Implementation verification for the DeEP tool

Tjalling Jager*

July 26, 2023

This documents is part of the DeEP-tool project (`https://deep-tox.info/`).

Contents

---

*DEBtox Research, Stevensweert, The Netherlands (tjalling@debtox.nl, `http://www.debtox.nl/`)

1 Software used and approach followed

The DEB-TKTD EPx Predictor (DeEP) is a user-friendly software to predict the effects of time-varying pesticide exposure profiles on individual organisms. The tool applies a moving time-window approach to predict EPx values across an entire profile (see [9, 1]). EPx is the factor by which an exposure profile (or in this case a time window of specific length) needs to be multiplied to yield x% effect on a specific endpoint (here: body length or cumulative reproduction) at the end of the profile. The DEB-TKTD model underlying DeEP is DEBtox2019, as published in [3]. To test the implementations of DeEP, I compared to two versions (standalone and web-based) to the Bring Your Own Model (BYOM) platform, which represents the original implementation of DEBtox2019.[1]

Software versions:

- DeEP standalone version 1.0.0 (made available by Rifcon on 20 June 2023).
- DeEP web version 1.0.0 (accessed on 24 July 2023).
- BYOM version 6.7 with DEBtox2019 package version 4.7 using the files in the `ERA-special` folder.

First, calibrations on a data set are performed in BYOM. Calibration is not strictly necessary for the comparison of the EPx calculations: as long as both DeEP and the BYOM EPx calculations depart from the same parameter values, the results can be meaningfully compared. However, calibrating to the data, with various setting for pMoA and feedback structure, makes sure that we obtain biologically and toxicologically relevant parameter sets. Further, re-calibration with different settings leads to rather similar EPx values for the same set of exposure profiles that I used for all predictions. This is helpful to keep the EPx estimates within the range where they are reported by the software. DeEP applies a cut-off for EPx (by default at 1000, which is also the hard-coded maximum allowed), so higher values would only be reported as $> x$. DeEP seems to use a minimum EPx of 0.001 (just like BYOM); lower values are reported as zero.

Notes for the calibrations with the BYOM DEBtox2019 package:

- This version of the DeEP tool cannot use a brood-pouch delay (see [8]), so such delays are ignored here (`Tlag=0`).
- This version of the DeEP tool cannot accommodate deviating growth curves, so such modification are turned off (`Lj=0` and `Lf=0`).

As main data set (Test A-E), I used the results from toxicity tests with *Folsomia candida* and cadmium, analysed earlier in [6]. Exposure is through food, but this does not matter for the verification. Several variations on the calibration were used, such as fixing k_d to a very high or very low value (fast and slow kinetics). Further, different PMoA/feedback configurations were used. The second data set used is also for *F. candida*,

---

[1] The BYOM platform and the DEBtox2019 package are available from: `http://www.debtox.info/byom.html`.

using chlorpyrifos. This data set was analysed earlier in [7], and used as case study in [3]. The same settings as for the case study in [3] were used, trying both PMoAs with direct effects on reproduction (Test F).[2]

After calibration, the BYOM file `convert_deep` was used to translate the MAT file produced by the calibration into the input file format for DeEP. This avoids errors that could arise from manually copying the parameter values and model settings. Next, EPx calculations are run. For the exposure profiles, I used the scenarios as applied in the EFSA Scientific Opinion [2] for demonstration of extrapolations with GUTS [4, 5] for the illustrative case study with propiconazole and gammarids. Obviously, these exposure profiles are completely irrelevant to the data sets that I used here for DEBtox2019 calibration. Nevertheless, they provide a rather typical range of exposure patterns that are well suited for code verification.

Settings:

- To ensure that most of the resulting EPx values fall into a range of 'interesting' values (1-1000) with the current test calibration, all profiles where multiplied by 50 for the *Folsomia* case study with cadmium. This does not affect the calculations, but limits the number of EPx values that are 'out of range' for reporting and plotting.
- Some of the default settings for DeEP were used for the extrapolations: 10% effect, timestep for moving time window of 1 day, and pass/fail of EPx at 10. The following modification were used for this comparison: a time window of 28 days was used, and thinning was turned off to show the entire time course of EPx in the plots. These settings were used for both software implementations.
- For BYOM, EPx is calculated for the endpoints body length and reproduction. The current version of DeEP does not use survival results for EPx calculation (which would usually be a less sensitive endpoint anyway).
- All exposure profiles were imported from text files. The functionality of importing TOXSWA out-files was not tested.
- This version of DeEP does not calculate confidence intervals on the estimates. Therefore, CIs were also not calculated with BYOM.

In comparing the plots produced by DeEP and BYOM, it is good to note that they use a different scale on the y-axis, and that DeEP plots EPx values that are too high (out of range) at the maximum allowed value (i.e., 1000 here). In contrast, in BYOM they are treated as 'not a number' and not plotted. The plots for EPx versus time can thus show gaps for BYOM, but not for DeEP.

---

[2] For both data sets, the parameter estimates will not be exactly the same as in the papers, which relates to slightly different model versions, and using different parts of the data set (e.g, I excluded survival data from the fit for these tests).

2 Test A

Calibration settings:

Calibration data set:	*Folsomia candida* exposed to cadmium in food.
Mode of action:	Assimilation ([10000]).
Feedbacks configuration:	No feedbacks ([0000]).

Parameter settings/estimates (* hitting bound of search range):

Param.	Calibrated value
L_0	0.1204
L_p	0.3954
L_m	0.6734
r_B	0.05865
R_m	15.60
f	1 (fixed)
k_d	0.1064
z_b	0.1066*
b_b	$9.695 \cdot 10^{-5}$
F_{BV}	0.008 (fixed)
K_{RV}	1 (fixed)
κ	0.8 (fixed)
y_P	0.64 (fixed)
L_m^{ref}	0.7 (fixed)

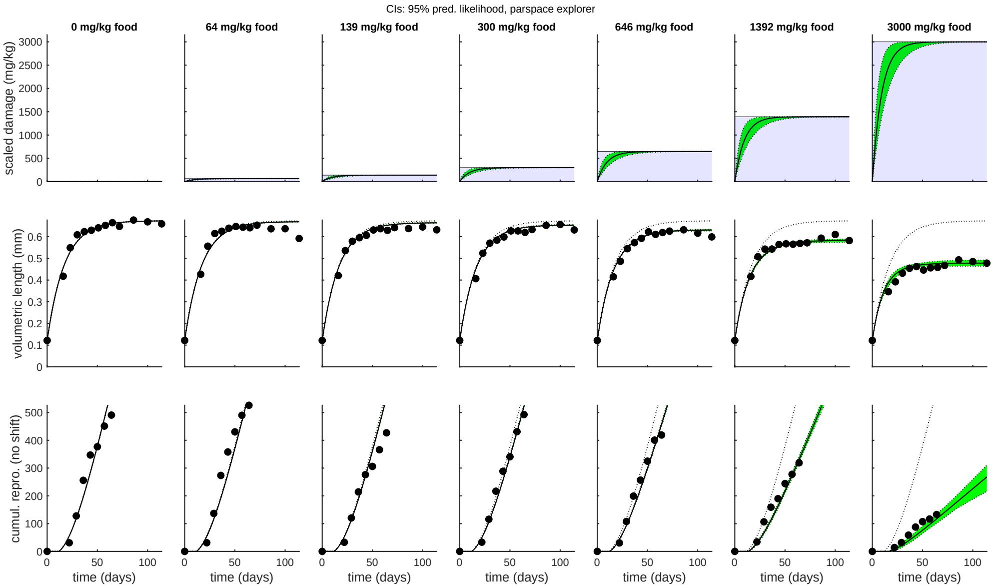

Figure 1: BYOM fit for test Test A. Note that the data set for reproduction is cut off at $t = 64$ days to avoid fitting on the ageing response that becomes obvious in the data set for longer exposure times.

Settings for EPx calculation:
% effect: 10
Window length: 28 days
Timestep: 1 day
Cut-off: 1000 in DeEP, 100000 in BYOM
Pass/fail: 10
Thinning: off
Modification of exposure profiles: all exposure values times 50.

For all profiles, DeEP identifies reproduction as the most sensitive endpoint (i.e., lower EP10 than for length). DeEP does not provide a value for the EP10 on length. BYOM confirms that the EP10 for length is indeed higher than the EP10 for reproduction, for all exposure profiles.

Table 1: EP10 values for test A. DeEP values in parentheses are for a previous version of the web application (run in May 2023), which accommodated a higher value for the cut-off.

Exposure scenario	DeEP stdalone	DeEP web	BYOM
apple R1 pond	5.979	5.979	5.98 ($t = 91$)
apple R2 stream	162.789	162.789	162.80 ($t = 92$)
cereal D1 ditch	2.435	2.435	2.43 ($t = 134$)
cereal D1 stream	13.262	13.262	13.26 ($t = 327$)
cereal D3 ditch	50.426	50.426	50.43 ($t = 126$)
cereal D4 pond	4.205	4.205	4.21 ($t = 150$)
cereal D4 stream	363.515	363.515	363.51 ($t = 336$)
cereal D5 pond	4.156	4.156	4.16 ($t = 131$)
cereal D5 stream	>1000	>1000 (1077.185)	1077.19 ($t = 122$)
cereal R4 stream	62.434	62.434	62.44 ($t = 72$)

The figures below compare the plots from the standalone to the plots from BYOM.

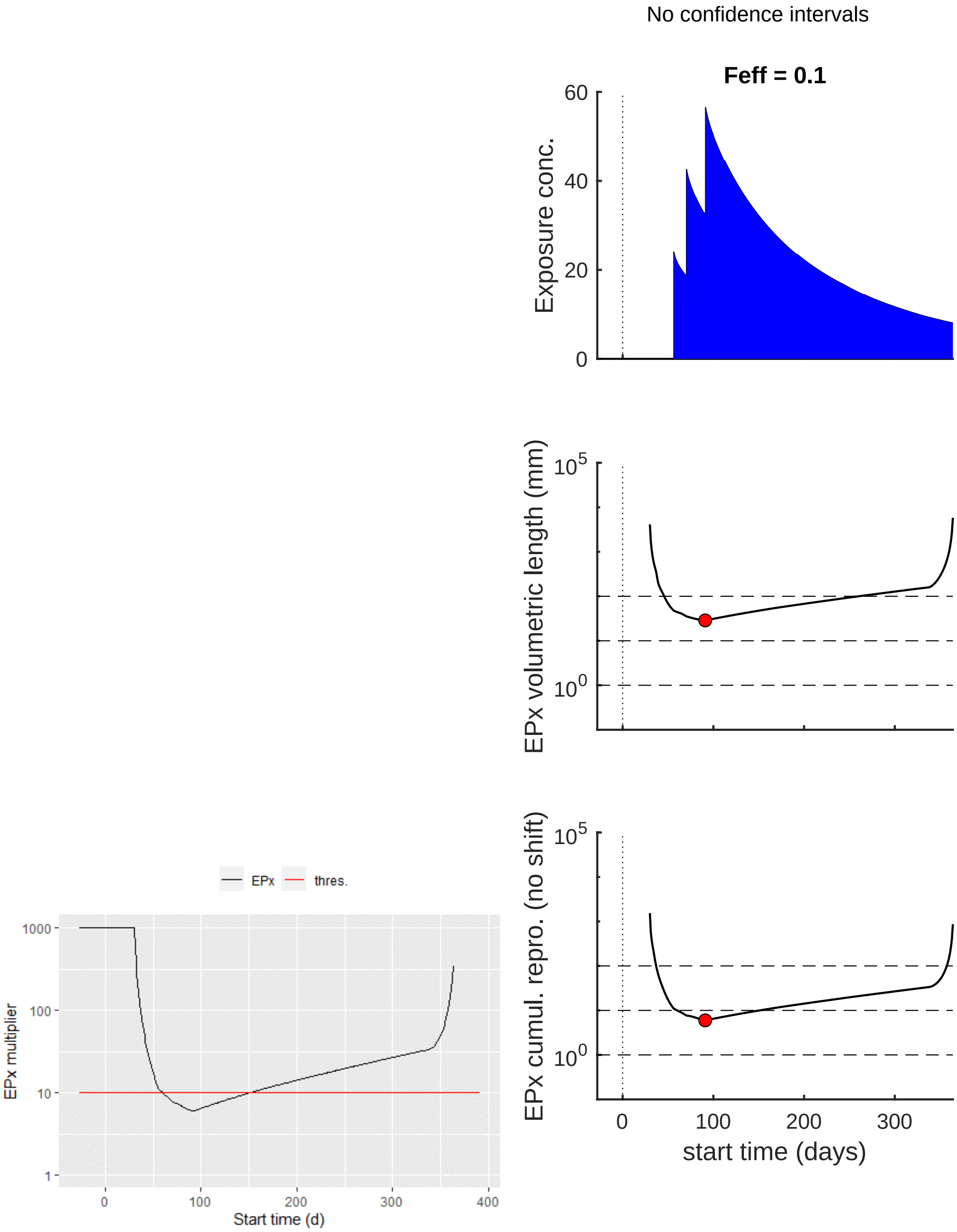

Figure 2: Result Test A, scenario apple R1 pond. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

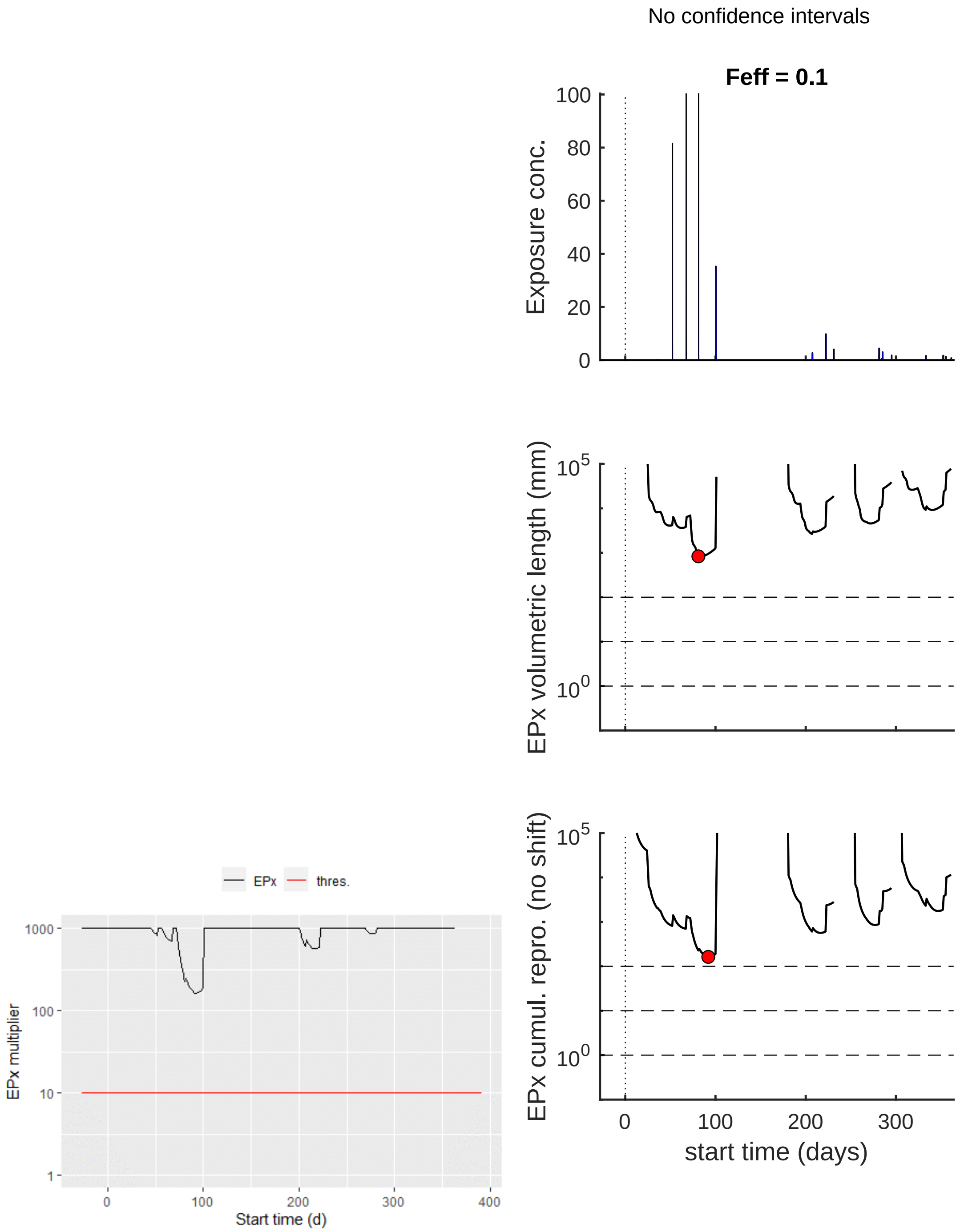

Figure 3: Result Test A, scenario apple R2 stream. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

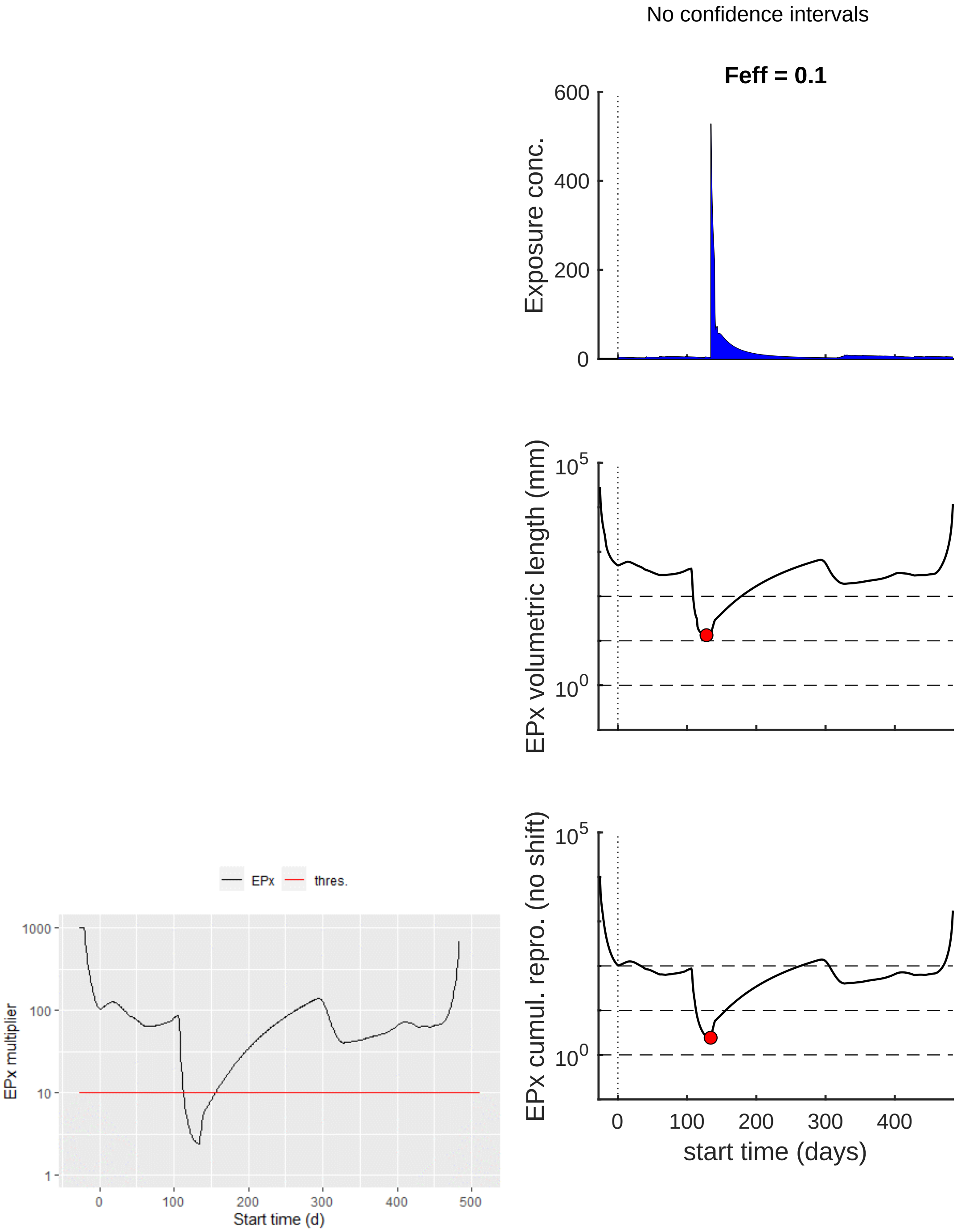

Figure 4: Result Test A, scenario cereal D1 ditch. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

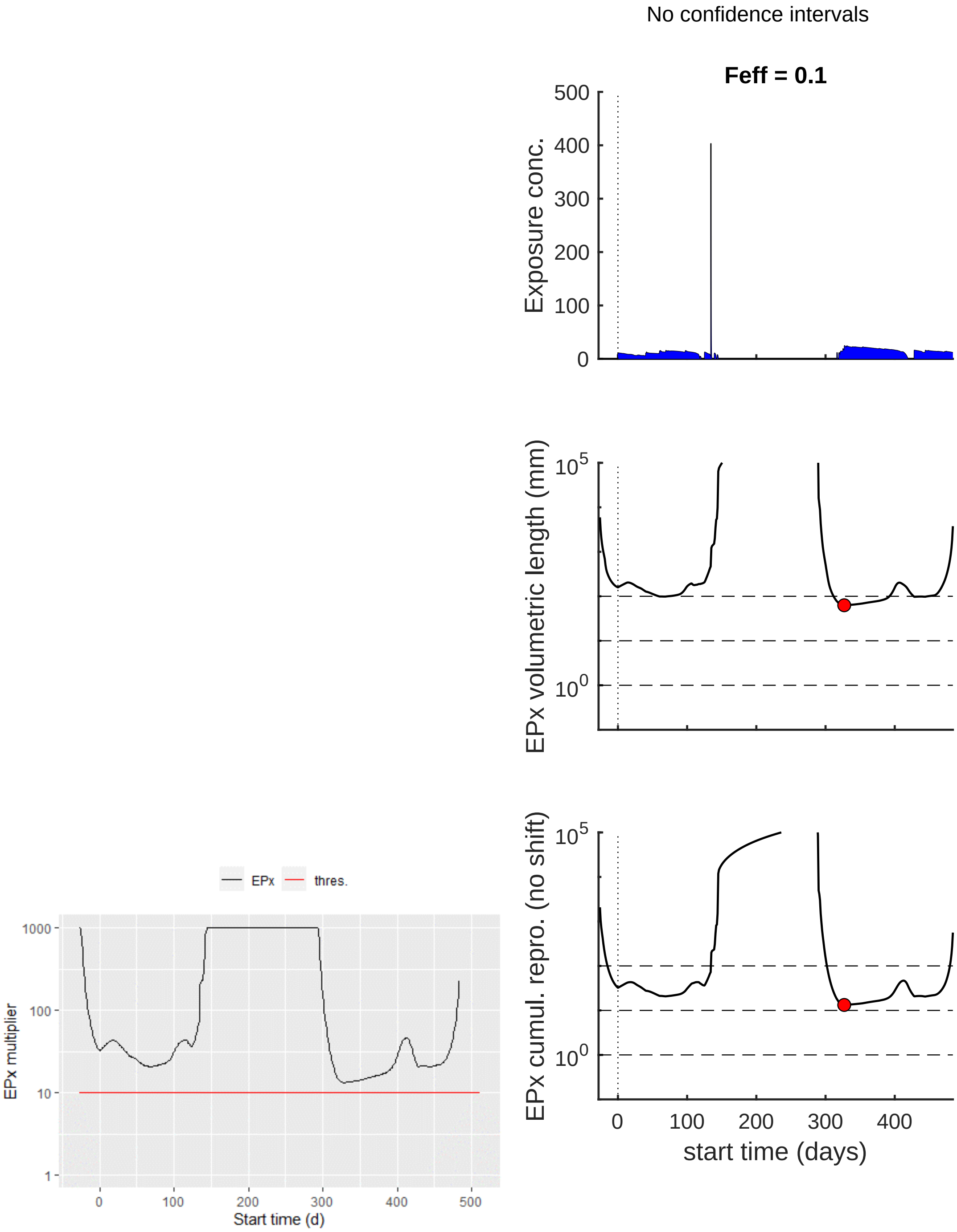

Figure 5: Result Test A, scenario cereal D1 stream. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

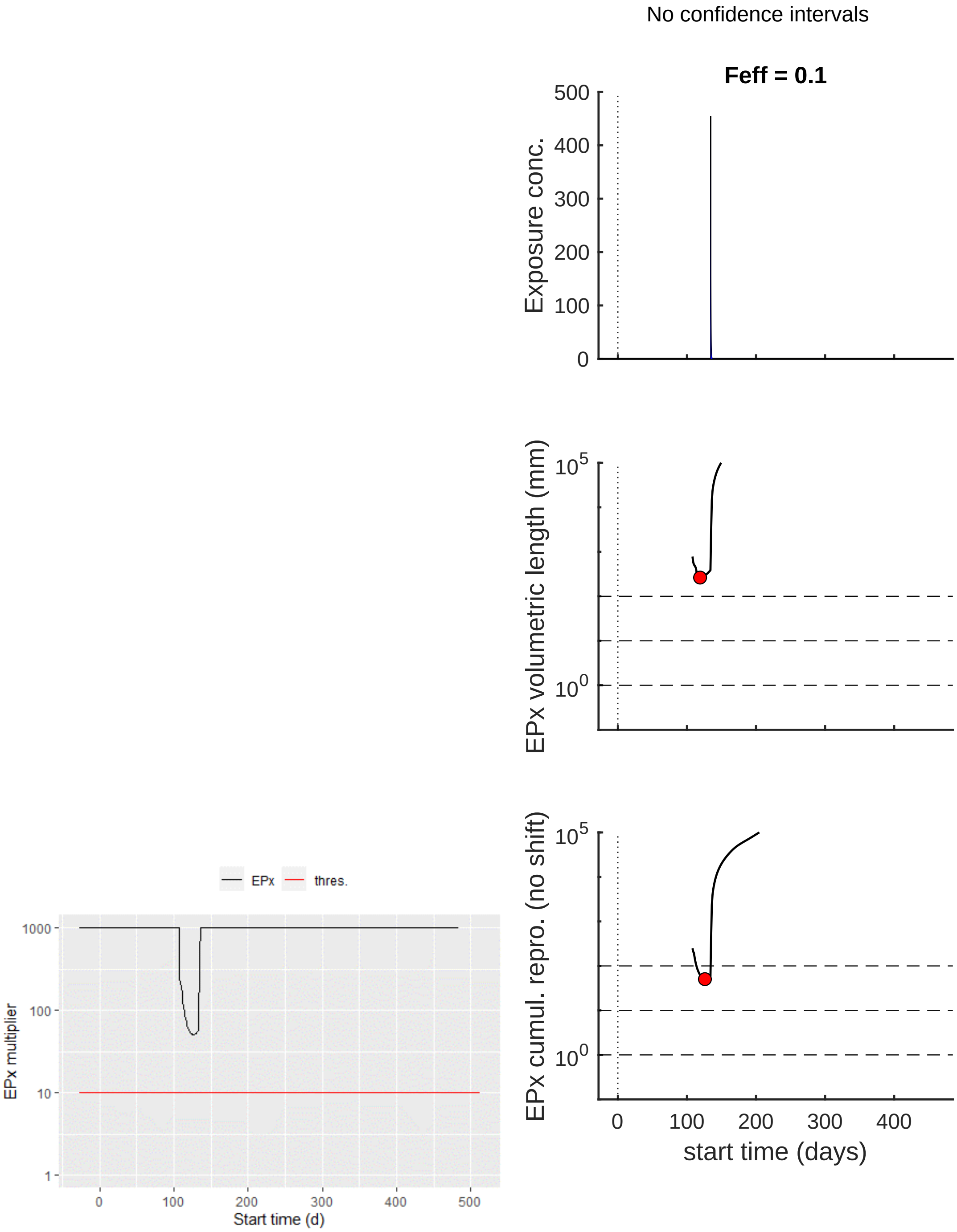

Figure 6: Result Test A, scenario cereal D3 ditch. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

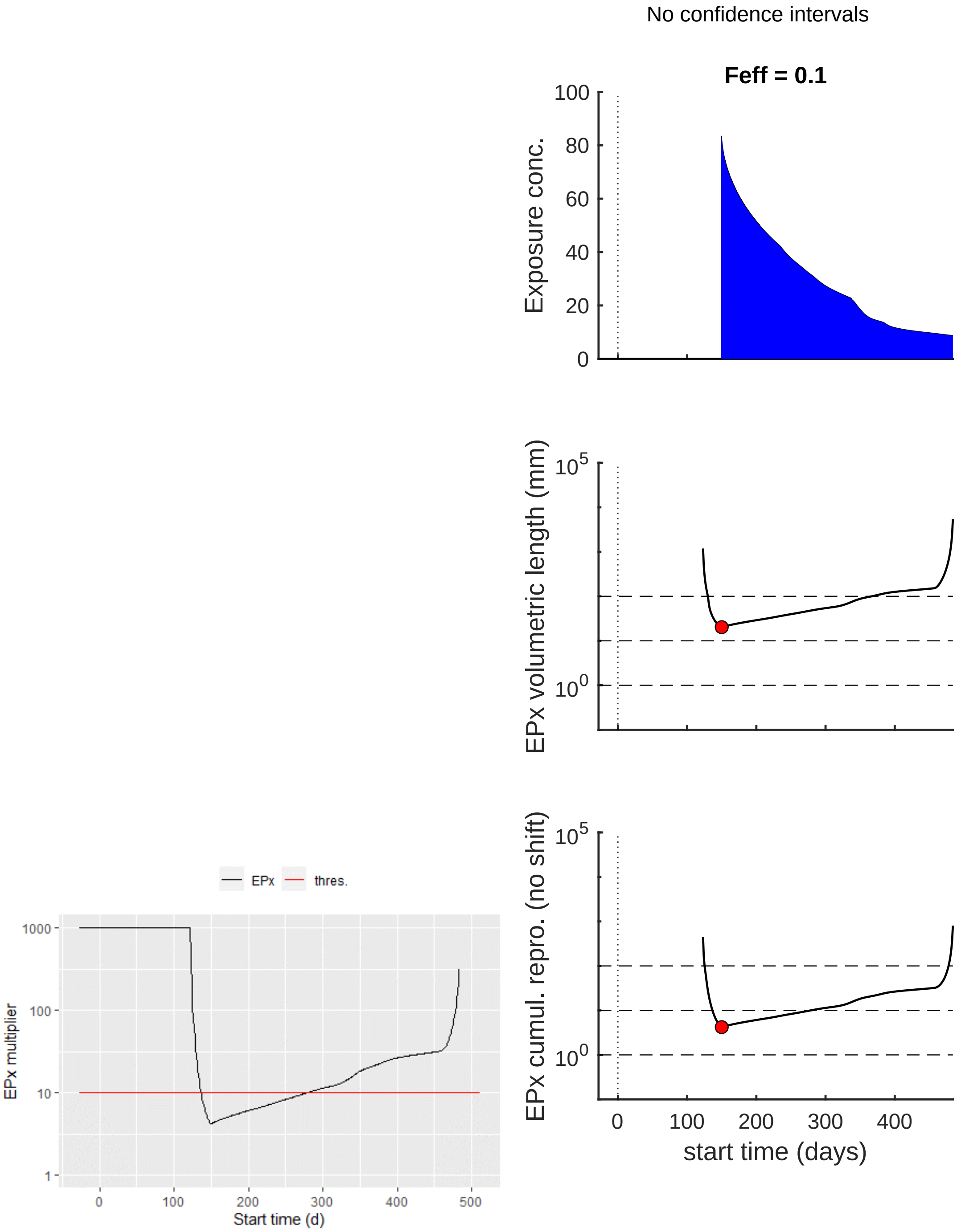

Figure 7: Result Test A, scenario cereal D4 pond. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

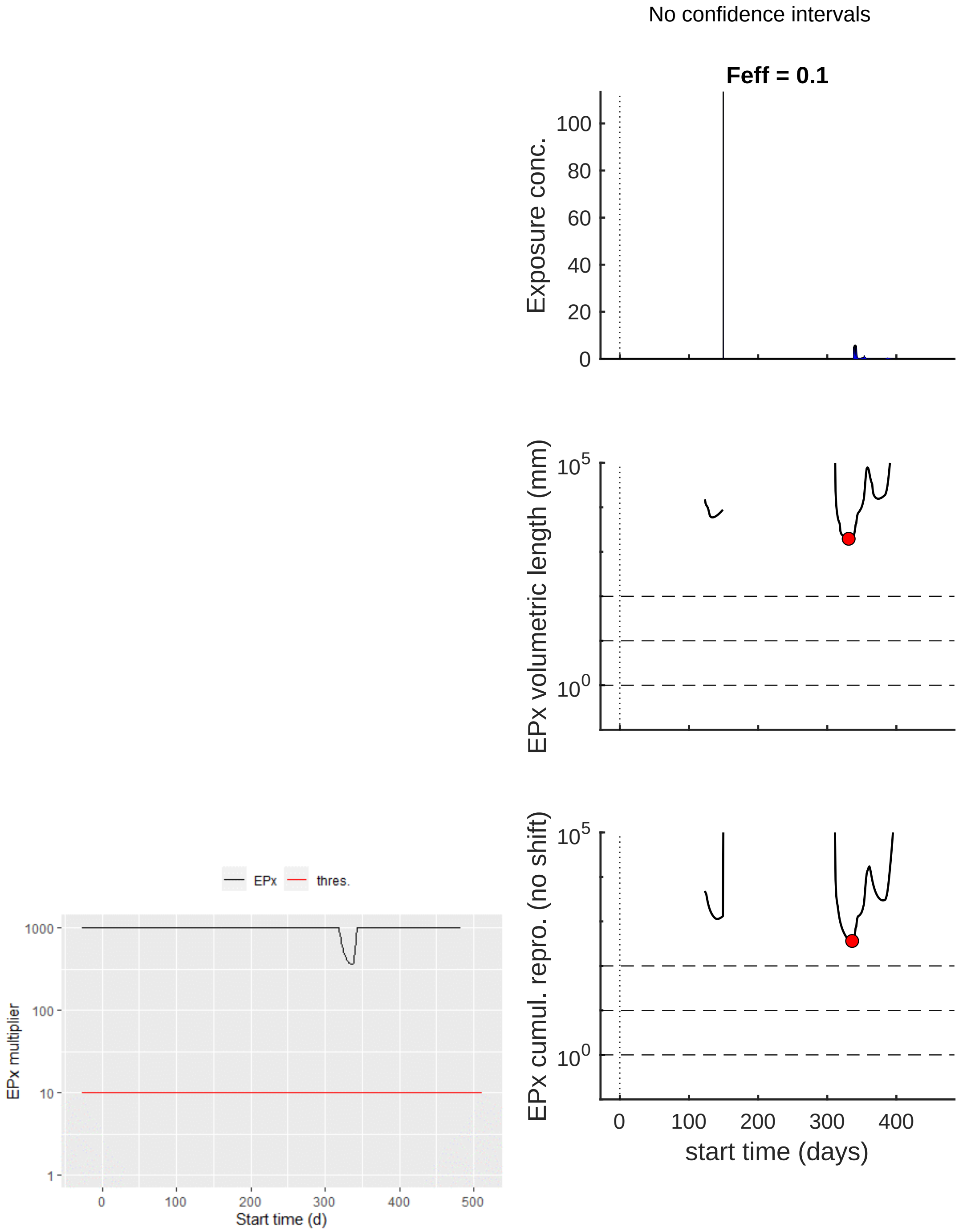

Figure 8: Result Test A, scenario cereal D4 stream. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

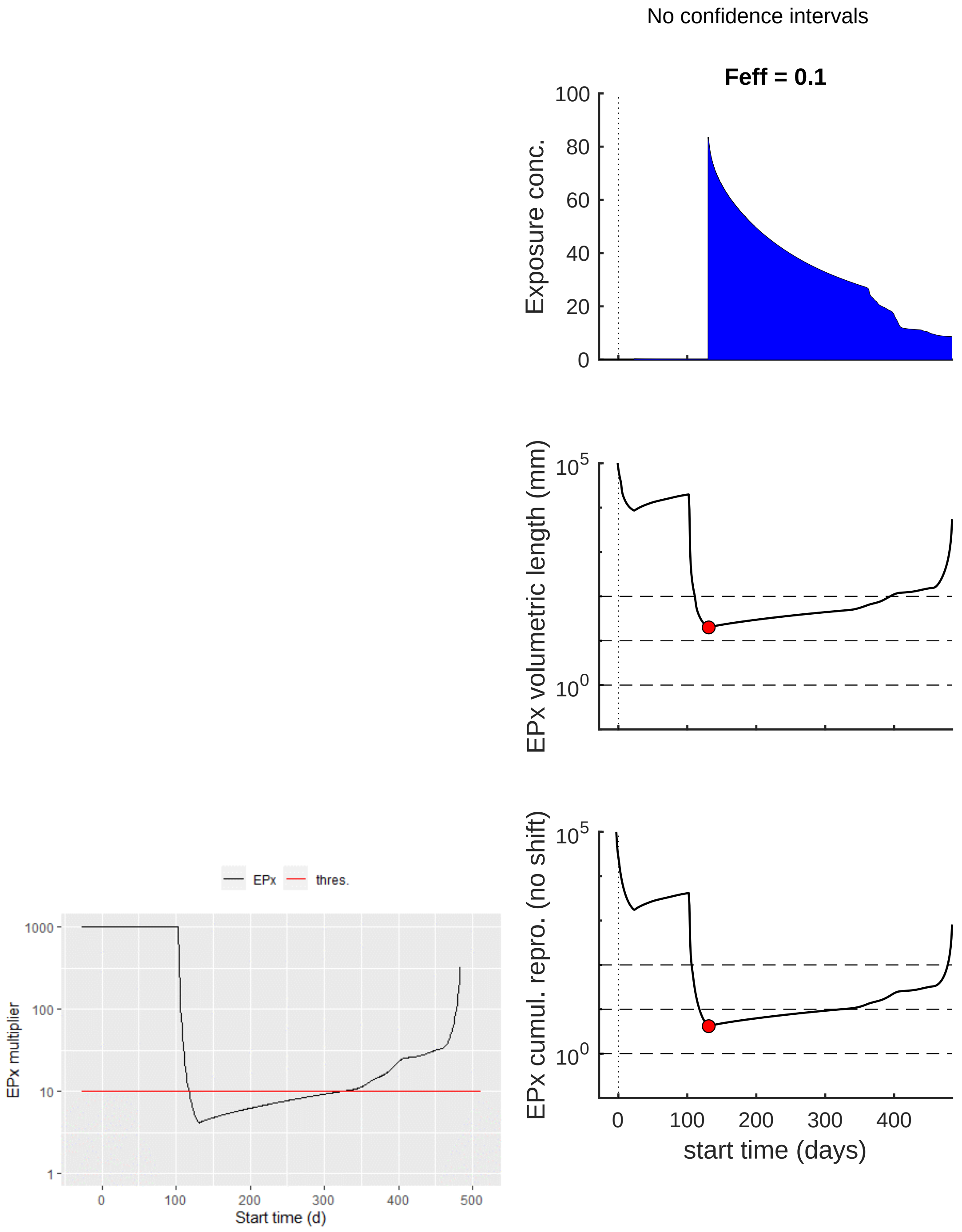

Figure 9: Result Test A, scenario cereal D5 pond. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

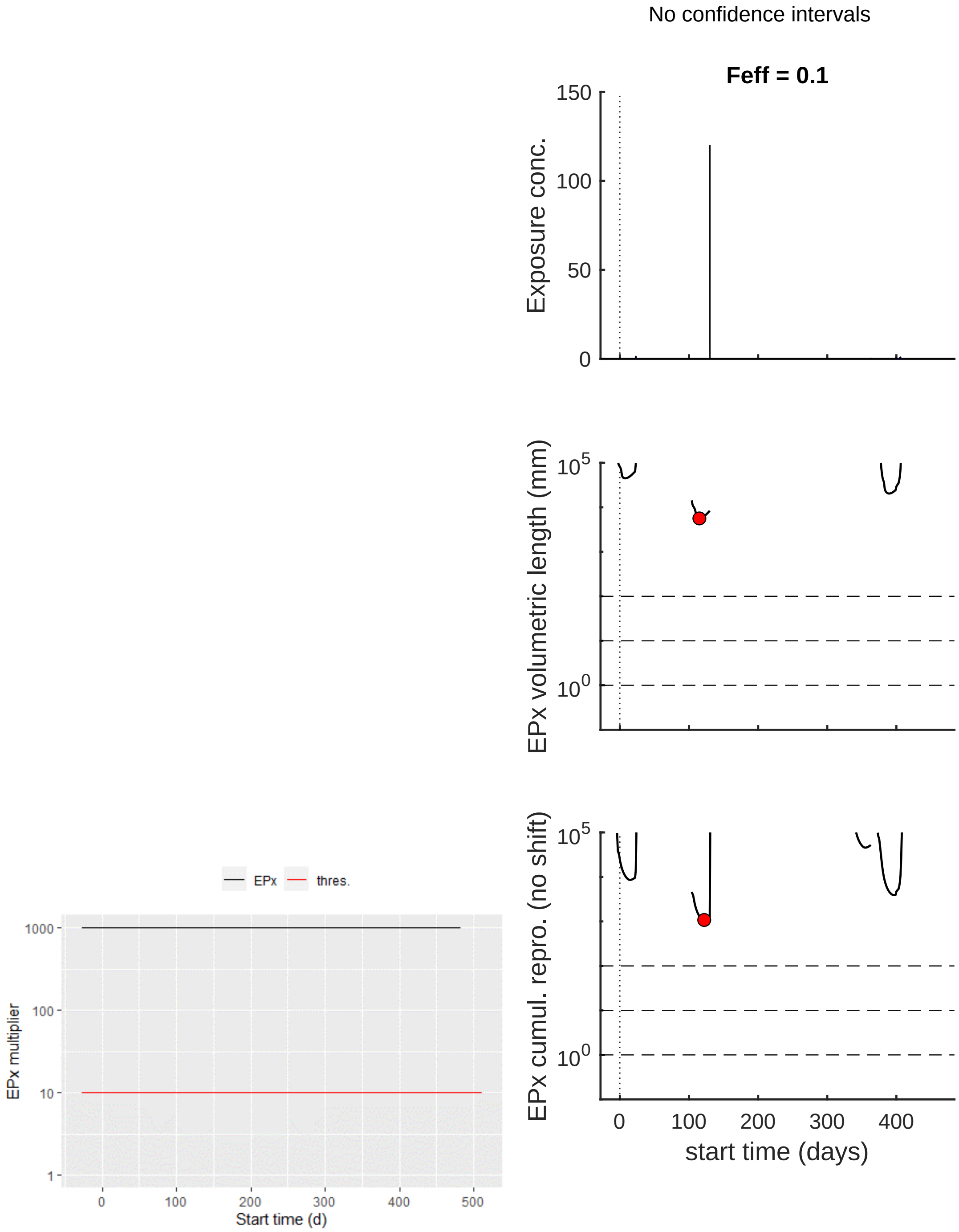

Figure 10: Result Test A, scenario cereal D5 stream. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

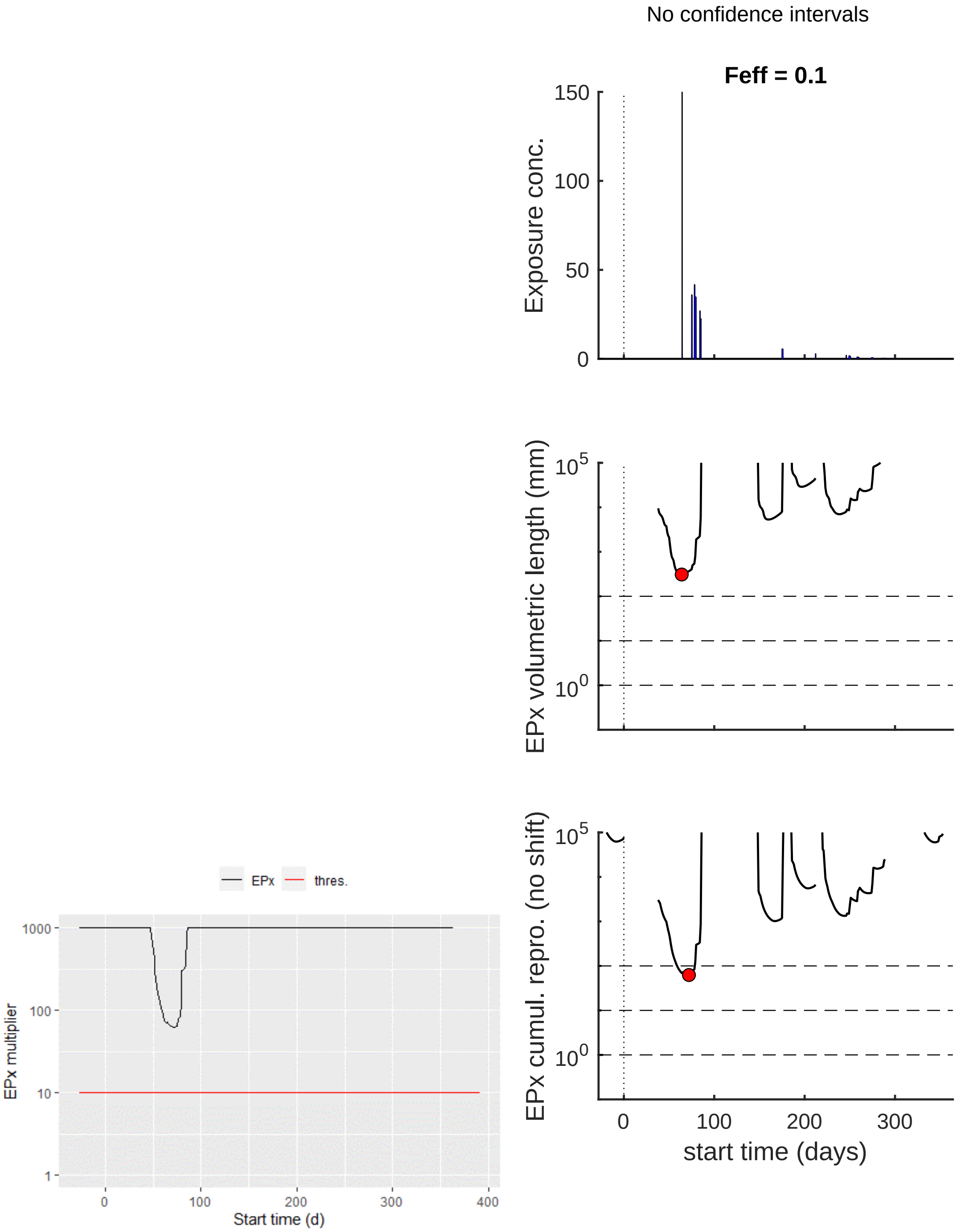

Figure 11: Result Test A, scenario cereal R4 stream. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

3 Test B: thinning and different x

3.1 Test B: thinning

This is the same calibration as Test A, and same settings for simulations, apart from one: thinning is now turned on. This is to test whether thinning does the same thing in all software implementations.

The DeEP standalone and web version yield the exact same EPx values as in the table for Test A. The number of thinned windows was the same between the two implementation. The BYOM EP10 values (and the times at which the lowest value occurs) are also the same as in the table for Test A. Therefore, the table is not repeated here. The figures below compare the plots from the standalone to the plots from BYOM.

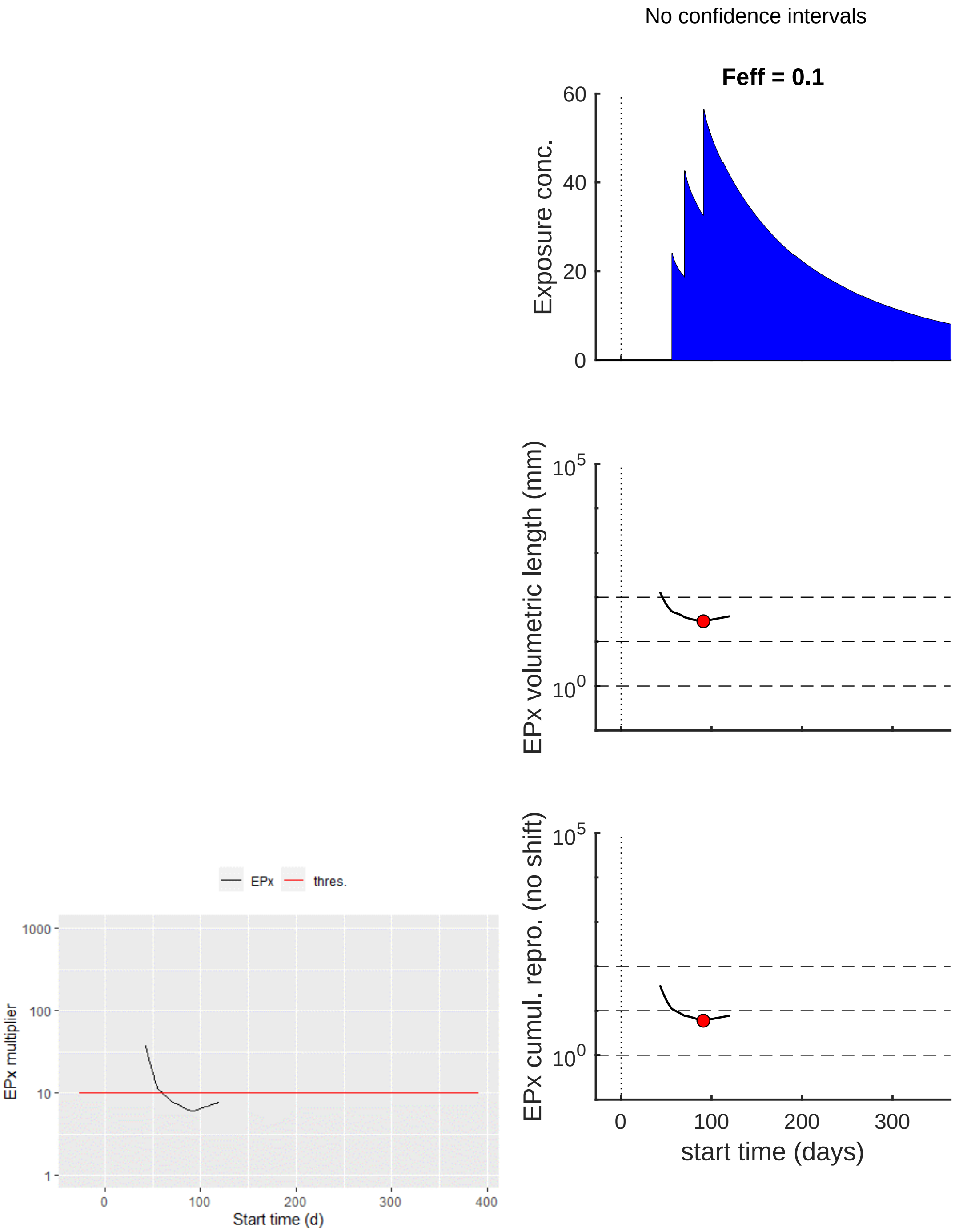

Figure 12: Result Test B, scenario apple R1 pond. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

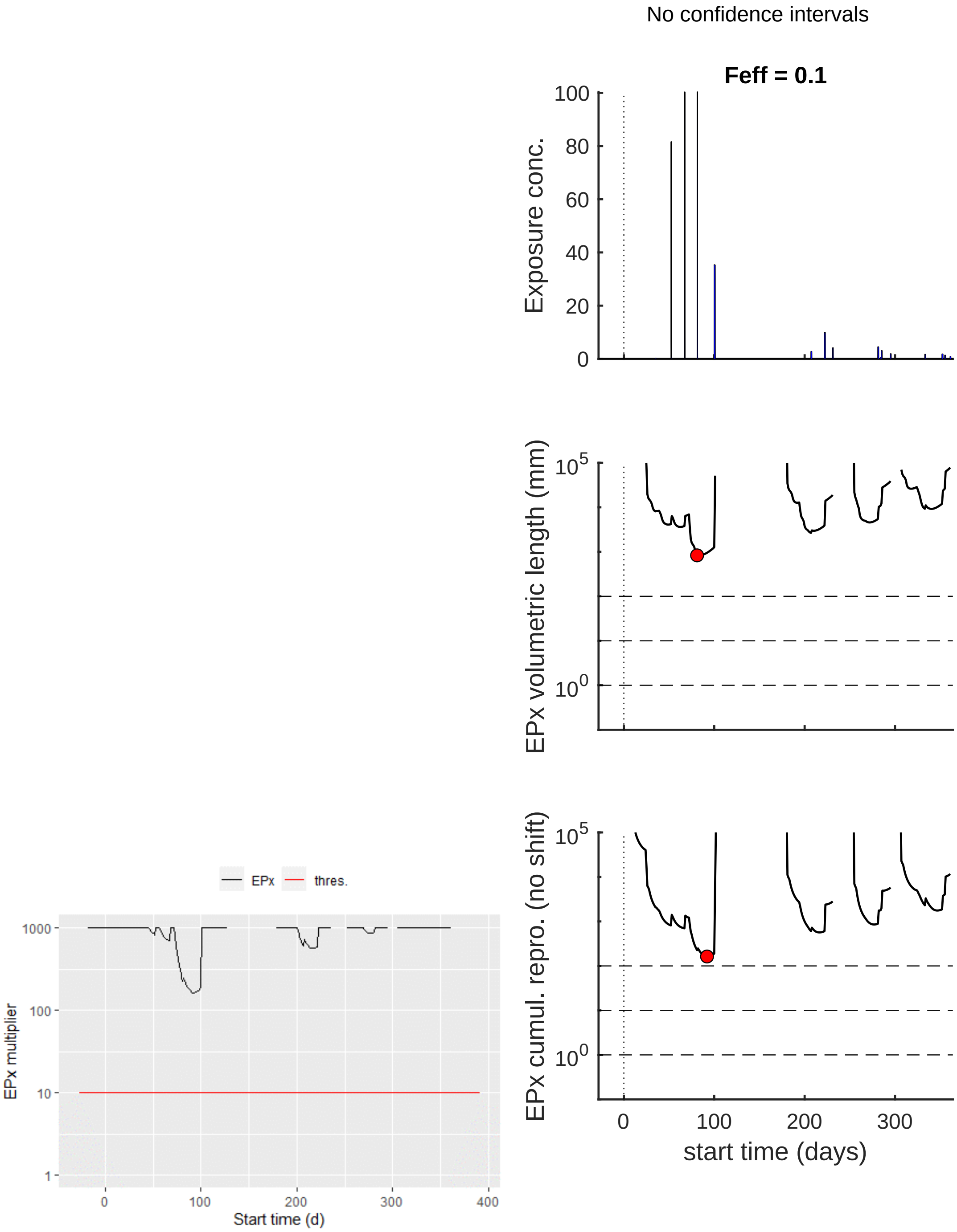

Figure 13: Result Test B, scenario apple R2 stream. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

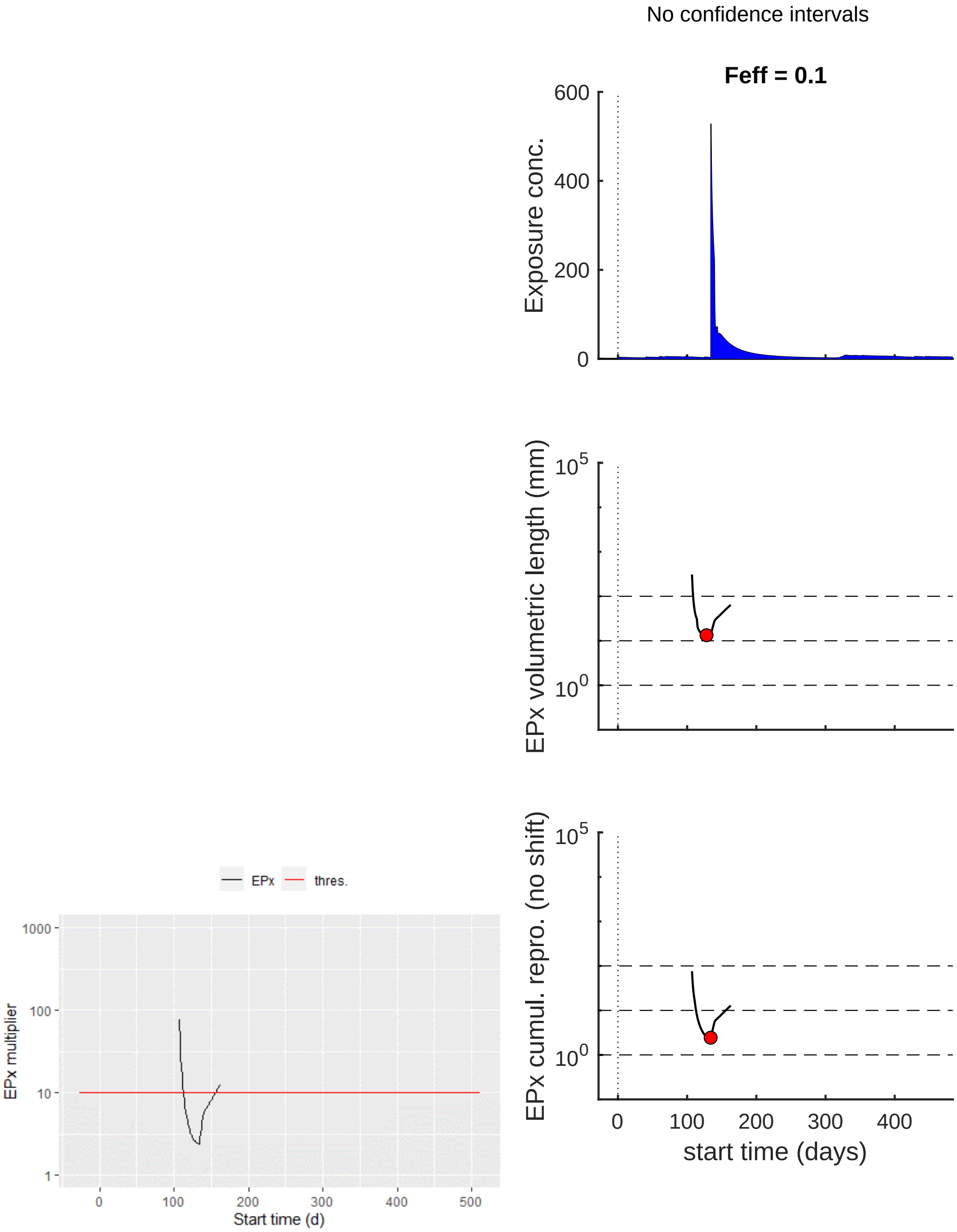

Figure 14: Result Test B, scenario cereal D1 ditch. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

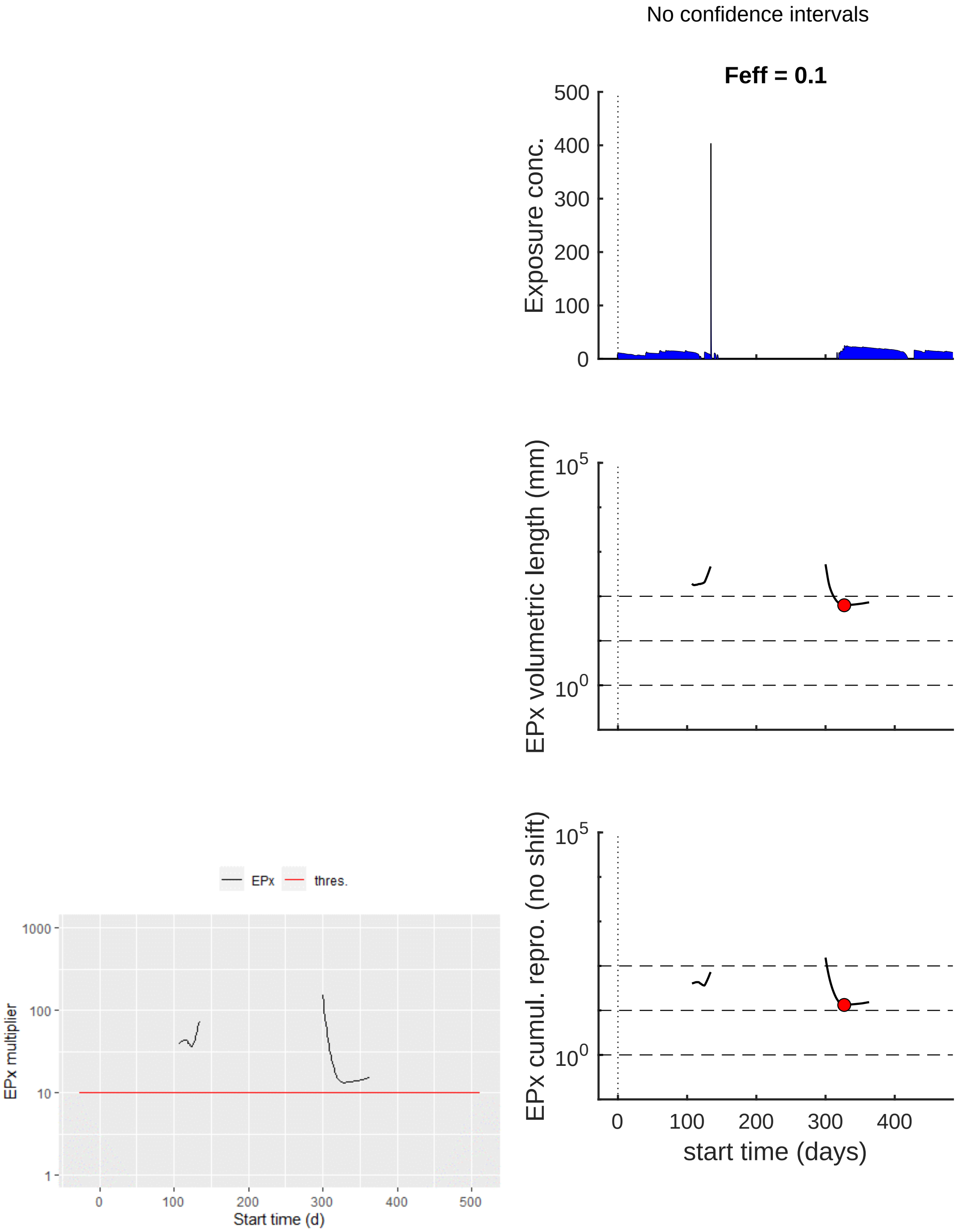

Figure 15: Result Test B, scenario cereal D1 stream. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

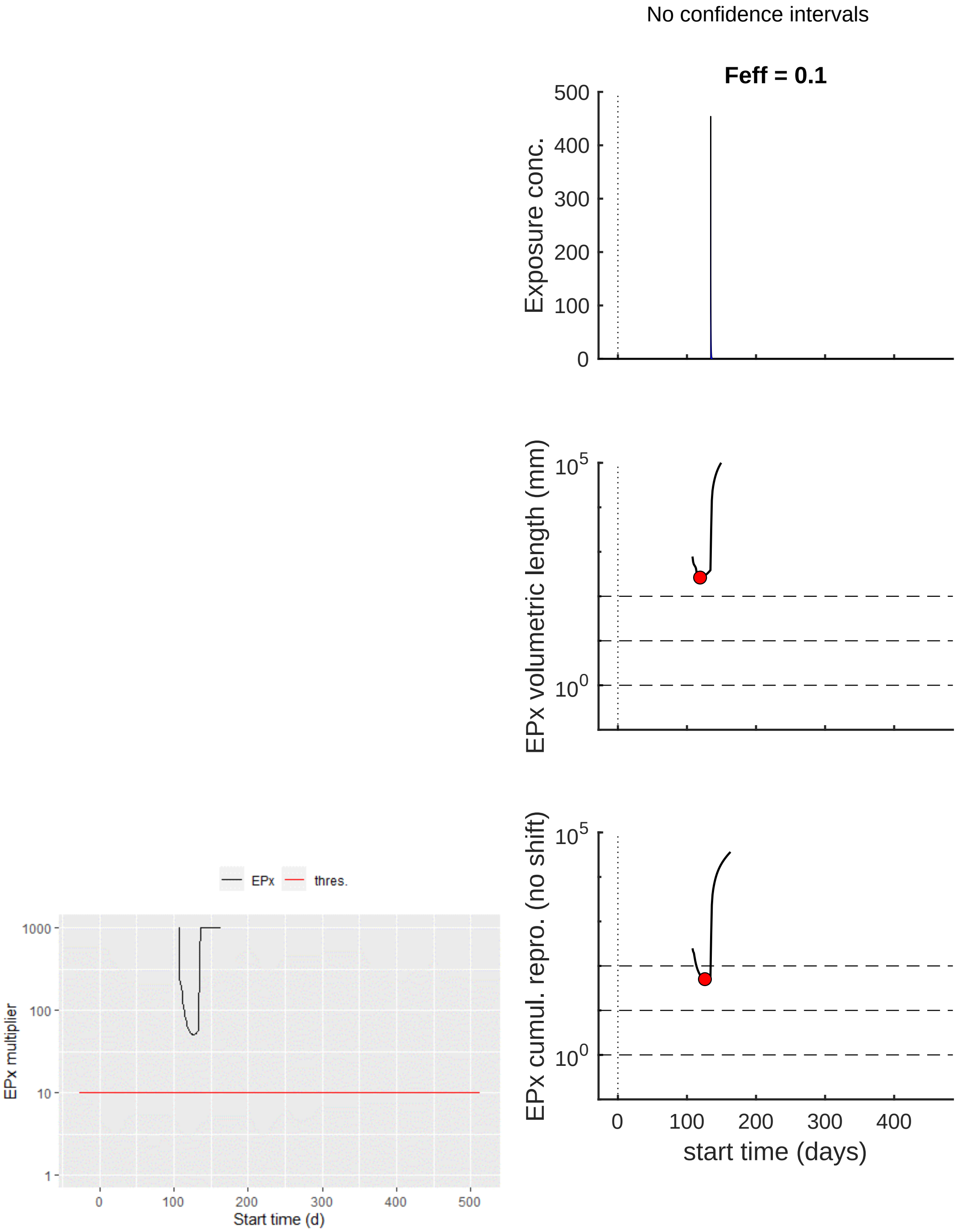

Figure 16: Result Test B, scenario cereal D3 ditch. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

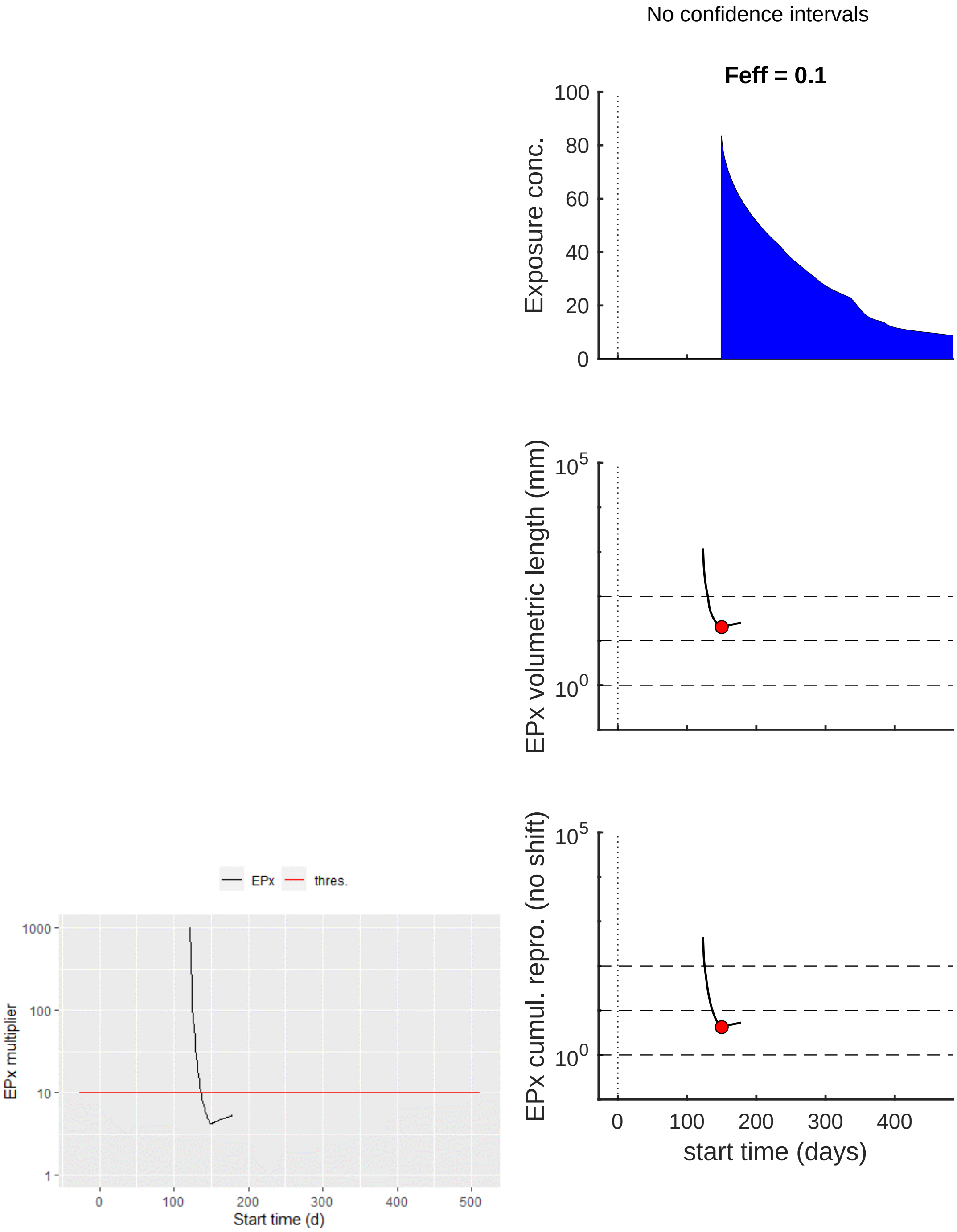

Figure 17: Result Test B, scenario cereal D4 pond. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

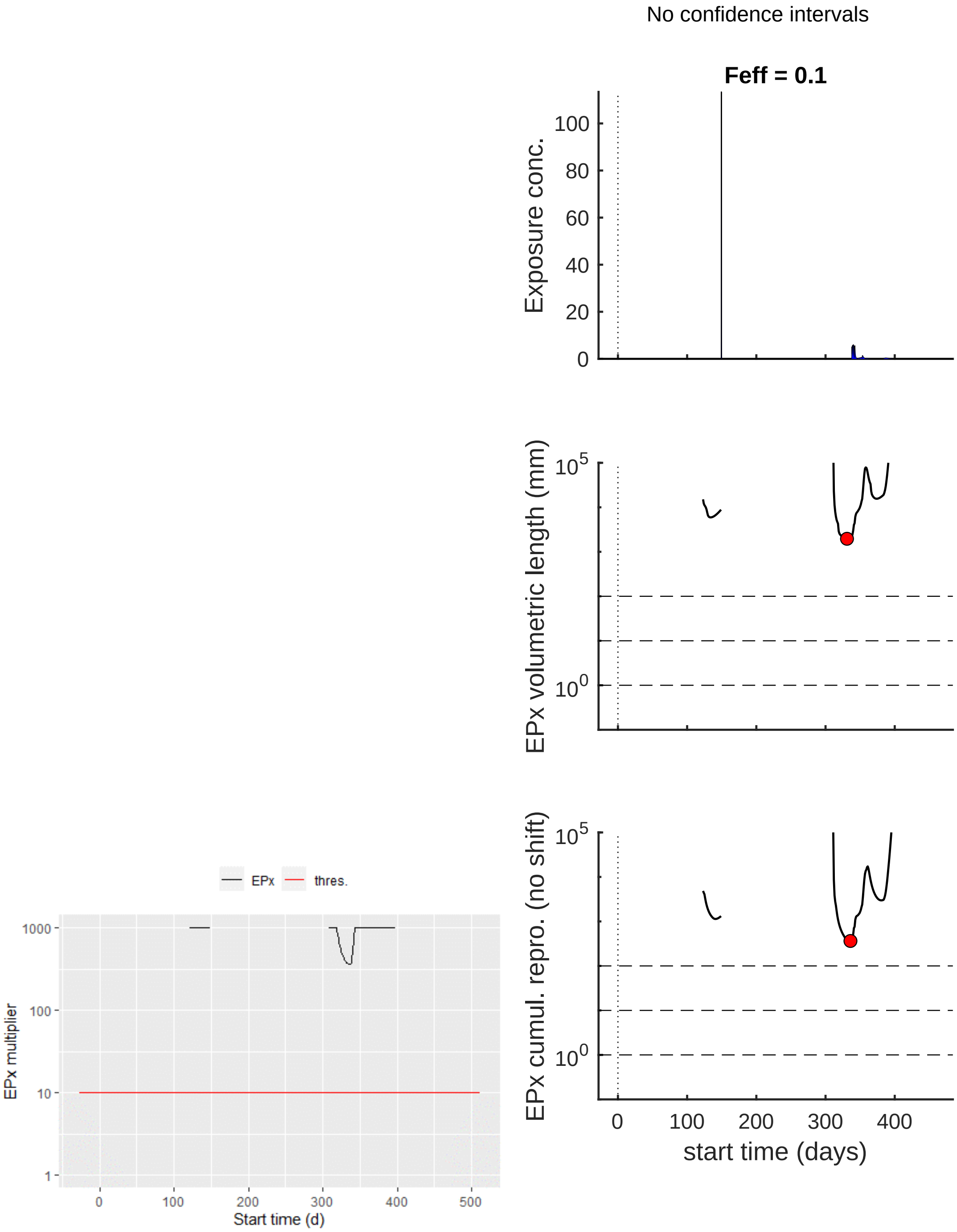

Figure 18: Result Test B, scenario cereal D4 stream. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

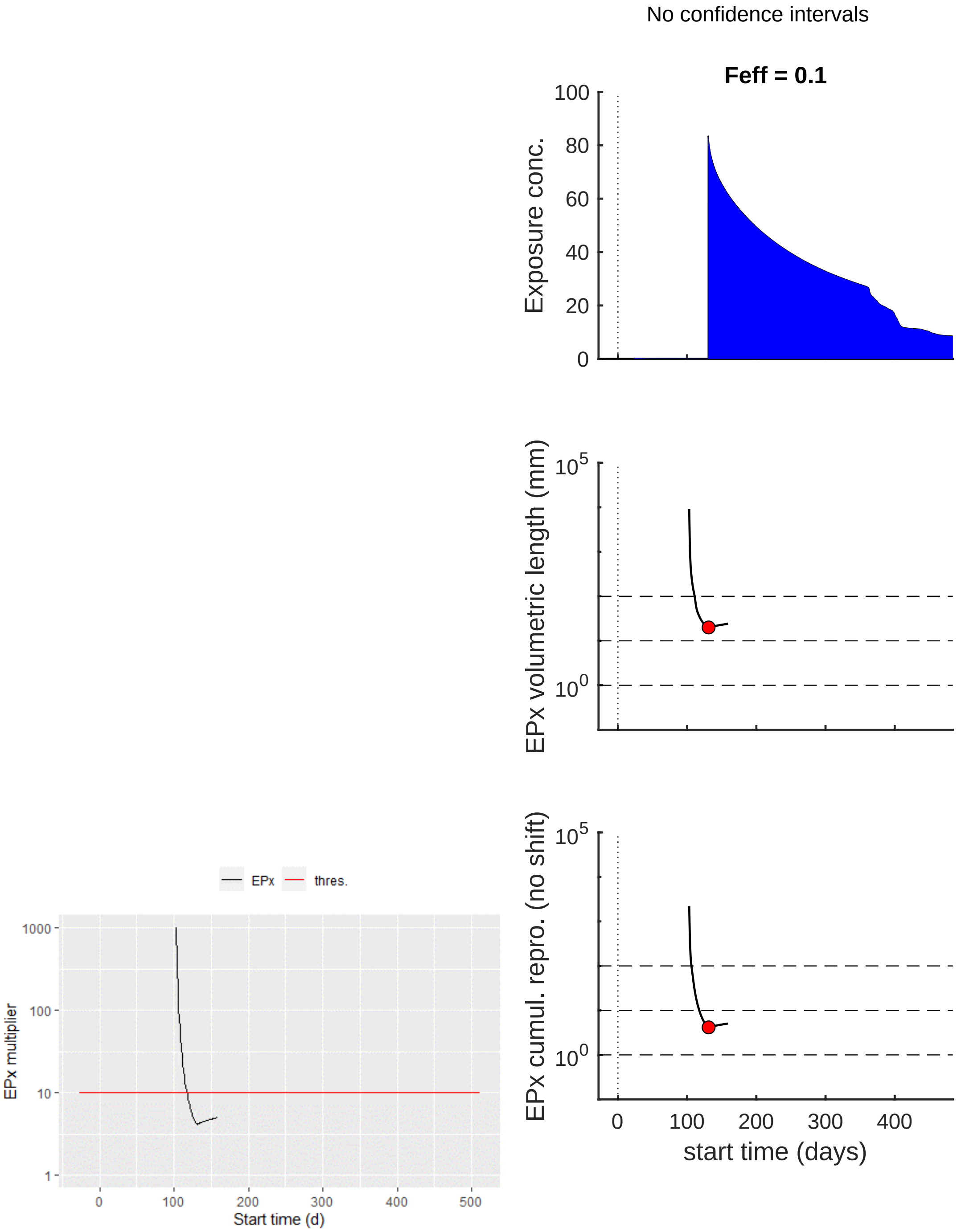

Figure 19: Result Test B, scenario cereal D5 pond. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

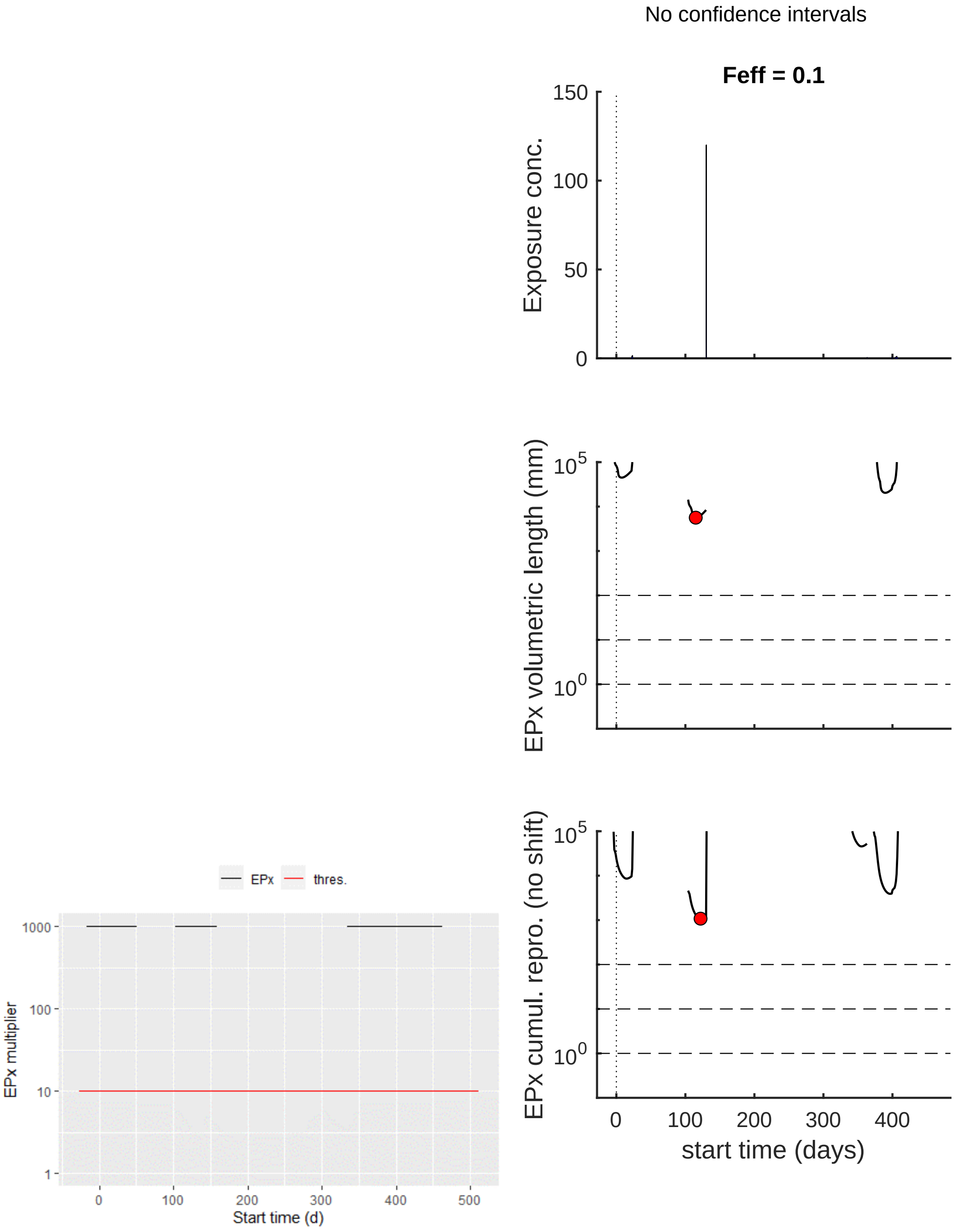

Figure 20: Result Test B, scenario cereal D5 stream. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

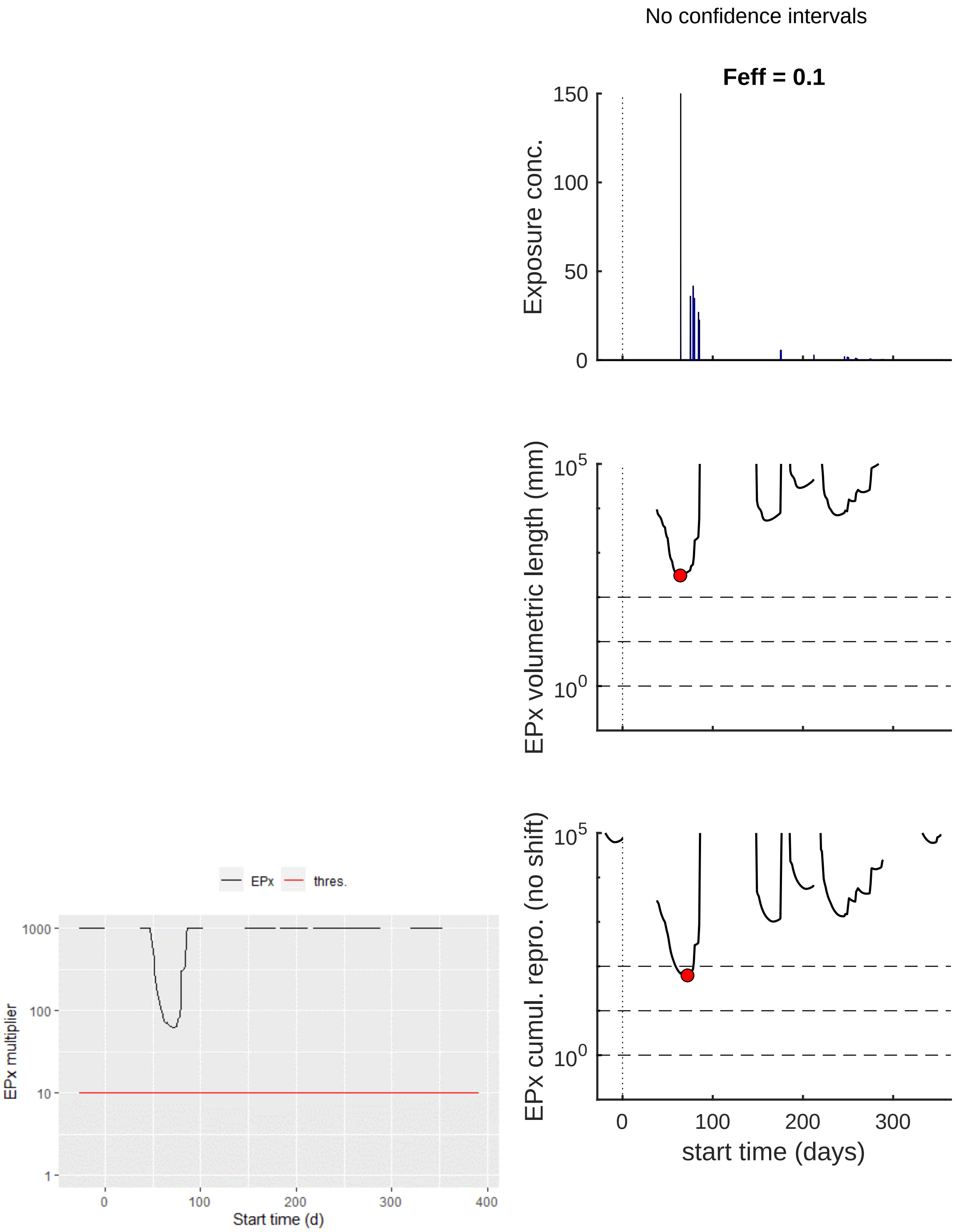

Figure 21: Result Test B, scenario cereal R4 stream. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

3.2 Test B: different x

This is the same calibration as Test A, and same settings for simulations, apart from one: instead of 10% effect, the software now looks for 50% effect. This is to test for issues with changing the effect percentage in the software implementations.

Settings for EPx calculation:

% effect:	50
Window length:	28 days
Timestep:	1 day
Cut-off:	1000 in DeEP, 100000 in BYOM
Pass/fail:	10
Thinning:	off
Modification of exposure profiles:	all exposure values times 50.

For all profiles, DeEP identifies reproduction as the most sensitive endpoint (i.e., lower EP50 than for length). DeEP does not provide a value for the EP50 on length. BYOM confirms that the EP50 for length is indeed higher than the EP50 for reproduction, for all exposure profiles. Note that all values are now higher than for the EP10 in Test A, as expected.

Table 2: EP50 values for test B with different x.

Exposure scenario	DeEP stdalone	DeEP web	BYOM
apple R1 pond	33.056	33.056	33.06 ($t = 91$)
apple R2 stream	909.722	909.722	909.72 ($t = 91$)
cereal D1 ditch	13.791	13.791	13.79 ($t = 133$)
cereal D1 stream	73.265	73.265	73.26 ($t = 327$)
cereal D3 ditch	281.663	281.663	281.66 ($t = 125$)
cereal D4 pond	23.257	23.257	23.26 ($t = 150$)
cereal D4 stream	>1000	>1000	2052.25 ($t = 335$)
cereal D5 pond	22.982	22.982	22.98 ($t = 131$)
cereal D5 stream	>1000	>1000	6022.44 ($t = 121$)
cereal R4 stream	348.888	348.888	348.89 ($t = 70$)

4 Test C: fast and slow kinetics

This is the same data set and settings as Test A, with one exception: the dominant rate constant k_d is fixed to either fast or slow kinetics. This provides a more extreme response to changes in exposure level, and places more strain on the ODE solvers.

4.1 Test C: fast kinetics

Calibration settings:

Calibration data set:	*Folsomia candida* exposed to cadmium in food.
Mode of action:	Assimilation ([10000]).
Feedbacks configuration:	No feedbacks ([0000]).

Parameter settings/estimates:

Param.	Calibrated value
L_0	0.1204
L_p	0.3954
L_m	0.6734
r_B	0.05865
R_m	15.60
f	1 (fixed)
k_d	10 (fixed)
z_b	51.91
b_b	$9.034 \cdot 10^{-5}$
F_{BV}	0.008 (fixed)
K_{RV}	1 (fixed)
κ	0.8 (fixed)
y_P	0.64 (fixed)
L_m^{ref}	0.7 (fixed)

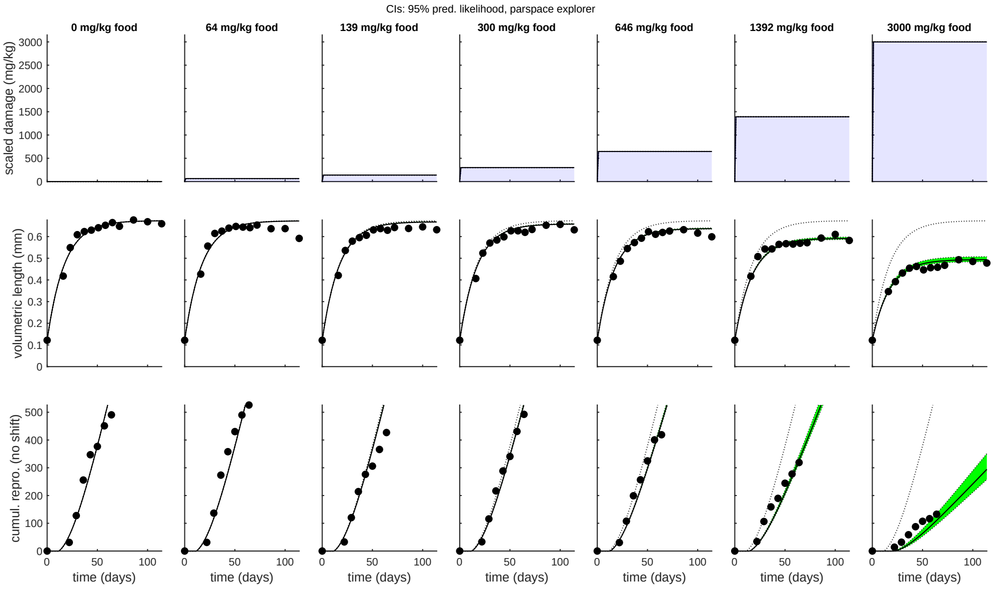

Figure 22: BYOM fit for test Test C with fast kinetics.

Settings for EPx calculation:

% effect:	10
Window length:	28 days
Timestep:	1 day
Cut-off:	1000 in DeEP, 100000 in BYOM
Pass/fail:	10
Thinning:	off
Modification of exposure profiles:	all exposure values times 50.

For all profiles, DeEP identifies reproduction as the most sensitive endpoint (i.e., lower EP10 than for length). DeEP does not provide a value for the EP10 on length. BYOM confirms that the EP10 for length is indeed higher than the EP10 for reproduction, for all exposure profiles.

Table 3: EP10 values for test C, fast kinetics. DeEP values in parentheses are for a previous version of the web application (run in May 2023), which accommodated a higher value for the cut-off.

Exposure scenario	DeEP stdalone	DeEP web	BYOM
apple R1 pond	5.816	5.816	5.82 ($t = 91$)
apple R2 stream	116.923	116.928	116.91 ($t = 74$)
cereal D1 ditch	2.538	2.538	2.54 ($t = 124$)
cereal D1 stream	12.627	12.627	12.63 ($t = 327$)
cereal D3 ditch	43.226	43.226	43.22 ($t = 123$)
cereal D4 pond	4.092	4.092	4.09 ($t = 149$)
cereal D4 stream	365.16	365.16	365.14 ($t = 328$)
cereal D5 pond	4.032	4.032	4.03 ($t = 130$)
cereal D5 stream	>1000	>1000 (2562.889)	2562.99 ($t = 119$)
cereal R4 stream	53.967	53.967	53.97 ($t = 64$)

The figures below compare the plots from the standalone to the plots from BYOM. Only the first two profiles are shown, which provide two distinct patterns: duration over a longer period of time and a few short pulses.

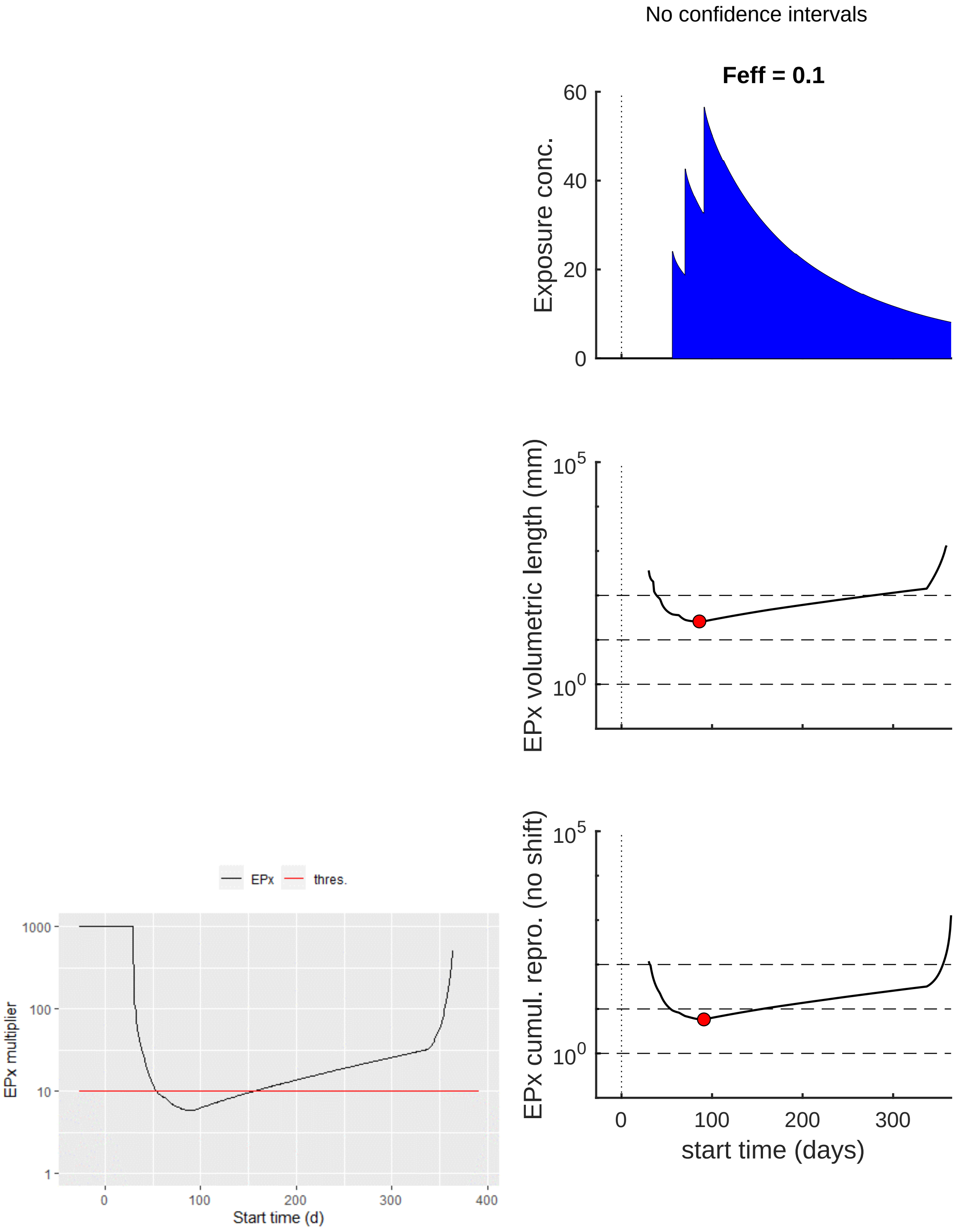

Figure 23: Result Test C-fast, scenario apple R1 pond. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

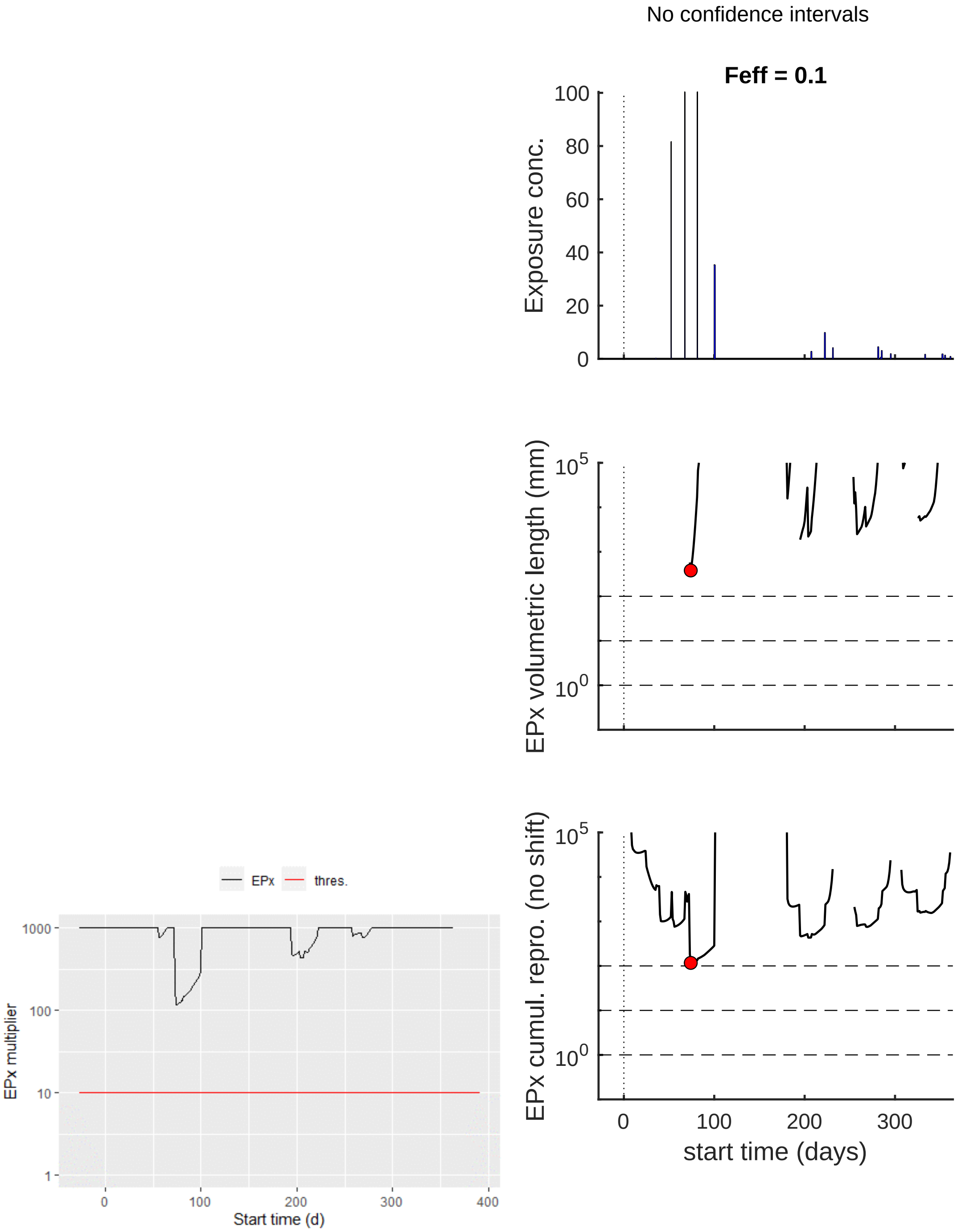

Figure 24: Result Test C-fast, scenario apple R2 stream. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

4.2 Test C: slow kinetics

The calibration, in this case, shows shrinking after long exposure. However, this shrinking occurs at very high effect levels, and around day 60, well after the length of the time window.

Calibration settings:

Calibration data set:	*Folsomia candida* exposed to cadmium in food.
Mode of action:	Assimilation ([10000]).
Feedbacks configuration:	No feedbacks ([0000]).

Parameter settings/estimates (* hitting bound of search range):

Param.	Calibrated value
L_0	0.1204
L_p	0.3954
L_m	0.6734
r_B	0.05865
R_m	15.60
f	1 (fixed)
k_d	0.01 (fixed)
z_b	0.1066*
b_b	$2.828 \cdot 10^{-4}$
F_{BV}	0.008 (fixed)
K_{RV}	1 (fixed)
κ	0.8 (fixed)
y_P	0.64 (fixed)
L_m^{ref}	0.7 (fixed)

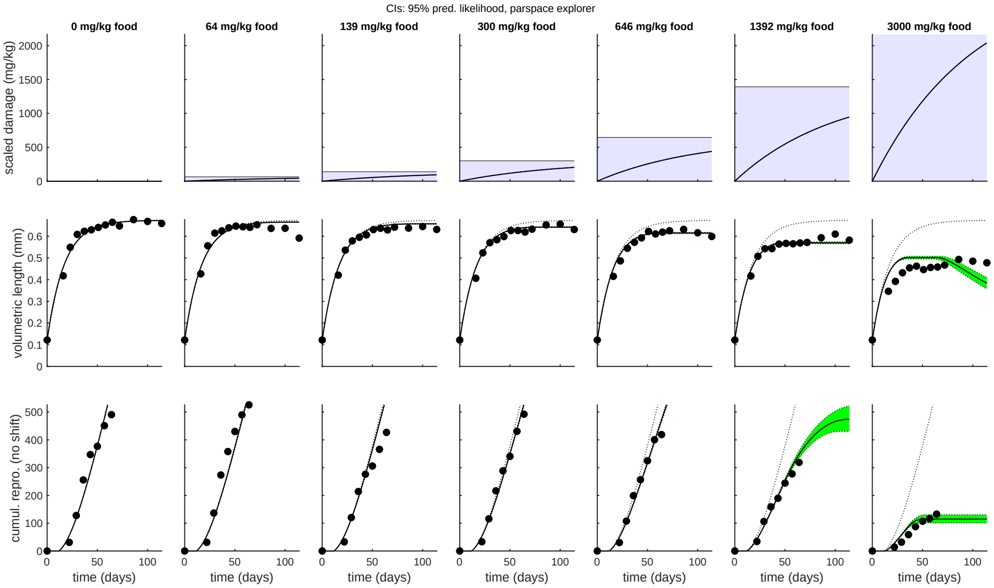

Figure 25: BYOM fit for test Test C with slow kinetics.

Settings for EPx calculation:

% effect:	10
Window length:	28 days
Timestep:	1 day
Cut-off:	1000 in DeEP, 100000 in BYOM
Pass/fail:	10
Thinning:	off
Modification of exposure profiles:	all exposure values times 50.

For all profiles, DeEP identifies reproduction as the most sensitive endpoint (i.e., lower EP10 than for length). DeEP does not provide a value for the EP10 on length. BYOM confirms that the EP10 for length is indeed higher than the EP10 for reproduction, for all exposure profiles.

Table 4: EP10 values for test C, slow kinetics. DeEP values in parentheses are for a previous version of the web application (run in May 2023), which accommodated a higher value for the cut-off.

Exposure scenario	DeEP stdalone	DeEP web	BYOM
apple R1 pond	10.784	10.784	10.78 ($t = 91$)
apple R2 stream	232.584	232.584	232.57 ($t = 100$)
cereal D1 ditch	3.711	3.711	3.71 ($t = 134$)
cereal D1 stream	24.145	24.145	24.14 ($t = 327$)
cereal D3 ditch	71.137	71.137	71.14 ($t = 134$)
cereal D4 pond	7.554	7.554	7.55 ($t = 150$)
cereal D4 stream	543.948	543.944	543.93 ($t = 338$)
cereal D5 pond	7.48	7.48	7.48 ($t = 131$)
cereal D5 stream	>1000	>1000 (1514.504)	1514.51 ($t = 130$)
cereal R4 stream	93.873	93.873	93.87 ($t = 75$)

The figures below compare the plots from the standalone to the plots from BYOM. Only the first two profiles are shown, which provide two distinct patterns: duration over a longer period of time and a few short pulses.

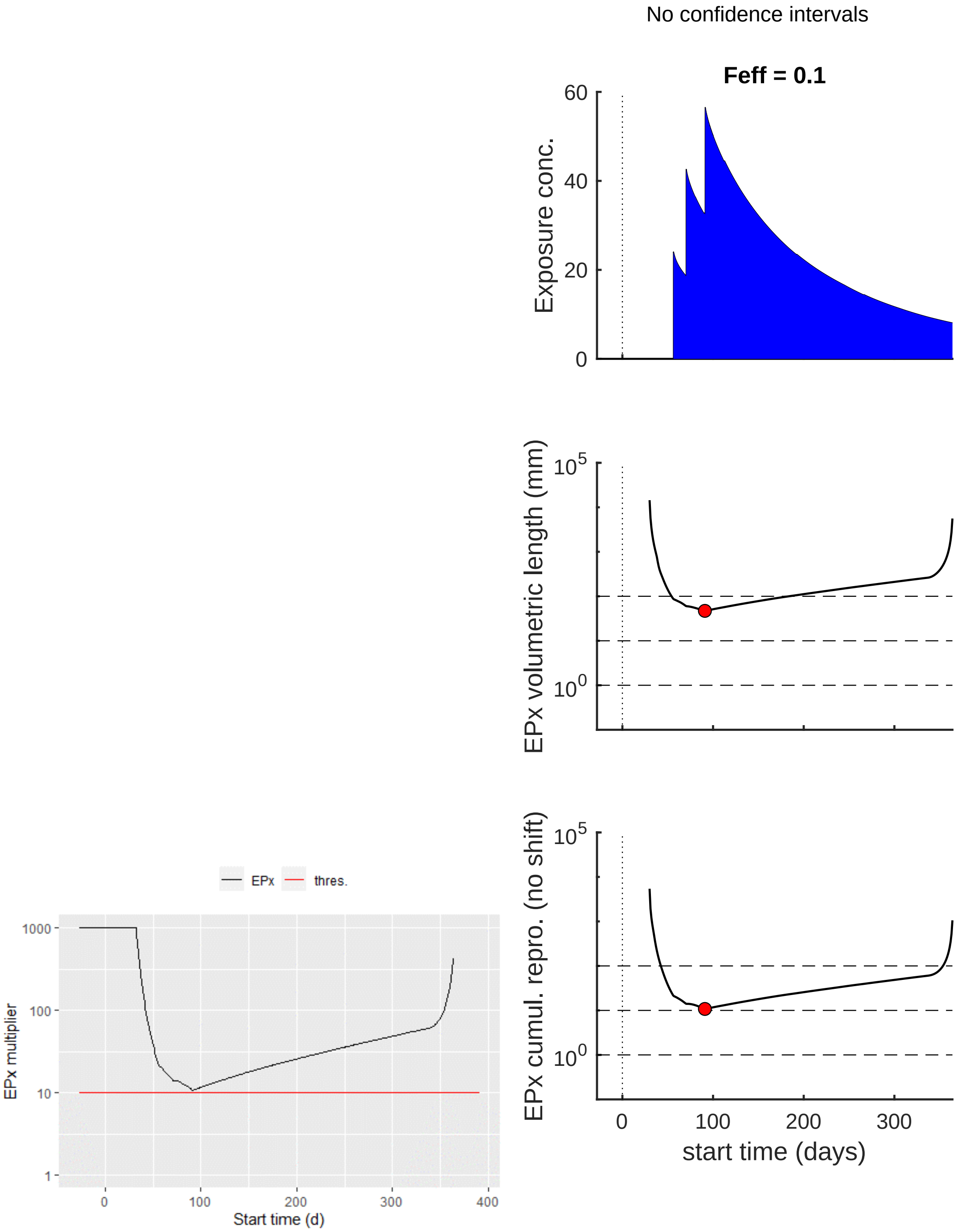

Figure 26: Result Test C-slow, scenario apple R1 pond. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

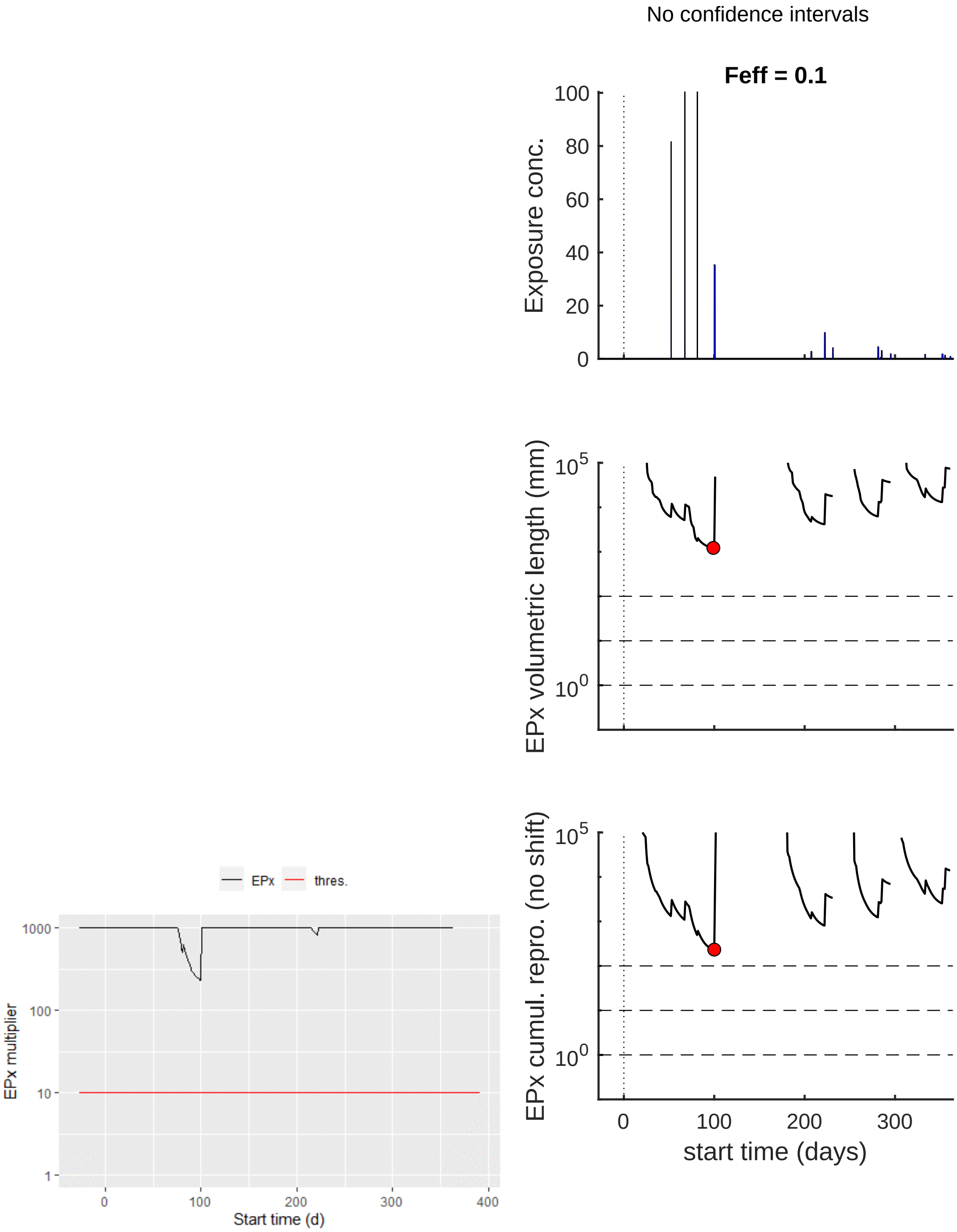

Figure 27: Result Test C-slow, scenario apple R2 stream. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

5 Test D: different feedback configurations

This is the same data set and settings as Test A, with one exception: different feedback configurations are used.

5.1 Test D: all feedbacks

Calibration settings:

Calibration data set:	*Folsomia candida* exposed to cadmium in food.
Mode of action:	Assimilation ([10000]).
Feedbacks configuration:	All feedbacks ([1111]).

Parameter settings/estimates (* hitting bound of search range):

Param.	Calibrated value
L_0	0.1204
L_p	0.3954
L_m	0.6734
r_B	0.05865
R_m	15.60
f	1 (fixed)
k_d	0.2112
z_b	0.1066*
b_b	$1.018 \cdot 10^{-4}$
F_{BV}	0.008 (fixed)
K_{RV}	1 (fixed)
κ	0.8 (fixed)
y_P	0.64 (fixed)
L_m^{ref}	0.7 (fixed)

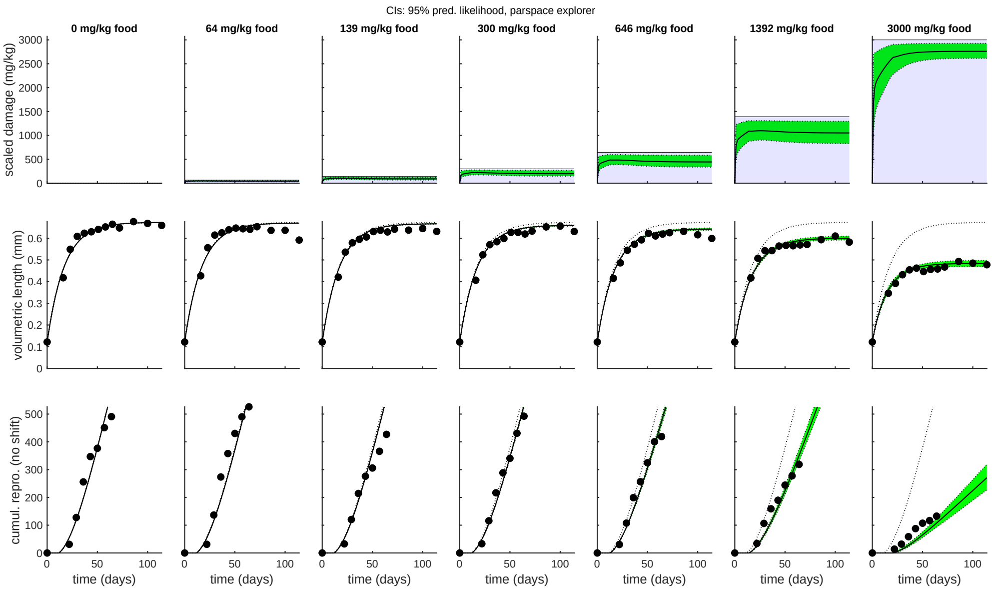

Figure 28: BYOM fit for test Test D, all feedbacks.

Settings for EPx calculation:

% effect:	10
Window length:	28 days
Timestep:	1 day
Cut-off:	1000 in DeEP, 100000 in BYOM
Pass/fail:	10
Thinning:	off
Modification of exposure profiles:	all exposure values times 50.

For all profiles, DeEP identifies reproduction as the most sensitive endpoint (i.e., lower EP10 than for length). DeEP does not provide a value for the EP10 on length. BYOM confirms that the EP10 for length is indeed higher than the EP10 for reproduction, for all exposure profiles.

Table 5: EP10 values for test D, all feedbacks. DeEP values in parentheses are for a previous version of the web application (run in May 2023), which accommodated a higher value for the cut-off.

Exposure scenario	DeEP stdalone	DeEP web	BYOM
apple R1 pond	5.854	5.854	5.85 ($t = 91$)
apple R2 stream	153.059	153.059	153.05 ($t = 89$)
cereal D1 ditch	2.499	2.499	2.50 ($t = 127$)
cereal D1 stream	12.826	12.826	12.83 ($t = 327$)
cereal D3 ditch	47.434	47.434	47.43 ($t = 123$)
cereal D4 pond	4.125	4.125	4.12 ($t = 149$)
cereal D4 stream	363.894	363.895	363.90 ($t = 330$)
cereal D5 pond	4.071	4.071	4.07 ($t = 130$)
cereal D5 stream	>1000	>1000 (1004.476)	1004.48 ($t = 119$)
cereal R4 stream	60.391	60.391	60.39 ($t = 64$)

5.2 Test D: classic feedbacks

Calibration settings:

Calibration data set:	*Folsomia candida* exposed to cadmium in food.
Mode of action:	Assimilation ([10000]).
Feedbacks configuration:	Classic feedbacks ([1110]).

Parameter settings/estimates (* hitting bound of search range):

Param.	Calibrated value
L_0	0.1204
L_p	0.3954
L_m	0.6734
r_B	0.05865
R_m	15.60
f	1 (fixed)
k_d	0.1361
z_b	0.1066*
b_b	$9.965 \cdot 10^{-5}$
F_{BV}	0.008 (fixed)
K_{RV}	1 (fixed)
κ	0.8 (fixed)
y_P	0.64 (fixed)
L_m^{ref}	0.7 (fixed)

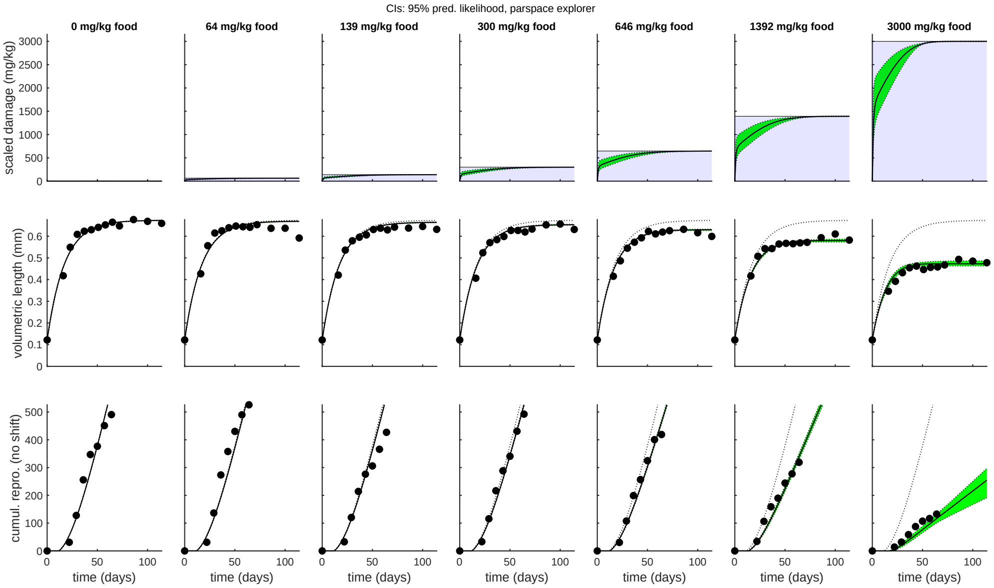

Figure 29: BYOM fit for test Test D, classic feedbacks.

Settings for EPx calculation:
% effect: 10
Window length: 28 days
Timestep: 1 day
Cut-off: 1000 in DeEP, 100000 in BYOM
Pass/fail: 10
Thinning: off
Modification of exposure profiles: all exposure values times 50.

For all profiles, DeEP identifies reproduction as the most sensitive endpoint (i.e., lower EP10 than for length). DeEP does not provide a value for the EP10 on length. BYOM confirms that the EP10 for length is indeed higher than the EP10 for reproduction, for all exposure profiles.

Additionally, for this test, a 'robust' calculation with BYOM is performed. Cases with a PMoA that affects growth, and feedbacks that include an effect on the elimination rate, have the (small) chance of yielding more than one EP10 value (see [10]). The robust calculation in BYOM ensure that the lower one is calculated, albeit with less precision (linear interpolation is used rather than zooming in with `fzero`). Nevertheless, for this case, the robust calculations yield almost identical results to the regular algorithm.[3] In this case, the EP10 from the standard calculation has thus indeed managed to locate the true lowest EP10.

Table 6: EP10 values for test D, classic feedbacks. Asterisk indicates that EP10 is higher than the highest value tried in the robust analysis (the time point is the first point at which this threshold was reached). DeEP values in parentheses are for a previous version of the web application (run in May 2023), which accommodated a higher value for the cut-off.

Exposure scenario	DeEP stdalone	DeEP web	BYOM	Robust
apple R1 pond	6.434	6.434	6.43 ($t = 91$)	6.43 ($t = 91$)
apple R2 stream	164.854	164.854	164.85 ($t = 88$)	164.85 ($t = 88$)
cereal D1 ditch	2.71	2.71	2.71 ($t = 126$)	2.71 ($t = 126$)
cereal D1 stream	14.107	14.107	14.11 ($t = 327$)	14.11 ($t = 327$)
cereal D3 ditch	51.057	51.057	51.06 ($t = 123$)	51.06 ($t = 123$)
cereal D4 pond	4.533	4.533	4.53 ($t = 149$)	4.53 ($t = 149$)
cereal D4 stream	396.47	396.469	396.48 ($t = 330$)	300* ($t = 123$)
cereal D5 pond	4.474	4.474	4.47 ($t = 130$)	4.47 ($t = 130$)
cereal D5 stream	>1000	>1000 (1086.104)	1086.11 ($t = 118$)	300* ($t = -4$)
cereal R4 stream	64.71	64.71	64.71 ($t = 64$)	64.71 ($t = 64$)

---

[3] On closer inspection, this is caused by the fact the that relationship between the multiplication factor and the effect on reproduction is almost linear in this effect range (which could relate to the extremely low threshold z_b).

5.3 Test D: growth dilution only

Calibration settings:

Calibration data set:	*Folsomia candida* exposed to cadmium in food.
Mode of action:	Assimilation ([10000]).
Feedbacks configuration:	Growth dilution ([0010]).

Parameter settings/estimates (* hitting bound of search range):

Param.	Calibrated value
L_0	0.1204
L_p	0.3954
L_m	0.6734
r_B	0.05865
R_m	15.60
f	1 (fixed)
k_d	0.3344
z_b	0.1066*
b_b	$9.732 \cdot 10^{-5}$
F_{BV}	0.008 (fixed)
K_{RV}	1 (fixed)
κ	0.8 (fixed)
y_P	0.64 (fixed)
L_m^{ref}	0.7 (fixed)

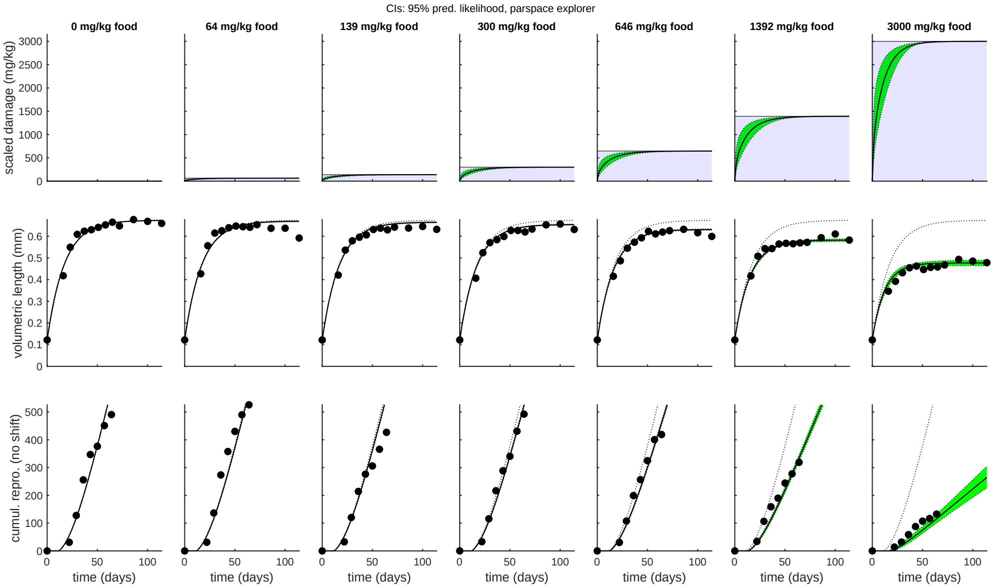

Figure 30: BYOM fit for test Test D, growth dilution only.

Settings for EPx calculation:

% effect:	10
Window length:	28 days
Timestep:	1 day
Cut-off:	1000 in DeEP, 100000 in BYOM
Pass/fail:	10
Thinning:	off
Modification of exposure profiles:	all exposure values times 50.

For all profiles, DeEP identifies reproduction as the most sensitive endpoint (i.e., lower EP10 than for length). DeEP does not provide a value for the EP10 on length. BYOM confirms that the EP10 for length is indeed higher than the EP10 for reproduction, for all exposure profiles.

Table 7: EP10 values for test D, growth dilution only. DeEP values in parentheses are for a previous version of the web application (run in May 2023), which accommodated a higher value for the cut-off.

Exposure scenario	DeEP stdalone	DeEP web	BYOM
apple R1 pond	6.249	6.249	6.25 ($t = 88$)
apple R2 stream	158.927	158.927	158.92 ($t = 86$)
cereal D1 ditch	2.631	2.631	2.63 ($t = 125$)
cereal D1 stream	13.671	13.671	13.67 ($t = 327$)
cereal D3 ditch	49.225	49.225	49.23 ($t = 120$)
cereal D4 pond	4.419	4.419	4.42 ($t = 149$)
cereal D4 stream	384.991	384.99	384.99 ($t = 329$)
cereal D5 pond	4.356	4.356	4.36 ($t = 130$)
cereal D5 stream	>1000	>1000 (1044.356)	1044.36 ($t = 116$)
cereal R4 stream	61.402	61.402	61.40 ($t = 64$)

5.4 Test D: elimination scaling only

Calibration settings:

Calibration data set:	*Folsomia candida* exposed to cadmium in food.
Mode of action:	Assimilation ([10000]).
Feedbacks configuration:	Elimination only ([0100]).

Parameter settings/estimates (* hitting bound of search range):

Param.	Calibrated value
L_0	0.1204
L_p	0.3954
L_m	0.6734
r_B	0.05865
R_m	15.60
f	1 (fixed)
k_d	0.2491
z_b	0.1066*
b_b	$1.379 \cdot 10^{-4}$
F_{BV}	0.008 (fixed)
K_{RV}	1 (fixed)
κ	0.8 (fixed)
y_P	0.64 (fixed)
L_m^{ref}	0.7 (fixed)

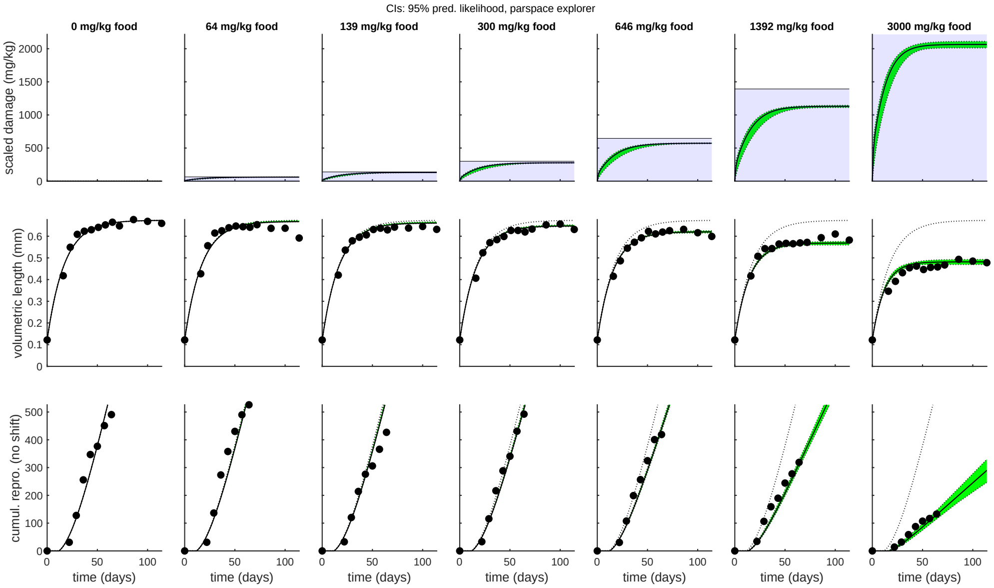

Figure 31: BYOM fit for test Test D, elimination feedback only.

Settings for EPx calculation:
% effect: 10
Window length: 28 days
Timestep: 1 day
Cut-off: 1000 in DeEP, 100000 in BYOM
Pass/fail: 10
Thinning: off
Modification of exposure profiles: all exposure values times 50.

For all profiles, DeEP identifies reproduction as the most sensitive endpoint (i.e., lower EP10 than for length). DeEP does not provide a value for the EP10 on length. BYOM confirms that the EP10 for length is indeed higher than the EP10 for reproduction, for all exposure profiles.

Table 8: EP10 values for test D, elimination scaling only. Asterisk indicates that EP10 is higher than the highest value tried in the robust analysis (the time point is the first point at which this threshold was reached).

Exposure scenario	DeEP stdalone	DeEP web	BYOM	Robust
apple R1 pond	5.694	5.694	5.69 ($t = 87$)	5.69 ($t = 87$)
apple R2 stream	144.727	144.727	144.72 ($t = 85$)	144.72 ($t = 85$)
cereal D1 ditch	2.372	2.372	2.37 ($t = 124$)	2.37 ($t = 124$)
cereal D1 stream	12.489	12.489	12.49 ($t = 326$)	12.49 ($t = 326$)
cereal D3 ditch	44.86	44.86	44.86 ($t = 119$)	44.86 ($t = 119$)
cereal D4 pond	4.046	4.046	4.05 ($t = 149$)	4.05 ($t = 149$)
cereal D4 stream	350.294	350.296	350.30 ($t = 328$)	300* ($t = 123$)
cereal D5 pond	3.986	3.986	3.99 ($t = 130$)	3.99 ($t = 130$)
cereal D5 stream	957.919	957.919	957.92 ($t = 115$)	300* ($t = -4$)
cereal R4 stream	55.054	55.054	55.05 ($t = 63$)	55.05 ($t = 63$)

6 Test E: other PMoAs affecting growth

This is the same data set and settings as Test A, with one exception: different modes of action (PMoA) are used. These mode of actions all affect growth (indirectly or directly).

6.1 Test E: maintenance

Calibration settings:

Calibration data set:	*Folsomia candida* exposed to cadmium in food.
Mode of action:	Assimilation ([01000]).
Feedbacks configuration:	Classic feedbacks ([1110]).

Parameter settings/estimates (* hitting bound of search range):

Param.	Calibrated value
L_0	0.1204
L_p	0.3954
L_m	0.6734
r_B	0.05865
R_m	15.60
f	1 (fixed)
k_d	10*
z_b	0.1066*
b_b	$1.369 \cdot 10^{-4}$
F_{BV}	0.008 (fixed)
K_{RV}	1 (fixed)
κ	0.8 (fixed)
y_P	0.64 (fixed)
L_m^{ref}	0.7 (fixed)

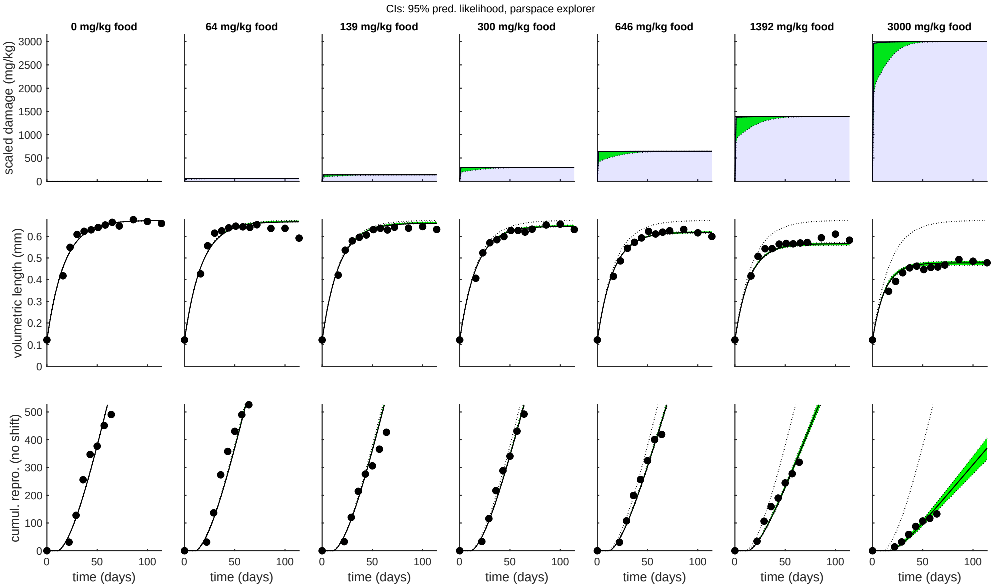

Figure 32: BYOM fit for test Test E, maintenance.

Settings for EPx calculation:

% effect:	10
Window length:	28 days
Timestep:	1 day
Cut-off:	1000 in DeEP, 100000 in BYOM
Pass/fail:	10
Thinning:	off
Modification of exposure profiles:	all exposure values times 50.

For all profiles, DeEP identifies reproduction as the most sensitive endpoint (i.e., lower EP10 than for length). DeEP does not provide a value for the EP10 on length. BYOM confirms that the EP10 for length is indeed higher than the EP10 for reproduction, for all exposure profiles.

Table 9: EP10 values for test E, maintenance.

Exposure scenario	DeEP stdalone	DeEP web	BYOM
apple R1 pond	6.242	6.242	6.24 ($t = 86$)
apple R2 stream	121.947	121.947	121.92 ($t = 74$)
cereal D1 ditch	2.539	2.539	2.54 ($t = 124$)
cereal D1 stream	13.691	13.691	13.69 ($t = 327$)
cereal D3 ditch	50.872	50.872	50.87 ($t = 116$)
cereal D4 pond	4.45	4.45	4.45 ($t = 149$)
cereal D4 stream	371.08	371.089	371.06 ($t = 328$)
cereal D5 pond	4.381	4.381	4.38 ($t = 130$)
cereal D5 stream	940.783	940.775	940.69 ($t = 119$)
cereal R4 stream	60.107	60.107	60.10 ($t = 64$)

6.2 Test E: growth and reproduction costs

Calibration settings:

Calibration data set:	*Folsomia candida* exposed to cadmium in food.
Mode of action:	Assimilation ([00110]).
Feedbacks configuration:	Classic feedbacks ([1110]).

Parameter settings/estimates (* hitting bound of search range):

Param.	Calibrated value
L_0	0.1204
L_p	0.3954
L_m	0.6734
r_B	0.05865
R_m	15.60
f	1 (fixed)
k_d	0.01*
z_b	111.9
b_b	$4.631 \cdot 10^{-3}$
F_{BV}	0.008 (fixed)
K_{RV}	1 (fixed)
κ	0.8 (fixed)
y_P	0.64 (fixed)
L_m^{ref}	0.7 (fixed)

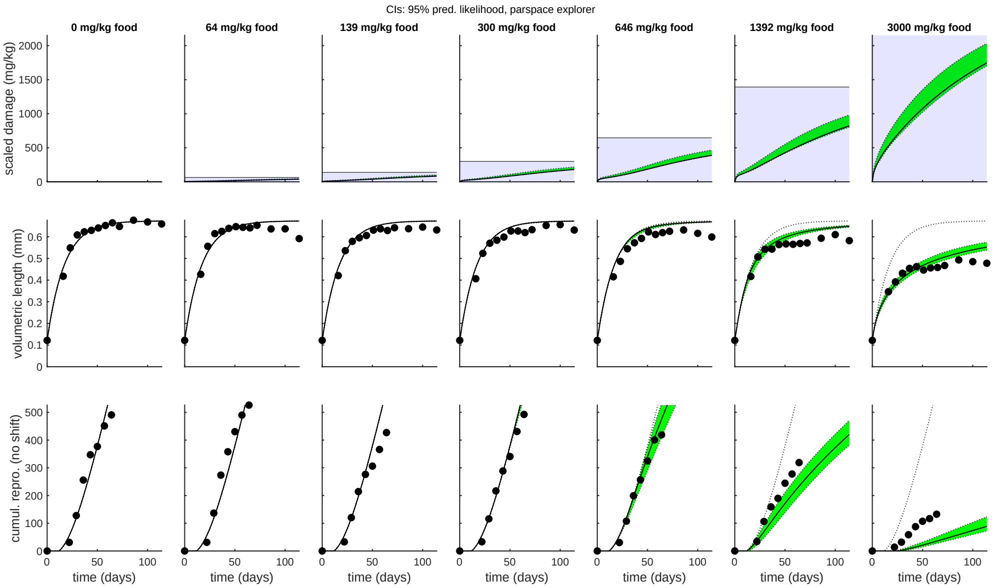

Figure 33: BYOM fit for test Test E, growth and repro costs.

Settings for EPx calculation:

% effect:	10
Window length:	28 days
Timestep:	1 day
Cut-off:	1000 in DeEP, 100000 in BYOM
Pass/fail:	10
Thinning:	off
Modification of exposure profiles:	all exposure values times 50.

For all profiles, DeEP identifies reproduction as the most sensitive endpoint (i.e., lower EP10 than for length). DeEP does not provide a value for the EP10 on length. BYOM confirms that the EP10 for length is indeed higher than the EP10 for reproduction, for all exposure profiles.

Table 10: EP10 values for test E, growth and reproduction. DeEP values in parentheses are for a previous version of the web application (run in May 2023), which accommodated a higher value for the cut-off.

Exposure scenario	DeEP stdalone	DeEP web	BYOM
apple R1 pond	19.56	19.56	19.56 ($t = 86$)
apple R2 stream	245.11	245.11	245.10 ($t = 100$)
cereal D1 ditch	7.053	7.053	7.05 ($t = 123$)
cereal D1 stream	43.25	43.25	43.25 ($t = 325$)
cereal D3 ditch	74.187	74.186	74.19 ($t = 134$)
cereal D4 pond	14.007	14.007	14.01 ($t = 147$)
cereal D4 stream	946.334	946.334	946.33 ($t = 338$)
cereal D5 pond	13.827	13.827	13.83 ($t = 129$)
cereal D5 stream	>1000	>1000 (1500.977)	1501.00 ($t = 130$)
cereal R4 stream	158.024	158.024	158.03 ($t = 62$)

6.3 Test E: growth costs

Calibration settings:

Calibration data set:	*Folsomia candida* exposed to cadmium in food.
Mode of action:	Assimilation ([00100]).
Feedbacks configuration:	Classic feedbacks ([1110]).

Parameter settings/estimates (* hitting bound of search range):

Param.	Calibrated value
L_0	0.1204
L_p	0.3954
L_m	0.6734
r_B	0.05865
R_m	15.60
f	1 (fixed)
k_d	0.01*
z_b	7.665
b_b	0.003334
F_{BV}	0.008 (fixed)
K_{RV}	1 (fixed)
κ	0.8 (fixed)
y_P	0.64 (fixed)
L_m^{ref}	0.7 (fixed)

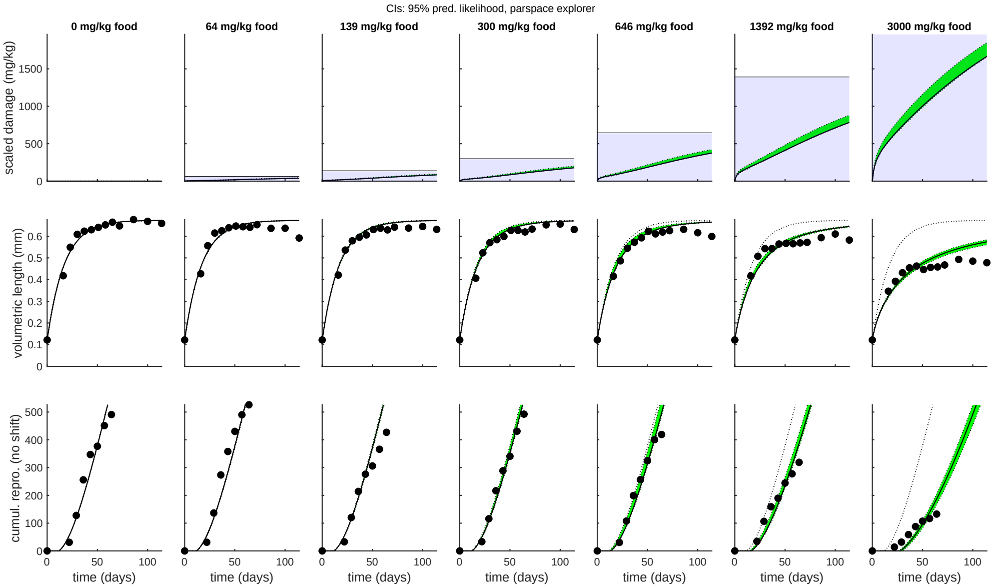

Figure 34: BYOM fit for test Test E, growth costs.

Settings for EPx calculation:

% effect:	10
Window length:	28 days
Timestep:	1 day
Cut-off:	1000 in DeEP, 100000 in BYOM
Pass/fail:	10
Thinning:	off
Modification of exposure profiles:	all exposure values times 50.

For all profiles, DeEP identifies reproduction as the most sensitive endpoint (i.e., lower EP10 than for length). DeEP does not provide a value for the EP10 on length. BYOM confirms that the EP10 for length is indeed higher than the EP10 for reproduction, for all exposure profiles.

Table 11: EP10 values for test E, growth costs.

Exposure scenario	DeEP stdalone	DeEP web	BYOM
apple R1 pond	6.199	6.199	6.20 ($t = 91$)
apple R2 stream	108.464	108.464	108.46 ($t = 96$)
cereal D1 ditch	1.88	1.88	1.88 ($t = 134$)
cereal D1 stream	14.062	14.062	14.06 ($t = 327$)
cereal D3 ditch	33.523	33.523	33.52 ($t = 131$)
cereal D4 pond	4.344	4.344	4.34 ($t = 150$)
cereal D4 stream	274.421	274.422	274.41 ($t = 337$)
cereal D5 pond	4.311	4.311	4.31 ($t = 131$)
cereal D5 stream	712.922	712.922	712.93 ($t = 126$)
cereal R4 stream	45.469	45.469	45.47 ($t = 75$)

7 Test F: direct effects on repro

For the tests in this section, a different data set is used, with effects on reproduction only. This allows testing the PMoA's with a direct effect on reproduction.

7.1 Test F: repro costs

Calibration settings:

Calibration data set:	*Folsomia candida* exposed to chlorpyrifos in food.
Mode of action:	Reproduction costs ([00010]).
Feedbacks configuration:	No feedbacks ([0000]).

Parameter settings/estimates (* hitting bound of search range):

Param.	Calibrated value
L_0	0.122 (fixed)
L_p	0.4037
L_m	0.6960
r_B	0.04499
R_m	24.04
f	1 (fixed)
k_d	8.215
z_b	8.766
b_b	2.325
F_{BV}	0.008 (fixed)
K_{RV}	1 (fixed)
κ	0.8 (fixed)
y_P	0.64 (fixed)
L_m^{ref}	0.7 (fixed)

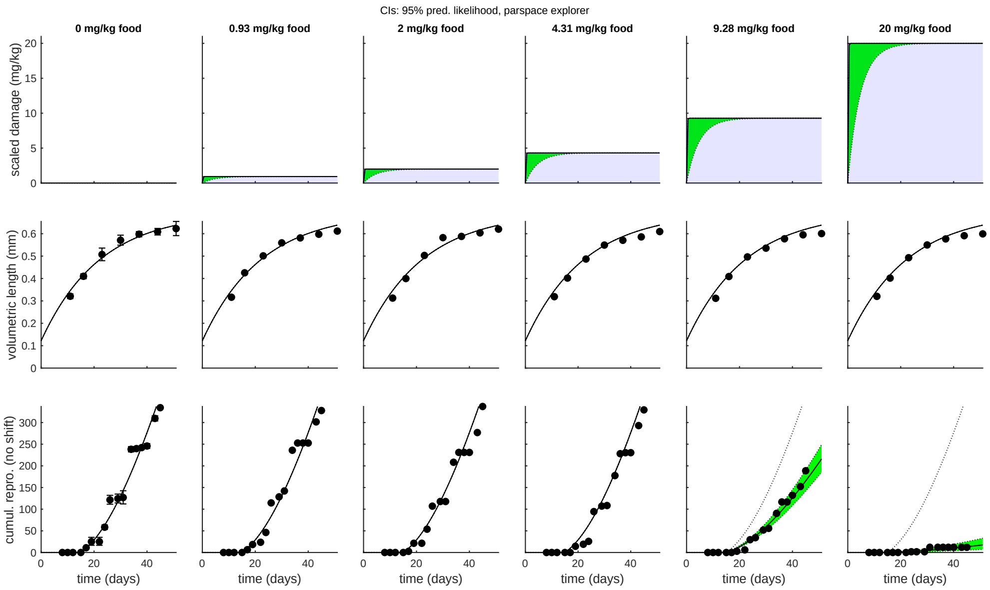

Figure 35: BYOM fit for test Test F, reproduction costs.

Settings for EPx calculation:

% effect:	10
Window length:	28 days
Timestep:	1 day
Cut-off:	1000 in DeEP, 100000 in BYOM
Pass/fail:	10
Thinning:	off
Modification of exposure profiles:	all exposure values times 1.

For all profiles, DeEP identifies reproduction as the most sensitive endpoint; there is no effect on body length with this pMoA.

Table 12: EP10 values for test F, reproduction costs. DeEP values in parentheses are for a previous version of the web application (run in May 2023), which accommodated a higher value for the cut-off.

Exposure scenario	DeEP stdalone	DeEP web	BYOM
apple R1 pond	8.218	8.218	8.22 ($t = 67$)
apple R2 stream	25.652	25.652	25.65 ($t = 74$)
cereal D1 ditch	1.185	1.185	1.18 ($t = 108$)
cereal D1 stream	19.059	19.059	19.06 ($t = 305$)
cereal D3 ditch	21.769	21.773	21.77 ($t = 108$)
cereal D4 pond	5.687	5.687	5.69 ($t = 125$)
cereal D4 stream	127.277	127.276	127.28 ($t = 315$)
cereal D5 pond	5.673	5.673	5.67 ($t = 106$)
cereal D5 stream	>1000	>1000 (1366.206)	1366.23 ($t = 379$)
cereal R4 stream	14.093	14.093	14.09 ($t = 52$)

The figures below compare the plots from the standalone to the plots from BYOM.

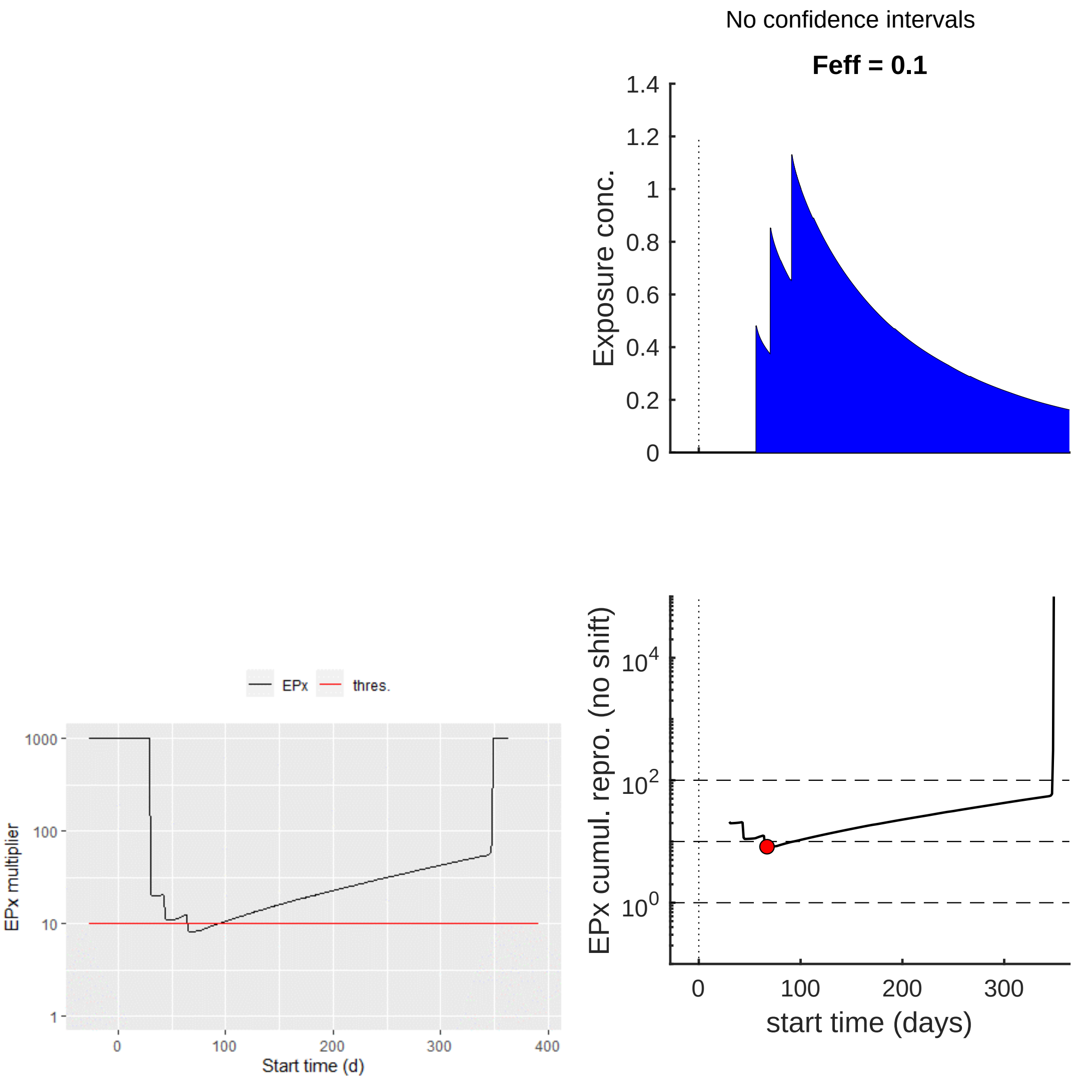

Figure 36: Result Test F, scenario apple R1 pond. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

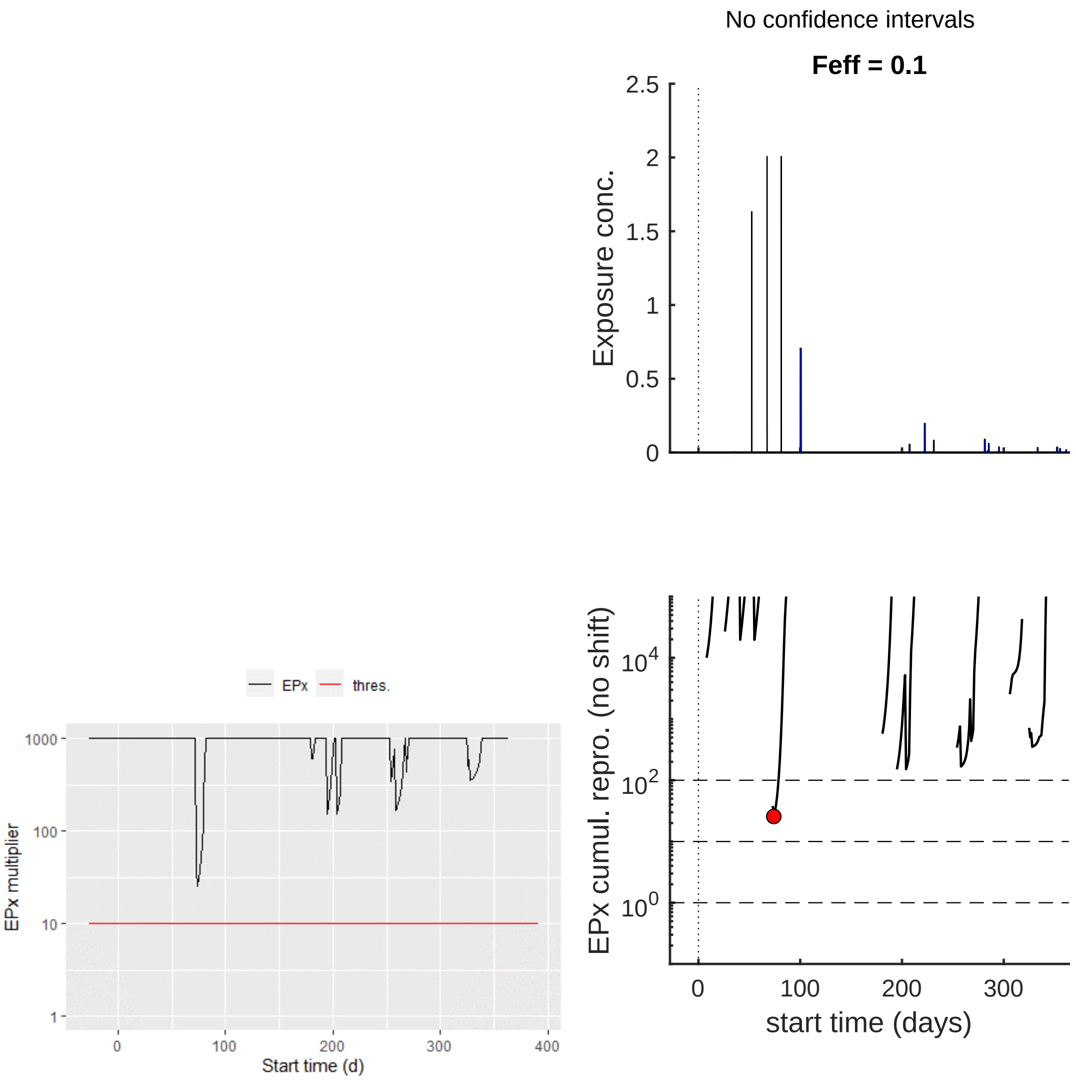

Figure 37: Result Test F, scenario apple R2 stream. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

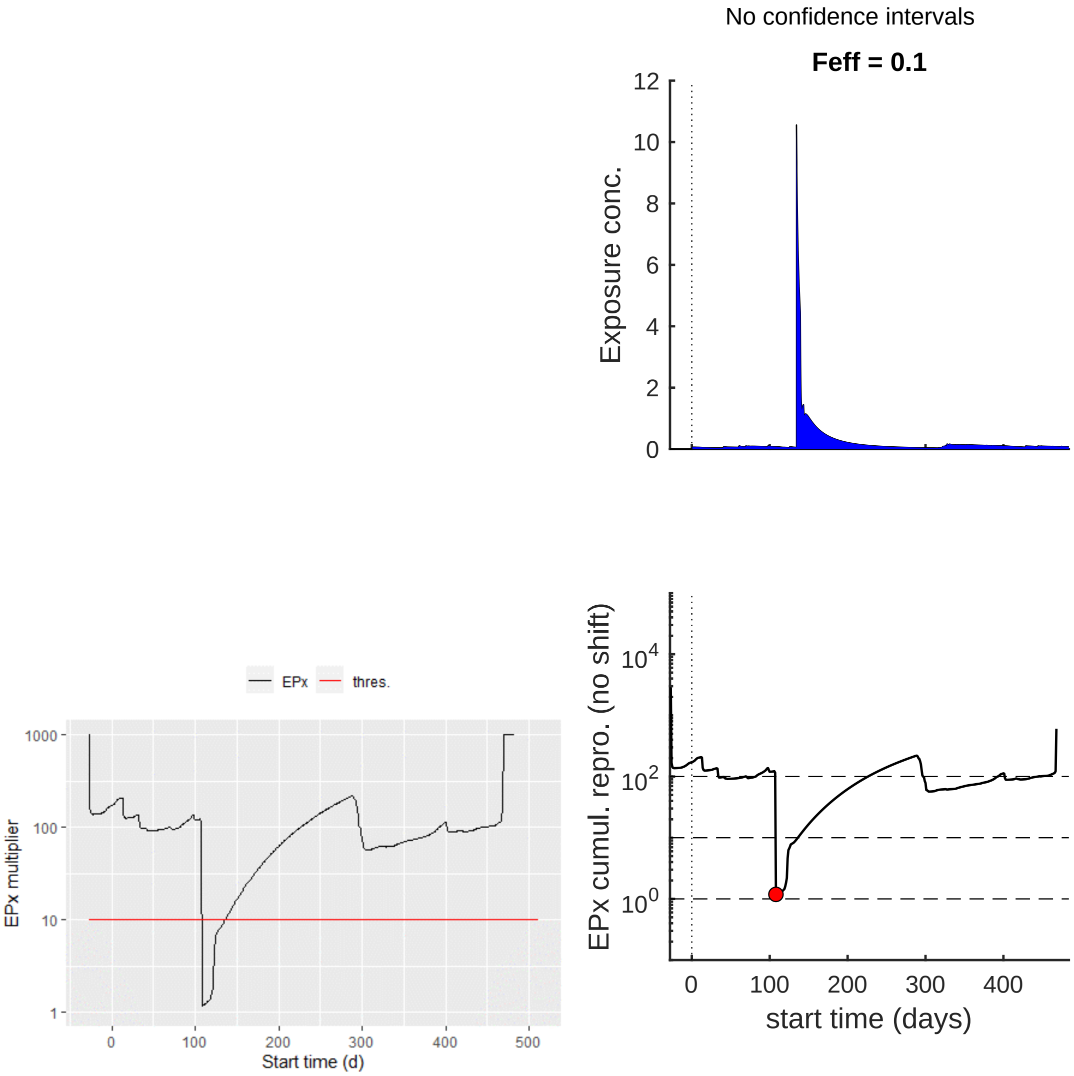

Figure 38: Result Test F, scenario cereal D1 ditch. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

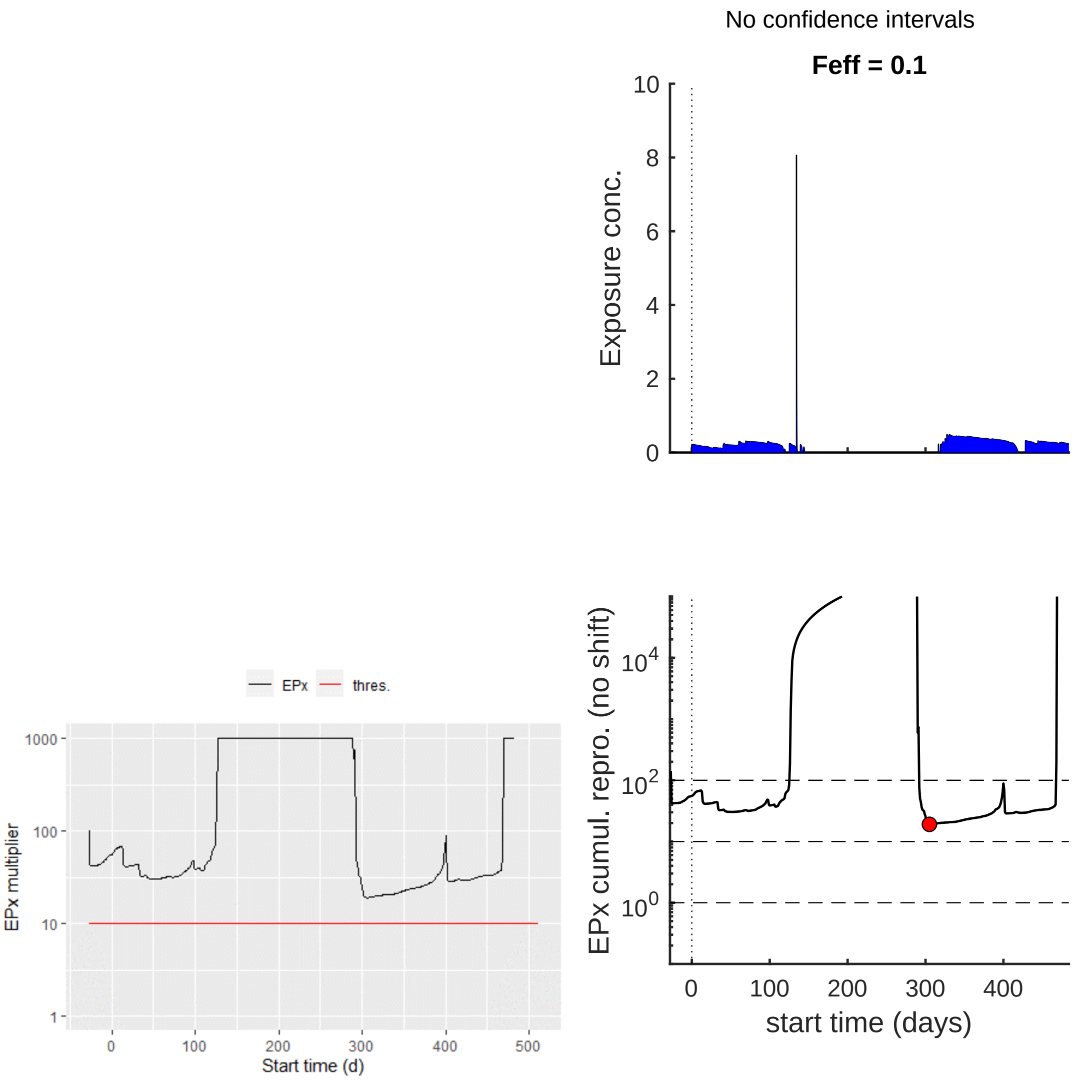

Figure 39: Result Test F, scenario cereal D1 stream. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

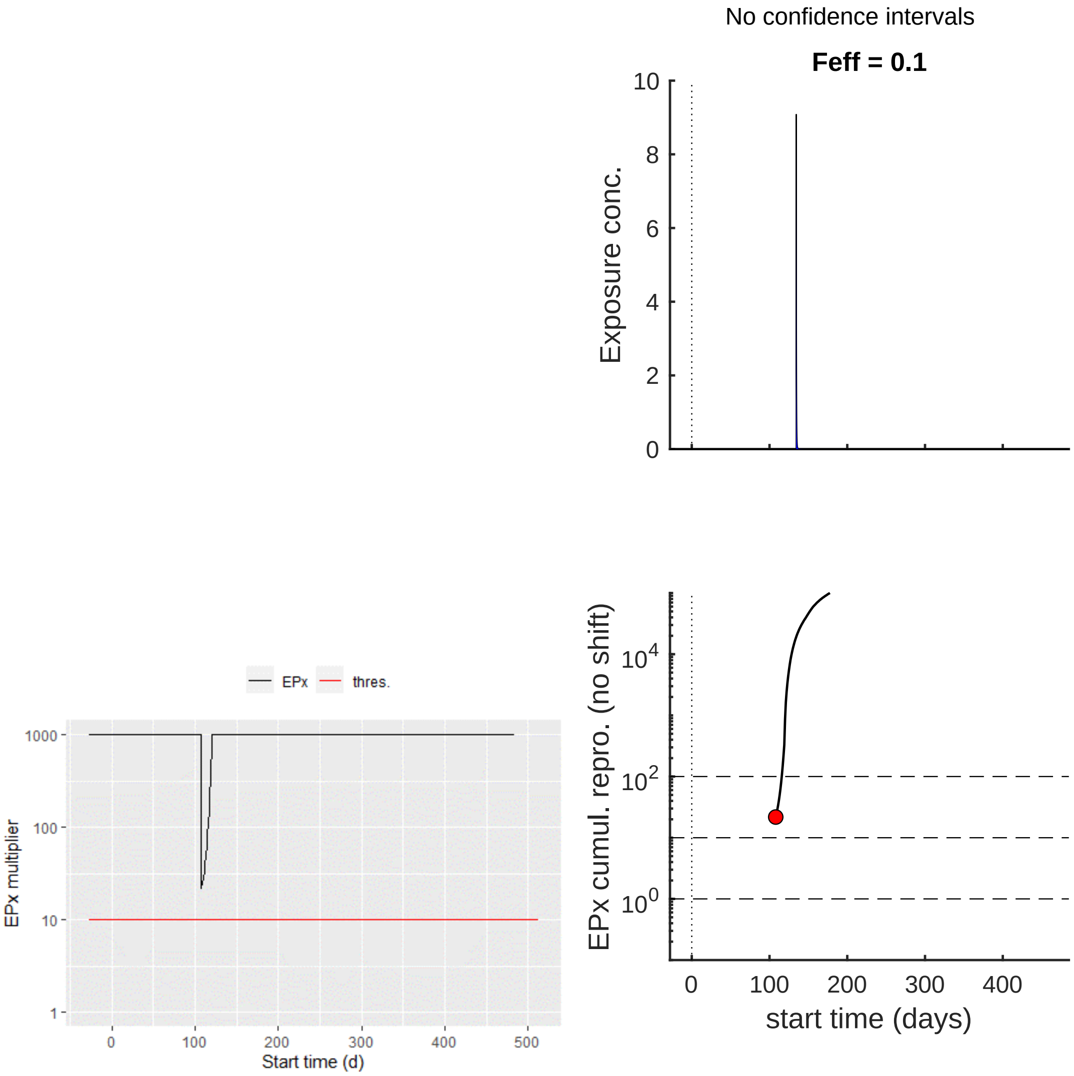

Figure 40: Result Test F, scenario cereal D3 ditch. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

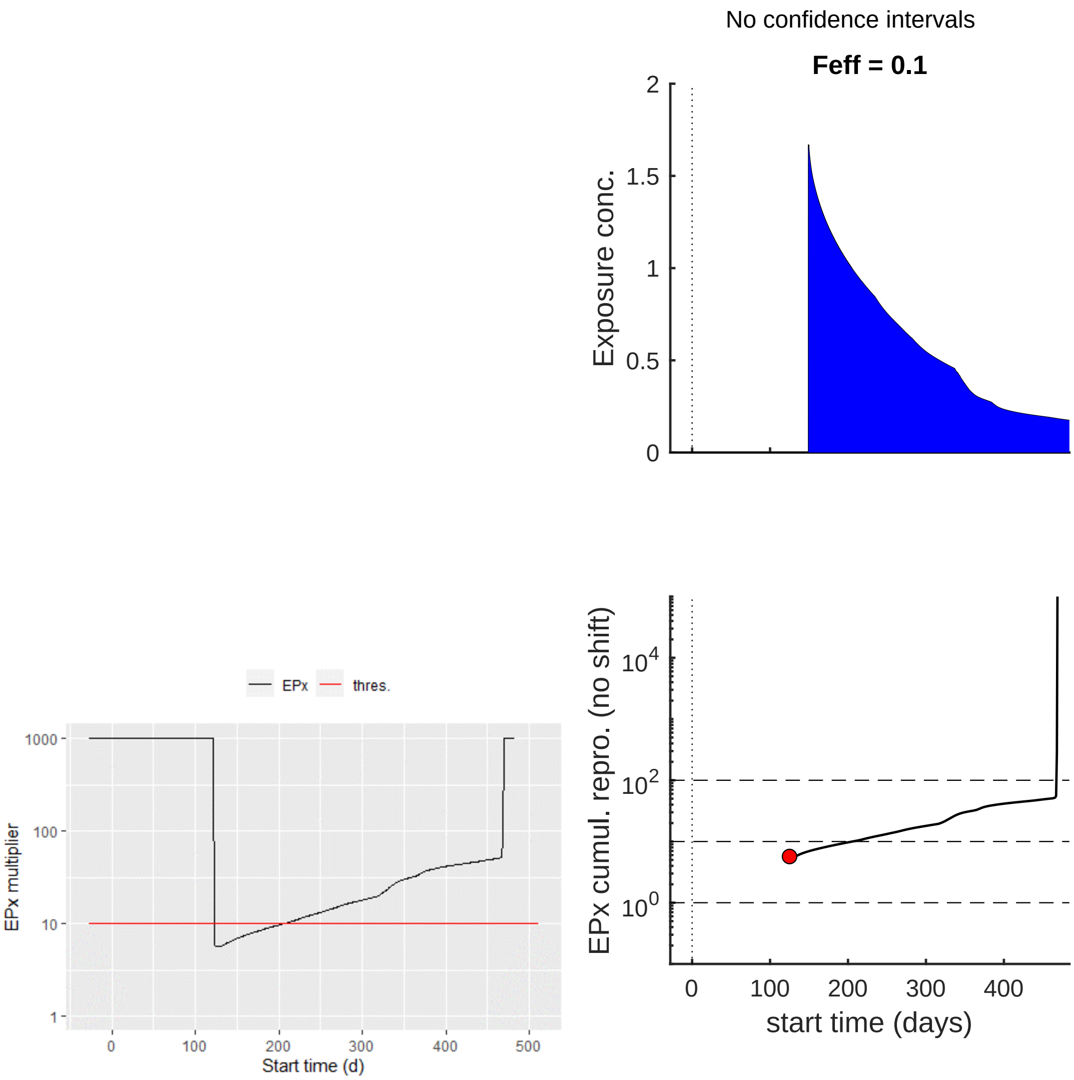

Figure 41: Result Test F, scenario cereal D4 pond. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

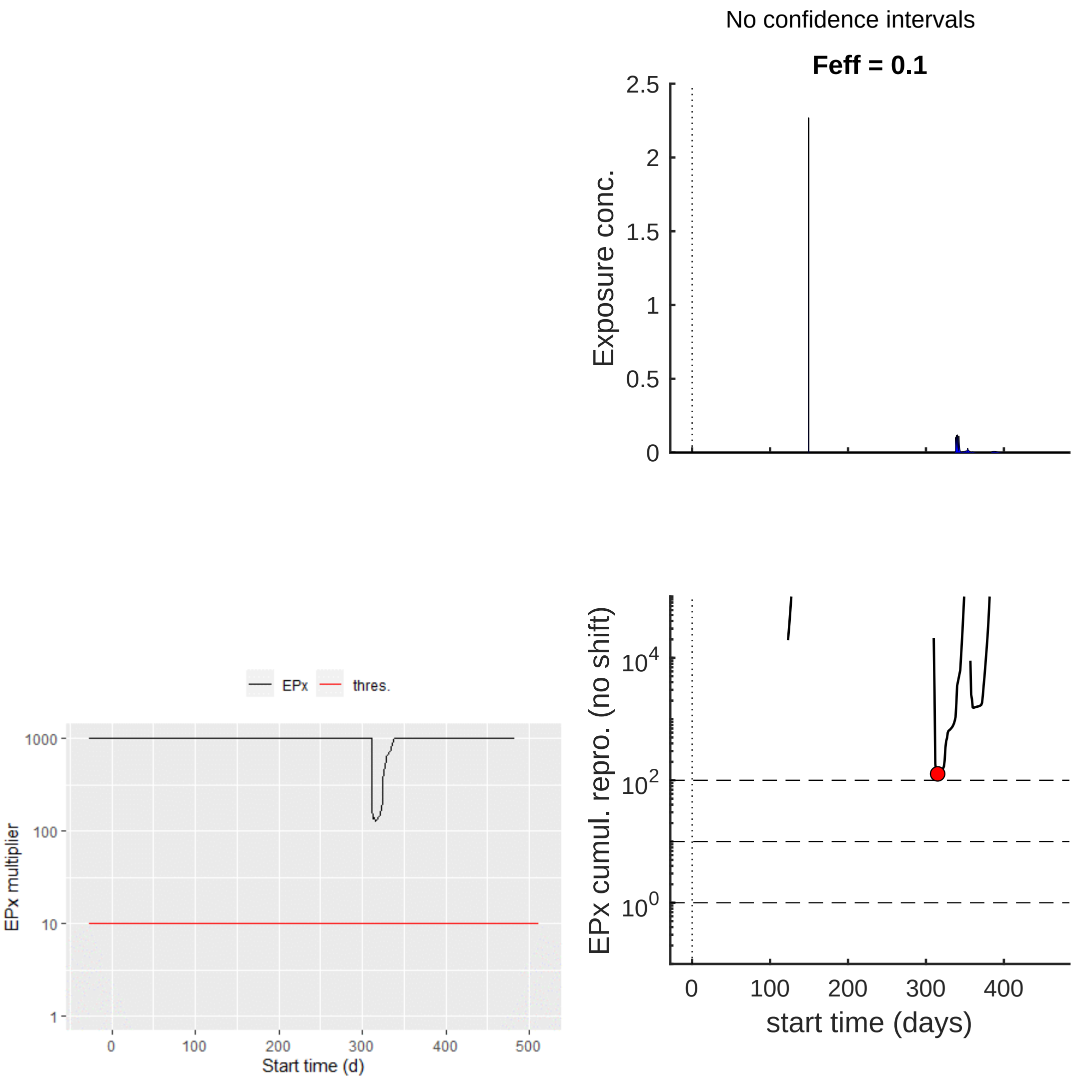

Figure 42: Result Test F, scenario cereal D4 stream. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

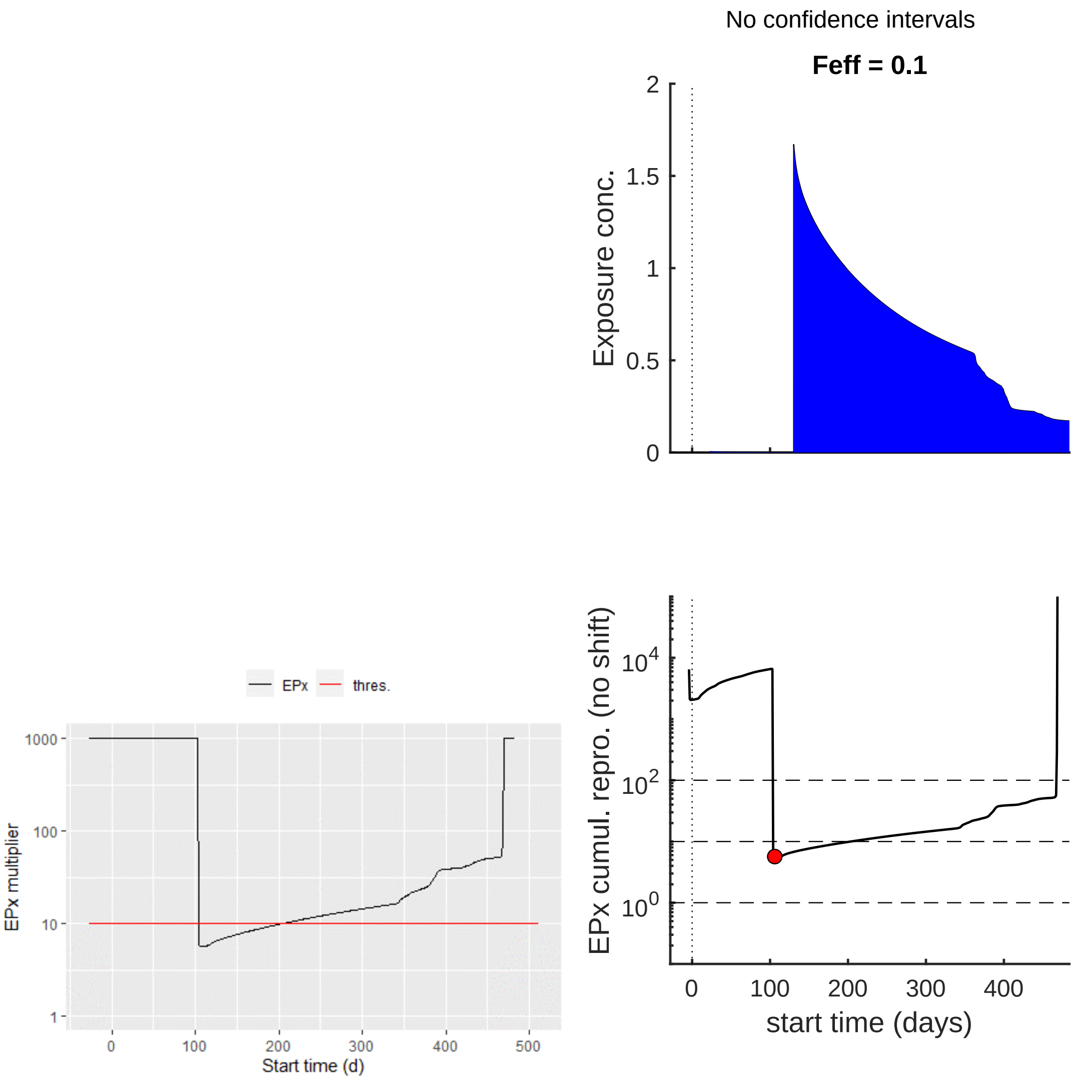

Figure 43: Result Test F, scenario cereal D5 pond. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

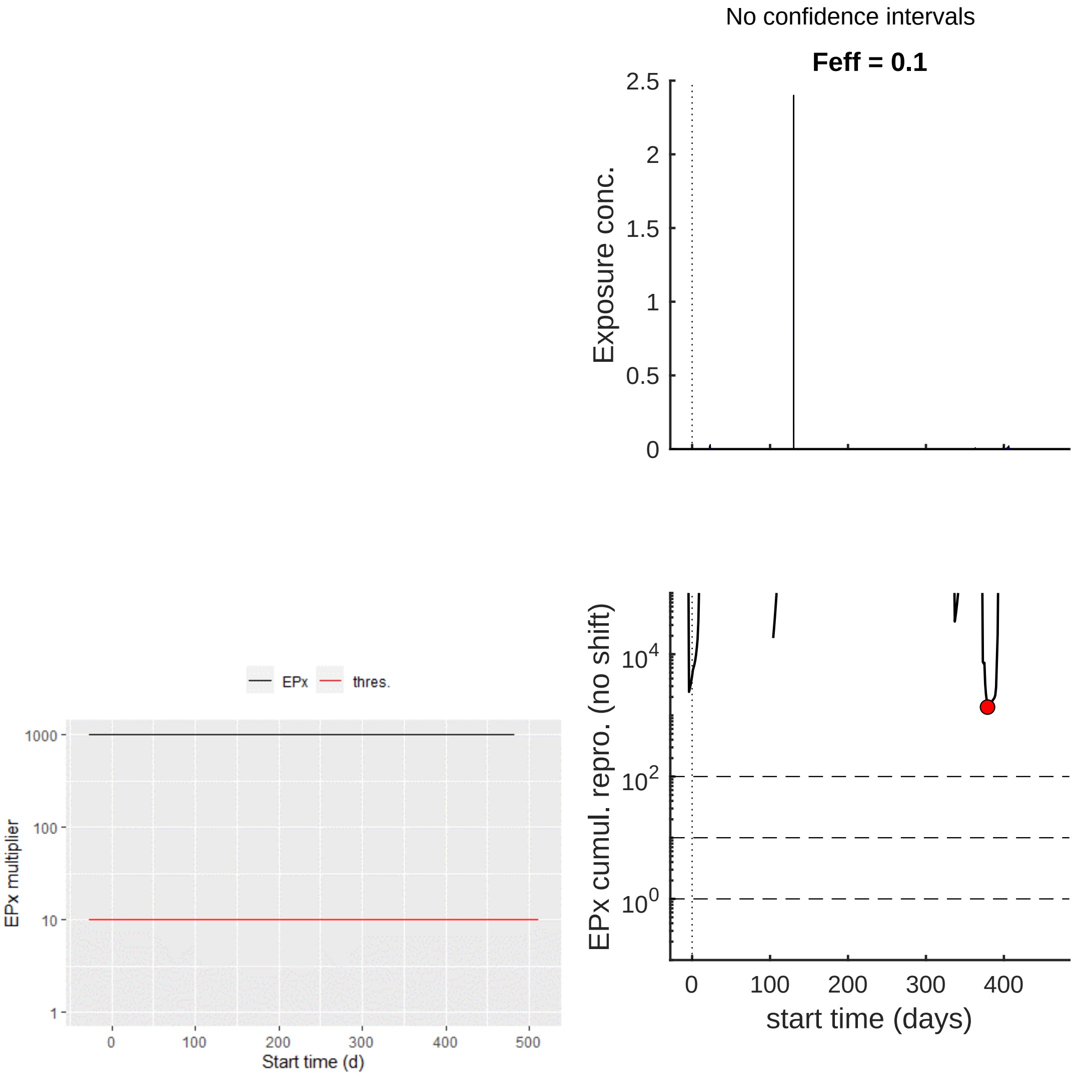

Figure 44: Result Test F, scenario cereal D5 stream. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

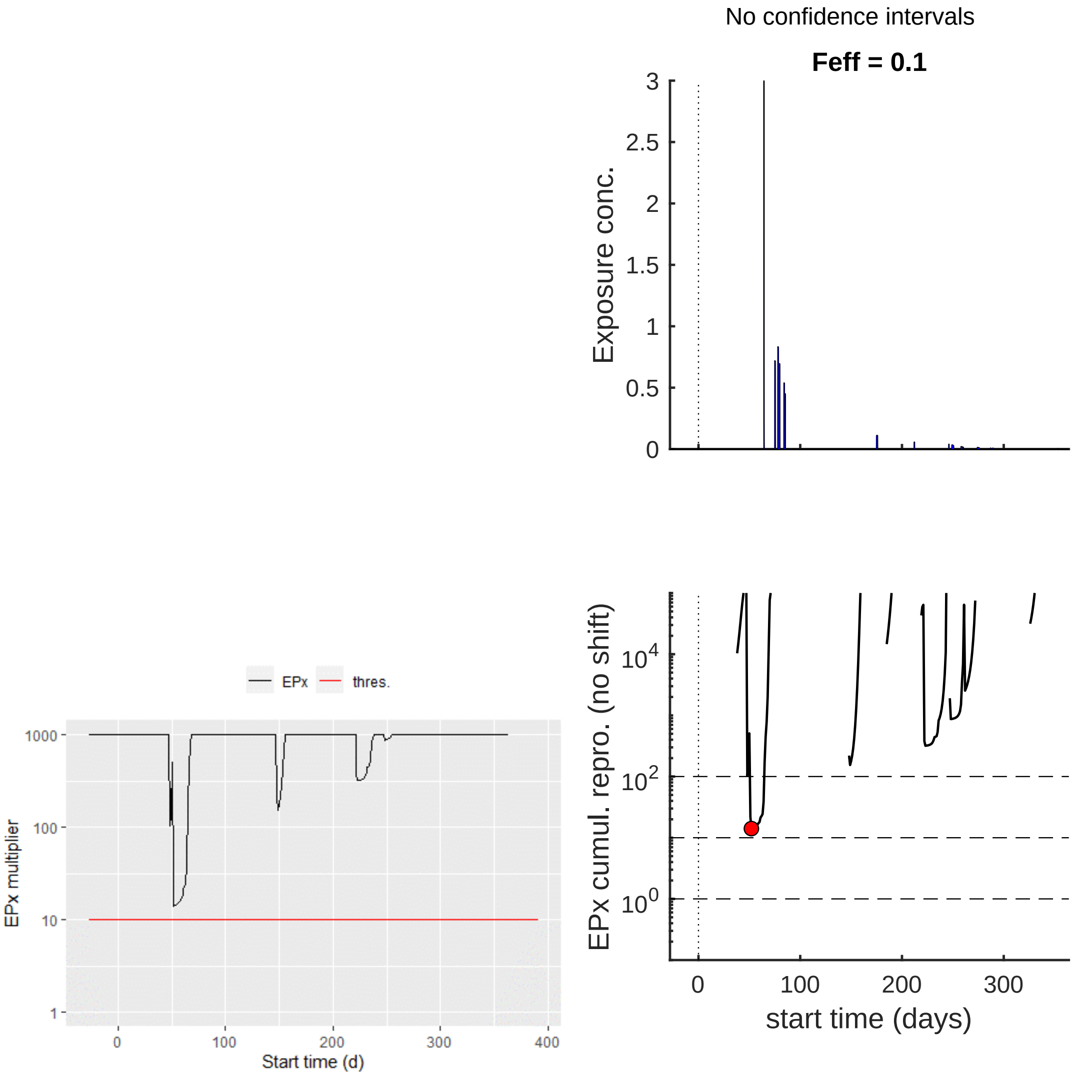

Figure 45: Result Test F, scenario cereal R4 stream. Left the DeEP plot and right the BYOM results. Note that the middle broken line in the right plots corresponds to the red line in the left plot (EP10=10).

7.2 Test F: repro hazards

Calibration settings:

Calibration data set:	*Folsomia candida* exposed to chlorpyrifos in food.
Mode of action:	Reproduction costs ([00001]).
Feedbacks configuration:	No feedbacks ([0000]).

Parameter settings/estimates (* hitting bound of search range):

Param.	Calibrated value
L_0	0.122 (fixed)
L_p	0.4037
L_m	0.6960
r_B	0.04499
R_m	24.04
f	1 (fixed)
k_d	3.586
z_b	3.682
b_b	0.1649
F_{BV}	0.008 (fixed)
K_{RV}	1 (fixed)
κ	0.8 (fixed)
y_P	0.64 (fixed)
L_m^{ref}	0.7 (fixed)

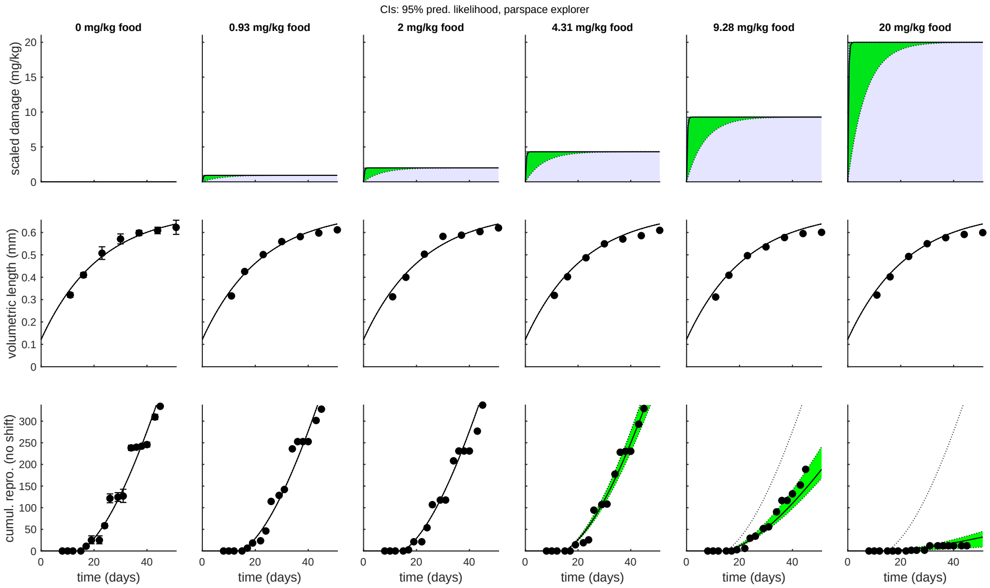

Figure 46: BYOM fit for test Test F, reproduction hazards.

Settings for EPx calculation:

% effect:	10
Window length:	28 days
Timestep:	1 day
Cut-off:	1000 in DeEP, 100000 in BYOM
Pass/fail:	10
Thinning:	off
Modification of exposure profiles:	all exposure values times 1.

For all profiles, DeEP identifies reproduction as the most sensitive endpoint; there is no effect on body length with this pMoA.

Table 13: EP10 values for test F, reproduction hazard.

Exposure scenario	DeEP stdalone	DeEP web	BYOM
apple R1 pond	4.248	4.249	4.25 ($t = 76$)
apple R2 stream	25.468	25.468	25.47 ($t = 74$)
cereal D1 ditch	0.815	0.815	0.82 ($t = 111$)
cereal D1 stream	9.681	9.681	9.68 ($t = 313$)
cereal D3 ditch	10.368	10.368	10.37 ($t = 108$)
cereal D4 pond	2.982	2.982	2.98 ($t = 134$)
cereal D4 stream	100.485	100.485	100.49 ($t = 315$)
cereal D5 pond	2.963	2.963	2.96 ($t = 115$)
cereal D5 stream	571.965	571.965	572.14 ($t = 104$)
cereal R4 stream	13.648	13.648	13.65 ($t = 53$)

8 Conclusions

- The DeEP standalone and web interface yield identical results in almost all cases. Only for a few cases, tiny differences are observed. These absolute differences are less than 0.01, and are thus meaningless for all practical purposes.

- The DeEP implementations yield almost identical results to the BYOM implementation. Some differences are seen, but they are extremely small, and thus also meaningless. Some differences between these implementations would be expected, as different ODE solvers (with different settings) would be used. In BYOM, `ode45` was applied for the calculations, with tolerances set very strict. Even for the extreme cases with fast and slow kinetics (test C) the results are almost identical to the DeEP implementations.

- In all cases, cumulative reproduction was a more sensitive endpoint than body length. This is not surprising, as, in a DEB context, small effects on body length will almost inevitably have a larger effect on reproduction.

- It is unlikely that the settings for starvation are important for EP10 calculation. Starvation can occur when the compound strongly affects assimilation or maintenance costs. Such strong effects on these processes will rapidly tend to produce large effects on reproduction (larger than the 10% that we are looking for). A starvation module is still important to deal with short and high exposure peaks, but prolonged starvation is unlikely for the multiplication factors that yield up to 10% effect.

- The DeEP calculations are considerably faster than the BYOM calculations (even with the parallel toolbox, running 8 cores in parallel). This is partly due to the cut-off used in DeEP: EPx values larger than 1000 are simply not calculated, while BYOM calculated exact EPx values up to 10^6.

References

[1] R. Ashauer, R. Kuhl, E. Zimmer, and M. Junghans. Effect modeling quantifies the difference between the toxicity of average pesticide concentrations and time-variable exposures from water quality monitoring. *Environmental Toxicology and Chemistry*, 39(11):2158–2168, 2020.

[2] EFSA. Scientific opinion on the state of the art of toxicokinetic/toxicodynamic (TKTD) effect models for regulatory risk assessment of pesticides for aquatic organisms. *EFSA journal*, 16(8):5377, 2018.

[3] T. Jager. Revisiting simplified DEBtox models for analysing ecotoxicity data. *Ecological Modelling*, 416:108904, 2020.

[4] T. Jager, C. Albert, T. G. Preuss, and R. Ashauer. General Unified Threshold model of Survival - a toxicokinetic-toxicodynamic framework for ecotoxicology. *Environmental Science & Technology*, 45:2529–2540, 2011.

[5] T. Jager and R. Ashauer. *Modelling survival under chemical stress. A comprehensive guide to the GUTS framework*. Toxicodynamics Ltd., York, UK. Available from Leanpub, `https://leanpub.com/guts_book`, Version 2.1, 21 November 2022, 2022.

[6] T. Jager, T. Crommentuijn, C. A. M. Van Gestel, and S. A. L. M. Kooijman. Simultaneous modeling of multiple endpoints in life-cycle toxicity tests. *Environmental Science & Technology*, 38:2894–2900, 2004.

[7] T. Jager, T. Crommentuijn, C. A. M. Van Gestel, and S. A. L. M. Kooijman. Chronic exposure to chlorpyrifos reveals two modes of action in the springtail *Folsomia candida*. *Environmental Pollution*, 145:452–458, 2007.

[8] T. Jager, M. Trijau, N. Sherborne, B. Goussen, and R. Ashauer. Considerations for using reproduction data in toxicokinetic-toxicodynamic modeling. *Integrated Environmental Assessment and Management*, 18(2):479–487, 2022.

[9] N. Sherborne, N. Galic, and R. Ashauer. Sublethal effect modelling for environmental risk assessment of chemicals: problem definition, model variants, application and challenges. *Science of The Total Environment*, 745:141027, 2020.

[10] N. Sherborne, T. Jager, B. Goussen, M. Trijau, and R. Ashauer. The application and limitations of exposure multiplication factors in sublethal effect modelling. *Scientific Reports*, 12(1):6031, 2022.