NPFC-2022-TWG CMSA06-IP01

Results of performance measures and metrics for performance evaluation

Momoko Ichinokawa (FRA)

2022-09-05

Rough overview of estimated and true values

- Simple comparison between estimated and true parameters would bring rough overview of each estimation model's characteristics
- In the following figures, estimated values are shown by colored boxplot and true values are shown by black points
- Results of KAFKA as estimation model was removed

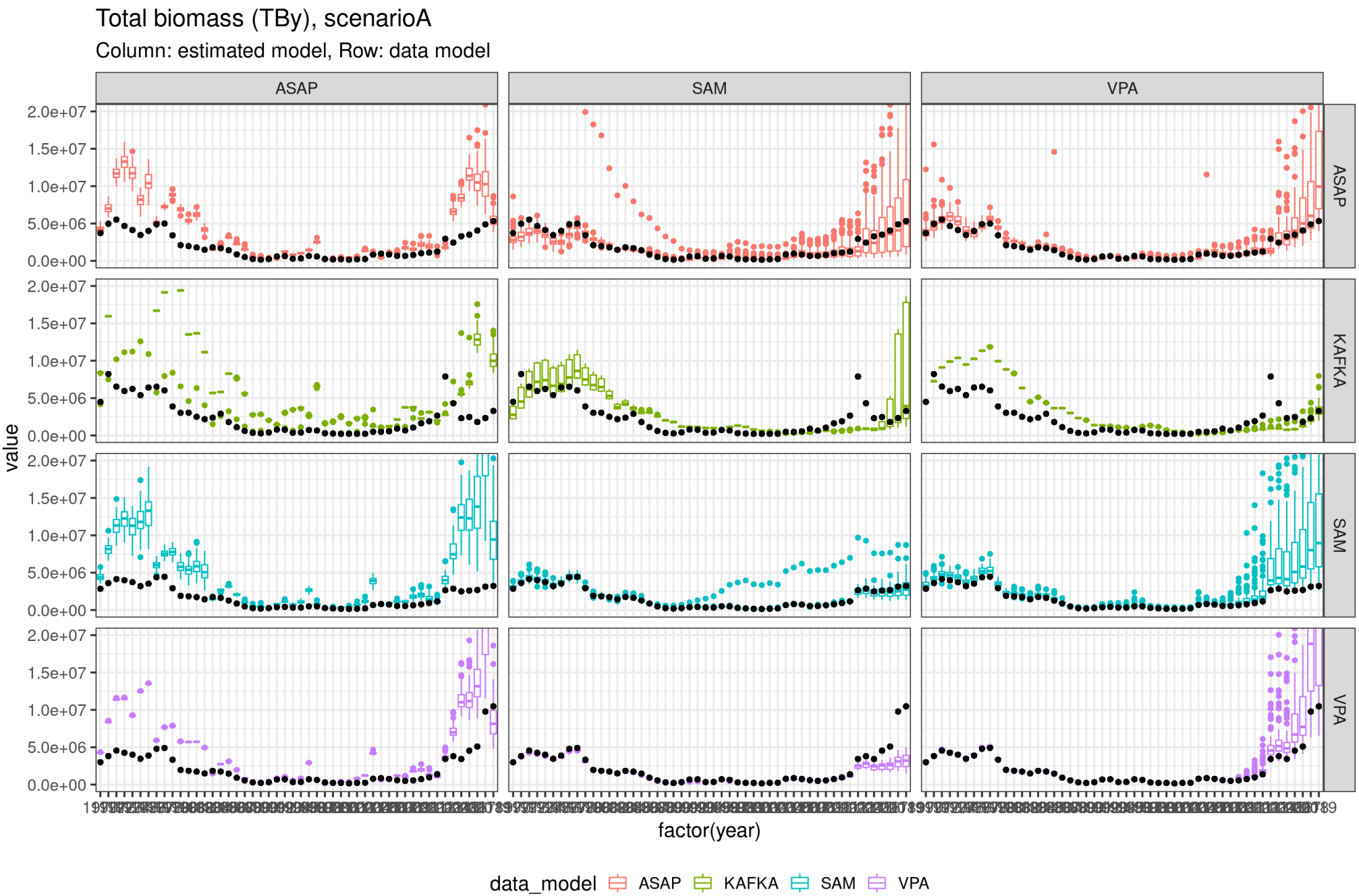
```
## Rows: 429760 Columns: 8
## ── Column specification ────────────────────────────────────────────────────────────────────────────────────────────────────────────
## Delimiter: ","
## chr (5): scenario, data_model, est_model, category, name
## dbl (3): sim_ID, value, value_orig
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 429760 Columns: 8
## ── Column specification ────────────────────────────────────────────────────────────────────────────────────────────────────────────
## Delimiter: ","
## chr (5): scenario, data_model, est_model, category, name
## dbl (3): sim_ID, value, value_orig
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 431392 Columns: 8
## ── Column specification ────────────────────────────────────────────────────────────────────────────────────────────────────────────
## Delimiter: ","
## chr (5): scenario, data_model, est_model, category, name
## dbl (3): sim_ID, value, value_orig
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 431392 Columns: 8
## ── Column specification ────────────────────────────────────────────────────────────────────────────────────────────────────────────
## Delimiter: ","
## chr (5): scenario, data_model, est_model, category, name
## dbl (3): sim_ID, value, value_orig
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 430304 Columns: 8
## ── Column specification ────────────────────────────────────────────────────────────────────────────────────────────────────────────
## Delimiter: ","
## chr (5): scenario, data_model, est_model, category, name
## dbl (3): sim_ID, value, value_orig
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

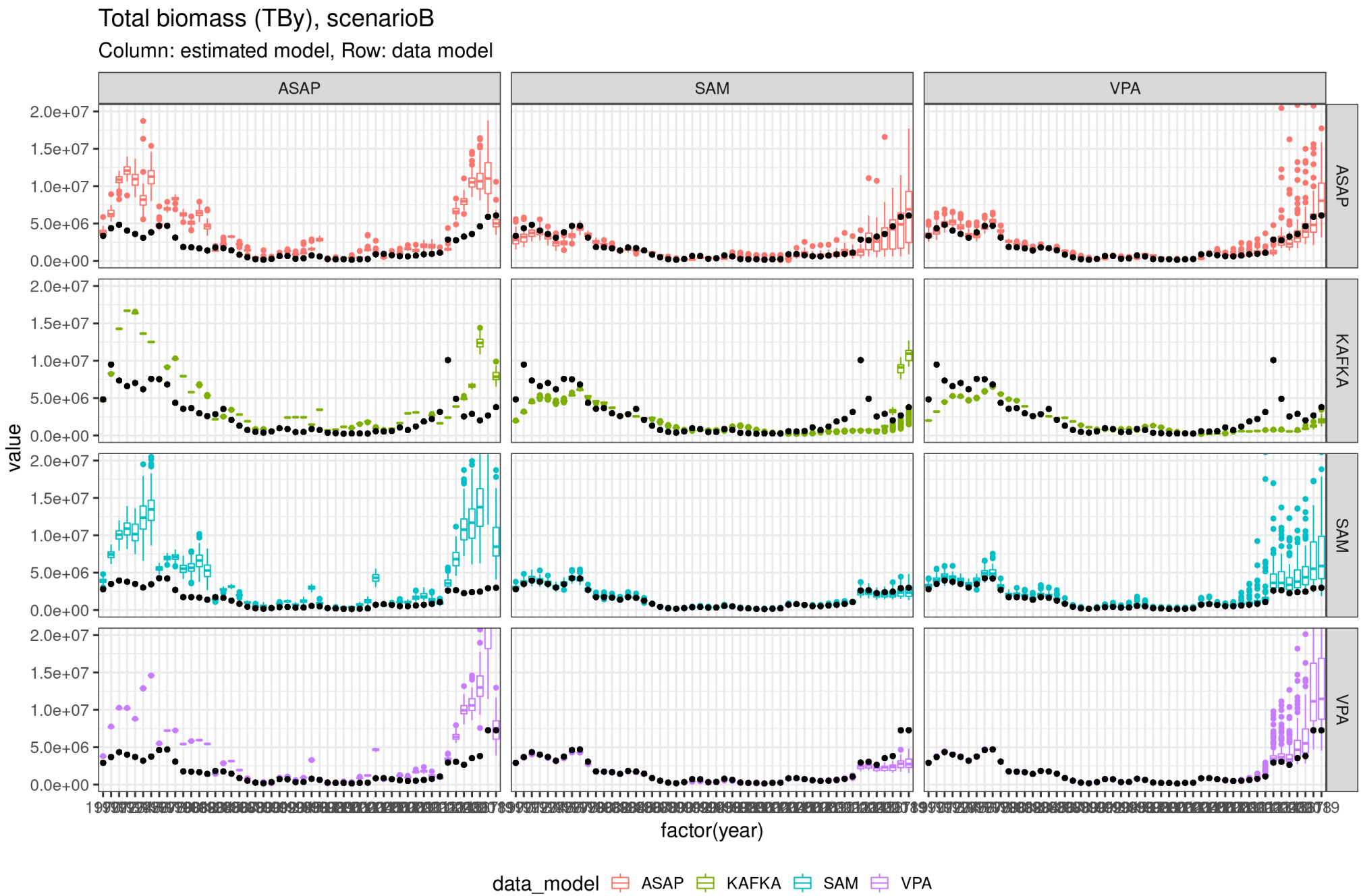
```
## Rows: 433024 Columns: 8
## ── Column specification ────────────────────────────────────────────────────────────────────────────────────────────────────────────
## Delimiter: ","
## chr (4): data_model, est_model, category, name
## dbl (3): sim_ID, value, value_orig
## lgl (1): scenario
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
## `summarise()` has grouped output by 'data_model', 'est_model', 'name'. You can override using the `.groups` argument.
## `summarise()` has grouped output by 'est_model'. You can override using the `.groups` argument.
```

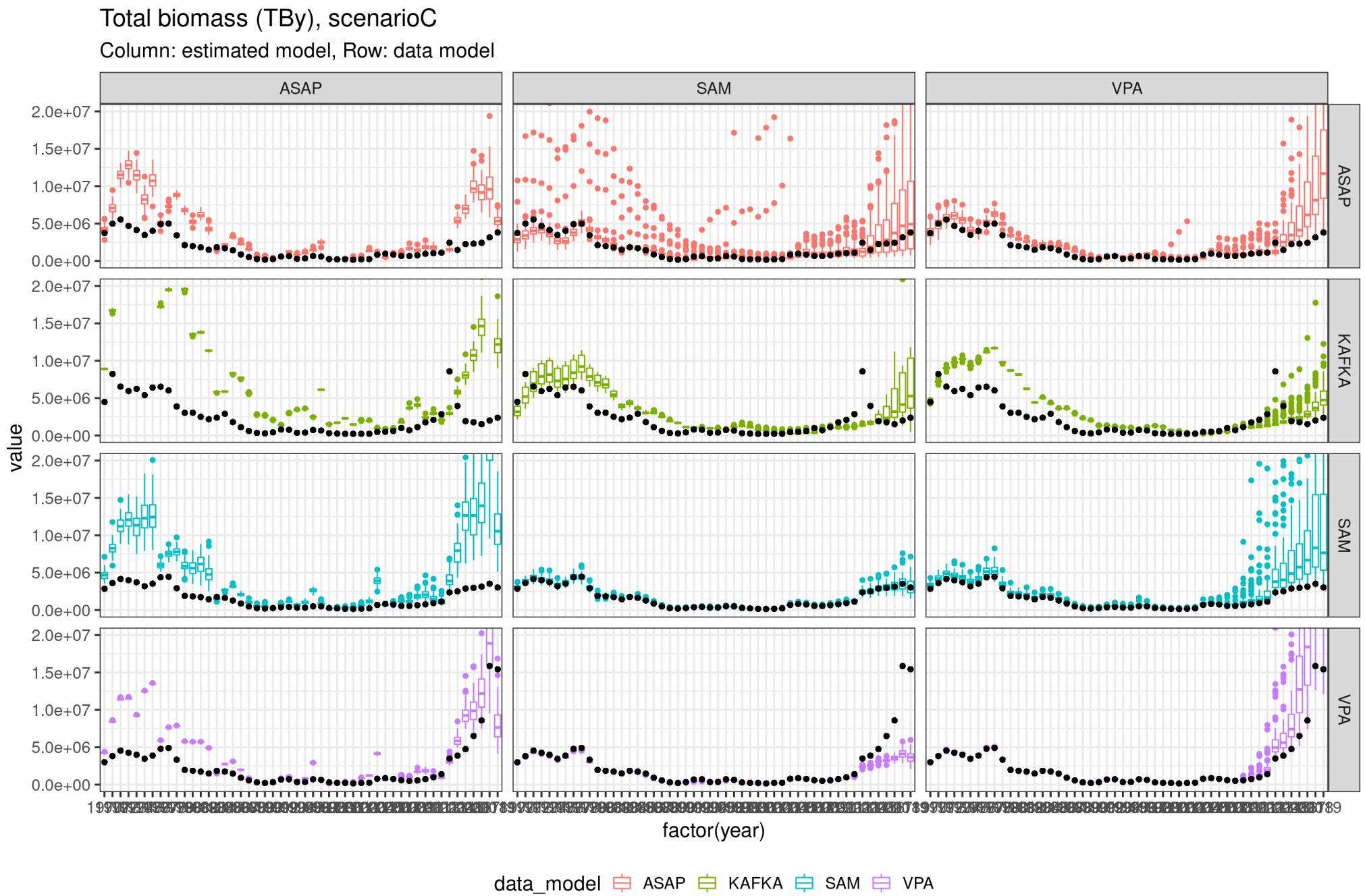
```
## # A tibble: 3 × 5
## # Groups:   est_model [3]
##   est_model     CV  MARB    MPRB  RMSE
##   <chr>      <dbl> <dbl>   <dbl> <dbl>
## 1 ASAP      0.0868 0.524 -0.0286  6.81
## 2 SAM       0.140  0.246 -0.0285  7.12
## 3 VPA       0.192  0.273  0.0101  7.70
```

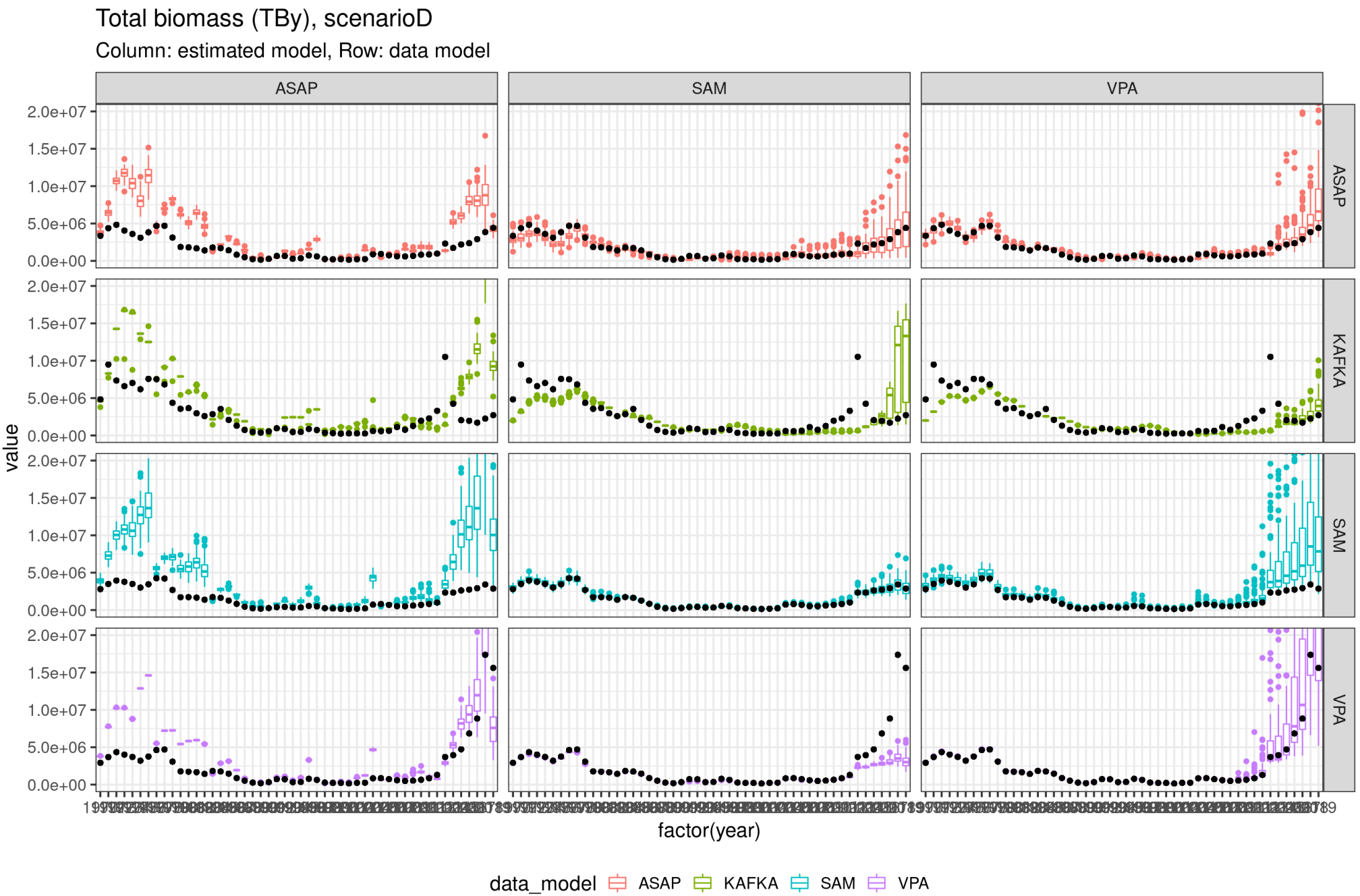
Total biomass (TBy)

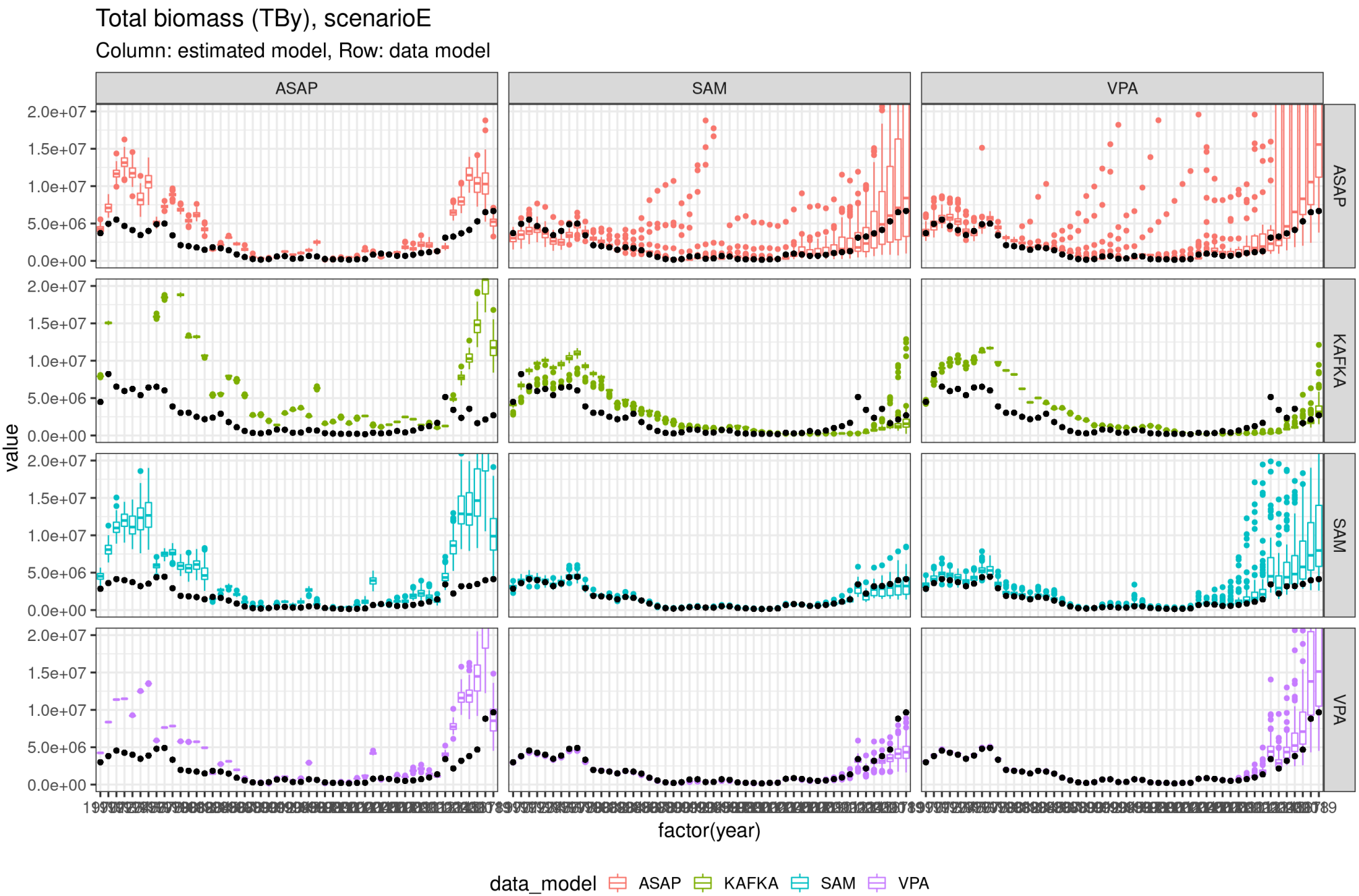
- ASAP: Always overestimate total biomass (TBy) especially when the stock level is historically hith (1970's and 2010's)
- SAM: Generally OK. In particular, there is no obvious bias in SAM and ASAP data, althogh and underestimation in recent years in VPA are observed.
- VPA: Varibilities in estimates are relatively large. Overestimation of recent TBy?

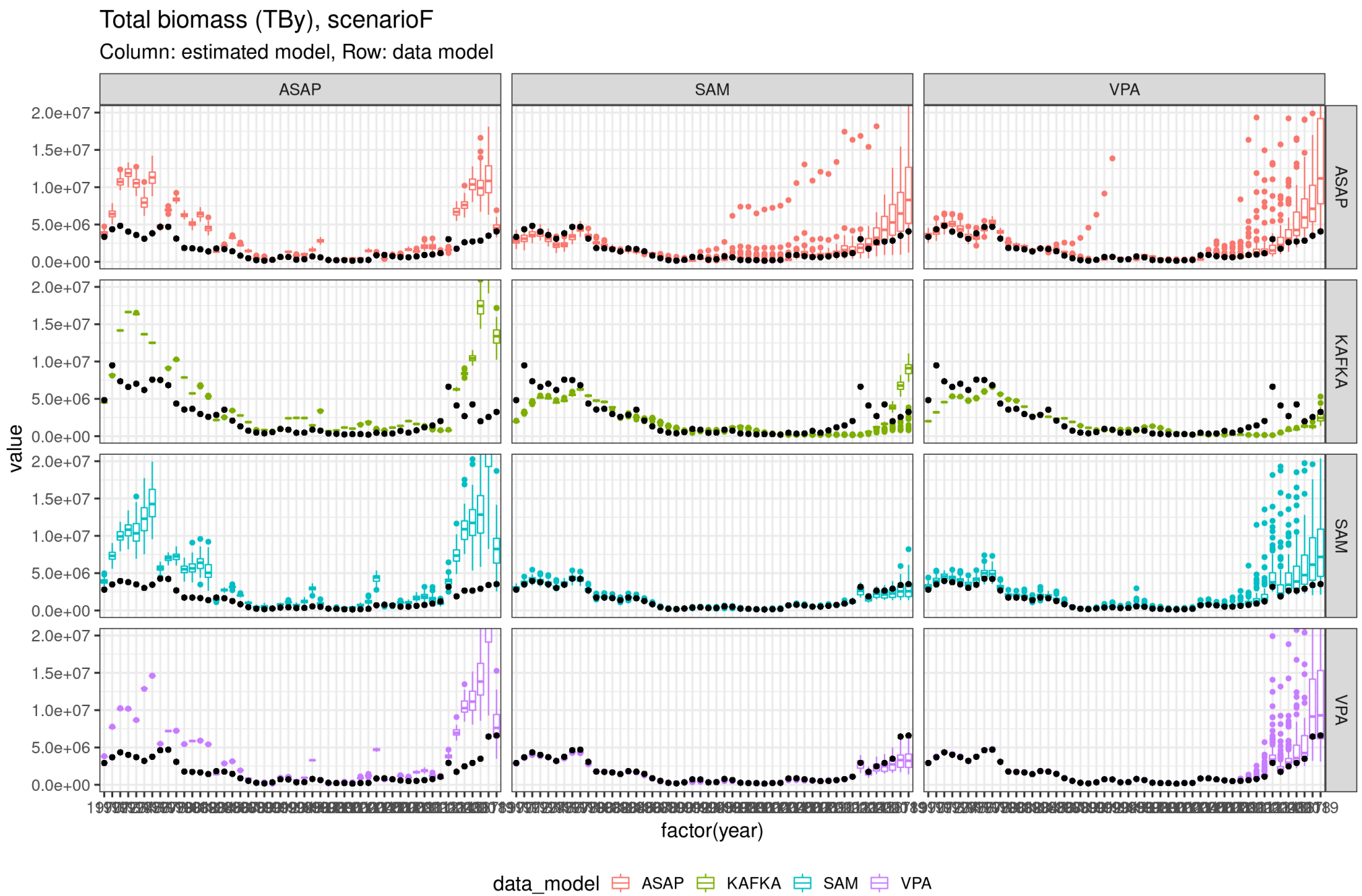

Total biomass (TBy), scenarioA
Column: estimated model, Row: data model
ASAP
SAM
VPA
ASAP
KAFKA
SAM
VPA
value
2.0e+07
1.5e+07
1.0e+07
5.0e+06
0.0e+00
factor(year)
data_model
ASAP
KAFKA
SAM
VPA


Total biomass (TBy), scenarioB
Column: estimated model, Row: data model
ASAP
SAM
VPA
KAFKA
value
2.0e+07
1.5e+07
1.0e+07
5.0e+06
0.0e+00
factor(year)
data_model
ASAP
KAFKA
SAM
VPA


Total biomass (TBy), scenarioC
Column: estimated model, Row: data model
ASAP
SAM
VPA
ASAP
KAFKA
SAM
VPA
value
2.0e+07
1.5e+07
1.0e+07
5.0e+06
0.0e+00
factor(year)
data_model
ASAP
KAFKA
SAM
VPA

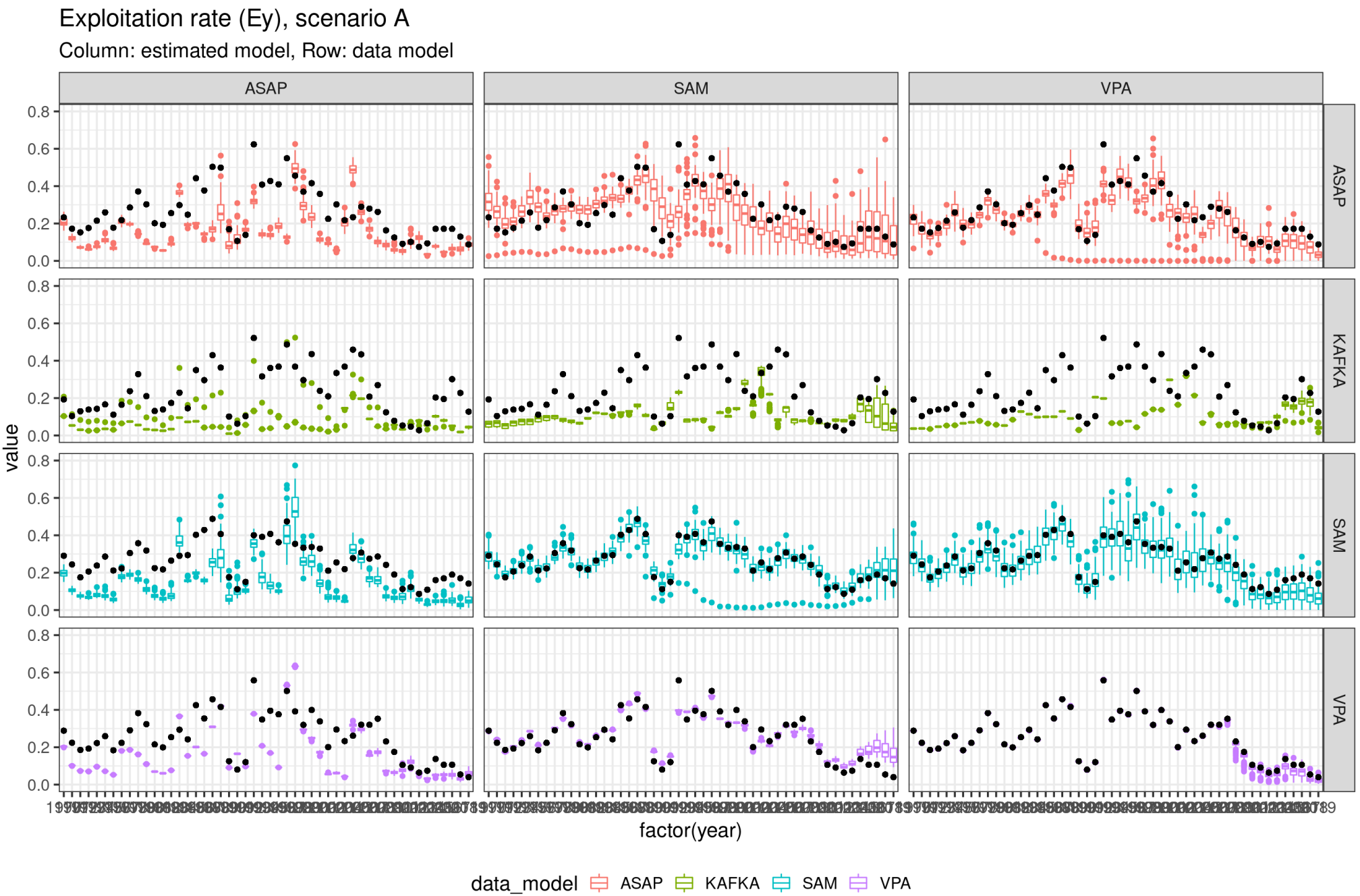

Total biomass (TBy), scenarioD
Column: estimated model, Row: data model
ASAP
SAM
VPA
KAFKA
value
factor(year)
data_model
ASAP
KAFKA
SAM
VPA

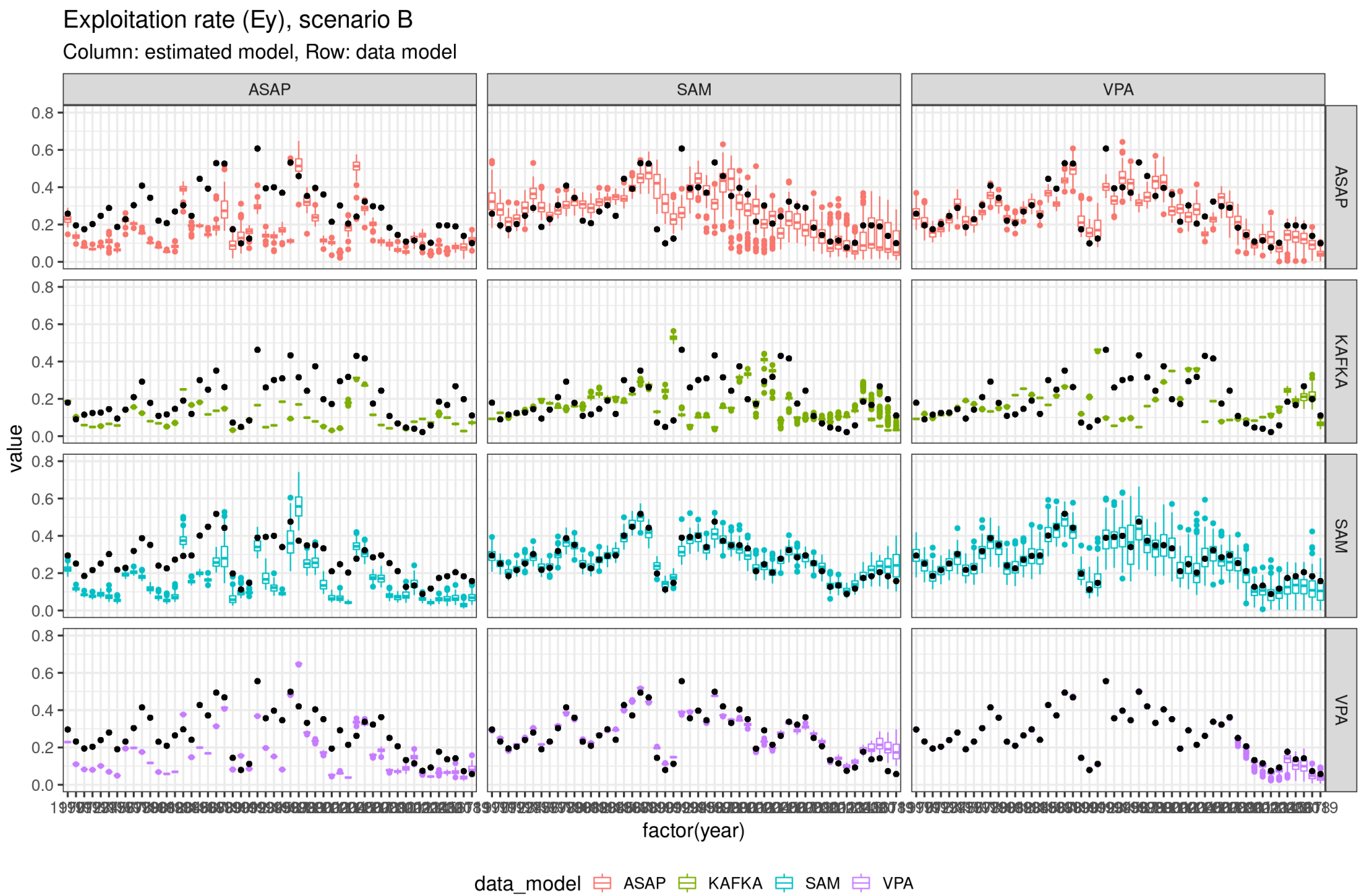

Total biomass (TBy), scenarioE
Column: estimated model, Row: data model
ASAP
SAM
VPA
KAFKA
value
2.0e+07
1.5e+07
1.0e+07
5.0e+06
0.0e+00
factor(year)
data_model
ASAP
KAFKA
SAM
VPA

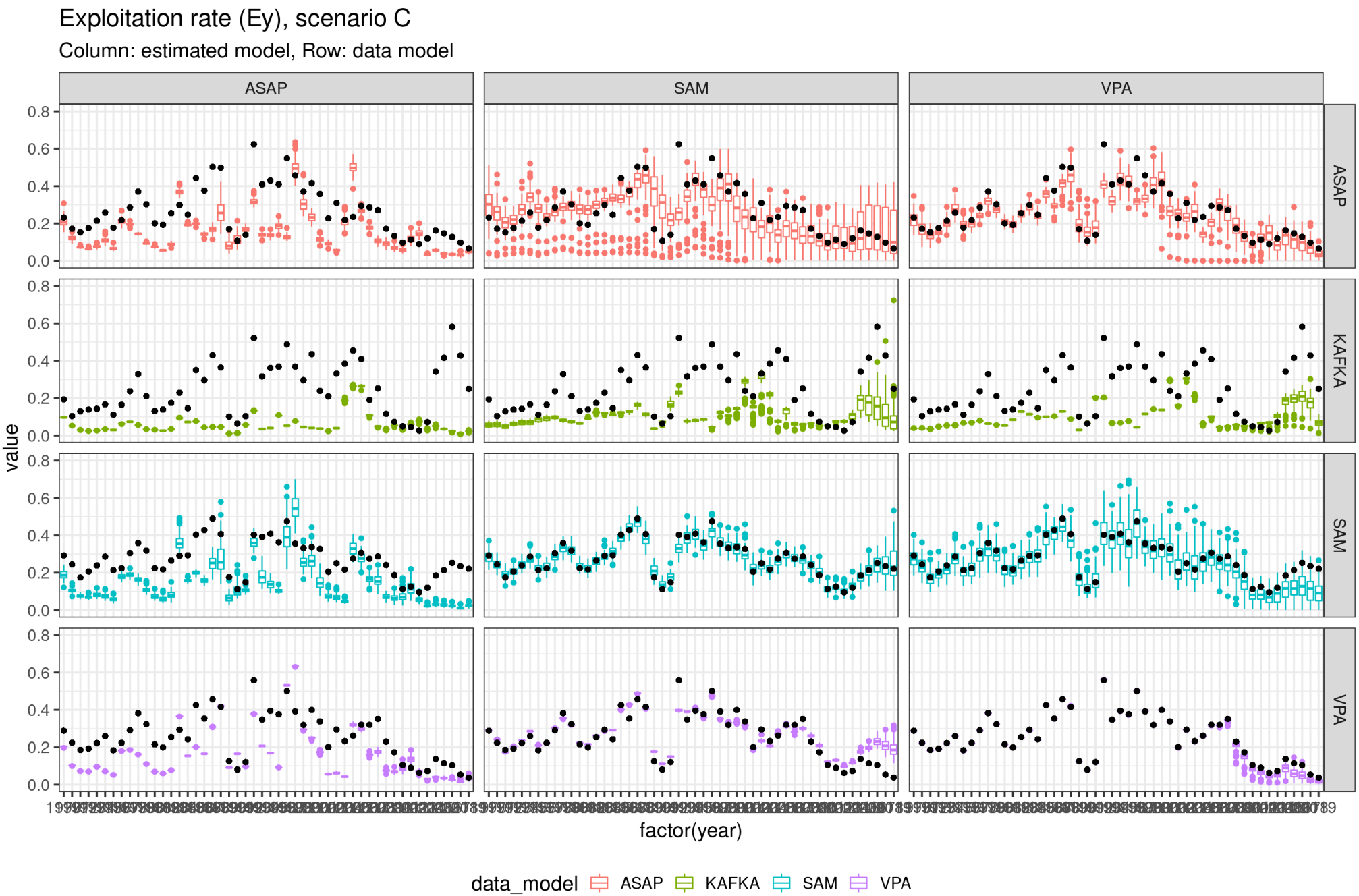

Total biomass (TBy), scenarioF
Column: estimated model, Row: data model
ASAP
SAM
VPA
ASAP
KAFKA
SAM
VPA
value
2.0e+07
1.5e+07
1.0e+07
5.0e+06
0.0e+00
factor(year)
data_model
ASAP
KAFKA
SAM
VPA

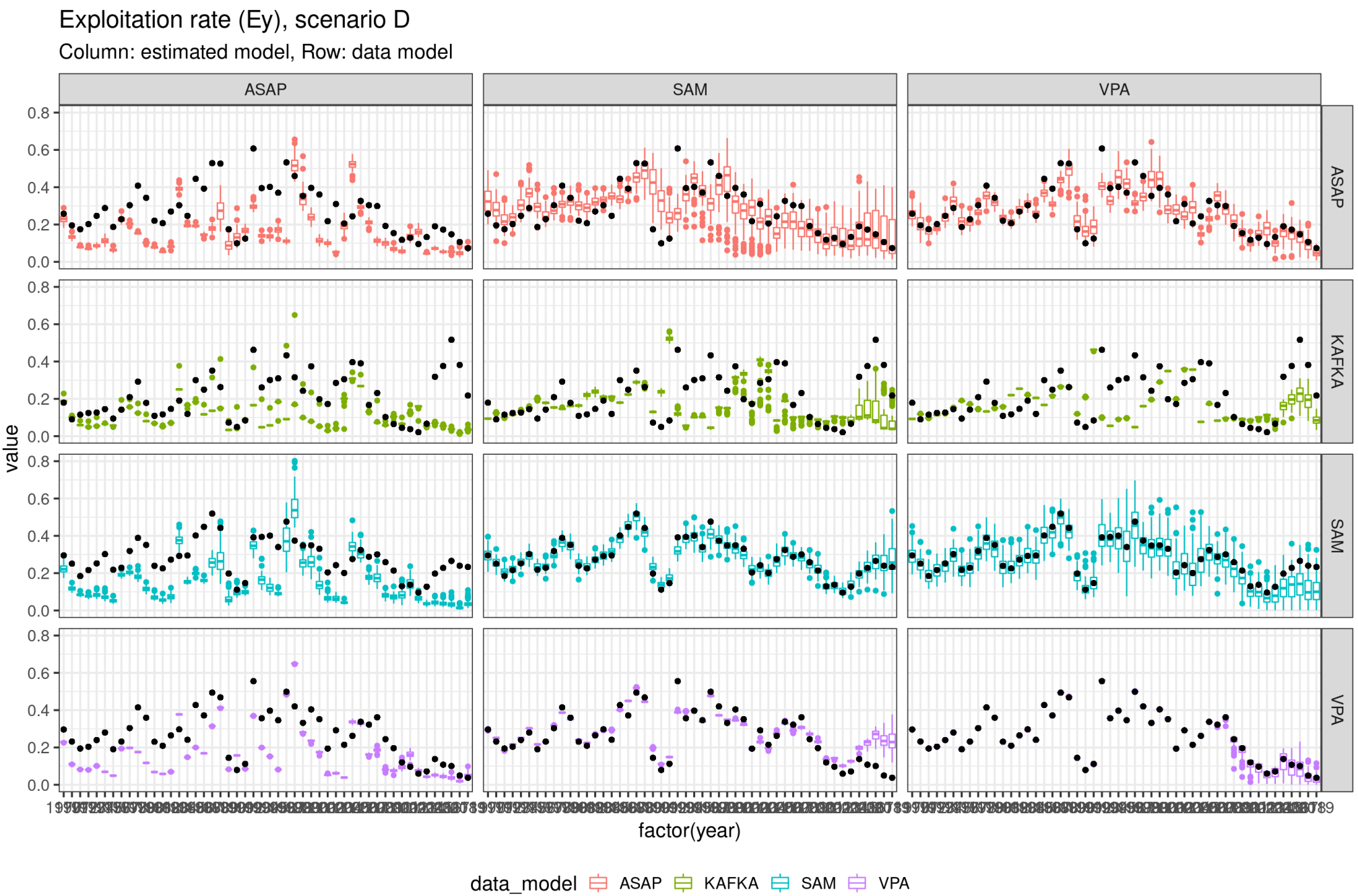
Exploitation rate (Ey)

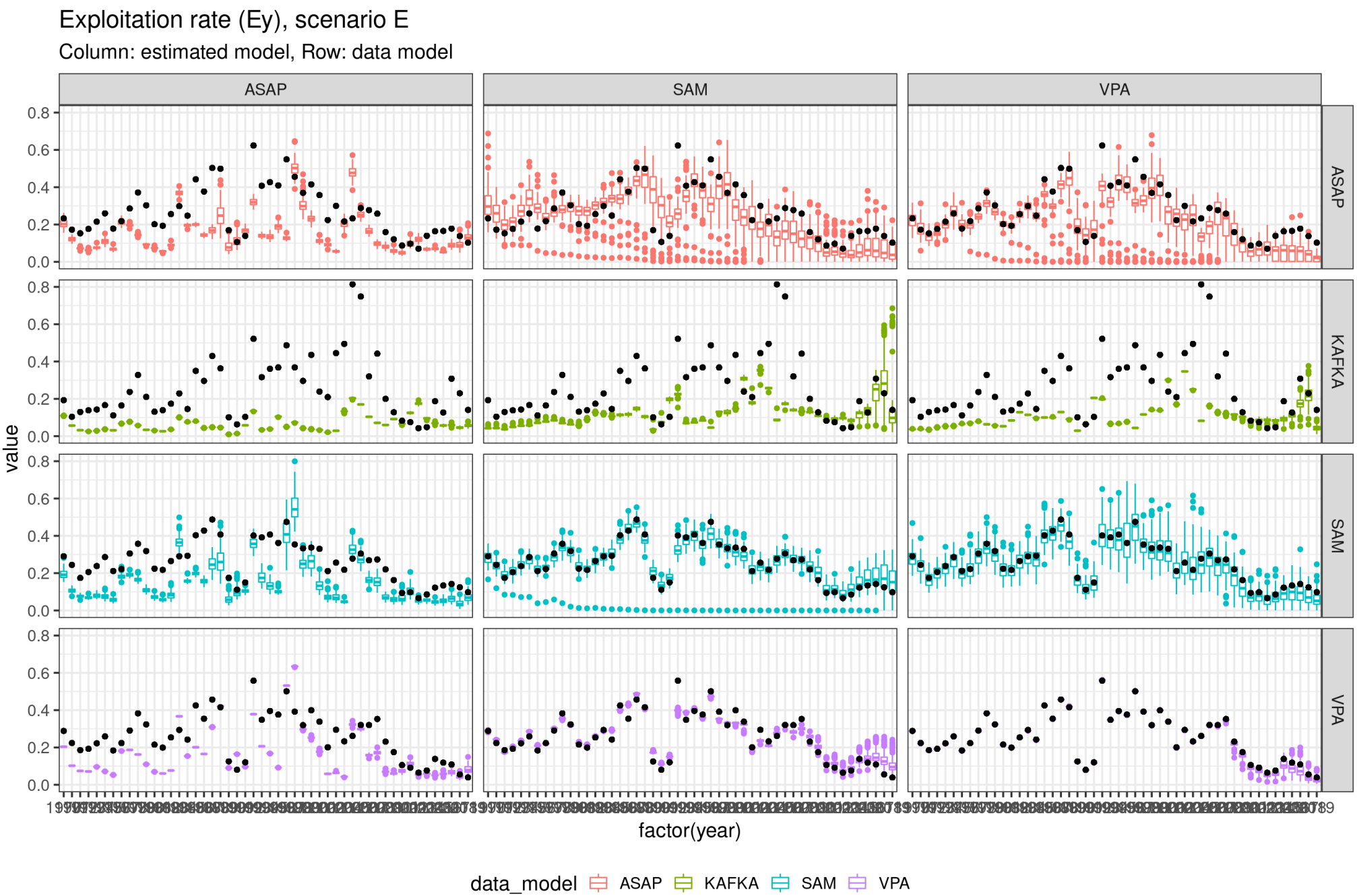
- Overall, there is opposite tendency to TBy (if TBy is overestimated, Ey is underestimated)

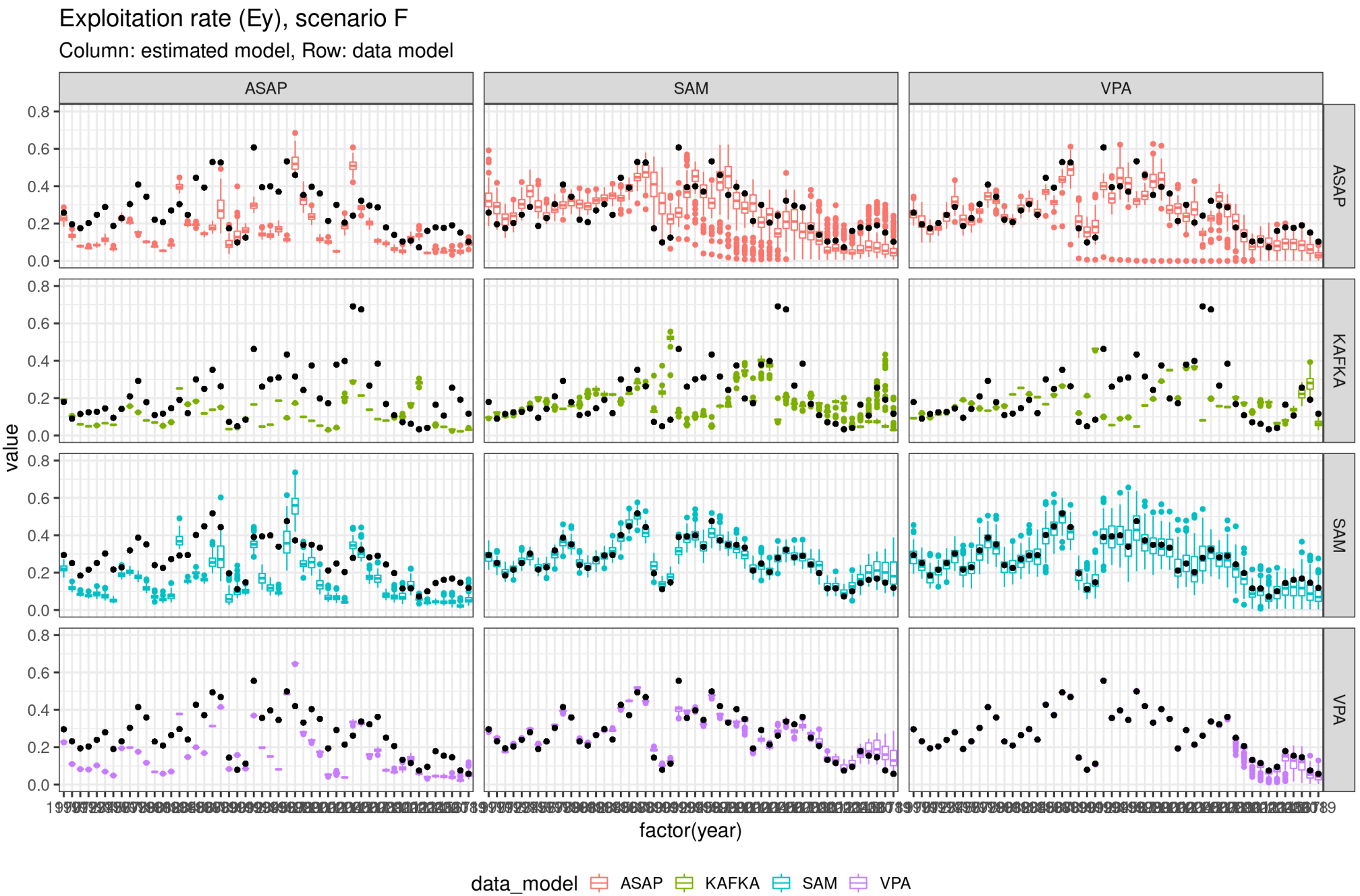

Exploitation rate (Ey), scenario A
Column: estimated model, Row: data model
ASAP
SAM
VPA
KAFKA
value
factor(year)
data_model
ASAP
KAFKA
SAM
VPA


Exploitation rate (Ey), scenario B
Column: estimated model, Row: data model
ASAP
SAM
VPA
ASAP
KAFKA
SAM
VPA
value
0.0
0.2
0.4
0.6
0.8
factor(year)
data_model
ASAP
KAFKA
SAM
VPA


Exploitation rate (Ey), scenario C
Column: estimated model, Row: data model
ASAP
SAM
VPA
ASAP
KAFKA
SAM
VPA
value
0.0
0.2
0.4
0.6
0.8
factor(year)
data_model
ASAP
KAFKA
SAM
VPA

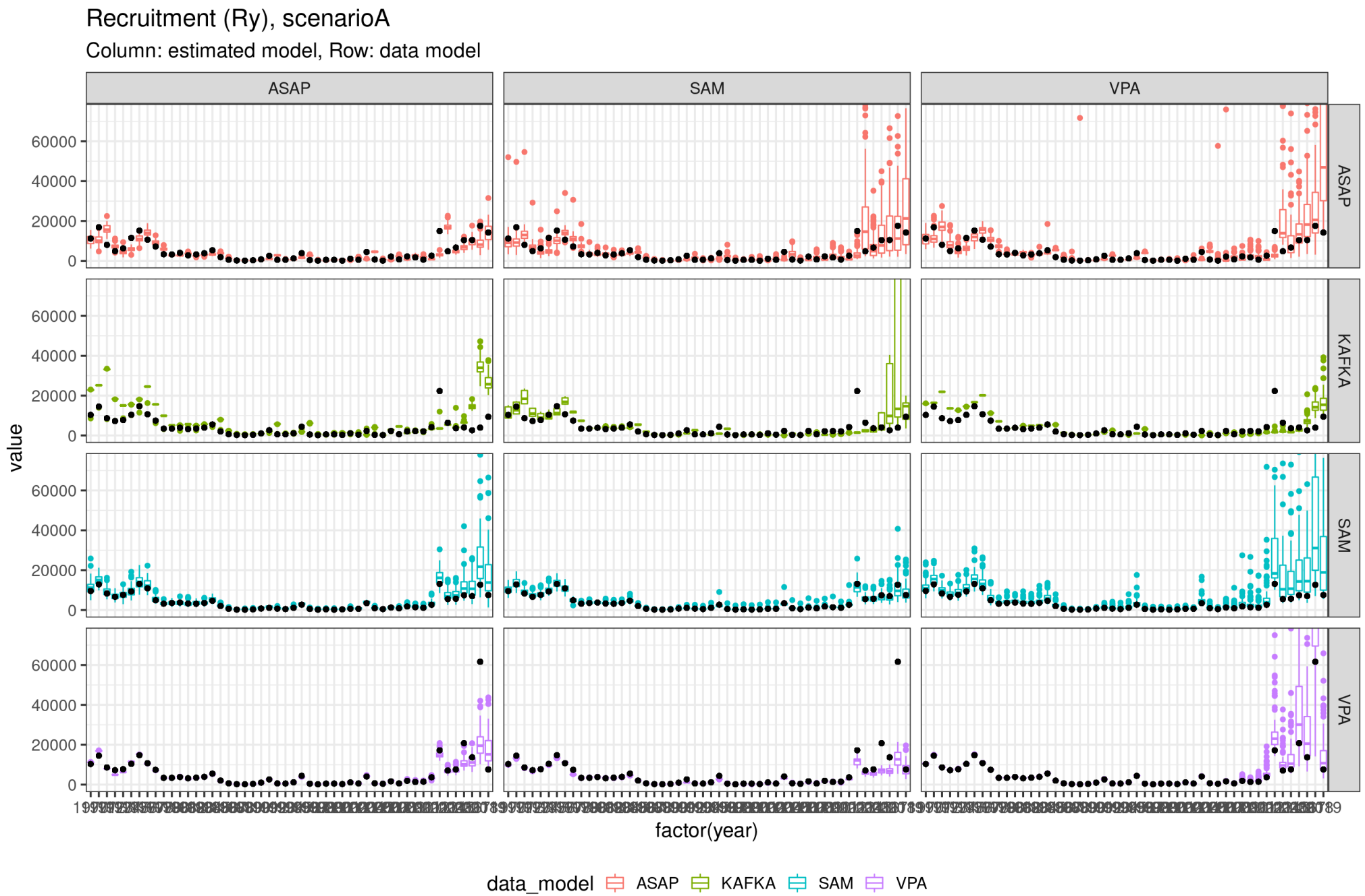

Exploitation rate (Ey), scenario D
Column: estimated model, Row: data model
ASAP
SAM
VPA
ASAP
KAFKA
SAM
VPA
value
0.0
0.2
0.4
0.6
0.8
factor(year)
data_model
ASAP
KAFKA
SAM
VPA

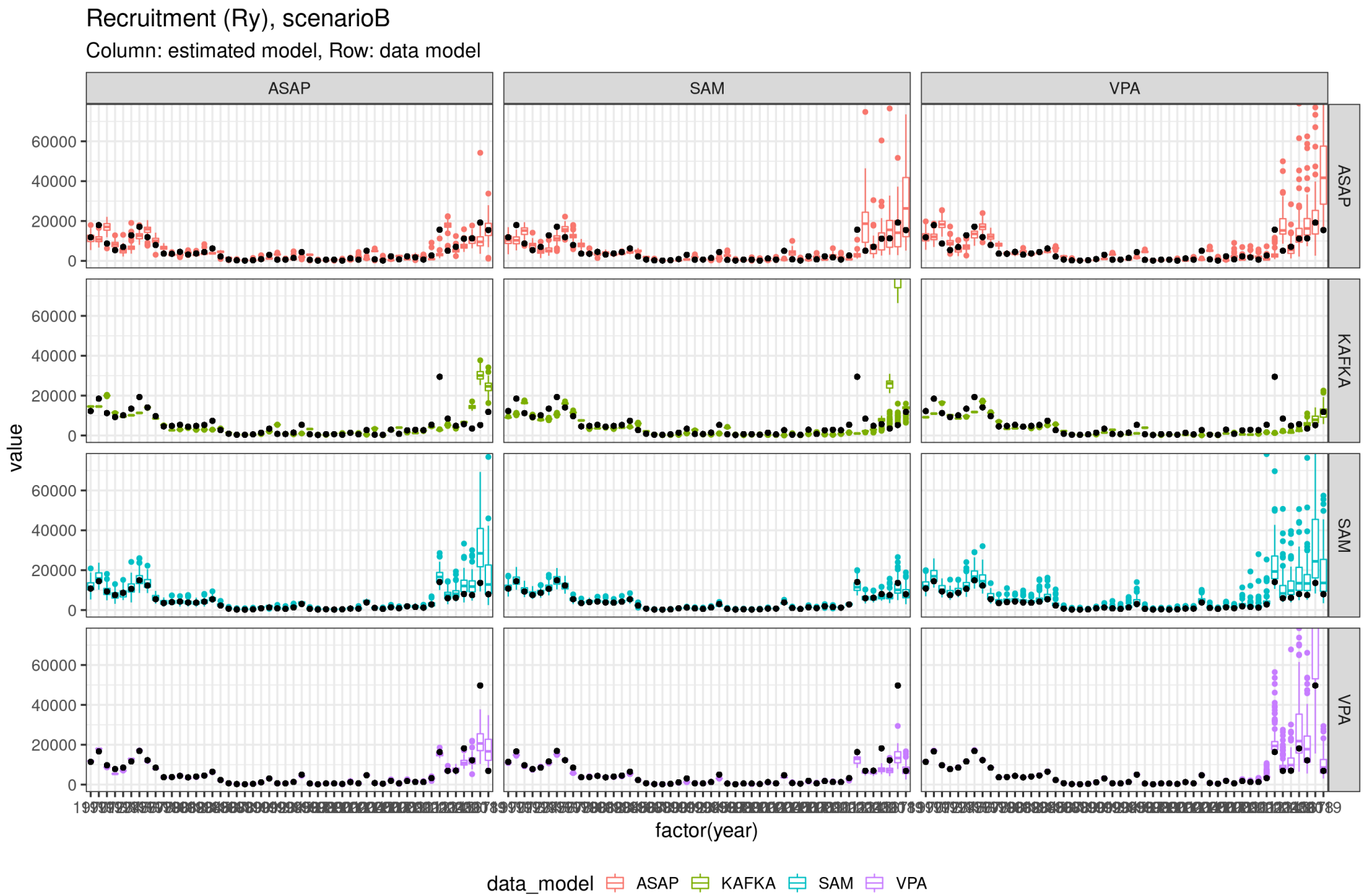

Exploitation rate (Ey), scenario E
Column: estimated model, Row: data model
ASAP
SAM
VPA
ASAP
KAFKA
SAM
VPA
value
0.0
0.2
0.4
0.6
0.8
factor(year)
data_model
ASAP
KAFKA
SAM
VPA

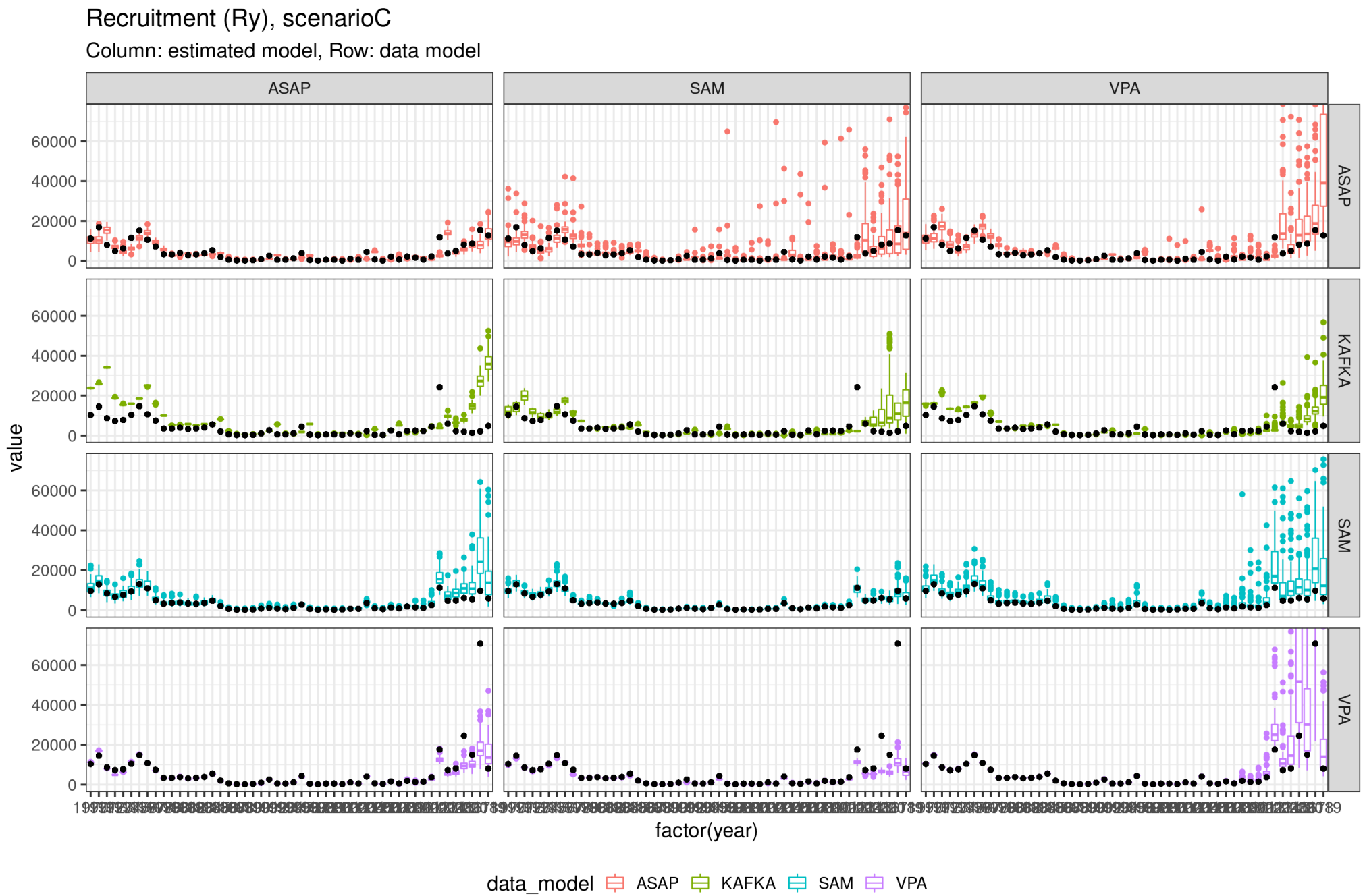

Exploitation rate (Ey), scenario F
Column: estimated model, Row: data model
ASAP
SAM
VPA
ASAP
KAFKA
SAM
VPA
value
0.0
0.2
0.4
0.6
0.8
factor(year)
data_model
ASAP
KAFKA
SAM
VPA

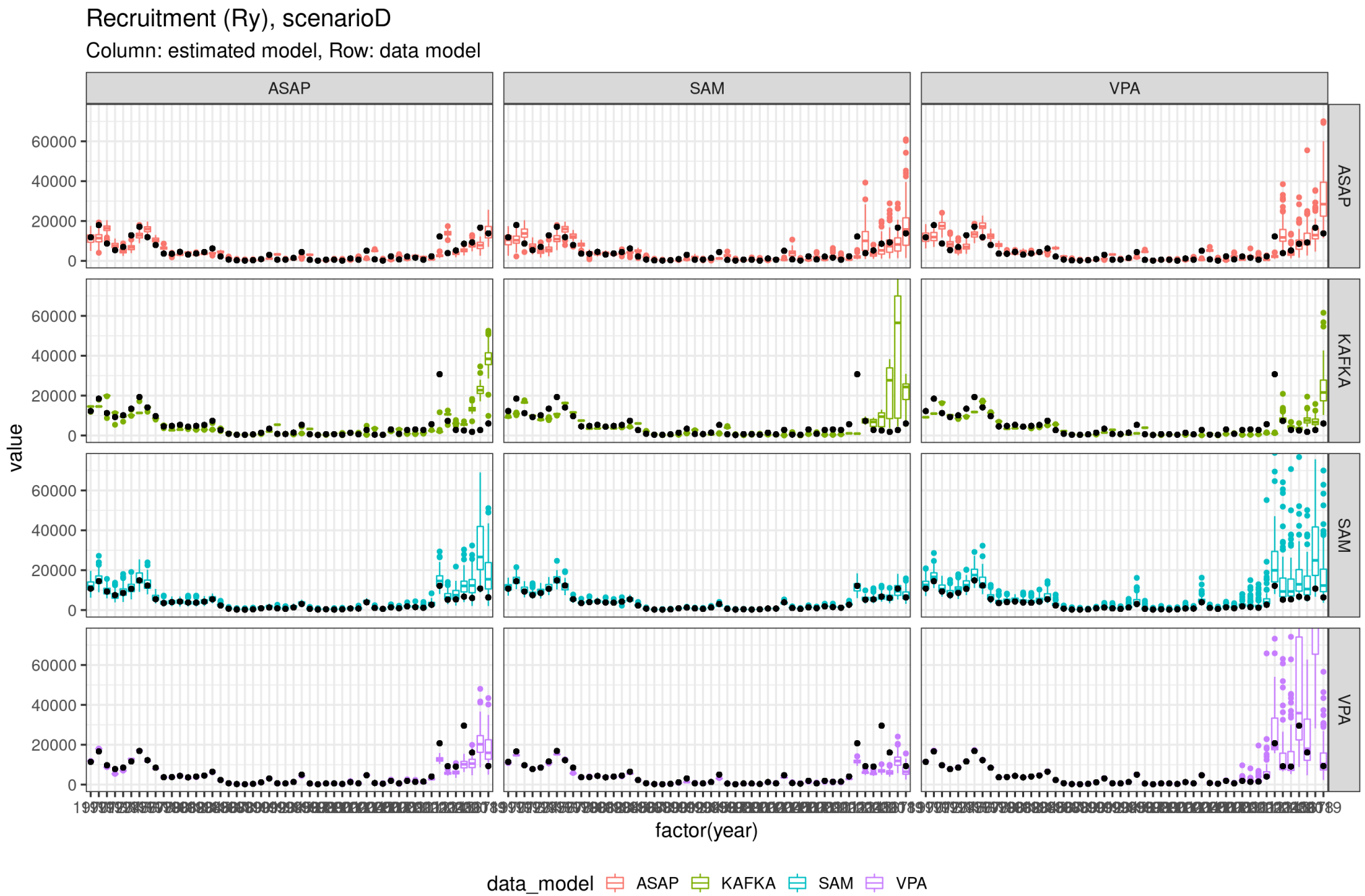
Recruitment (Ry)

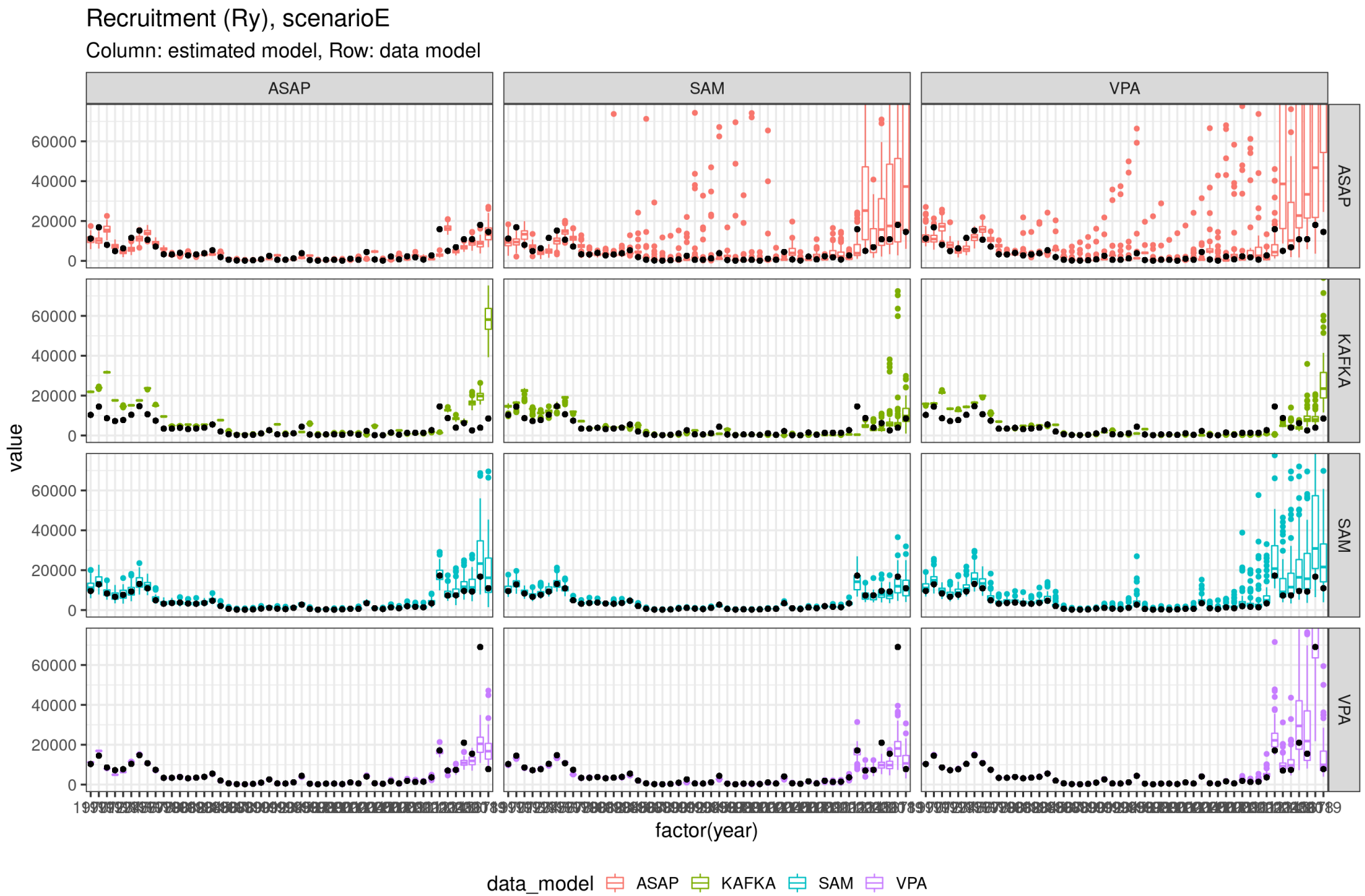
- It it curious that ASAP estimates number of recruits relatively well especically for the past years, while there are obvious overestination of total biomass in ASAP (where the recruitment have gone?).

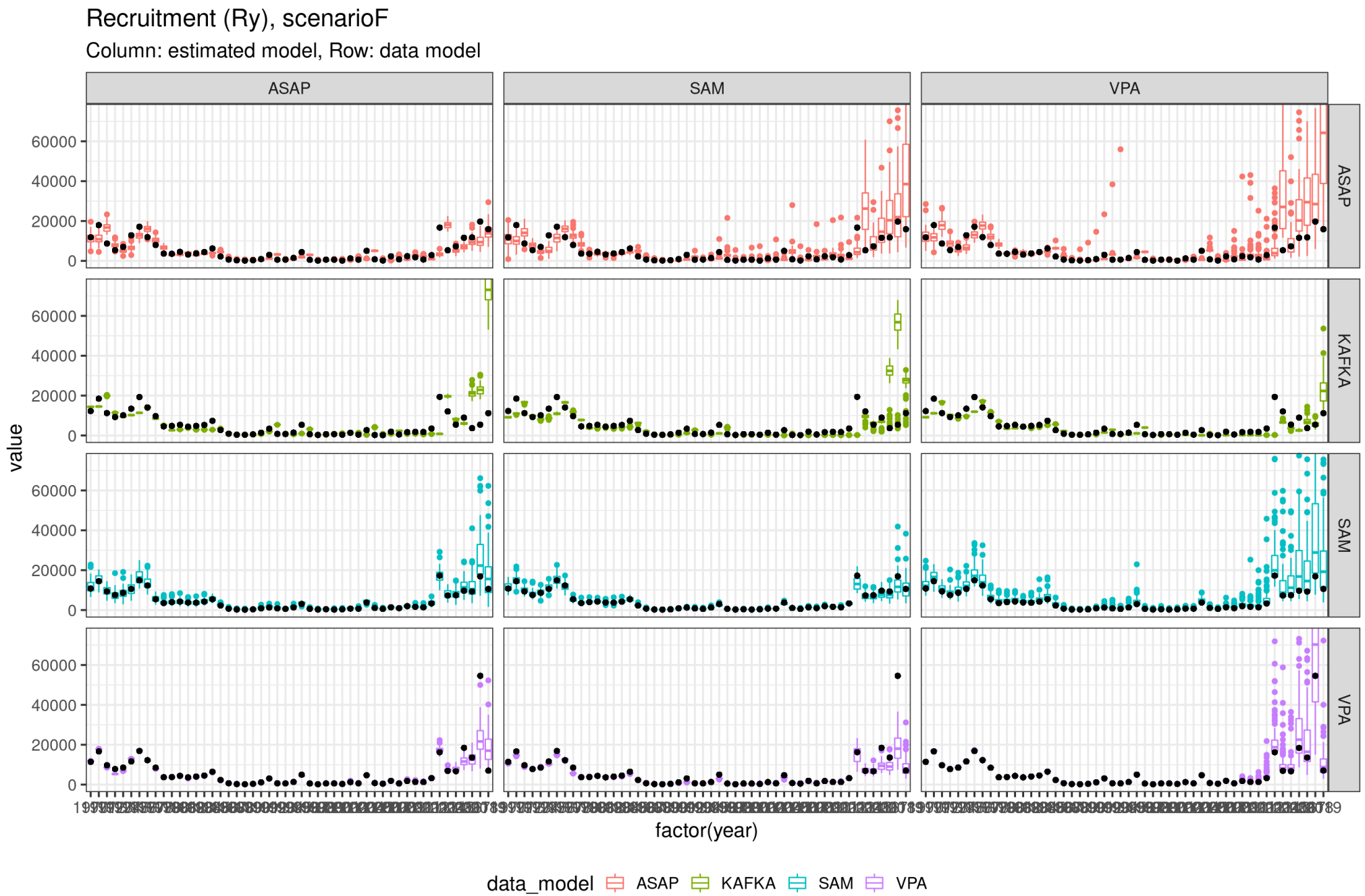

Recruitment (Ry), scenarioA
Column: estimated model, Row: data model
ASAP
SAM
VPA
ASAP
KAFKA
SAM
VPA
value
0
20000
40000
60000
factor(year)
data_model
ASAP
KAFKA
SAM
VPA


Recruitment (Ry), scenarioB
Column: estimated model, Row: data model
ASAP
SAM
VPA
KAFKA
value
factor(year)
data_model
ASAP
KAFKA
SAM
VPA


Recruitment (Ry), scenarioC
Column: estimated model, Row: data model
ASAP
SAM
VPA
ASAP
KAFKA
SAM
VPA
value
0
20000
40000
60000
factor(year)
data_model
ASAP
KAFKA
SAM
VPA

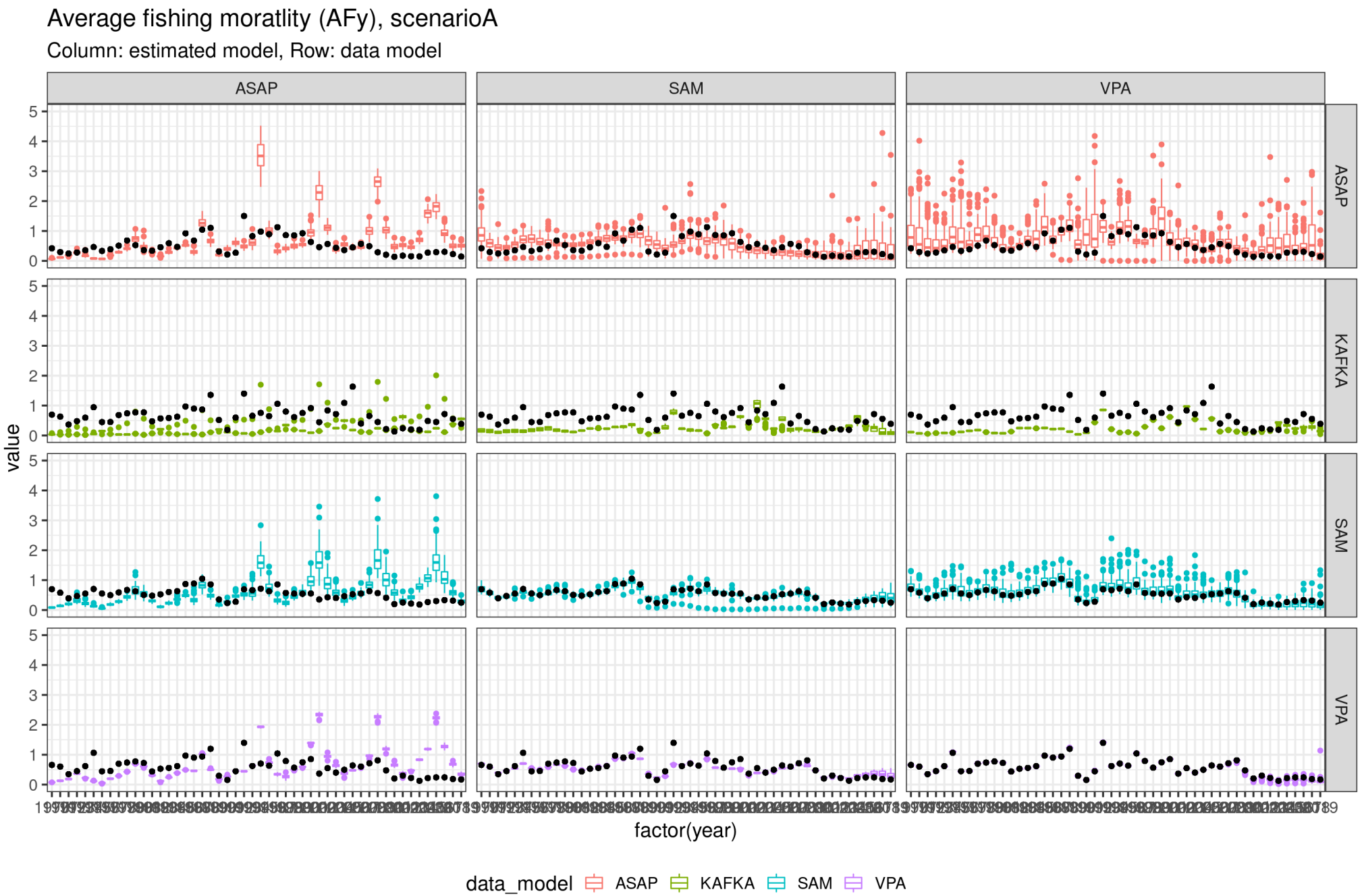

Recruitment (Ry), scenarioD
Column: estimated model, Row: data model
ASAP
SAM
VPA
KAFKA
value
factor(year)
data_model
ASAP
KAFKA
SAM
VPA

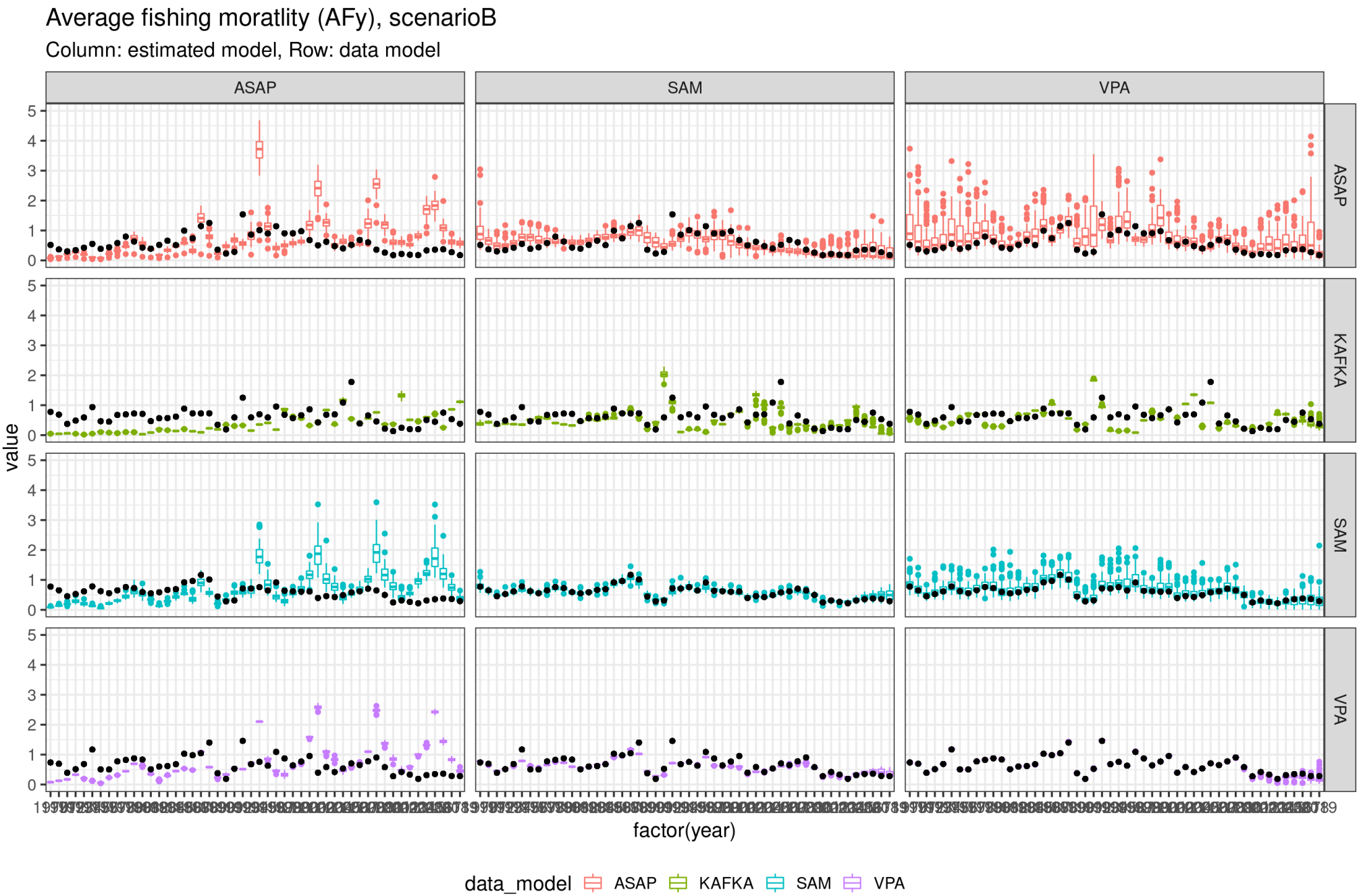

Recruitment (Ry), scenarioE
Column: estimated model, Row: data model
ASAP
SAM
VPA
KAFKA
value
factor(year)
data_model
ASAP
KAFKA
SAM
VPA

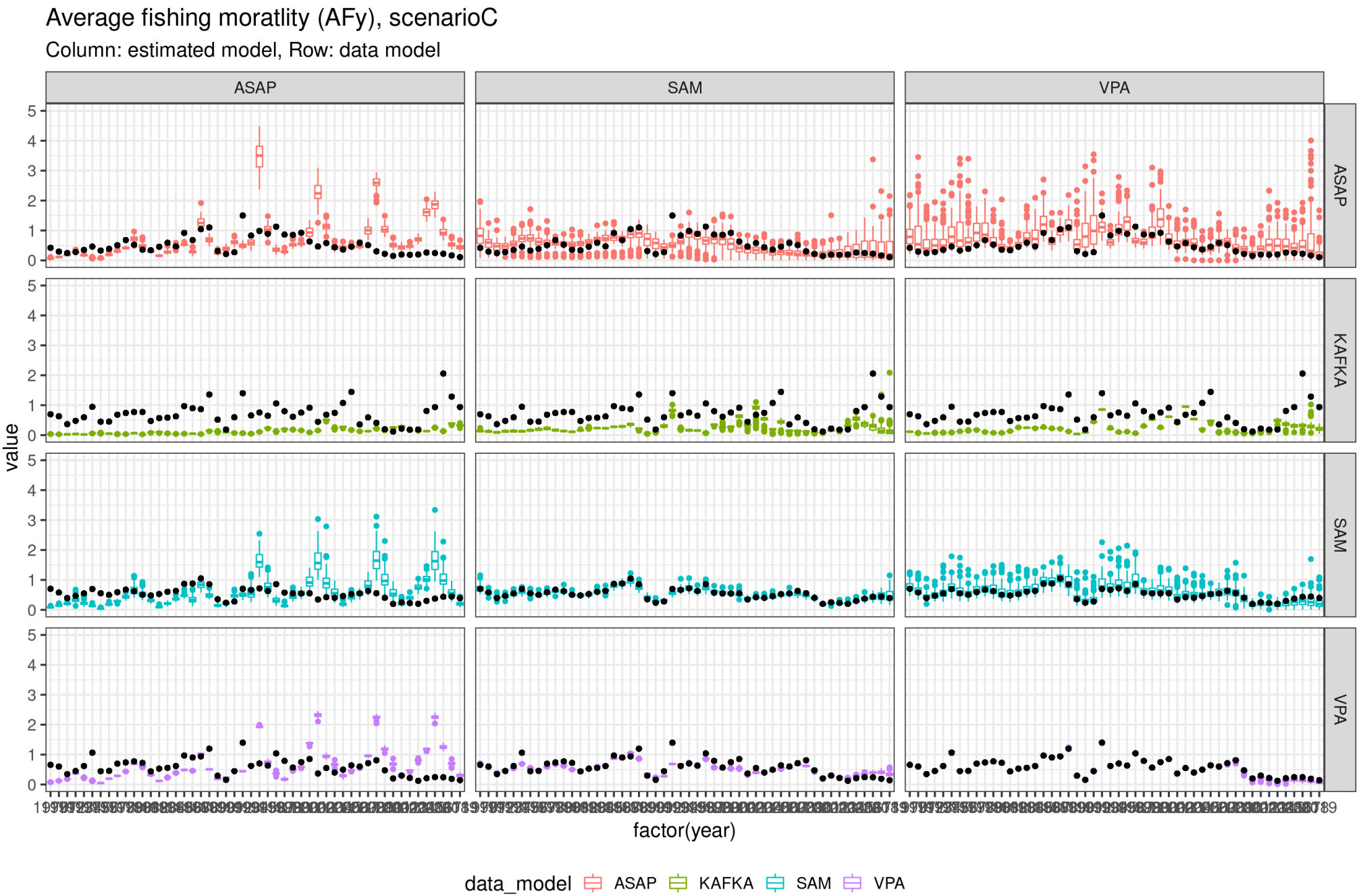

Recruitment (Ry), scenarioF
Column: estimated model, Row: data model
ASAP
SAM
VPA
KAFKA
value
factor(year)
data_model
ASAP
KAFKA
SAM
VPA

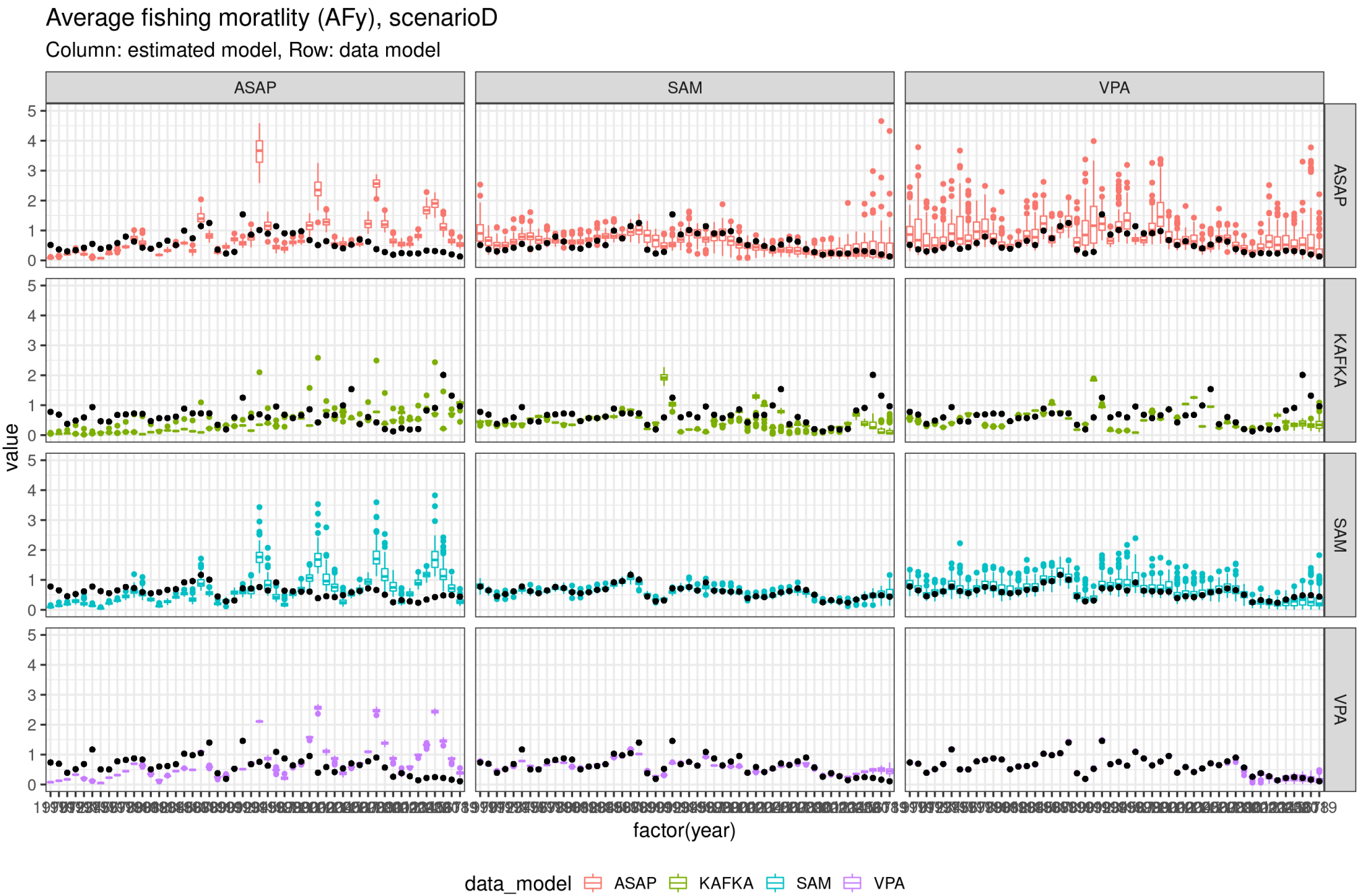
Average fishing mortality (AFy)

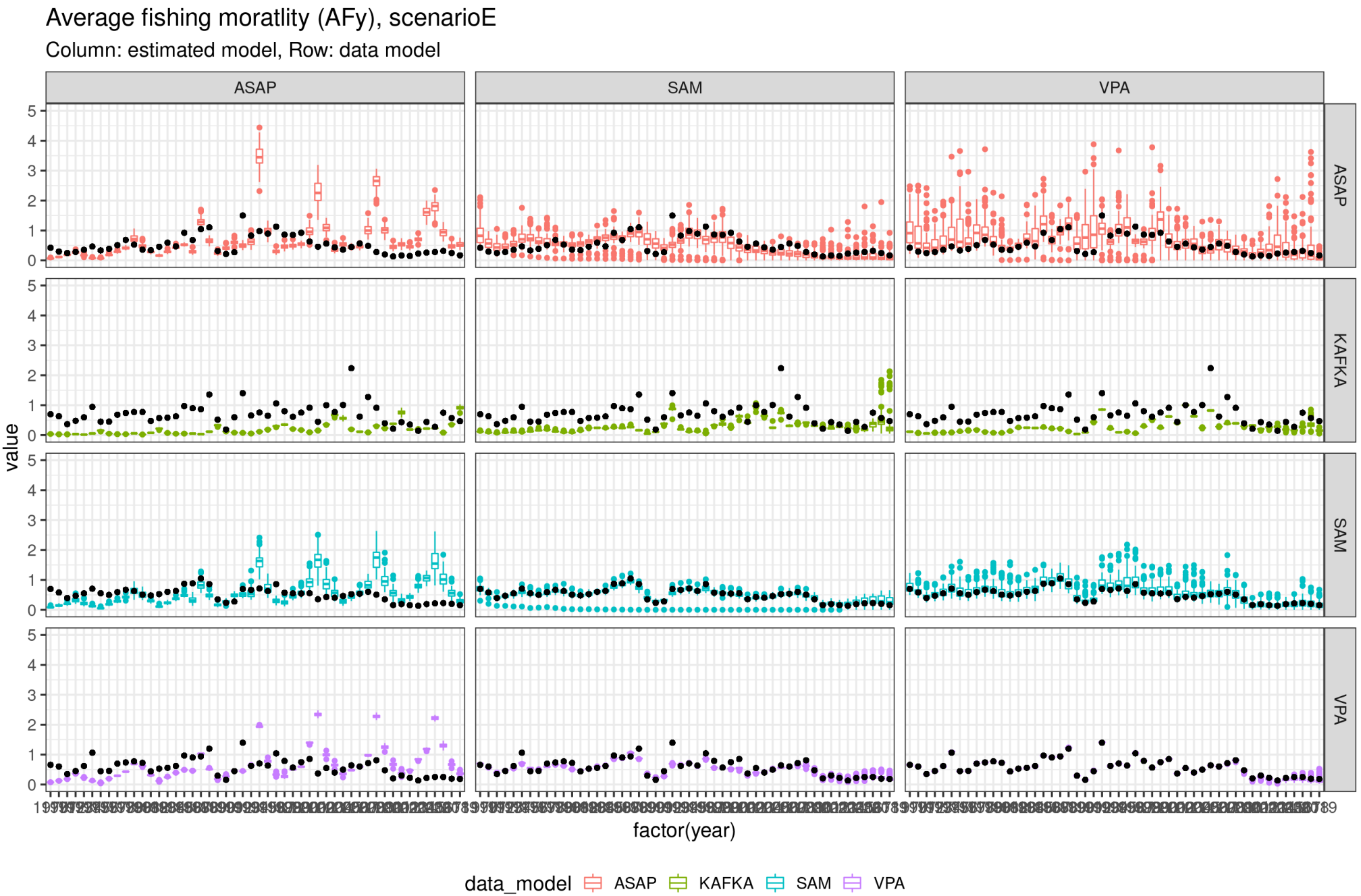
- It is curious that periodic patterns are observed in ASAP

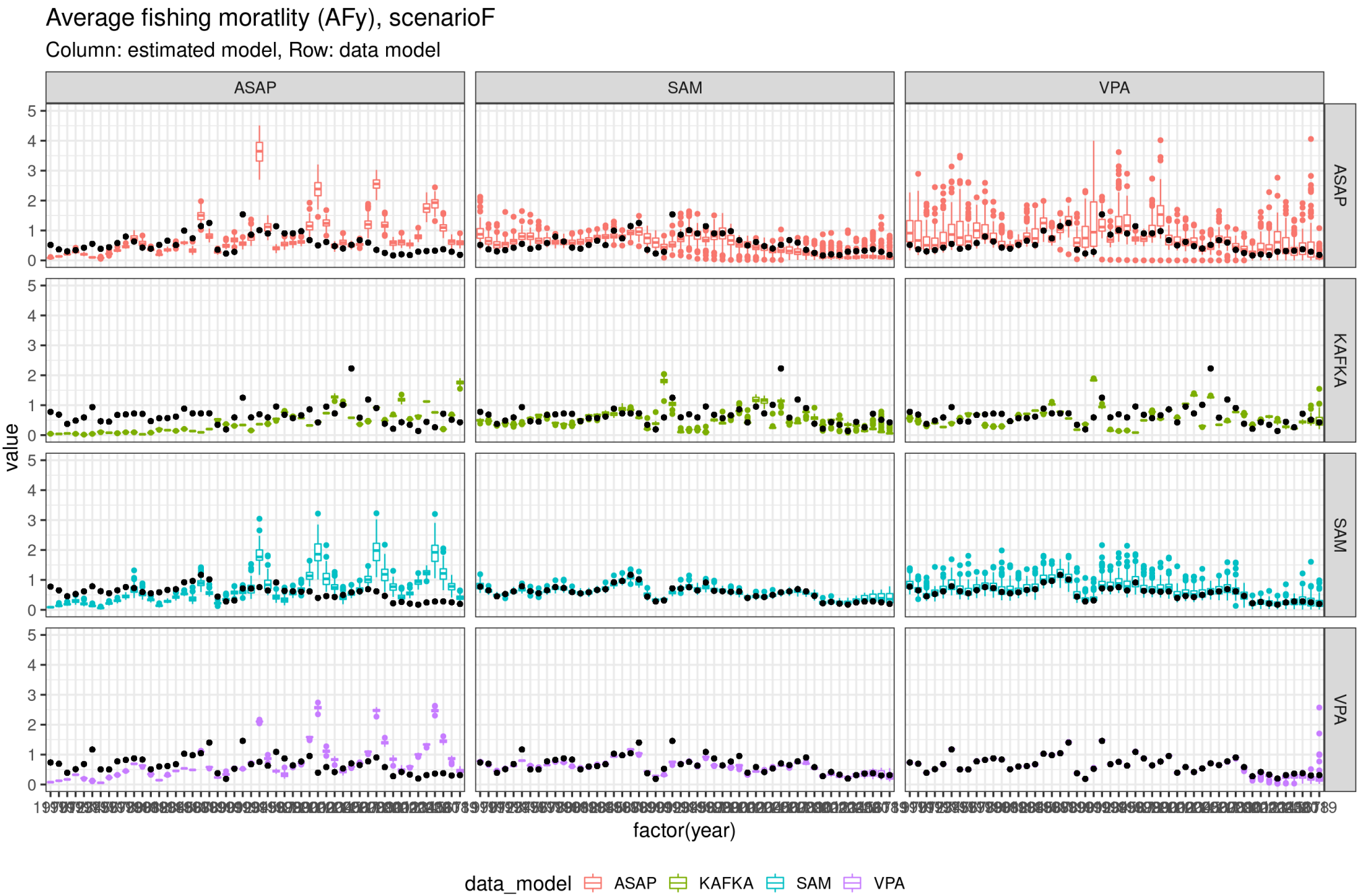

Average fishing moratlity (AFy), scenarioA
Column: estimated model, Row: data model
ASAP
SAM
VPA
ASAP
KAFKA
SAM
VPA
value
factor(year)
data_model
ASAP
KAFKA
SAM
VPA


Average fishing moratlity (AFy), scenarioB
Column: estimated model, Row: data model
ASAP
SAM
VPA
ASAP
KAFKA
SAM
VPA
value
factor(year)
data_model
ASAP
KAFKA
SAM
VPA


Average fishing moratlity (AFy), scenarioC
Column: estimated model, Row: data model
ASAP
SAM
VPA
ASAP
KAFKA
SAM
VPA
value
factor(year)
data_model
ASAP
KAFKA
SAM
VPA


Average fishing moratlity (AFy), scenarioD
Column: estimated model, Row: data model
ASAP
SAM
VPA
ASAP
KAFKA
SAM
VPA
value
factor(year)
data_model
ASAP
KAFKA
SAM
VPA


Average fishing moratlity (AFy), scenarioE
Column: estimated model, Row: data model
ASAP
SAM
VPA
ASAP
KAFKA
SAM
VPA
value
factor(year)
data_model
ASAP
KAFKA
SAM
VPA


Average fishing moratlity (AFy), scenarioF
Column: estimated model, Row: data model
ASAP
SAM
VPA
ASAP
KAFKA
SAM
VPA
value
factor(year)
data_model
ASAP
KAFKA
SAM
VPA

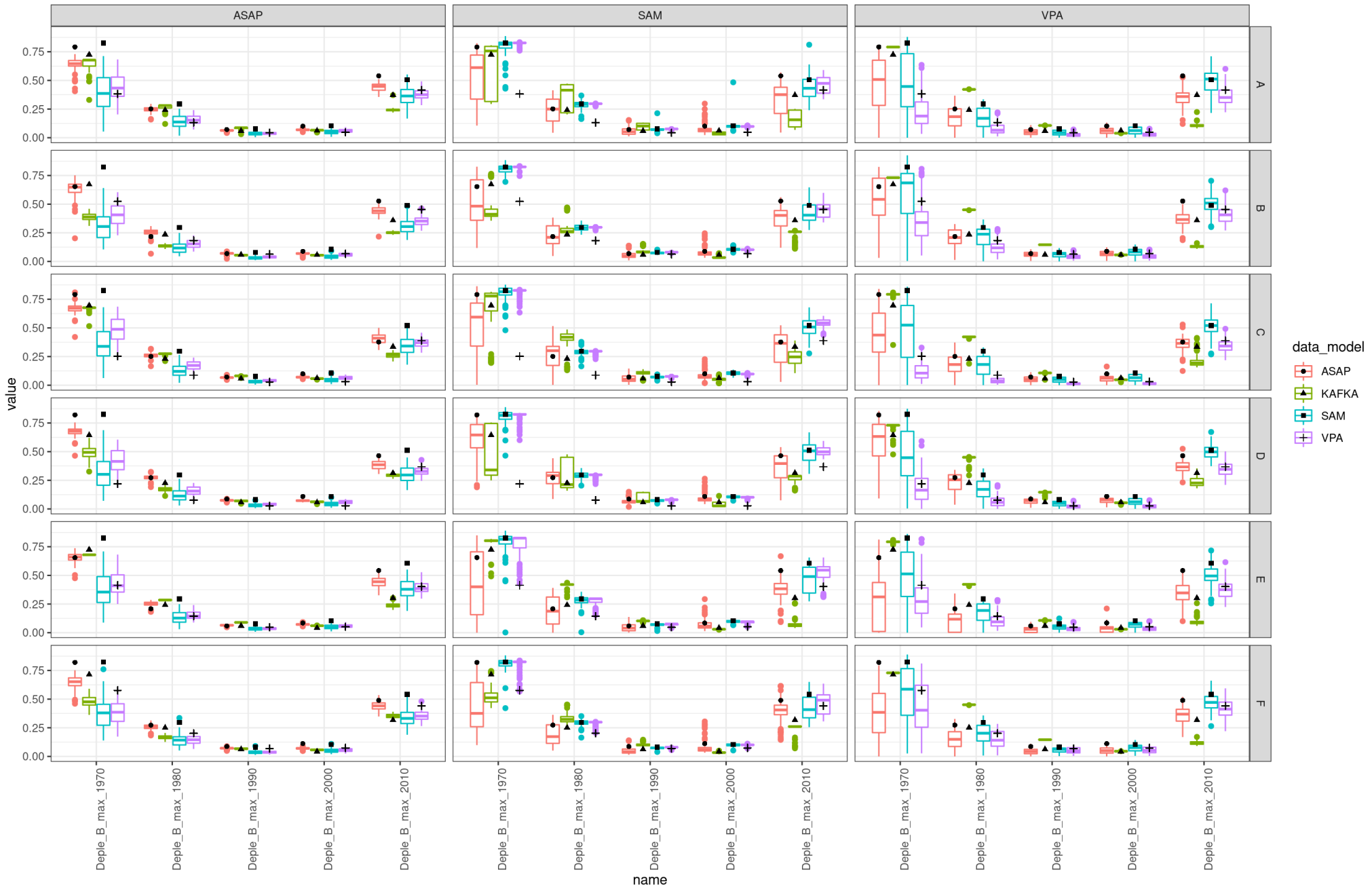
Summary of state variables

- ASAP generally **overestimate total biomass** while there is no much bias in recruimtnet estimation
- VPA especially overestimate recent year's biomass
- SAM is generally good, but underestimate total biomass of VPA data

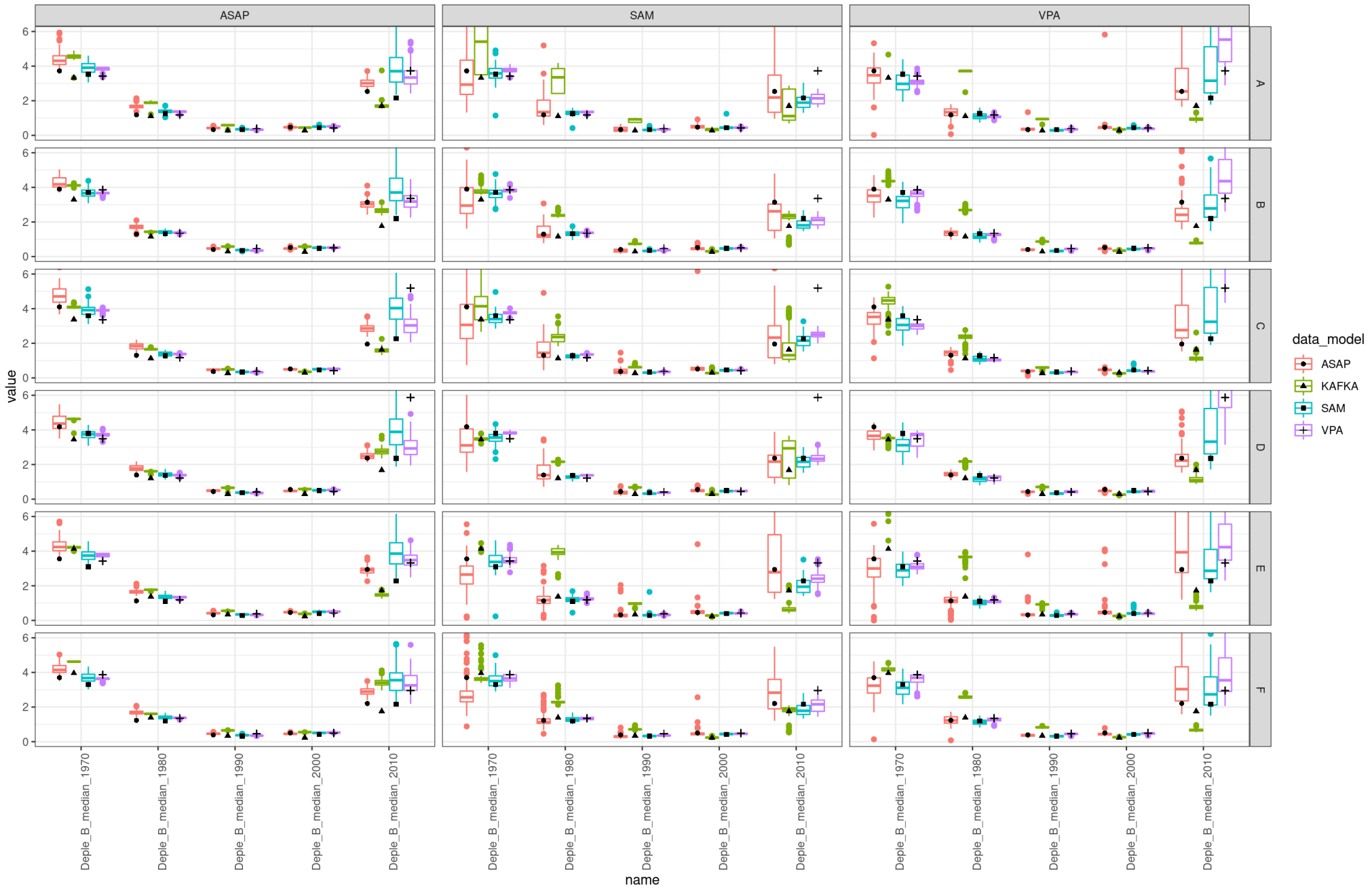
Depletion statistics

- Variance of recent year's from VPA is higher in Deple_B_median than Deple_B_max, probably because VPA estimates the most recent total biomass as the maximum

Depletion statistics



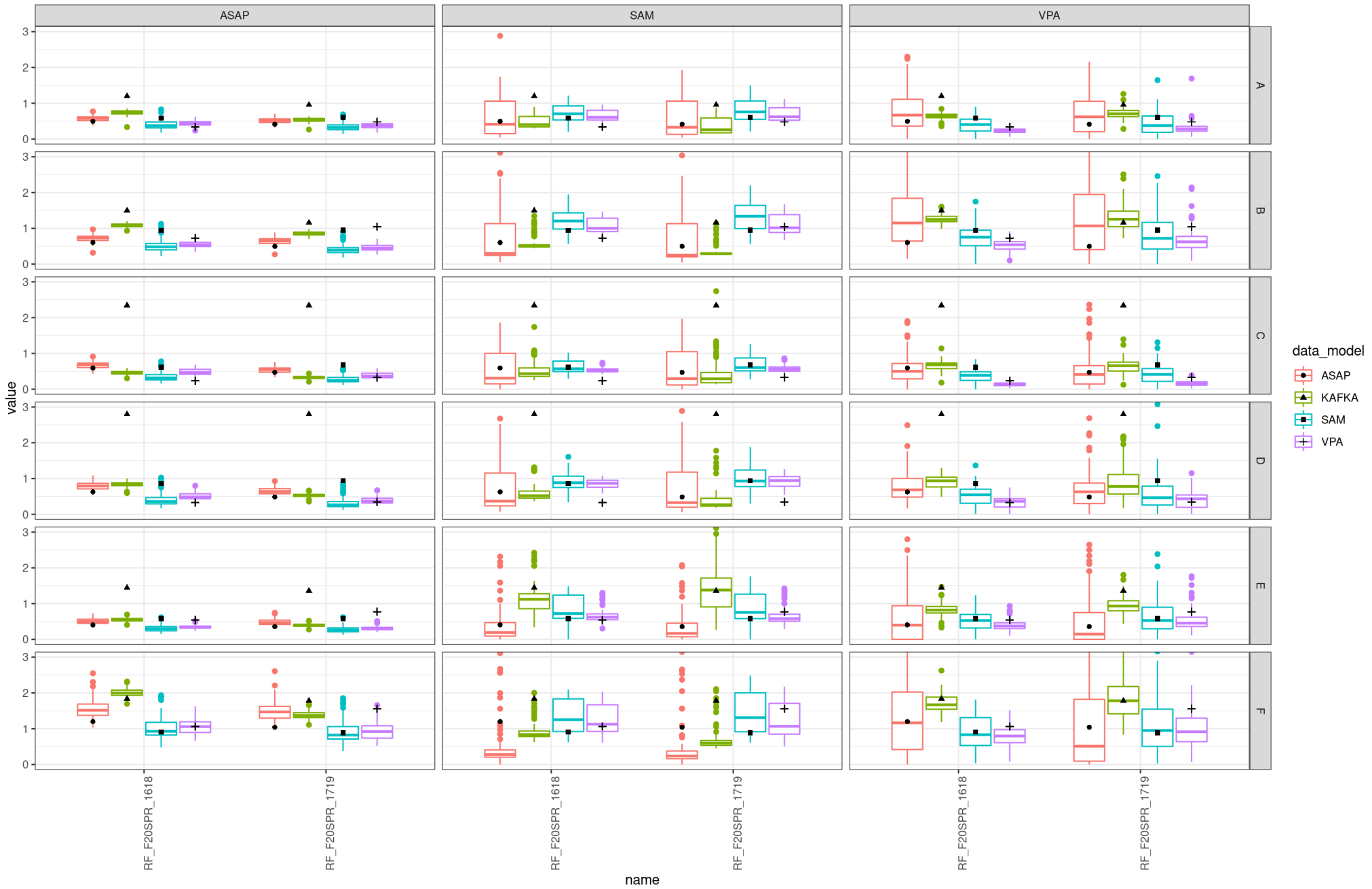
Depletion statistics



F/F%SPR (relative value to F2016-2018 and F2017-2019)

- Comparison between RF_F20SPR_1618 & RF_F20SPR_1719: no large difference but larger varitaion is shown in VPA in 1719 than 1618
- Comparison among estimation models
  - *Variation: ASAP > SAM > VPA (probably because of the assumption of selectivity)*
  - *Bias: Difficult to see?*

F/F%SPR


ASAP
SAM
VPA
A
B
C
D
E
F
value
0
1
2
3
RF_F20SPR_1618
RF_F20SPR_1719
name
data_model
ASAP
KAFKA
SAM
VPA

F/F0.1 & F/Fmax (relative value to F2016-2018 and F2017-2019 when assuminig HS)

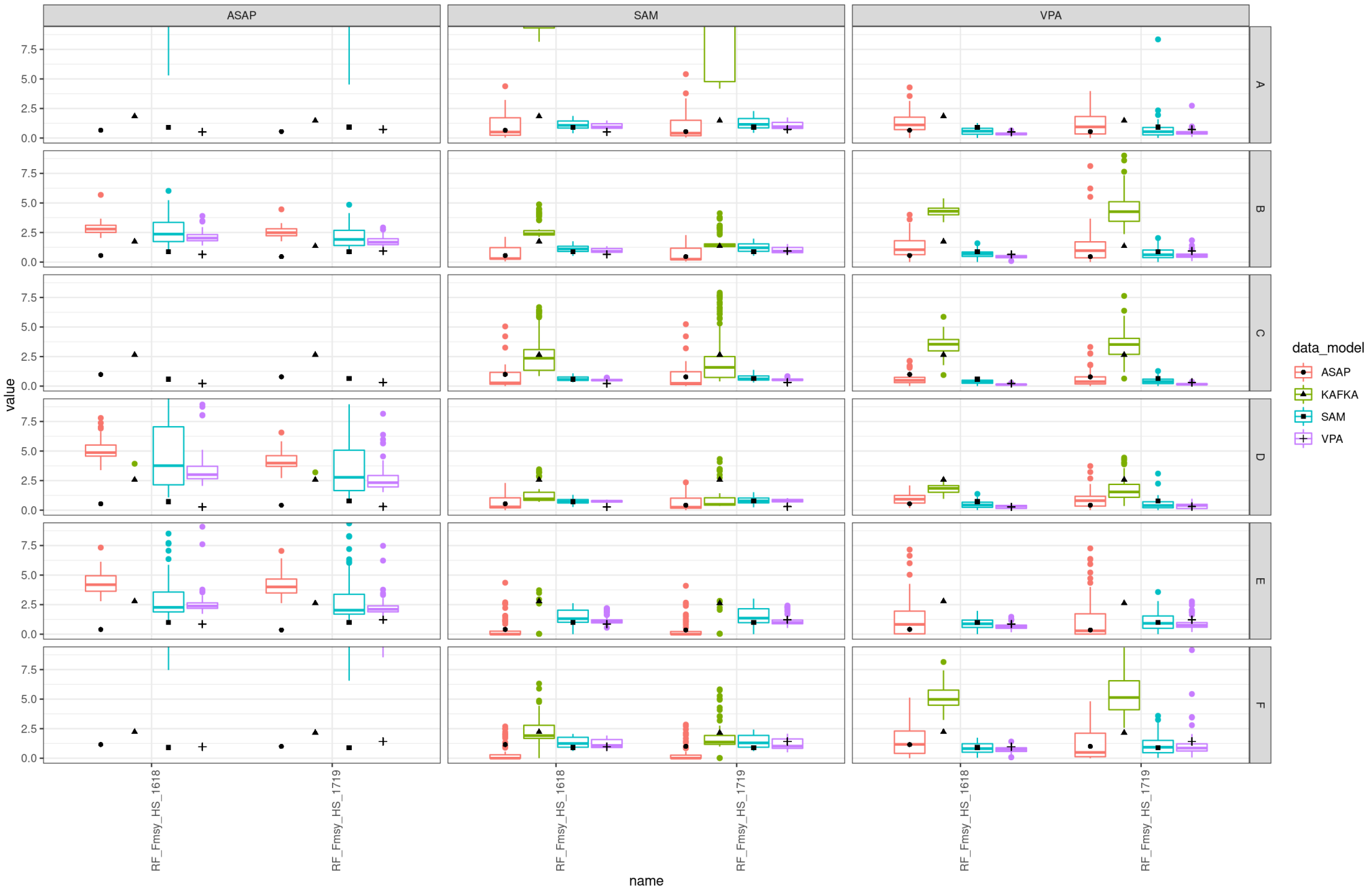
- Pettern of variances are similar with F20%
- F0.1 is more uncertain than Fmax especially in VPA

F/Fmax of F0.1


ASAP
SAM
VPA
A
B
C
D
E
F
value
name
RF_F0.1_1618
RF_F0.1_1719
RF_Fmax_1618
RF_Fmax_1719
data_model
ASAP
KAFKA
SAM
VPA

F/Fmsy based on HS (relative value to F2016-2018 and F2017-2019 when assuminig HS)

F/Fmsy (HS)


ASAP
SAM
VPA
A
B
C
D
E
F
value
0.0
2.5
5.0
7.5
RF_Fmsy_HS_1618
RF_Fmsy_HS_1719
name
data_model
ASAP
KAFKA
SAM
VPA

F/Fmsy based on BH (relative value to F2016-2018 and F2017-2019 when assuminig HS)

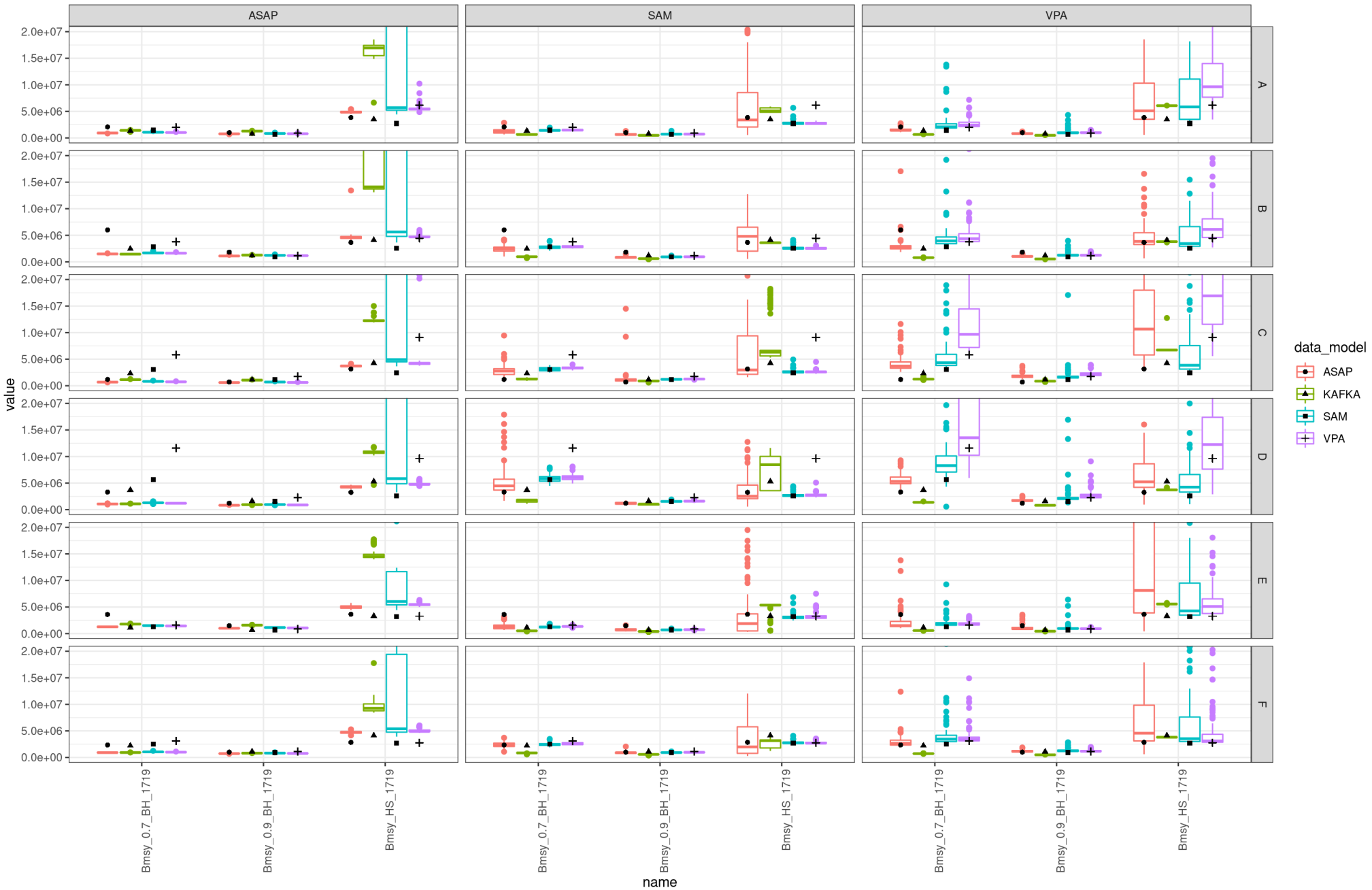
- More stable than F/Fmsy based on HS (probably because steepness parameter is not estimated)

F/Fmsy (BH)


ASAP
SAM
VPA
A
B
C
D
E
F
value
0.0
2.5
5.0
7.5
RF_Fmsy_0.7_BH_1618
RF_Fmsy_0.7_BH_1719
RF_Fmsy_0.9_BH_1618
RF_Fmsy_0.9_BH_1719
name
data_model
ASAP
KAFKA
SAM
VPA

Bmsy

Bmsy (HS) & Bmsy (BH)



Summary of reference points

- Variation in F1719 is tend to be larger than F1618 especially in VPA
- Fmsy and Bmsy from HS is unstable probably because steepness is estimated, which suggest dificulties in estimating stock-recruitment relationship (h) under CM population dynamics
- The smallest variance is shown in ASAP probably because of assumption of selectivitiy (constant throgh all years)

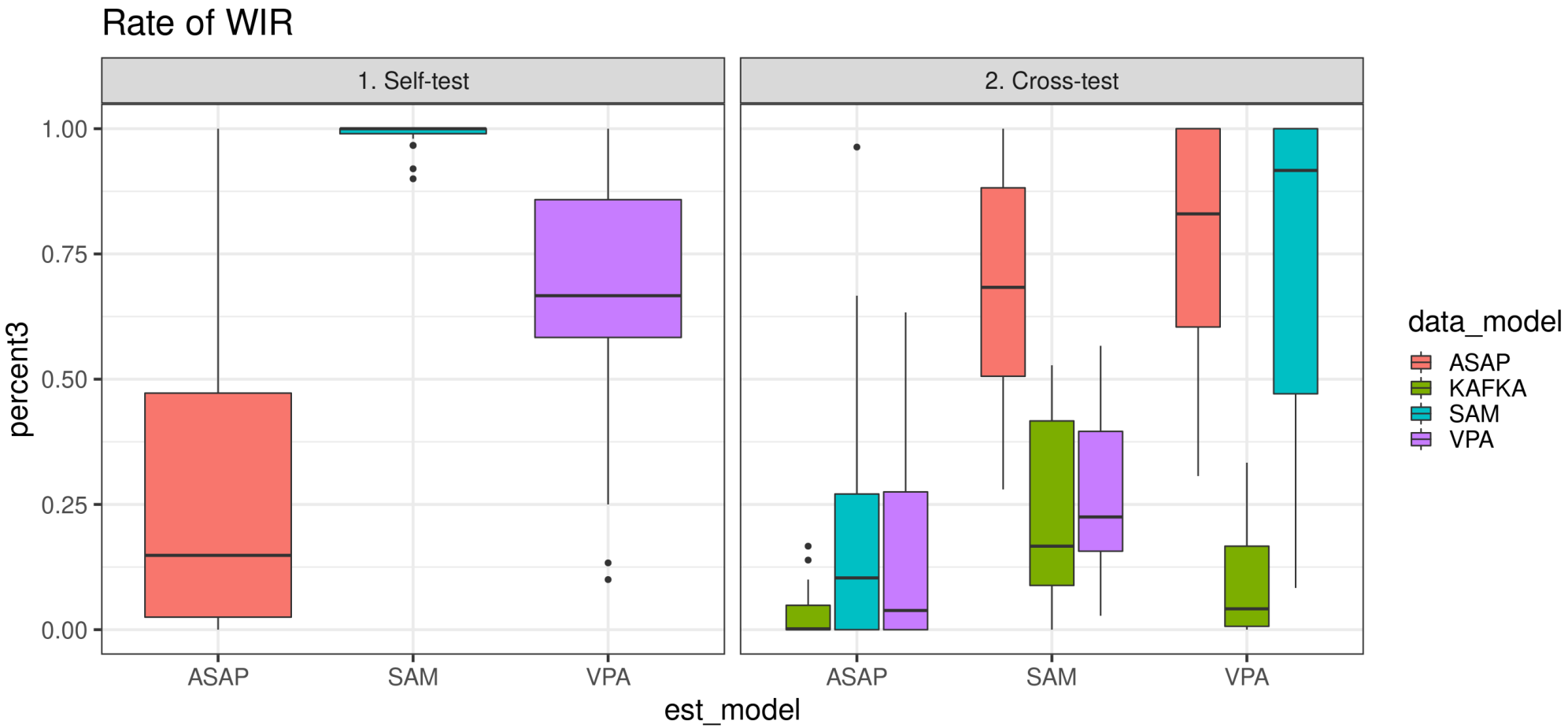
One idea: the index for evaluation of model: WIR (within range) indicator

- Count the following scores
  - *1: quantile(est_value, 0.1) < true_value < quantile(est_value, 0.9)*
  - *0: otherwise*
- Then, the rate of positive (score=1) is calculated by performance measure categories (such as AFy, TBy, RF_FMSY etc..)
- High rate of positive means good performance
- **Potential cavet: the modesl with higher variances can take an advantage**

```
## [1] TRUE
```

Rough comparison of the total score rate

- We separate “self test”/“cross test”
  - *“self test”: need to show high performance because underline assumption is same betwee data generation and estimation models*
  - *“cross test”: check for robustness of the estimation model even thogh underline assumption of population dynamics is mis-specified*



comparison by categories

- As observed in the rough comparison between est and true, ASAP perform well for reference points, but not in state variables
- SAM generally perform well


Rate of WIR3 by category
AFy
Bmsy_._BH_
Bmsy_HS_
Deple_B_max_
Deple_B_median_
Deple_SSB_max_
Deple_SSB_median_
Ey
F._
Fmax_
Fmsy_._BH_
Fmsy_HS_
FSPR_
RF_F._
RF_Fmax_
RF_Fmsy_._BH_
RF_Fmsy_HS_
RF_FSPR_
Ry
SBmsy_._BH_
SBmsy_HS_
TBy
percent3
1.00
0.75
0.50
0.25
0.00
ASAP
SAM
VPA
est_model
data_model
ASAP
KAFKA
SAM
VPA
est_model
ASAP
SAM
VPA