

# creekchubRivLargerStrmDistWithin5km model summary

## Contents

<b>Overview</b> . . . . .	2
<b>Variables:</b> . . . . .	2
<b>Components:</b> . . . . .	5
<b>Model equation:</b> . . . . .	5
<b>Global sensitivity and uncertainty analysis:</b> . . . . .	5
<b>Model uncertainty</b> . . . . .	5
<b>Model sensitivity</b> . . . . .	8
Original model . . . . .	9
Arithmetic mean model . . . . .	10
Geometric mean model . . . . .	11
Limiting factor model . . . . .	12
Multiplicative model . . . . .	13
Summary of influential variables . . . . .	14
References . . . . .	15

## Overview

This document summarizes the results of a global sensitivity and uncertainty analysis for the **creekchubRivLargerStrmDistWithin5km** habitat suitability index (HSI) model for *Semotilus atromaculatus*. Metadata for the model is stored in the `ecorest` package in R.

The original documentation for this model can be found [here](#)<sup>1</sup>.

Sub-model: **Riverine habitats where access to larger water streams occurs within 5 km of study area (SIV4)**

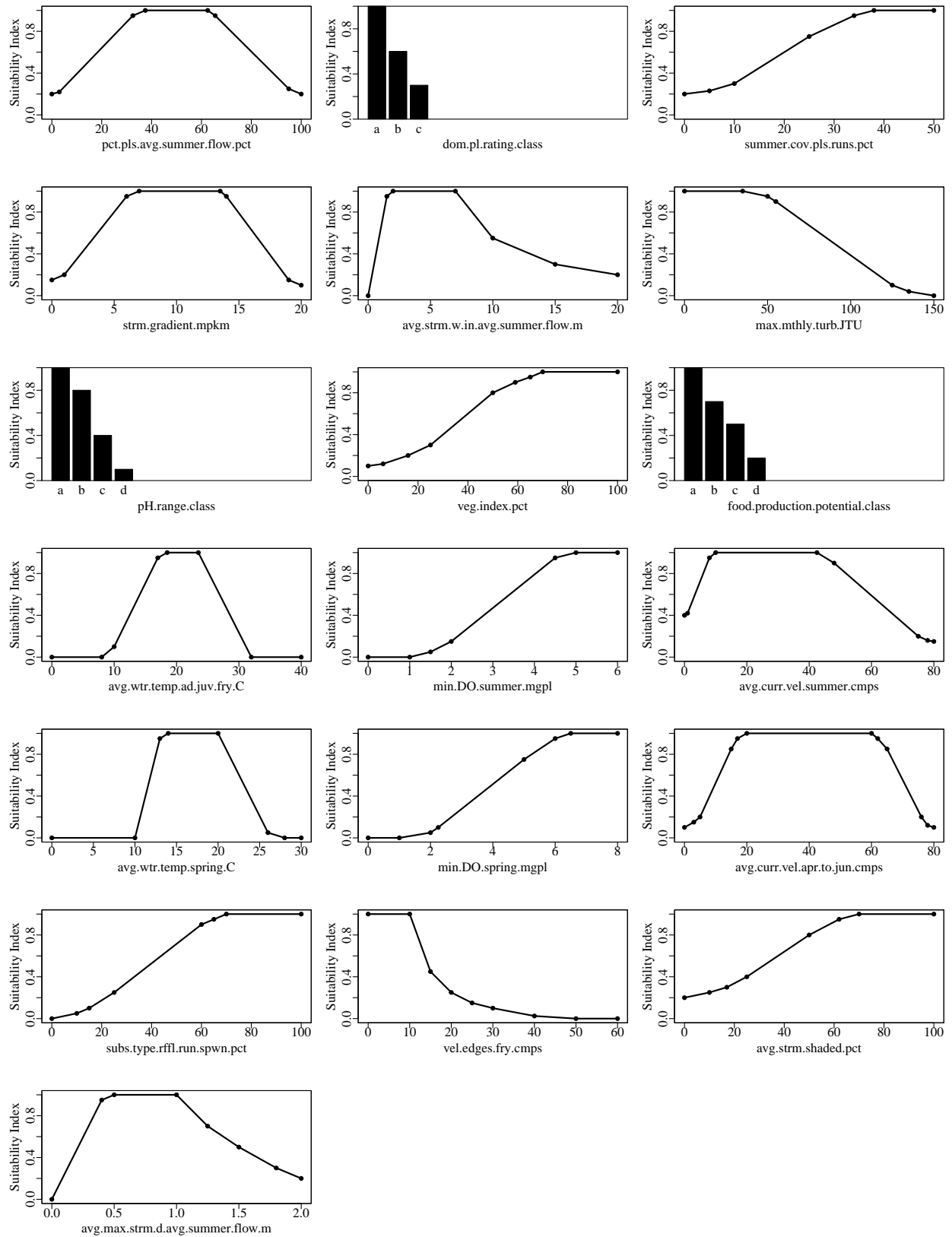
The `creekchubRivLargerStrmDistWithin5km` model is comprised of **19** variables and **5** components.

## Variables:

**Table 1.** SIV variables included in the `creekchubRivLargerStrmDistWithin5km` model. Type indicates whether a variable is numeric or categorical and breakpoints indicates the number of distinct breakpoints in suitability graphs.

	Variable name	Type	Breakpoints
SIV1	pct.pls.avg.summer.flow.SIV	numeric	8
SIV2	dom.pl.rating.SIV	categorical	3
SIV3	summer.cov.pls.runs.SIV	numeric	7
SIV5	strm.gradient.SIV	numeric	8
SIV6	avg.strm.w.in.avg.summer.flow.SIV	numeric	7
SIV7	max.mthly.turb.SIV	numeric	7
SIV8	pH.range.SIV	categorical	4
SIV9	veg.index.SIV	numeric	9
SIV10	food.production.potential.SIV	categorical	4
SIV11	avg.wtr.temp.ad.juv.fry.SIV	numeric	8
SIV12	min.DO.summer.SIV	numeric	7
SIV13	avg.curr.vel.summer.SIV	numeric	9
SIV14	avg.wtr.temp.spring.SIV	numeric	8
SIV15	min.DO.spring.SIV	numeric	8
SIV16	avg.curr.vel.apr.to.jun.SIV	numeric	12
SIV17	subs.type.rffl.run.spwn.SIV	numeric	8
SIV18	vel.edges.fry.SIV	numeric	9
SIV19	avg.strm.shaded.SIV	numeric	8
SIV20	avg.max.strm.d.avg.summer.flow.SIV	numeric	8

<sup>1</sup><https://ecolibrary.sec.usace.army.mil/resource/970f46b1-07f2-47d1-a908-1e149928055c>



**Figure 1.** Suitability index graphs for variables included in the creekchubRivLargerStrmDistWithin5km

model in ecoREST.

## Components:

**Table 2.** Components included in the creekchubRivLargerStrmDistWithin5km model in ecoRest.

	Component	Equation
CF	Food component	$(SIV9 + SIV10) / 2$
CC	Cover component	$(SIV1 * SIV2 * SIV3 * \min(((SIV1 * SIV2 * SIV3)^{(1/3)} + 0.2), 1.0) * SIV13 * SIV18)^{(1/6)}$
CWQ	Water quality component	$\text{ifelse}(SIV8 \leq 0.4   SIV11 \leq 0.4   SIV12 \leq 0.4, \min(SIV8, SIV11, SIV12, (SIV7 * SIV8 * SIV11 * SIV12 * SIV19)^{(1/5)}), (SIV7 * SIV8 * SIV11 * SIV12 * SIV19)^{(1/5)})$
CR	Reproduction component	$\text{ifelse}(SIV14 \leq 0.4   SIV15 \leq 0.4, \min(SIV14, SIV15, (SIV14 * SIV15 * SIV16 * (SIV17^2))^{(1/5)}), (SIV14 * SIV15 * SIV16 * (SIV17^2))^{(1/5)})$
COT	Other component	$(SIV5 + SIV6 + SIV20) / 3$

## Model equation:

The equation to calculate an overall HSI index for the creekchubRivLargerStrmDistWithin5km model is:

$$\text{ifelse}(CC \leq 0.4 | CWQ \leq 0.4 | CR \leq 0.4, \min(CC, CWQ, CR, (CF * CC * CWQ * CR * COT)^{(1/5)}), (CF * CC * CWQ * CR * COT)^{(1/5)})$$

According to our classification, this model's format is: **author-specified**

## Global sensitivity and uncertainty analysis:

We ran global sensitivity and uncertainty analyses on the creekchubRivLargerStrmDistWithin5km model using the sensobol package in R (Puy et al. 2022). The following parameters were used for the sensobol analysis:

**Table 3.** Parameters and settings used for sensobol sensitivity and uncertainty analyses.

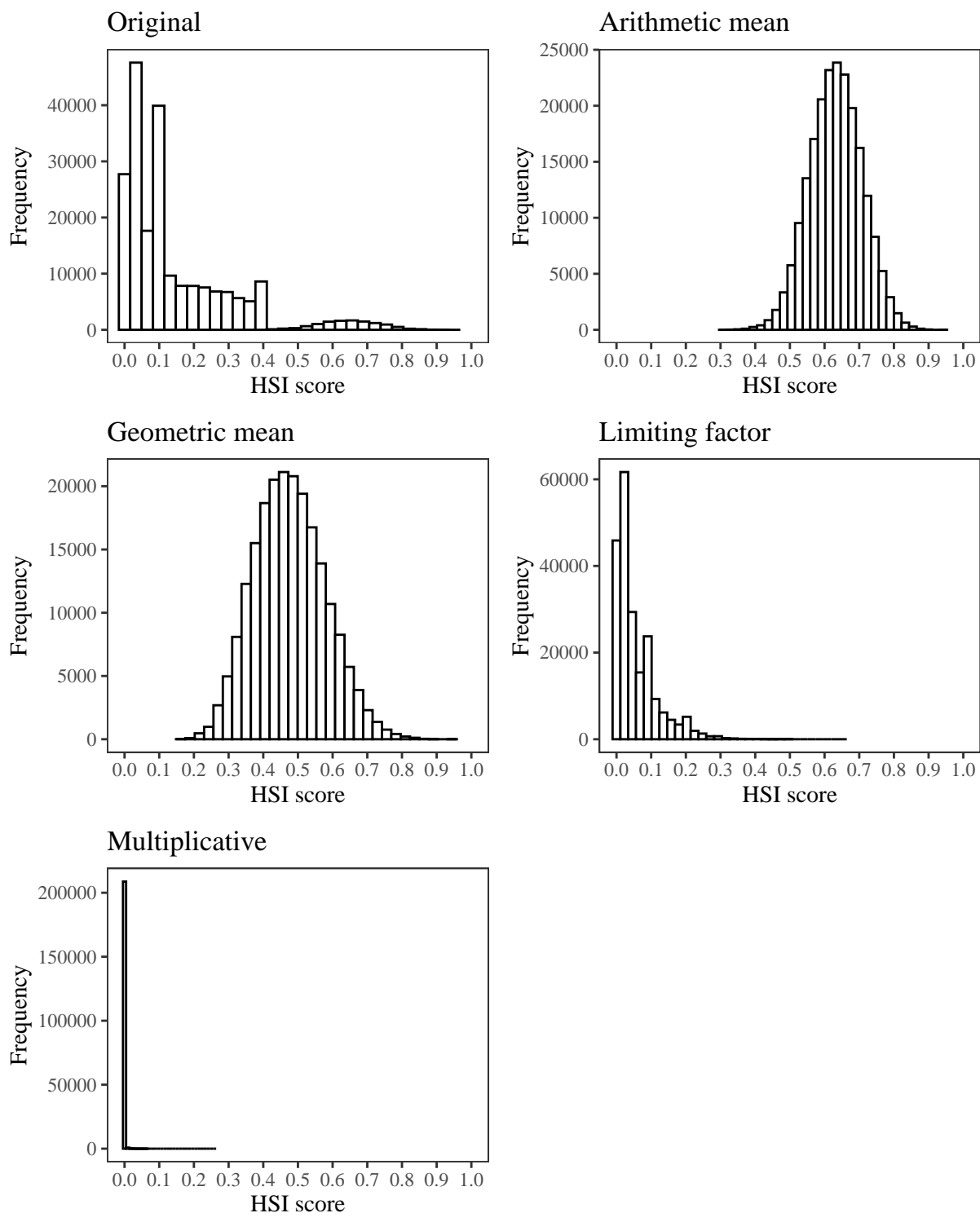
Parameter	Equation	Value
Number of input variables (M)	-	19
Base sample size (n)	-	10000
Number of model evaluations (N)	$n * (M + 2)$	210000
First order estimator	See Puy et al. (2022)	Saltelli
Total order estimator	See Puy et al. (2022)	Jansen
Number of bootstrap replications	-	1000
Sampling scheme	-	Quasi-random
Matrices	-	A, B, AB

We ran a sensitivity and uncertainty analysis for the creekchubRivLargerStrmDistWithin5km model using the original equation outlined in the documentation from McMahon (1982) and using arithmetic mean, geometric mean, limiting factor, and multiplicative equations to contrast the results across different equation structures.

## Model uncertainty

We ran the creekchubRivLargerStrmDistWithin5km model using 210000 combinations of its SIV variables, which were sampled from a uniform distribution spanning the range of possible values listed in the

creekchubRivLargerStrmDistWithin5km documentation. We limited the range of possible values for each parameter to the range in which the SIV values were greater than zero to prevent HSI score distributions with primarily zero values.



**Figure 2.** Empirical distributions of HSI scores for the creekchubRivLargerStrmDistWithin5km model using the original author-specified model equation from McMahon (1982), and an arithmetic mean, geometric

mean, limiting factor, and multiplicative structure incorporating all SIV variables. Note differences in the y axis.

We assumed a uniform distribution for all parameters because we evaluated all ecorest models in batch. Should you decide to run your own sensitivity analysis, this assumption should be evaluated independently for each parameter in the model.

**Table 4.** Quantiles from the empirical distribution of HSI scores for the original creekchubRivLargerStrmDistWithin5km model structure, an arithmetic mean equation, a geometric mean equation, a limiting factor equation, and a multiplicative equation structure.

	1%	2.5%	5%	25%	50%	75%	95%	97.5%	99%	100%
Original	0.00	0.00	0.01	0.03	0.10	0.21	0.53	0.66	0.73	0.95
Arithmetic	0.46	0.48	0.51	0.58	0.63	0.68	0.76	0.78	0.80	0.94
Geometric	0.26	0.29	0.32	0.40	0.47	0.54	0.65	0.69	0.73	0.93
Limiting	0.00	0.00	0.00	0.01	0.03	0.09	0.19	0.21	0.26	0.65
Multiplicative	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.26

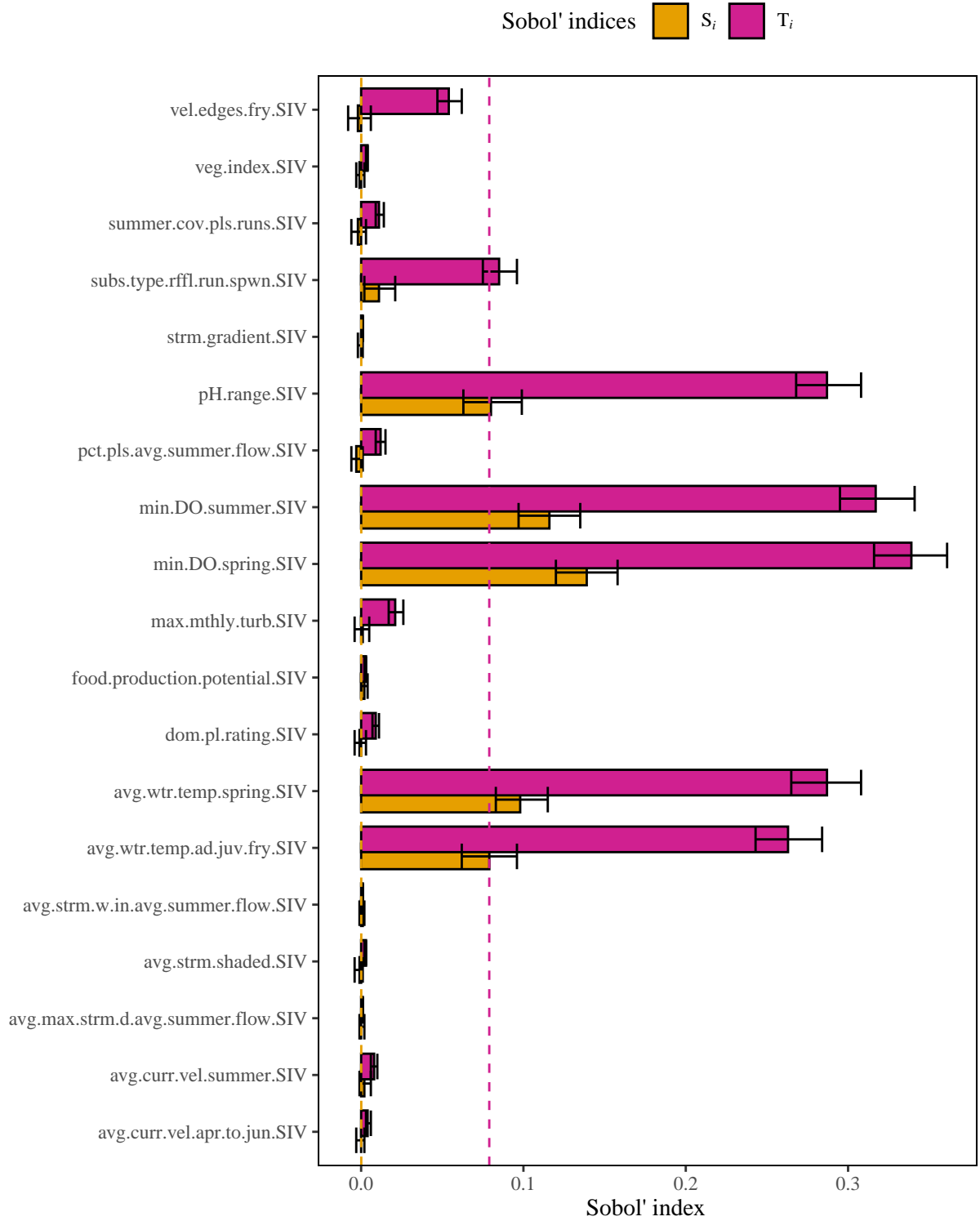
The empirical distribution of the original creekchubRivLargerStrmDistWithin5km model has a coefficient of variation (CV) of **1.109**, while the arithmetic mean model has a CV of **0.12**, the geometric mean model has a CV of **0.216**, the limiting factor model has a CV of **1.075**, and the multiplicative model has a CV of **10.912**. Hence, the **Multiplicative** model is the most uncertain, while the **Arithmetic mean** model is the least uncertain.

### Model sensitivity

Below are the results of the global sensitivity analysis for the creekchubRivLargerStrmDistWithin5km model using the original equation, an arithmetic mean, a geometric mean, a limiting factor, and a multiplicative model structure. The sensobol package uses variance-based sensitivity metrics, so the model’s sensitivity to a given parameter is a measure of how much variance in the HSI score decreases in response to that parameter being fixed (Puy et al. 2022). For each parameter, the observed changes in the variance of the HSI score can be described with a first order sensitivity index ( $S_i$ ) that accounts for the influence of a single parameter of interest on variance in HSI, or with a total order index ( $T_i$ ) that accounts for the influence of a single parameter on its own and in combination with all other parameters (*i.e.*, interactions) (Puy et al. 2022). We can compare the 95% confidence intervals for the first and total order indices to a dummy parameter, which represents a parameter that has no influence on the variance in a model’s output. While an uninfluential variable should theoretically have an  $S_i$  and  $T_i$  of zero, small approximation errors can lead variables to have a non-zero influence on a model’s output (Puy et al. 2022). If the confidence interval of the  $S_i$  and  $T_i$  index for a given parameter overlaps the confidence interval of the dummy parameter, we can deduce that the parameter has a negligible effect on variance in HSI scores, both on its own and in combination with all other variables.

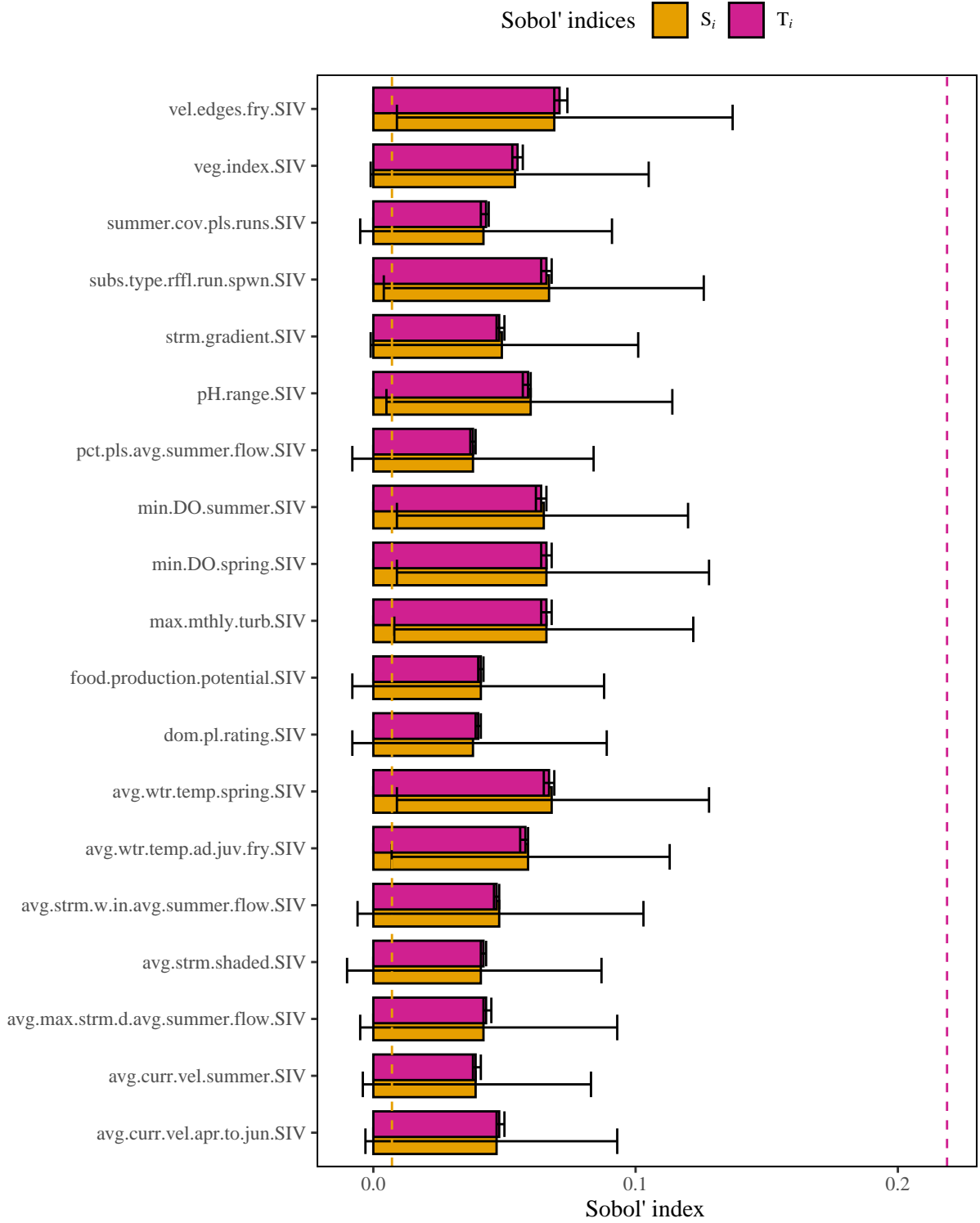


## Original model



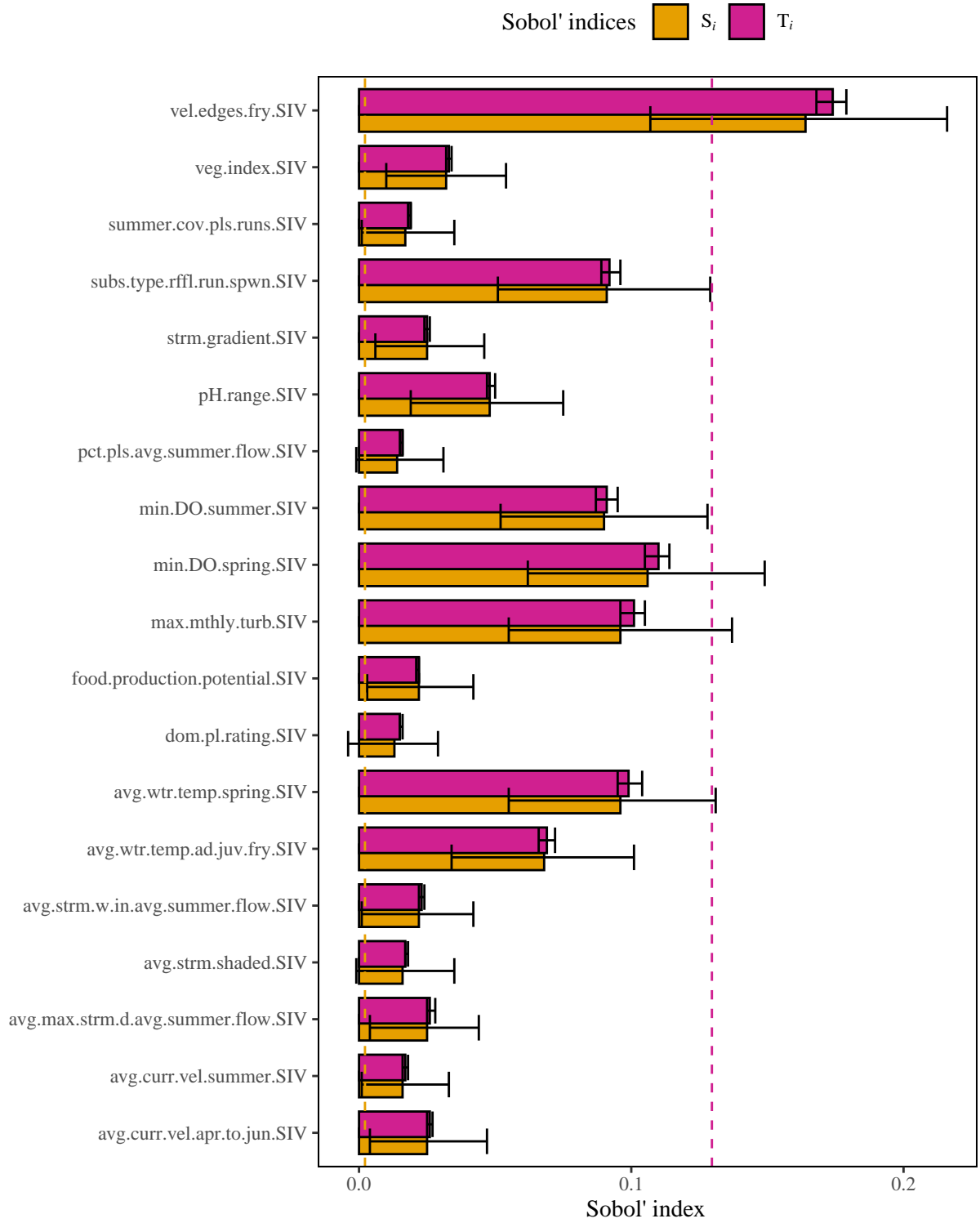
**Figure 3.** Results of a sensitivity analysis for the creekchubRivLargerStrmDistWithin5km model based on the original author-specified model outlined in McMahon (1982). Dashed lines represent baseline numerical approximation error for  $S_i$  and  $T_i$  (*i.e.*, dummy variables).

# Arithmetic mean model



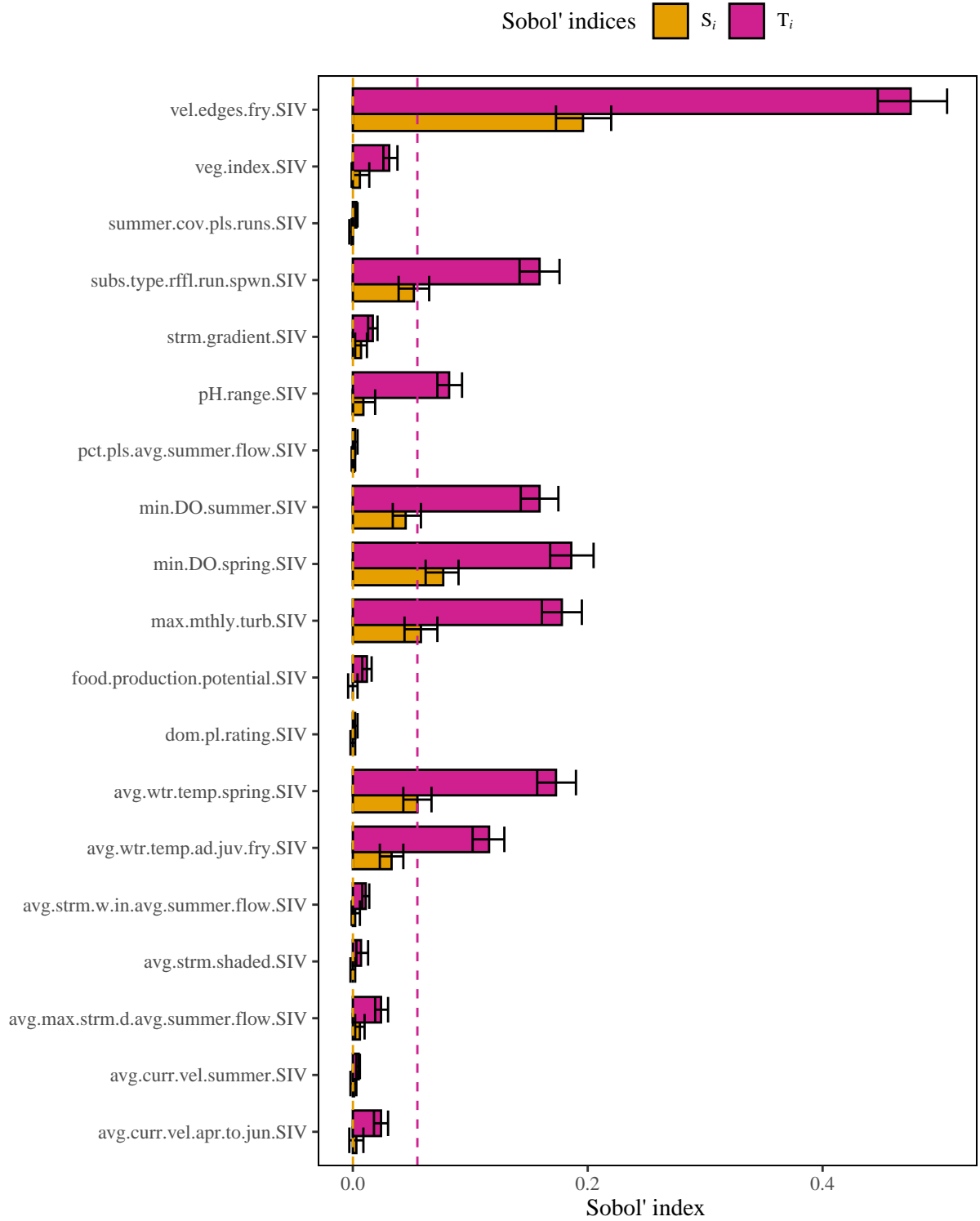
**Figure 4.** Results of a sensitivity analysis for the creekchubRivLargerStrmDistWithin5km model based on an arithmetic mean structure. Dashed lines represent baseline numerical approximation error for  $S_i$  and  $T_i$  (*i.e.*, dummy variables).

## Geometric mean model



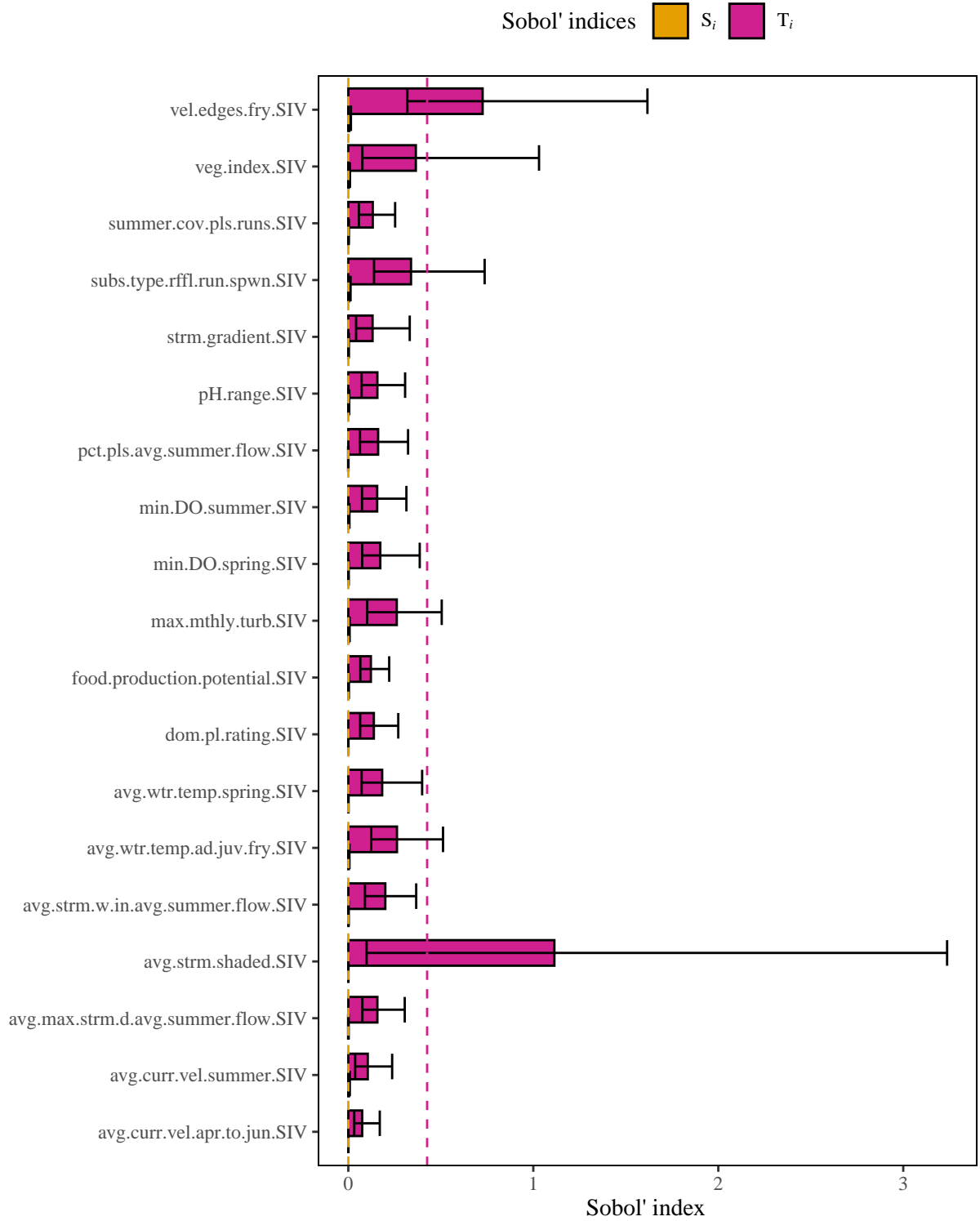
**Figure 5.** Results of a sensitivity analysis for the creekhubRivLargerStrmDistWithin5km model based on a geometric mean structure. Dashed lines represent baseline numerical approximation error for  $S_i$  and  $T_i$  (*i.e.*, dummy variables).

## Limiting factor model



**Figure 6.** Results of a sensitivity analysis for the creekhubRivLargerStrmDistWithin5km model based on a limiting factor structure. Dashed lines represent baseline numerical approximation error for  $S_i$  and  $T_i$  (*i.e.*, dummy variables).

## Multiplicative model



**Figure 7.** Results of a sensitivity analysis for the creekchubRivLargerStrmDistWithin5km model based on a multiplicative mean structure. Dashed lines represent baseline numerical approximation error for  $S_i$  and  $T_i$  (*i.e.*, dummy variables).

## Summary of influential variables

**Original model** In the **original creekchubRivLargerStrmDistWithin5km** model, **6 of 19** variables are influential and **min.DO.spring.SIV** has the highest first order sensitivity. In addition, **min.DO.spring.SIV** has the highest total order sensitivity.

### Un-influential variables in original model:

pct.pls.avg.summer.flow.SIV  
dom.pl.rating.SIV  
summer.cov.pls.runs.SIV  
strm.gradient.SIV  
avg.strm.w.in.avg.summer.flow.SIV  
max.mthly.turb.SIV  
veg.index.SIV  
food.production.potential.SIV  
avg.curr.vel.summer.SIV  
avg.curr.vel.apr.to.jun.SIV  
vel.edges.fry.SIV  
avg.strm.shaded.SIV  
avg.max.strm.d.avg.summer.flow.SIV

**Arithmetic mean model** In the **arithmetic mean creekchubRivLargerStrmDistWithin5km** model, **5 of 19** variables are influential and **vel.edges.fry.SIV** has the highest first order sensitivity. In addition, **vel.edges.fry.SIV** has the highest total order sensitivity.

### Un-influential variables in arithmetic mean model:

pct.pls.avg.summer.flow.SIV  
dom.pl.rating.SIV  
summer.cov.pls.runs.SIV  
strm.gradient.SIV  
avg.strm.w.in.avg.summer.flow.SIV  
pH.range.SIV  
veg.index.SIV  
food.production.potential.SIV  
avg.wtr.temp.ad.juv.fry.SIV  
avg.curr.vel.summer.SIV  
avg.curr.vel.apr.to.jun.SIV  
subs.type.rffl.run.spwn.SIV  
avg.strm.shaded.SIV  
avg.max.strm.d.avg.summer.flow.SIV

**Geometric mean model** In the **geometric mean creekchubRivLargerStrmDistWithin5km** model, **13 of 19** variables are influential and **vel.edges.fry.SIV** has the highest first order sensitivity. In addition, **vel.edges.fry.SIV** has the highest total order sensitivity.

### Un-influential variables in geometric mean model:

pct.pls.avg.summer.flow.SIV  
dom.pl.rating.SIV  
summer.cov.pls.runs.SIV  
avg.strm.w.in.avg.summer.flow.SIV  
avg.curr.vel.summer.SIV  
avg.strm.shaded.SIV

**Limiting factor model** In the **limiting factor creekchubRivLargerStrmDistWithin5km** model, **10 of 19** variables are influential and **vel.edges.fry.SIV** has the highest first order sensitivity. In addition, **vel.edges.fry.SIV** has the highest total order sensitivity.

**Un-influential variables in limiting factor mean model:**

pct.pls.avg.summer.flow.SIV  
dom.pl.rating.SIV  
summer.cov.pls.runs.SIV  
avg.strm.w.in.avg.summer.flow.SIV  
veg.index.SIV  
food.production.potential.SIV  
avg.curr.vel.summer.SIV  
avg.curr.vel.apr.to.jun.SIV  
avg.strm.shaded.SIV

**Multiplicative model** In the **multiplicative mean creekchubRivLargerStrmDistWithin5km** model, **3 of 19** variables are influential and **vel.edges.fry.SIV** has the highest first order sensitivity. In addition, **avg.strm.shaded.SIV** has the highest total order sensitivity.

**Un-influential variables in multiplicative model:**

pct.pls.avg.summer.flow.SIV  
dom.pl.rating.SIV  
summer.cov.pls.runs.SIV  
strm.gradient.SIV  
avg.strm.w.in.avg.summer.flow.SIV  
pH.range.SIV  
veg.index.SIV  
food.production.potential.SIV  
avg.wtr.temp.ad.juv.fry.SIV  
avg.curr.vel.summer.SIV  
avg.wtr.temp.spring.SIV  
min.DO.spring.SIV  
avg.curr.vel.apr.to.jun.SIV  
subs.type.rffl.run.spwn.SIV  
avg.strm.shaded.SIV  
avg.max.strm.d.avg.summer.flow.SIV

## References

1. McKay S, D Hernandez-Abrams, and K Cushway. 2024. ecorest: conducts analyses informing ecosystem restoration decisions. R package version 2.0.0, <https://CRAN.R-project.org/package=ecorest>.
2. McMahon, TE. 1982. Habitat suitability index models: Creek chub. U.S.D.I. Fish and Wildlife Service. FWS/OBS-82/10.4 23 pp.
3. Puy, A, S Lo Piano, A Saltelli, and SA Levin. 2022. sensobol: an R package to compute variance based sensitivity indices. Journal of Statistical Software 102(5):1-37. doi: 10.18637/jss.v102.i05