Bucknell University

sas072@bucknell.edu

### File list (files found within DataS1.zip)

check\_reproducibility.R

figure1.R

install\_proj\_packages.R

MetadataS1.docx

modified\_source\_code

svisitFormula.R

svocc.fit.R

svocc.R

overall\_code.R

paper\_figures

fig1a.jpeg

fig1b.jpeg

fig1c.jpeg

fig1d.jpeg

fig1e.jpeg

fig1f.jpeg

fig2a.jpeg

fig2b.jpeg

fig2c.jpeg

fig2d.jpeg

fig2e.jpeg

fig2f.jpeg

fig2g.jpeg

fig3a.jpeg

fig3b.jpeg

fig4a.jpeg

fig4b.jpeg

fig4c.jpeg

fig4d.jpeg

fig4e.jpeg

fig4f.jpeg

fig4g.jpeg

fig4h.jpeg

fig4i.jpeg

fig4j.jpeg

fig4k.jpeg

fig4l.jpeg

fig4m.jpeg

fig4n.jpeg

fig4o.jpeg

fig4p.jpeg

presence\_absence\_only

prevalence\_large.R

prevalence\_medium.R

prevalence\_small.R

prevalence\_smallNoSpline.R

raw\_data

popaData\_large.RData

popaData\_medium.RData

popaData\_small.RData

results

paResults\_small.RData

paSplineResults\_largeMoreKnots.RData

paSplineResults\_mediumMoreKnots.RData

paSplineResults\_smallMoreKnots.RData

poResults\_small.RData

poSplineResults\_largeMoreKnots.RData

poSplineResults\_mediumMoreKnots.RData

poSplineResults\_smallMoreKnots.RData

single\_double\_visit

occurDetect\_large.R

occurDetect\_medium.R

occurDetect\_small.R

occurDetect\_small\_noSpline.R

occurDetectDataSetup.R

raw\_data

svdvOccurData\_largeNoSaturateLine.RData

svdvOccurData\_mediumNoSaturateLine.RData

svdvOccurData\_smallNoSaturateLine.RData

results

dvOccurResults\_smallNoSaturateLine.RData

dvOccurSplineResults\_largeNoSaturateLine.RData

dvOccurSplineResults\_mediumNoSaturateLine.RData

dvOccurSplineResults\_smallNoSaturateLine.RData

svOccurResults\_smallNoSaturateLine.RData

svOccurSplineResults\_largeNoSaturateLine.RData

svOccurSplineResults\_mediumNoSaturateLine.RData

svOccurSplineResults\_smallNoSaturateLine.RData

single\_double\_visit\_abundance

abunDetect\_large.R

abunDetect\_medium.R

abunDetect\_small.R

abunDetect\_small\_noSpline.R

abunDetectDataSetup.R

raw\_data

abunData\_large.RData

abunData\_medium.RData

abunData\_small.RData

results

dvAbundanceLarge.RData

dvAbundanceMedium.RData

dvAbundanceSmall\_noSpline.RData

dvAbundanceSmall.RData

svAbundanceSmall.RData

svAbundanceSmall\_noSpline.RData

svAbundanceMedium.RData

svAbundanceLarge.RData

**Description**

check\_reproducibility.R – reproducibility checks using fertile package

figure1.R – all code needed to produce Figure 1 in the manuscript

install\_proj\_packages.R – code to install all required packages needed to run the code in this repository, recommended to run as the first step, this file also notes package and R versions

MetadataS1.docx – description of files in supplementary code

modified\_source\_code

svisitFormula.R – tweaked code for svisitFormula function found in the detect package for extra tracking

svocc.fit.R - tweaked code for svocc.fit function found in the detect package for extra tracking

svocc.R - tweaked code for svocc function found in the detect package for extra tracking

overall\_code.R- high level script that calls helper scripts to execute all of the analysis in the manuscript and supplement and make the remaining plots, run after installing the requisite packages

paper\_figures – figures in the manuscript

fig1a.jpeg

fig1b.jpeg

fig1c.jpeg

fig1d.jpeg

fig1e.jpeg

fig1f.jpeg

fig2a.jpeg

fig2b.jpeg

fig2c.jpeg

fig2d.jpeg

fig2e.jpeg

fig2f.jpeg

fig2g.jpeg

fig3a.jpeg

fig3b.jpeg

fig4a.jpeg

fig4b.jpeg

fig4c.jpeg

fig4d.jpeg

fig4e.jpeg

fig4f.jpeg

fig4g.jpeg

fig4h.jpeg

fig4i.jpeg

fig4j.jpeg

fig4k.jpeg

fig4l.jpeg

fig4m.jpeg

fig4n.jpeg

fig4o.jpeg

fig4p.jpeg

presence\_absence\_only

prevalence\_large.R – code to fit presence-only and presence-absence models for large sample size, make corresponding plots

prevalence\_medium.R - code to fit presence-only and presence-absence models for medium sample size, make corresponding plots

prevalence\_small.R - code to fit presence-only and presence-absence models for small sample size, make corresponding plots

prevalence\_smallNoSpline.R- code to fit presence-only and presence-absence models for small sample size without spline, make corresponding plots

prevalenceExampleDataSetup.R – get simulated presence-only and presence-absence data to fit models to

raw\_data

popaData\_large.RData – large presence-only and presence-absence data generated in prevalenceExampleDataSetup.R

popaData\_medium.RData – medium presence-only and presence-absence data generated in prevalenceExampleDataSetup.R

popaData\_small.RData - – small presence-only and presence-absence data generated in prevalenceExampleDataSetup.R

results

paResults\_small.RData – presence-absence results for small data without splines

paSplineResults\_largeMoreKnots.RData – presence-absence results for large data (not stored on GitHub because too large, generate yourself using the provided code)

paSplineResults\_mediumMoreKnots.RData - presence-absence results for medium data

paSplineResults\_smallMoreKnots.RData - presence-absence results for small data with splines

poResults\_small.RData - presence-only results for small data without splines

poSplineResults\_largeMoreKnots.RData - presence-only results for large data (not stored on GitHub because too large, generate yourself using the provided code)

poSplineResults\_mediumMoreKnots.RData - presence-only results for medium data

poSplineResults\_smallMoreKnots.RData - presence-only r esults for small data with splines

single\_double\_visit

occurDetect\_large.R - code to fit single-visit and double-visit models for large sample size, make corresponding plots

occurDetect\_medium.R - code to fit single-visit and double-visit models for medium sample size, make corresponding plots

occurDetect\_small.R - code to fit single-visit and double-visit models for small sample size, make corresponding plots

occurDetect\_small\_noSpline.R - code to fit single-visit and double-visit models for small sample size without spline, make corresponding plots

occurDetectDataSetup.R - get simulated single-visit and double-visit data to fit models to

raw\_data

svdvOccurData\_largeNoSaturateLine.RData - large single-visit and double-visit data generated in occurDetectDataSetup.R (not stored on GitHub because too large, generate yourself using the provided code)

svdvOccurData\_mediumNoSaturateLine.RData- medium single-visit and double-visit data generated in occurDetectDataSetup.R

svdvOccurData\_smallNoSaturateLine.RData- small single-visit and double-visit data generated in occurDetectDataSetup.R

results

dvOccurResults\_smallNoSaturateLine.RData – double-visit results for small data without splines

dvOccurSplineResults\_largeNoSaturateLine.RData - double-visit results for large data (not stored on GitHub because too large, generate yourself using the provided code)

dvOccurSplineResults\_mediumNoSaturateLine.RData - double-visit results for medium data

dvOccurSplineResults\_smallNoSaturateLine.RData –

double-visit results for small data with splines

svOccurResults\_smallNoSaturateLine.RData - single-visit results for small data without splines

svOccurSplineResults\_largeNoSaturateLine.RData - single-visit results for large data (not stored on GitHub because too large, generate yourself using the provided code)

svOccurSplineResults\_mediumNoSaturateLine.RData – single-visit results for medium data

svOccurSplineResults\_smallNoSaturateLine.RData - single-visit results for small data with splines

single\_double\_visit\_abundance

abunDetect\_large.R - code to fit single-visit and double-visit abundance models for large sample size, make corresponding plots

abunDetect\_medium.R - code to fit single-visit and double-visit abundance models for medium sample size, make corresponding plots

abunDetect\_small.R - code to fit single-visit and double-visit abundance models for small sample size, make corresponding plots

abunDetect\_small\_noSpline.R - code to fit single-visit and double-visit abundance models for small sample size without spline, make corresponding plots

abunDetectDataSetup.R - get simulated single-visit and double-visit abundance data to fit models to

raw\_data

abunData\_large.RData - large single-visit and double-visit abundance data generated in abunDetectDataSetup.R (not stored on GitHub because too large, generate yourself using the provided code)

abunData\_medium.RData - medium single-visit and double-visit abundance data generated in abunDetectDataSetup.R

abunData\_small.RData - small single-visit and double-visit abundance data generated in abunDetectDataSetup.R

results

dvAbundanceLarge.RData – double-visit results for large data (not stored on GitHub because too large, generate yourself using the provided code)

dvAbundanceMedium.RData - double-visit results for medium data

dvAbundanceSmall\_noSpline.RData - double-visit results for small data without splines

dvAbundanceSmall.RData - double-visit results for small data with splines

svAbundanceLarge.RData - single-visit results for large data (not stored on GitHub because too large, generate yourself using the provided code)

svAbundanceMedium.RData - single-visit results for medium data

svAbundanceSmall\_noSpline.RData - single-visit results for small data without splines

svAbundanceSmall.RData - single-visit results for small data with splines