

1 Trend analysis for USFWS species status assessment for
2 bull trout (*Salvelinus confluentus*)

3 Mark D. Scheuerell^{1,*}

4 ¹ U.S. Geological Survey Washington Cooperative Fish and Wildlife Research Unit, School
5 of Aquatic and Fishery Sciences, University of Washington, Seattle, WA

6 * Correspondence: [Mark D. Scheuerell <scheuerl@uw.edu>](mailto:scheuerl@uw.edu)

7 This draft is distributed solely for purposes of internal review. Its content is deliberative and
8 predecisional, so it must not be disclosed or released. Because it has not yet been approved
9 for publication by the U.S. Geological Survey (USGS), it does not represent any official
10 USGS finding or policy.

Background

Bull trout (*Salvelinus confluentus*) in the western U.S. were listed as threatened under the U.S. Endangered Species Act in 1998. The purpose of this analysis is to estimate trends in the abundance (counts) of bull trout within predefined core areas spread across Oregon, Washington, Idaho, and Montana, as part of the current Species Status Assessment (SSA).

Data

I provided a MS Excel template for the desired data in a “tidy” format, which consisted of the following 8 fields (columns):

- **dataset** (i.e., integer value for unique ID)
- **recovery unit** (e.g., Mid Columbia)
- **core area** (e.g., South Fork Clearwater)
- **popn/stream** (e.g., Crooked River)
- **metric** (e.g., abundance)
- **method** (e.g., redd survey)
- **year**
- **value** (i.e., counts)

The data coordinators also provided me with some metadata indicating which of the data specific to a location were generally for adults versus juveniles. Data files were subsequently provided to me by data coordinators from each of the four states:

- Oregon: Stephanie Gunckel (ODFW)
- Washington: Marie Winkowski (WDFW)
- Idaho: Brett Bowersox (IDFG)
- Montana: Dan Brewer (USFWS)

The data from Montana came via biologists with the USGS (Clint Muhlfeld, Tim Cline), and did not conform to the template file I had provided. Thus, those data were subjected to additional cleaning prior to their inclusion with the data from other states (see below).

Prior to analysis, all of the count data were log-transformed and standardized to a mean of zero and variance of one, which aided in model fitting and, in particular, the estimation of the process and observation (sampling) variances.

Modeling framework

Population model

I fit discrete-time versions of exponential models for population growth (decline), such that the abundance of bull trout (N) is a function of the initial population size N_0 , time (t), the population growth rate (u), and a time-dependent stochastic effect of the environment (w). Specifically, in continuous time the model is

$$N(t) = N_0 \exp(u) \exp(wt). \quad (1)$$

In discrete time, with a time step of 1 unit (e.g., a year), the model becomes

$$N_t = N_{t-1} \exp(u + w_t). \quad (2)$$

If we take the logarithm of both sides and define $x_t = \log(N_t)$, we have

$$x_t = x_{t-1} + u + w_t. \quad (3)$$

Further defining $w_t \sim N(0, q)$ leads us to a so-called “biased random walk” model, where u is the tendency for the population to increase or decrease each time step (i.e., the bias), and w_t is some unknown stochastic aspect of the environment that partially drives population dynamics.

Observation model

The data available to us rarely come from complete censuses, and instead are typically derived from partial counts. Furthermore, mistakes may occur when counting individuals or redds. Thus, we should account for these possible sampling or observation errors with a so-called “data model.”

In this case, we assume that the data in hand are a somewhat distorted view of the “true state of nature,” such that the logarithm of the observed count at time t (y_t) equals that of the true count plus or minus some error. Specifically, we can write this as

$$y_t = x_t + a + v_t \quad (4)$$

where a is an offset to account for partial sampling, and $v_t \sim N(0, r)$.

61 State-space model

62 We can combine equations (3) and (4), along with a definition for the initial state (x_0), into
 63 a so-called “state-space model,” where

$$\begin{aligned} y_t &= x_t + a + v_t \\ x_t &= x_{t-1} + u + w_t \\ x_0 &\sim N(\mu, \sigma) \end{aligned} \tag{5}$$

64 Multiple populations

65 Here we want to estimate the annual change in population size for each of the many different
 66 core areas across the four states. Furthermore, some core areas comprise several different
 67 populations/locations, so we need to frame our state-space model in a multivariate context.

68 Observation model

69 If we have n different populations within a core area, then our observation model becomes

$$y_{i,t} = x_{i,t} + a_i + v_{i,t} \tag{6}$$

70 where $y_{i,t}$ is the log-count for population i and year t , a_i is an offset to account for partial
 71 sampling in population i , and $v_{i,t} \sim N(0, r_i)$ ¹. We can combine each of the population specific
 72 observation models into a matrix form, such that

$$\begin{aligned} y_{1,t} &= x_{1,t} + a_1 + v_{1,t} \\ y_{2,t} &= x_{2,t} + a_2 + v_{2,t} \\ &\vdots \\ y_{n,t} &= x_{n,t} + a_n + v_{n,t} \end{aligned} \tag{7}$$

73 becomes

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_t = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}_t + \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}_t, \tag{8}$$

¹Here the variance of the observation errors is assumed to be population specific, but it might be reasonable to assume that each survey/census type might have the same variance, such that $v_{i,t} \sim N(0, r)$.

74 or more compactly in matrix notation as

$$\mathbf{y}_t = \mathbf{x}_t + \mathbf{a} + \mathbf{v}_t. \quad (9)$$

75 where \mathbf{y}_t , \mathbf{x}_t , \mathbf{a} , \mathbf{v}_t are all $n \times 1$ vectors, and $\mathbf{w}_t \sim \text{MVN}(\mathbf{0}, \mathbf{Q})$.

76 **Population model**

77 Just as we did for the observation model, we can write the models for population dynamics
78 as

$$x_{j,t} = x_{j,t-1} + u_j + w_{j,t} \quad (10)$$

79 where u_i is the bias, which is unique to each population², and $w_{j,t} \sim \text{N}(0, q)\}$.

80 We can again express all of the population models in matrix form, such that

$$\begin{aligned} x_{1,t} &= x_{1,t-1} + u_1 + w_{1,t} \\ x_{2,t} &= x_{2,t-1} + u_2 + w_{2,t} \\ &\vdots \\ x_{n,t} &= x_{n,t-1} + u_3 + w_{n,t} \end{aligned} \quad (11)$$

81 becomes

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}_t = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}_t + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}_t, \quad (12)$$

82 or more compactly in matrix notation as

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{u} + \mathbf{w}_t \quad (13)$$

83 where \mathbf{x}_t , \mathbf{x}_{t-1} , \mathbf{u} , \mathbf{w}_t are all $n \times 1$ vectors, and $\mathbf{w}_t \sim \text{MVN}(\mathbf{0}, \mathbf{Q})$.

²It might be reasonable to assume that some/all of the populations have the same bias, given their membership within a core area.

State-space forms

At this point, however, we are assuming that the monitoring data for each population is telling us something about only the specific population itself, rather than contributing information to the population trend at the larger scale of their core area, which is the really the scale of interest here. Thus, we need to modify our equations to accommodate this hierarchical framework.

For example, assume that we have $p = 2$ core areas (call them A and B), each with data from 2 representative populations. In this case, $n = 4$, but the number of states (i.e., the number of rows in \mathbf{x}_t) is 2, so we need a way to “map” each of the observed time series onto its respective core area. We begin by writing out the equations for the observations in long matrix form akin to equation (8), such that

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}_t = \begin{bmatrix} x_A \\ x_A \\ x_B \\ x_B \end{bmatrix}_t + \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix}_t, \quad (14)$$

Because both x_A and x_B appear twice in equation (14), we can use a 4×2 matrix of 1’s and 0’s as our map. Specifically, we have

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}_t = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_A \\ x_B \end{bmatrix}_t + \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix}_t, \quad (15)$$

We can write equation (15) more compactly in matrix notation as

$$\mathbf{y}_t = \mathbf{Z}\mathbf{x}_t + \mathbf{a} + \mathbf{v}_t. \quad (16)$$

where \mathbf{y}_t , \mathbf{a} , and \mathbf{v}_t are all $n \times 1$ vectors, \mathbf{Z} is an $n \times k$ matrix, and \mathbf{x}_t is a $k \times 1$ vector.

The equation for the population dynamics in each of the 2 core areas then becomes

$$\begin{bmatrix} x_A \\ x_B \end{bmatrix}_t = \begin{bmatrix} x_A \\ x_B \end{bmatrix}_{t-1} + \begin{bmatrix} u_A \\ u_B \end{bmatrix} + \begin{bmatrix} w_A \\ w_B \end{bmatrix}_t, \quad (17)$$

which can be written more compactly in matrix notation as

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{u} + \mathbf{w}_t, \quad (18)$$

101 and combined with equation (16) to form the full multivariate state-space model

$$\begin{aligned}\mathbf{y}_t &= \mathbf{Z}\mathbf{x}_t + \mathbf{a} + \mathbf{v}_t \\ \mathbf{x}_t &= \mathbf{x}_{t-1} + \mathbf{u} + \mathbf{w}_t.\end{aligned}\tag{19}$$

102 Thus, by simply altering the dimensions of \mathbf{Z} , and the locations of 1's and 0's within it, we
103 can evaluate any number of different hypotheses about how the population dynamics are
104 structured spatially. For example, if we set \mathbf{Z} equal to an $n \times n$ identity matrix, where

$$\mathbf{Z} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix},\tag{20}$$

105 then each of the time series of data is assumed to represent a unique state of nature. If, on
106 the other hand, we set \mathbf{Z} equal to an $n \times 1$ column vector of 1's, such that

$$\mathbf{Z} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix},\tag{21}$$

107 then each of the time series of data is assumed to represent a sample from a single state of
108 nature.

109 Variance specification

110 The multivariate state-space model allows us to be quite specific about how the observation
111 errors (\mathbf{v}_t) and process errors (\mathbf{w}_t) are related to one another, if at all. In the most simple
112 case, the errors could be independent and identically distributed (IID), such that (for the
113 observation variance)

$$\mathbf{R} = \begin{bmatrix} r & 0 & \cdots & 0 \\ 0 & r & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r \end{bmatrix}.\tag{22}$$

114 Alternatively, the errors might be independent, but not identically distributed

$$\mathbf{R} = \begin{bmatrix} r_1 & 0 & \cdots & 0 \\ 0 & r_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_n \end{bmatrix}, \quad (23)$$

or identically distributed, but not independent

$$\mathbf{R} = \begin{bmatrix} r & c & \cdots & c \\ c & r & \cdots & c \\ \vdots & \vdots & \ddots & \vdots \\ c & c & \cdots & r \end{bmatrix}. \quad (24)$$

Model fitting

The adult data come from 242 populations within 62 core area spread across the 6 recovery domains. Thus, \mathbf{y}_t is a 242×1 vector of transformed observed counts in year t , and \mathbf{x}_t is a 62×1 vector of underlying states in year t . The matrix \mathbf{Z} is a 242×62 matrix of 1's and 0's that maps each of the respective observations onto their respective states.

Because the source of the data varies by location, I estimated different variance terms for each of the methods (e.g., screw trap, snorkel survey, weir). Specifically, the covariance matrix is given by

$$\mathbf{R} = \begin{bmatrix} r_i & 0 & \cdots & 0 \\ 0 & r_i & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_i \end{bmatrix}. \quad (25)$$

where i indicates the source of the data.

In this particular case, the data were standardized prior to model fitting, so I assumed that the process errors ($w_{j,t}$) were independent and identically distributed, such that

$$\mathbf{Q} = \begin{bmatrix} q & 0 & \cdots & 0 \\ 0 & q & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & q \end{bmatrix}. \quad (26)$$

All models were fit using the `{MARSS}` package (Holmes *et al.* 2012; 2021b, a) for the \mathbf{R} computing software (R Core Team 2020). I estimated 90% confidence intervals (CI's) around

each of the estimated bias terms (u_j) using the `MARSSparamCIs()` function based upon an asymptotic variance assumption. All of the data and code necessary to reproduce the results of the analysis can be found online at <https://github.com/mdscheuerell/bulltrout>.

Results

Estimated trends in abundance were rather mixed (Tables 1-4), such that 37 out of 62 core areas showed decreasing trends, but only 3 of trends had 90% CI's that did not overlap 0. The remaining 25 core areas had decreasing trends, but only 5 of them had 90% CI's that did not overlap 0. Fitted trend lines and their associated CI's are provided at the end of this report.

References

- Holmes EE, Scheuerell MD, and Ward EJ. 2021a. Analysis of multivariate time-series using the MARSS package. Version 3.11.4.
- Holmes EE, Ward EJ, Scheuerell MD, and Wills K. 2021b. MARSS: Multivariate autoregressive state-space modeling.
- Holmes EE, Ward EJ, and Wills K. 2012. MARSS: Multivariate autoregressive state-space models for analyzing time-series data. *The R Journal* **4**: 30.
- R Core Team. 2020. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Table 1: Estimated trends and uncertainty for core areas in Montana. The trend column indicates if the estimated trend was significantly positive (+), negative (-), or neutral (0).

| Core area | Lower CI | Mean | Upper CI | Trend |
|----------------------------|----------|---------|----------|-------|
| CANADA | 0.0139 | 0.0864 | 0.1590 | + |
| Big Salmon Lake | -0.1507 | -0.0651 | 0.0206 | 0 |
| Bitterroot River | -0.0748 | -0.0021 | 0.0706 | 0 |
| Blackfoot River | -0.0863 | -0.0216 | 0.0431 | 0 |
| Bowman | -0.0355 | 0.1017 | 0.2390 | 0 |
| Bull Lake | -0.1479 | -0.0730 | 0.0019 | 0 |
| Cerulean/Quartz Mid Lakes | -0.0796 | 0.0278 | 0.1352 | 0 |
| Clearwater | -0.1364 | -0.0439 | 0.0487 | 0 |
| Cyclone Lake | -0.1255 | -0.0464 | 0.0327 | 0 |
| Flathead Lake | -0.0792 | -0.0165 | 0.0463 | 0 |
| Harrison Lake | -0.2067 | -0.0785 | 0.0498 | 0 |
| Holland Lake | -0.1207 | -0.0494 | 0.0220 | 0 |
| Hungry Horse Reservoir | -0.0505 | 0.0161 | 0.0826 | 0 |
| Kootenai River | -0.0884 | -0.0240 | 0.0405 | 0 |
| Lake Koocanusa | -0.1008 | -0.0321 | 0.0366 | 0 |
| Lake Pend Oreille | -0.0800 | -0.0146 | 0.0509 | 0 |
| Lindbergh Lake | -0.1436 | -0.0583 | 0.0271 | 0 |
| Logging | -0.2333 | -0.1196 | -0.0059 | - |
| Lower Quartz Lake | -0.1355 | -0.0182 | 0.0992 | 0 |
| Middle Clark Fork River | -0.1096 | -0.0277 | 0.0541 | 0 |
| Rock Creek | -0.1020 | -0.0352 | 0.0317 | 0 |
| Swan Lake | -0.0829 | -0.0198 | 0.0434 | 0 |
| Upper Clark Fork River | -0.0671 | 0.0077 | 0.0825 | 0 |
| Upper Stillwater Lake | -0.0078 | 0.0657 | 0.1392 | 0 |
| Upper Whitefish Lake | 0.0118 | 0.0908 | 0.1698 | + |
| West Fork Bitterroot River | -0.0738 | 0.0057 | 0.0852 | 0 |
| Whitefish Lake | -0.0374 | 0.0376 | 0.1125 | 0 |
| St. Mary River | -0.1054 | -0.0216 | 0.0622 | 0 |

Table 2: Estimated trends and uncertainty for core areas in Idaho. The trend column indicates if the estimated trend was significantly positive (+), negative (-), or neutral (0).

| Core area | Lower CI | Mean | Upper CI | Trend |
|---------------------------|----------|---------|----------|-------|
| Couer d’Alene | -0.1028 | -0.0356 | 0.0316 | 0 |
| Kootenai | -0.2246 | -0.1210 | -0.0175 | - |
| LPO | -0.0565 | 0.0069 | 0.0704 | 0 |
| Priest Lakes | -0.0617 | 0.0062 | 0.0741 | 0 |
| North Fork Clearwater | -0.1874 | -0.0869 | 0.0136 | 0 |
| Lemhi River | -0.1397 | -0.0224 | 0.0948 | 0 |
| Little-Lower Salmon River | -0.0111 | 0.0614 | 0.1339 | 0 |
| South Fork Salmon River | -0.0626 | 0.0337 | 0.1299 | 0 |
| Upper Salmon | -0.0489 | 0.0185 | 0.0858 | 0 |

Table 3: Estimated trends and uncertainty for core areas in Washington. The trend column indicates if the estimated trend was significantly positive (+), negative (-), or neutral (0).

| Core area | Lower CI | Mean | Upper CI | Trend |
|--------------------------------|----------|---------|----------|-------|
| Chester Morse Lake | -0.1567 | -0.0611 | 0.0346 | 0 |
| Lewis River | 0.0125 | 0.0821 | 0.1517 | + |
| Lower Skagit River | -0.0864 | -0.0166 | 0.0533 | 0 |
| Puyallup River | -0.0222 | 0.0447 | 0.1116 | 0 |
| Skokomish River | -0.0889 | -0.0055 | 0.0779 | 0 |
| Snohomish and Skykomish Rivers | -0.0783 | -0.0062 | 0.0659 | 0 |
| Stillaguamish River | -0.2515 | -0.1447 | -0.0379 | - |
| Asotin Creek | -0.1385 | -0.0276 | 0.0832 | 0 |
| Entiat River | -0.0400 | 0.0305 | 0.1010 | 0 |
| Methow River | -0.0314 | 0.0432 | 0.1178 | 0 |
| Touchet River | -0.0970 | -0.0260 | 0.0449 | 0 |
| Tucannon River | -0.0717 | 0.0016 | 0.0750 | 0 |
| Wenatchee River | -0.0801 | -0.0142 | 0.0518 | 0 |
| Yakima River | -0.0793 | -0.0129 | 0.0535 | 0 |

Table 4: Estimated trends and uncertainty for core areas in Oregon. The trend column indicates if the estimated trend was significantly positive (+), negative (-), or neutral (0).

| Core area | Lower CI | Mean | Upper CI | Trend |
|------------------------------|----------|---------|----------|-------|
| Hood River | -0.1562 | -0.0295 | 0.0971 | 0 |
| Lower Deschutes River | 0.0038 | 0.0727 | 0.1415 | + |
| Odell Lake | -0.0537 | 0.0233 | 0.1003 | 0 |
| Upper Willamette River | 0.0094 | 0.0819 | 0.1544 | + |
| Imnaha River | -0.1385 | -0.0509 | 0.0366 | 0 |
| Lookingglass/Wenaha | -0.0482 | 0.0409 | 0.1301 | 0 |
| Pine-Indian-Wildhorse Creeks | -0.1738 | -0.0837 | 0.0063 | 0 |
| Umatilla River | -0.1264 | -0.0495 | 0.0274 | 0 |
| Upper Grande Ronde | -0.0378 | 0.0445 | 0.1268 | 0 |
| Walla Walla River | -0.0856 | -0.0153 | 0.0551 | 0 |
| Wallowa/Minam | -0.0607 | 0.0248 | 0.1103 | 0 |