Forecast probabilities - work in progress

Nicolas Blöchliger, Institute of Medical Microbiology, University of Zurich

07/06/2016

Contents

1 Goal **1**

2 Materials and methods **1**

2.1 Model 1

2.2 Data 2

2.3 Probabilities of misclassification errors 3

3 Results **4**

3.1 Beta-lactams 5

3.2 Aminoglycosides 17

3.3 Quinolones 19

3.4 Tetracyclines 22

4 Appendix **26**

4.1 Clinical breakpoints 26

4.2 Distributions for beta-lactams 26

1 Goal

Compute the probability that a strain is susceptible according to official breakpoints given an observed diameter y.

2 Materials and methods

2.1 Model

We assume:

- The distribution of the true diameter X is a mixture of $n \geq 2$ components with weights $w_i = p(C = i)$, where C encodes the component. The true diameter is 6 mm for the first component and normally distributed for the other components:

$$p_i(x) = f_X(x|C = i) = \begin{cases} \delta_6(x) & \text{if } i = 1, \\ \phi(x; \mu_i, \sigma_i^2) & \text{else,} \end{cases}$$

where

$$\delta_6(x) = \begin{cases} \infty & \text{if } x = 6 \text{ mm}, \\ 0 & \text{else.} \end{cases}$$

Thus,

$$f_X(x) = w_1\delta_6 + \sum_{i=2}^{n} w_i\phi(x; \mu_i, \sigma_i^2).$$

- We observe $Y = X + E$, where E models technical error. E is zero for the first component and normally distributed and independent of X with mean $\mu_E = 0$ and constant variance σ_E^2 for the other components. The conditional density of Y given the component C is therefore

$$f_Y(y|C = i) = \begin{cases} \delta_6(y) & \text{if } i = 1, \\ p_i * \phi(\cdot; 0, \sigma_E^2) = \phi(y; \mu_i, \sigma_i^2 + \sigma_E^2) & \text{else.} \end{cases}$$

Thus,

$$f_Y(y) = w_1\delta_6 + \sum_{i=2}^{n} w_i\phi(y; \mu_i, \sigma_i^2 + \sigma_E^2).$$

Note that we do not account for the fact that the observed data are rounded to integer values.

2.1.1 Estimation of model parameters

- We estimate w_1 as the fraction of data points in the sample that are equal to 6 mm.
- For $n = 2, \ldots, 10$, the parameters of the normally distributed components of Y, i.e. w_i, μ_i, and $\sigma_i^2 + \sigma_E^2$, are estimated by fitting a normal mixture model of $n - 1$ components to the data in the sample with diameters greater than 6 mm. Based on the Bayesian information criterion, the best best value of n is selected. Values of n that resulted in models that did not satisfy $\sigma_i > 0$ were excluded. We use the R package `mclust`.
- Estimates for the variance of the error σ_E^2 will be taken from independent work in order to obtain σ_i^2. For the time being we use $\sigma_E^2 = 1$ mm.

2.1.2 Limitations of the model

- The model does not account for the fact that $X \geq 6$ mm. As long as the means of the the normally distributed components are sufficiently large (say $\mu_i - 6 \text{ mm} > 2\sigma$), this should not cause problems.
- The model does not account for the fact that $Y \leq 40$ mm. As long as the means of the normally distributed components are sufficiently small (say $40 \text{ mm} - \mu_i > 2\sigma$), this should not cause problems.
- The error is assumed to be normally distributed with constant variance. This assumption is obviously violated if X is close to 6 or 40 mm. It is also violated for antibiotics like CPD, for which diameters are distorted in order to avoid additional laborious tests.
- The distributions of all but the first component are assumed to be normal.

2.2 Data

Raw data for the beta-lactams was obtained from Giorgia Valsesia on 11.03.2016. It contains measurements for 9766 *E. coli* strains collected from January 2010 until March 2014.

Raw data for the aminoglycosides was obtained from Giorgia Valsesia on 15.03.2016. It contains measurements for 3521 *E. coli* strains.

Raw data for the quinolones was obtained from Giorgia Valsesia on 18.05.2015. It contains measurements for 9761 *E. coli* strains collected from February 2011 until May 2014.

Raw data for the tetracyclines was obtained from Giorgia Valsesia on 18.05.2015. It contains measurements for 10662 *E. coli* strains collected from January 2010 until May 2014.

The following 24 antibiotics were used:

antibiotic	abbreviation
ampicillin	AM10
cephalothin	KF
cefoxitin	FOX
cefpodoxime	CPD
amoxicillin - clavulanic acid	AMC
piperacillin - tazobactam	TPZ
cefuroxime	CXM
cefotaxime	CTX
ceftazidime	CAZ
ceftriaxone	CRO
cefepime	FEP
ertapenem	ETP
imipenem	IPM
meropenem	MEM
kanamycin	KAN
gentamicin	GEN
tobramycin	TOB
nalidixic acid	NAL
norfloxacin	NOR
ciprofloxacin	CIP
levofloxacin	LEV
minocycline	MI
tetracycline	TE
tigecycline	TGC

2.3 Probabilities of misclassification errors

The probabilities of very major and major misclassification errors based on offical CBPs are given by

$$p(\text{very major error}) = p(X < \text{CBP}_\text{R} \text{ and } Y \geq \text{CBP}_\text{S}) \text{ and}$$

$$p(\text{major error}) = p(X \geq \text{CBP}_\text{S} \text{ and } Y < \text{CBP}_\text{R}).$$

Additionally, we compute probabilities of misclassification errors if the prediction of S is more conservative:

$$p(\text{very major error}) = p(X < \text{CBP}_\text{R} \text{ and } Y \geq \text{CBP}_\text{S} + 2 \text{ mm}) \text{ and}$$

$$p(\text{major error}) = p(X \geq \text{CBP}_\text{S} \text{ and } Y < \text{CBP}_\text{R}).$$

Finally, if strains with observed diameters in the ZTU are classified as I, probabilities of misclassification errors are given by

$$p(\text{very major error}) = p(X < Z_R \text{ and } Y \geq Z_S) \text{ and}$$

$$p(\text{major error}) = p(X \geq Z_S \text{ and } Y < Z_R),$$

where Z_R is the minimum of CBP~R and the left boundary of the ZTU and Z_S is the maximum of CBP~S and the right boundary of the ZTU.

3 Results

In the figures on the next pages, dashed vertical lines indicate CBPs according to http://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Breakpoint_tables/v_6.0_Breakpoint_table.xls (13.05.2016, see Appendix). The figures are organized as follows.

- Top-left: Histogram of sample and the estimated density of Y (black) and its components (coloured). The contribution from the first component (δ_6) is visualised as a uniform distribution with support [5.5 mm, 6.5 mm].
- Middle-left: Empirical cumulative distribution function (cdf) of Y (grey), its estimate (black) and estimated cdfs for the components of Y (coloured).
- Bottom-left: Q-Q plot. If the estimated density of Y explained the data perfectly, all point would lie on the identity line (grey).
- Top-right: Histogram of sample and the estimated density of X for various values of σ_E.
- Bottom-right: $p(X \geq \text{CBP}_\text{S} | Y = y)$, i.e. the probability that the true diameter is above the CBP defining susceptibility given an observed diameter y. Observed diameters for which this probability is in [0.025,0.975] are defined to be in a zone of technical uncertainty (ZTU). The ZTU is highlighted in red, and its properties are summarised in Tab. 38.

All analysis is restricted to the antibiotics for which CBPs are available.

3.1 Beta-lactams

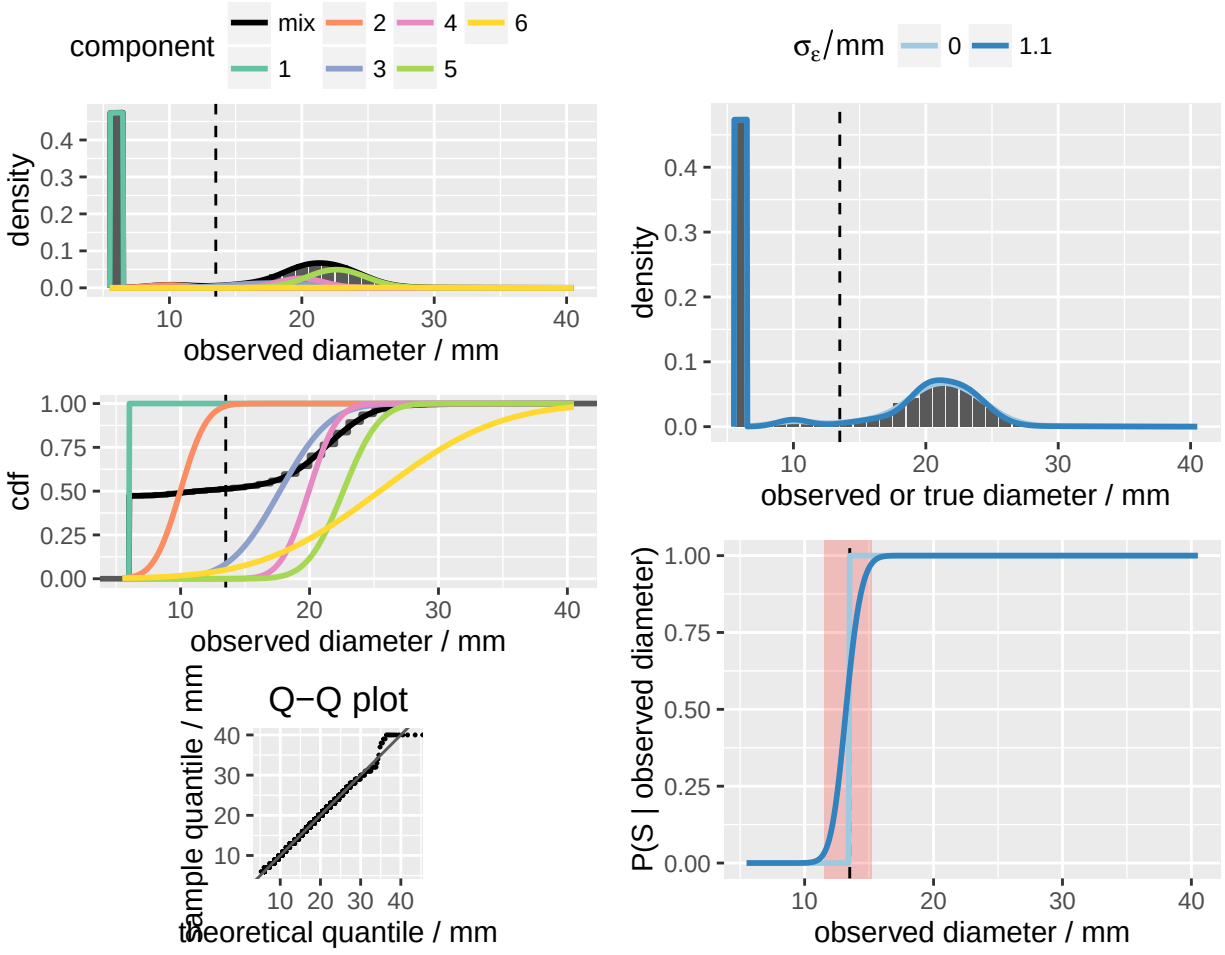
3.1.1 AM10

component	mean/mm	sd/mm	weight
1	6.00	0.00	0.47
2	9.97	1.62	0.03
3	17.70	3.07	0.10
4	19.95	1.73	0.11
5	22.62	2.20	0.27
6	25.45	7.32	0.02

Official CBPs: 14 mm and 14 mm. ZTU: [11.5 mm, 15.2 mm].

Table 3: Probabilities of misclassification errors.

basis for prediction	major error	very major error
official CBPs	3.0e-03	2.0e-03
CBP_S increased by 2mm	3.0e-03	6.7e-05
I for ZTU	8.9e-05	1.3e-04



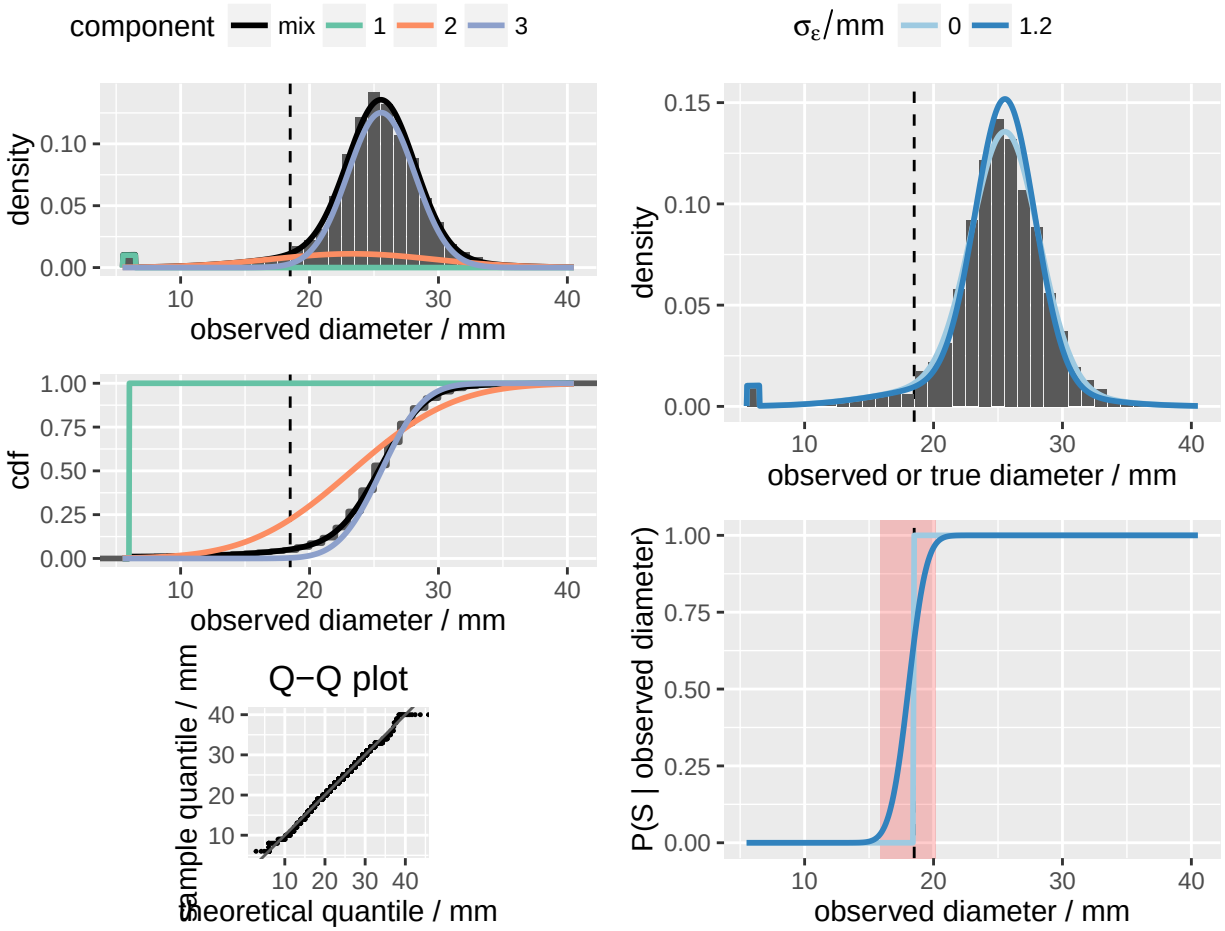
3.1.2 FOX

component	mean/mm	sd/mm	weight
1	6.0	0.0	0.01
2	23.3	6.3	0.18
3	25.6	2.6	0.81

Official CBPs: 19 mm and 19 mm. ZTU: [15.9 mm, 20.2 mm].

Table 5: Probabilities of misclassification errors.

basis for prediction	major error	very major error
official CBPs	6.4e-03	3.9e-03
CBP_S increased by 2mm	6.4e-03	2.1e-04
I for ZTU	7.2e-05	3.7e-04



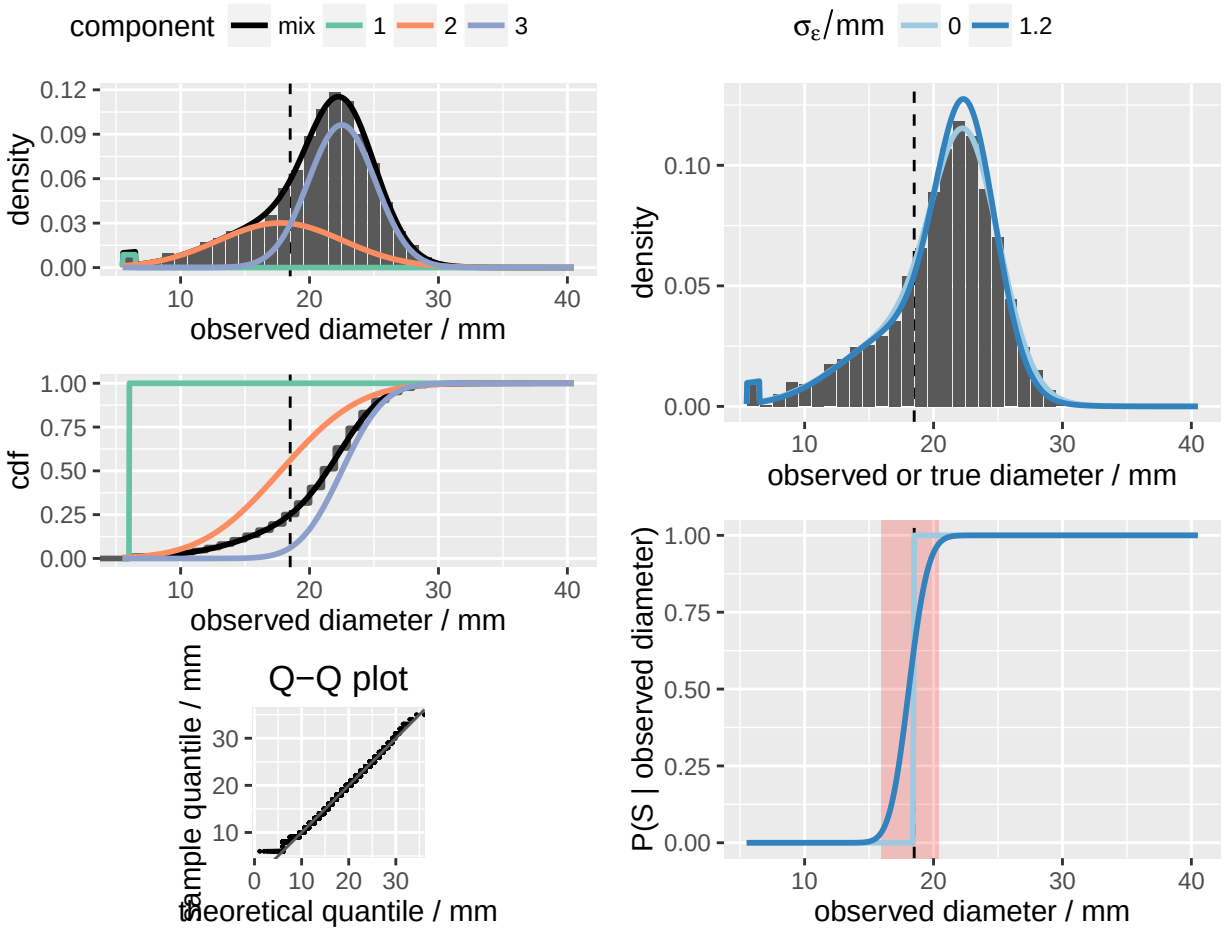
3.1.3 AMC

component	mean/mm	sd/mm	weight
1	6.0	0.00	0.01
2	17.8	4.82	0.36
3	22.5	2.60	0.63

Official CBPs: 19 mm and 19 mm. ZTU: [15.9 mm, 20.4 mm].

Table 7: Probabilities of misclassification errors.

basis for prediction	major error	very major error
official CBPs	3.4e-02	2.2e-02
CBP_S increased by 2mm	3.4e-02	1.2e-03
I for ZTU	4.0e-04	1.4e-03



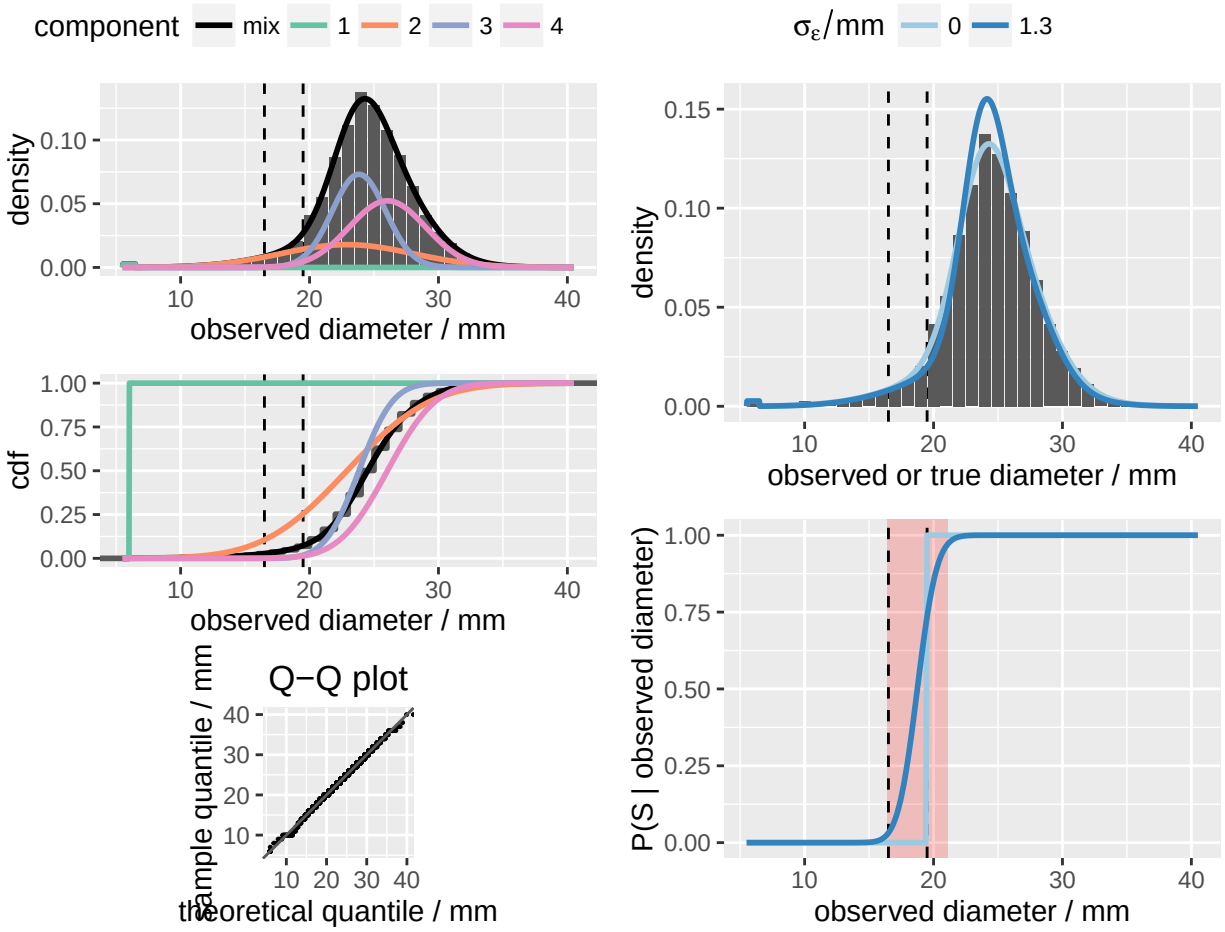
3.1.4 TPZ

component	mean/mm	sd/mm	weight
1	6.0	0.00	0.00
2	22.9	5.12	0.23
3	23.8	2.07	0.38
4	26.0	2.96	0.39

Official CBPs: 17 mm and 20 mm. ZTU: [16.4 mm, 21.1 mm].

Table 9: Probabilities of misclassification errors.

basis for prediction	major error	very major error
official CBPs	1.2e-04	3.4e-05
CBP_S increased by 2mm	1.2e-04	1.4e-07
I for ZTU	9.3e-05	4.9e-07



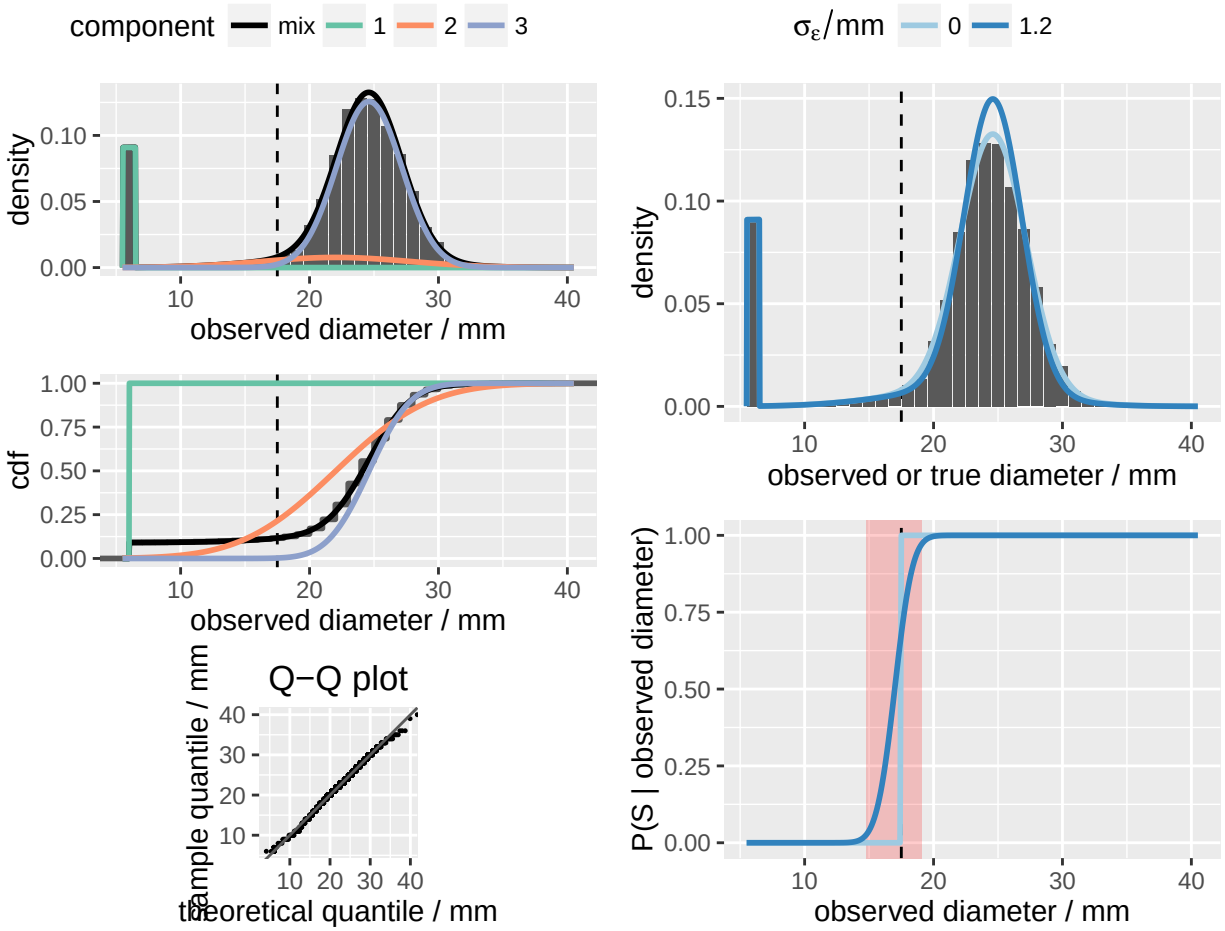
3.1.5 CXM

component	mean/mm	sd/mm	weight
1	6.0	0.00	0.09
2	22.0	5.75	0.11
3	24.6	2.54	0.80

Official CBPs: 18 mm and 18 mm. ZTU: [14.8 mm, 19.1 mm].

Table 11: Probabilities of misclassification errors.

basis for prediction	major error	very major error
official CBPs	4.6e-03	2.6e-03
CBP_S increased by 2mm	4.6e-03	1.4e-04
I for ZTU	3.9e-05	2.9e-04



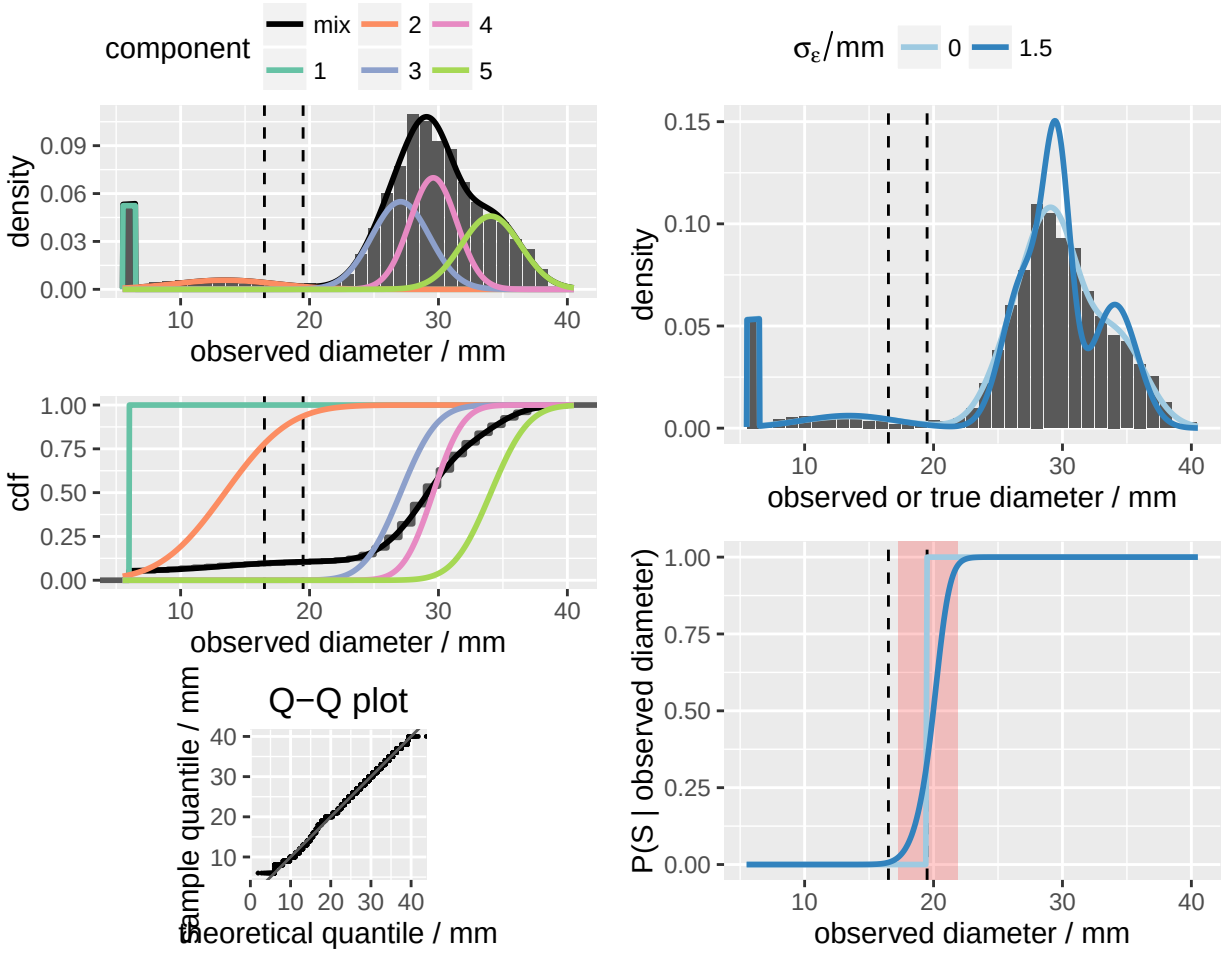
3.1.6 CTX

component	mean/mm	sd/mm	weight
1	6.0	0.00	0.05
2	13.4	3.96	0.06
3	27.1	2.24	0.31
4	29.6	1.82	0.32
5	34.1	2.30	0.26

Official CBPs: 17 mm and 20 mm. ZTU: [17.3 mm, 21.9 mm].

Table 13: Probabilities of misclassification errors.

basis for prediction	major error	very major error
official CBPs	1.6e-05	6.1e-05
CBP_S increased by 2mm	1.6e-05	7.9e-07
I for ZTU	1.6e-05	2.7e-07



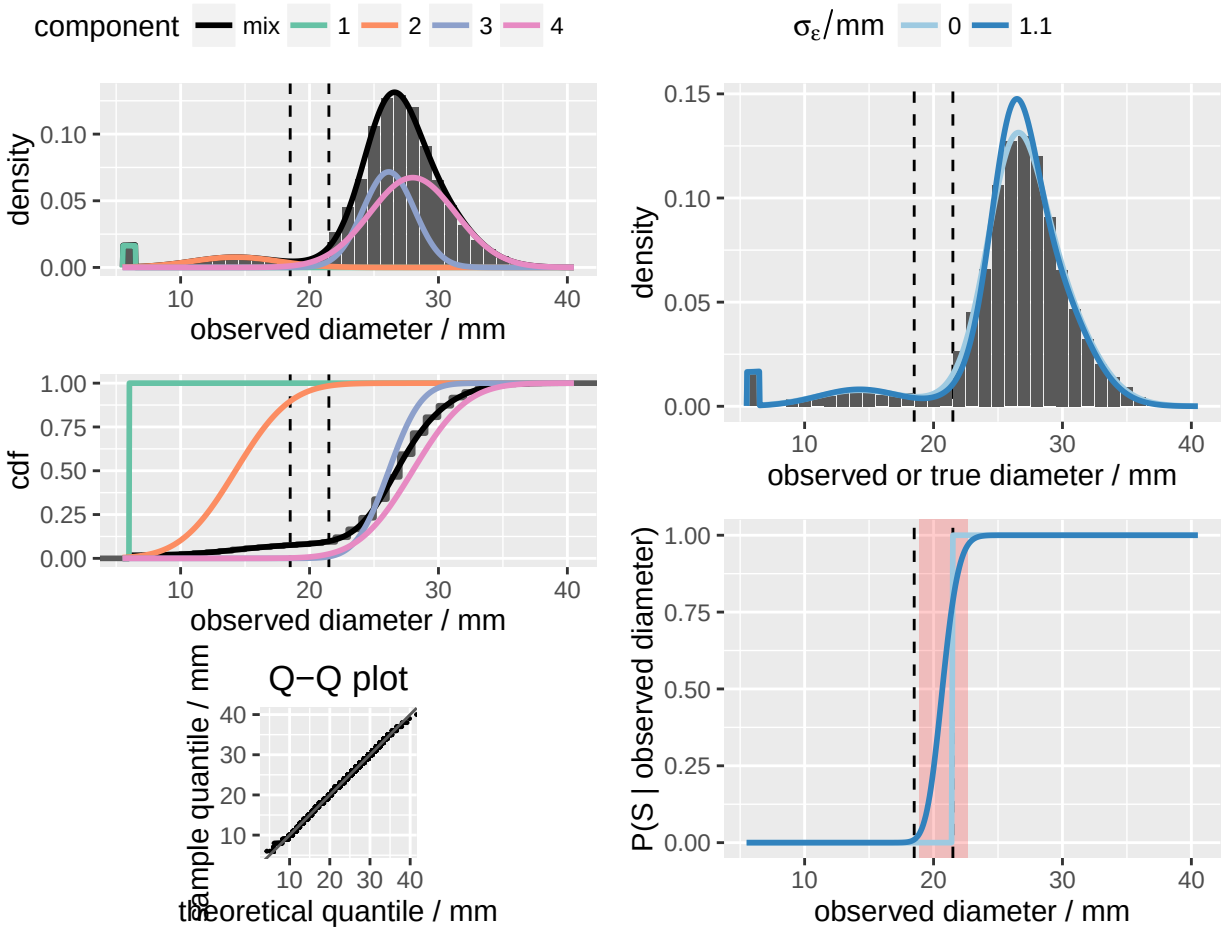
3.1.7 CAZ

component	mean/mm	sd/mm	weight
1	6.0	0.00	0.02
2	14.2	3.36	0.06
3	26.1	1.99	0.36
4	28.0	3.32	0.56

Official CBPs: 19 mm and 22 mm. ZTU: [18.9 mm, 22.7 mm].

Table 15: Probabilities of misclassification errors.

basis for prediction	major error	very major error
official CBPs	1.5e-05	4.6e-06
CBP_S increased by 2mm	1.5e-05	2.6e-09
I for ZTU	1.5e-05	7.4e-08



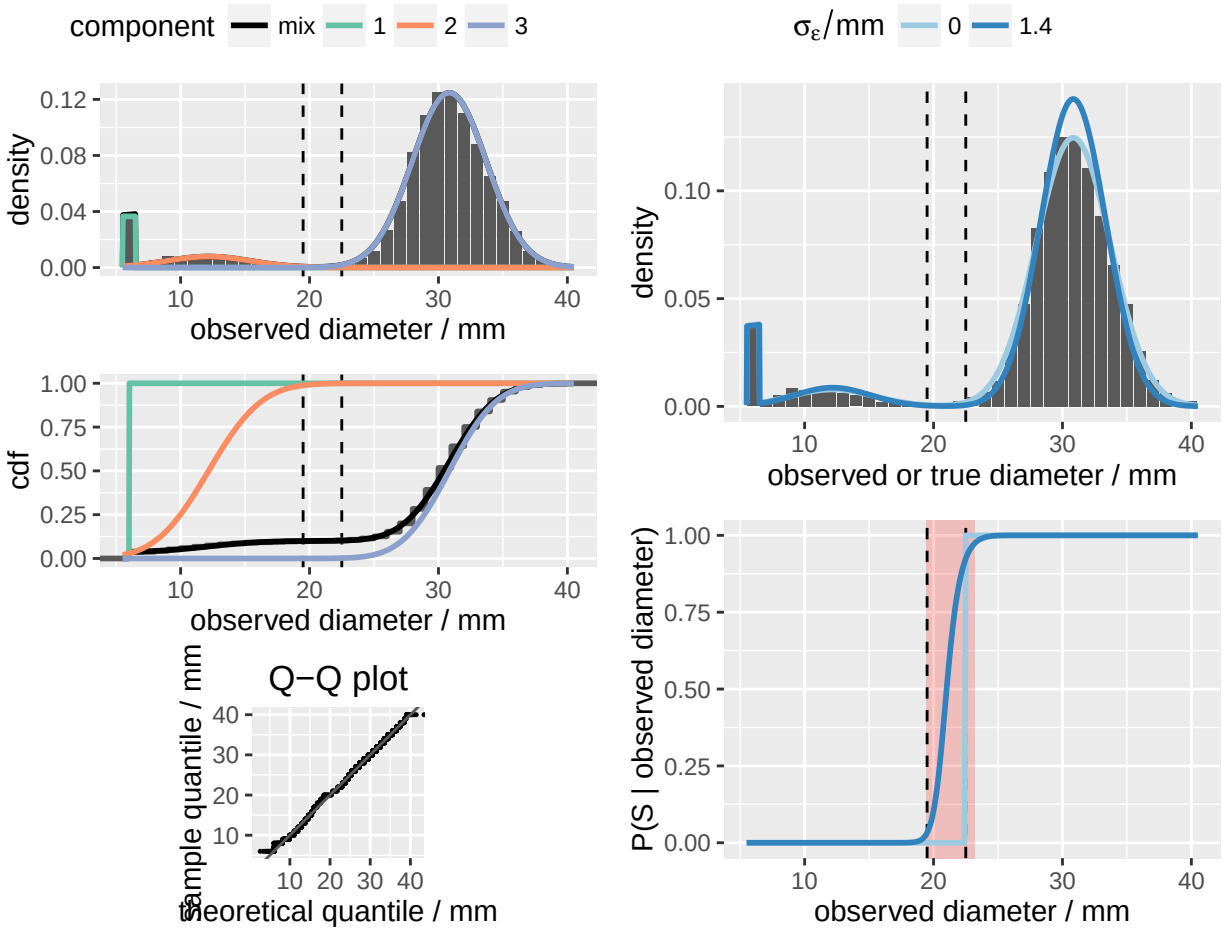
3.1.8 CRO

component	mean/mm	sd/mm	weight
1	6.0	0.00	0.04
2	12.2	3.24	0.06
3	30.8	2.88	0.90

Official CBPs: 20 mm and 23 mm. ZTU: [19.4 mm, 23.2 mm].

Table 17: Probabilities of misclassification errors.

basis for prediction	major error	very major error
official CBPs	1.0e-05	4.7e-06
CBP_S increased by 2mm	1.0e-05	3.2e-08
I for ZTU	8.1e-06	9.9e-07



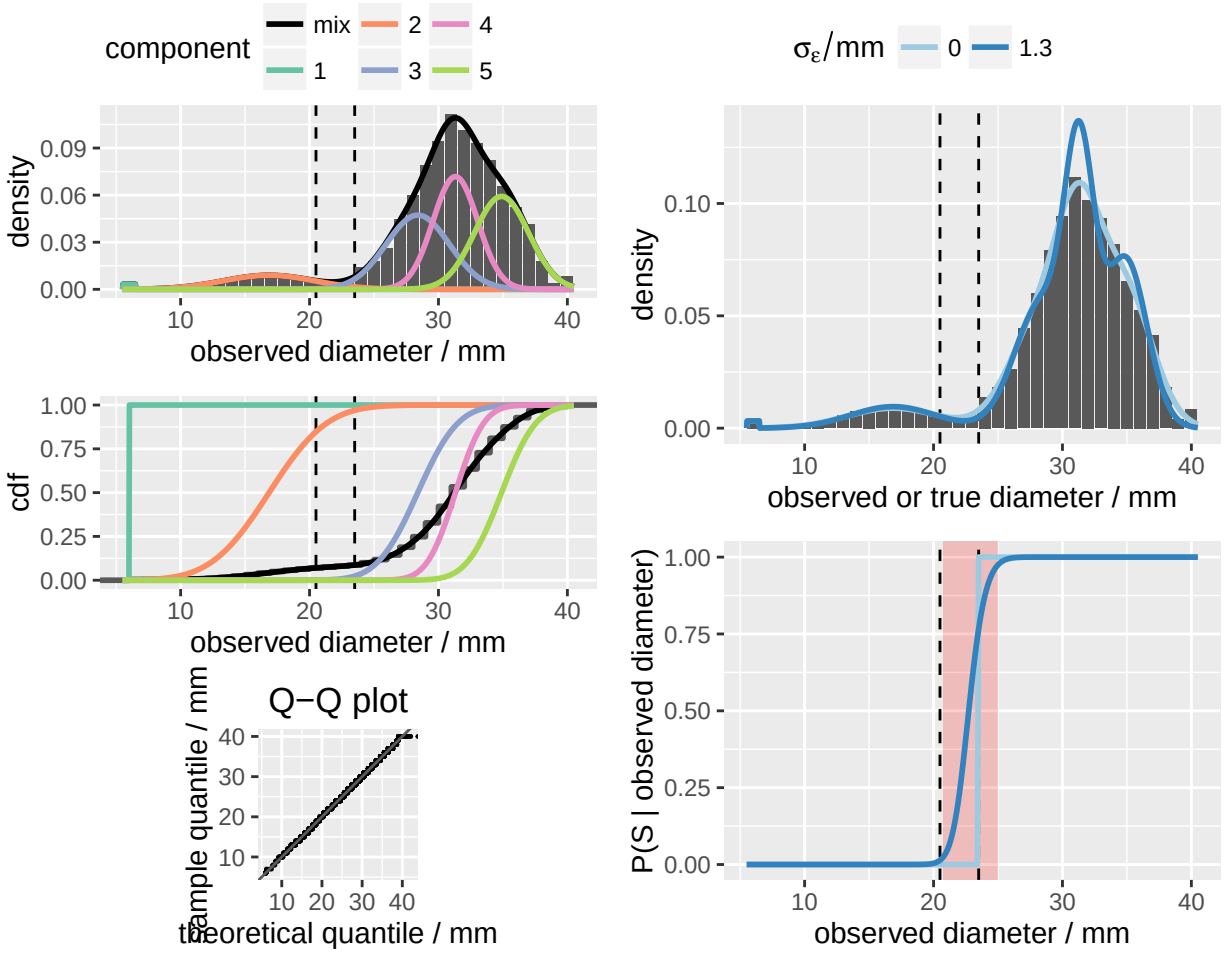
3.1.9 FEP

component	mean/mm	sd/mm	weight
1	6.0	0.00	0.00
2	16.9	3.54	0.08
3	28.4	2.47	0.29
4	31.3	1.74	0.31
5	34.9	2.11	0.31

Official CBPs: 21 mm and 24 mm. ZTU: [20.8 mm, 25 mm].

Table 19: Probabilities of misclassification errors.

basis for prediction	major error	very major error
official CBPs	3.3e-05	2.8e-05
CBP_S increased by 2mm	3.3e-05	1.0e-07
I for ZTU	3.3e-05	5.1e-07



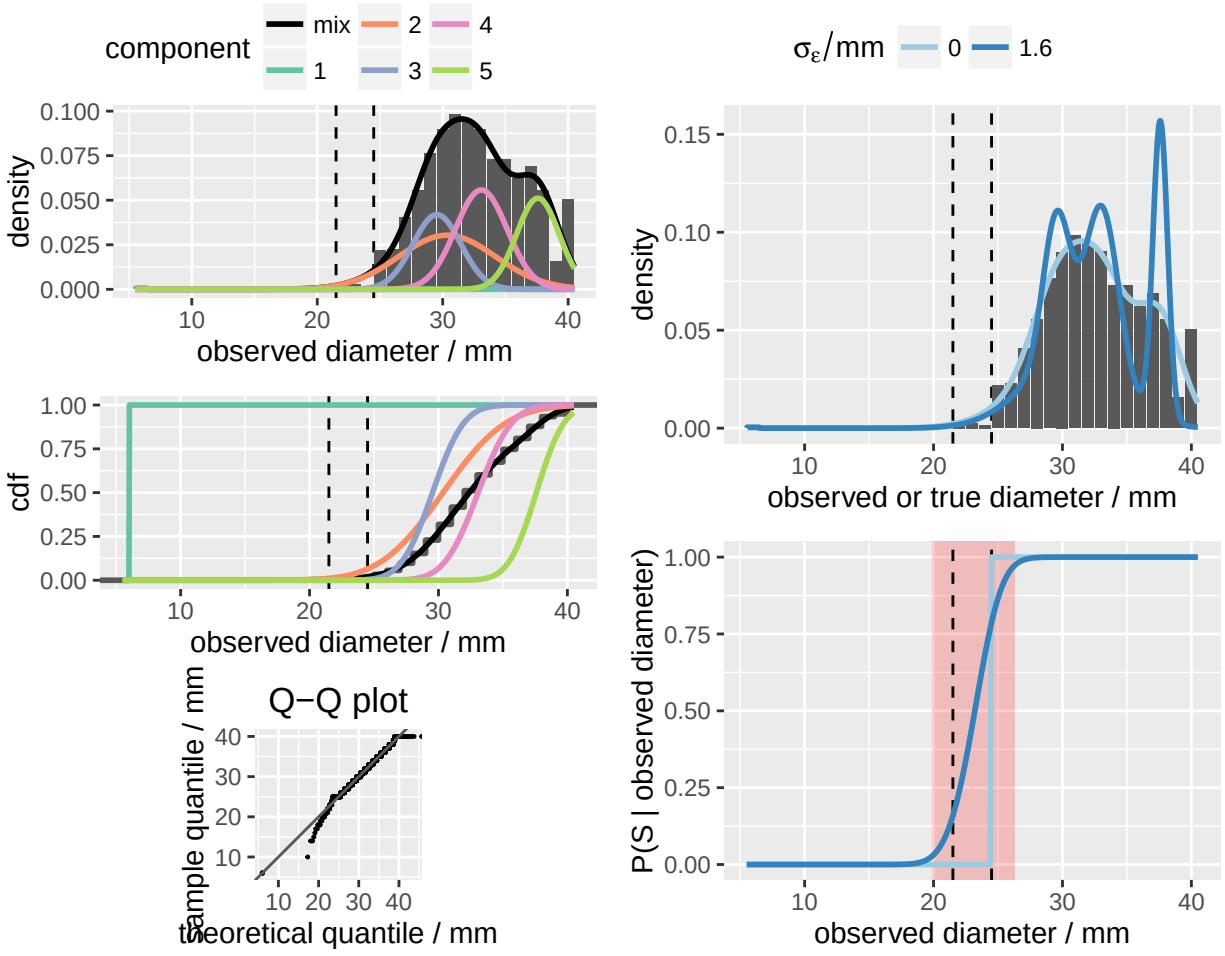
3.1.10 ETP

component	mean/mm	sd/mm	weight
1	6.0	0.00	0.00
2	30.4	3.85	0.29
3	29.5	1.92	0.20
4	33.1	2.06	0.29
5	37.6	1.70	0.22

Official CBPs: 22 mm and 25 mm. ZTU: [19.9 mm, 26.3 mm].

Table 21: Probabilities of misclassification errors.

basis for prediction	major error	very major error
official CBPs	2.1e-04	1.8e-05
CBP_S increased by 2mm	2.1e-04	4.1e-07
I for ZTU	9.7e-06	6.3e-07



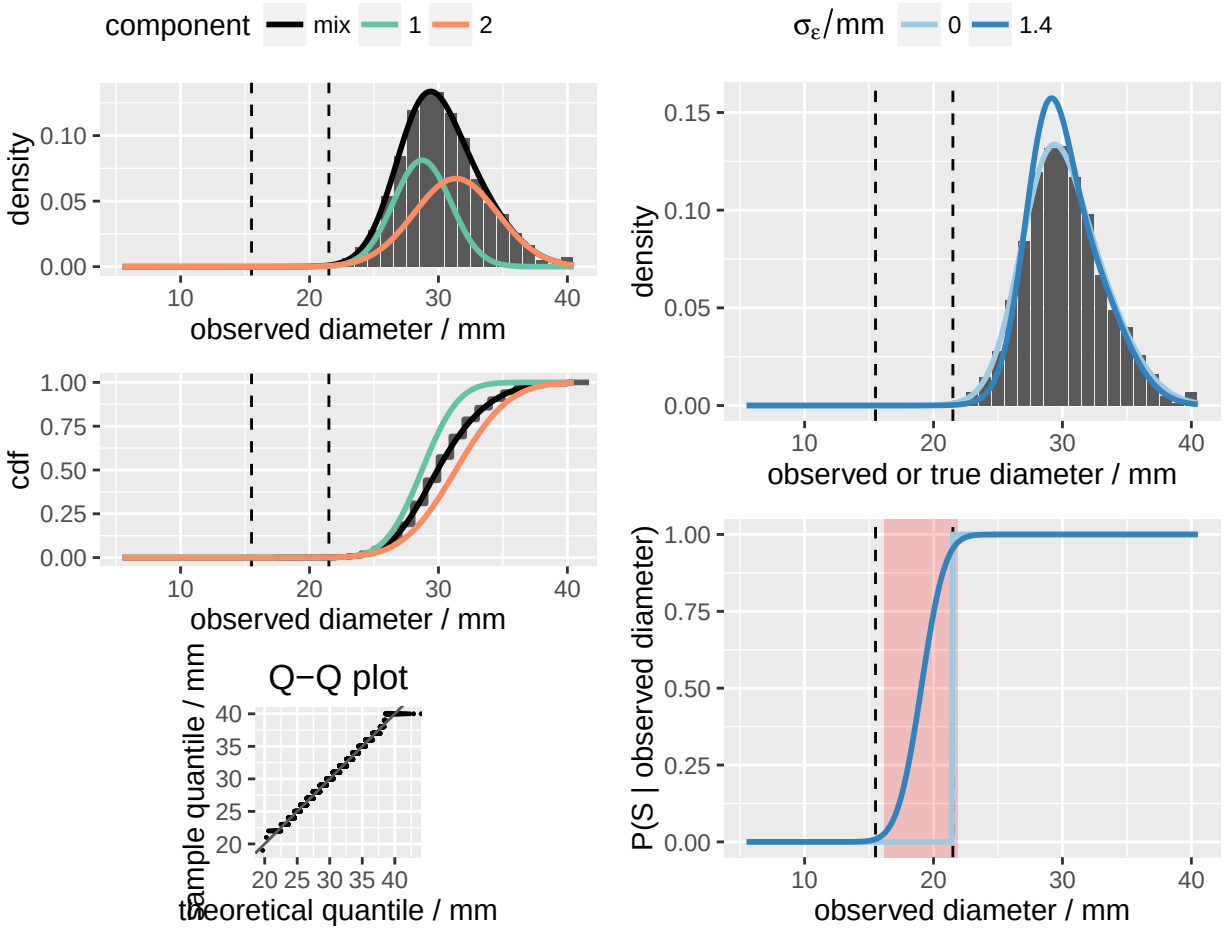
3.1.11 IPM

component	mean/mm	sd/mm	weight
1	28.7	2.23	0.45
2	31.4	3.24	0.55

Official CBPs: 16 mm and 22 mm. ZTU: [16.2 mm, 21.9 mm].

Table 23: Probabilities of misclassification errors.

basis for prediction	major error	very major error
official CBPs	1.1e-09	5.6e-14
CBP_S increased by 2mm	1.1e-09	2.8e-17
I for ZTU	1.1e-09	1.4e-14



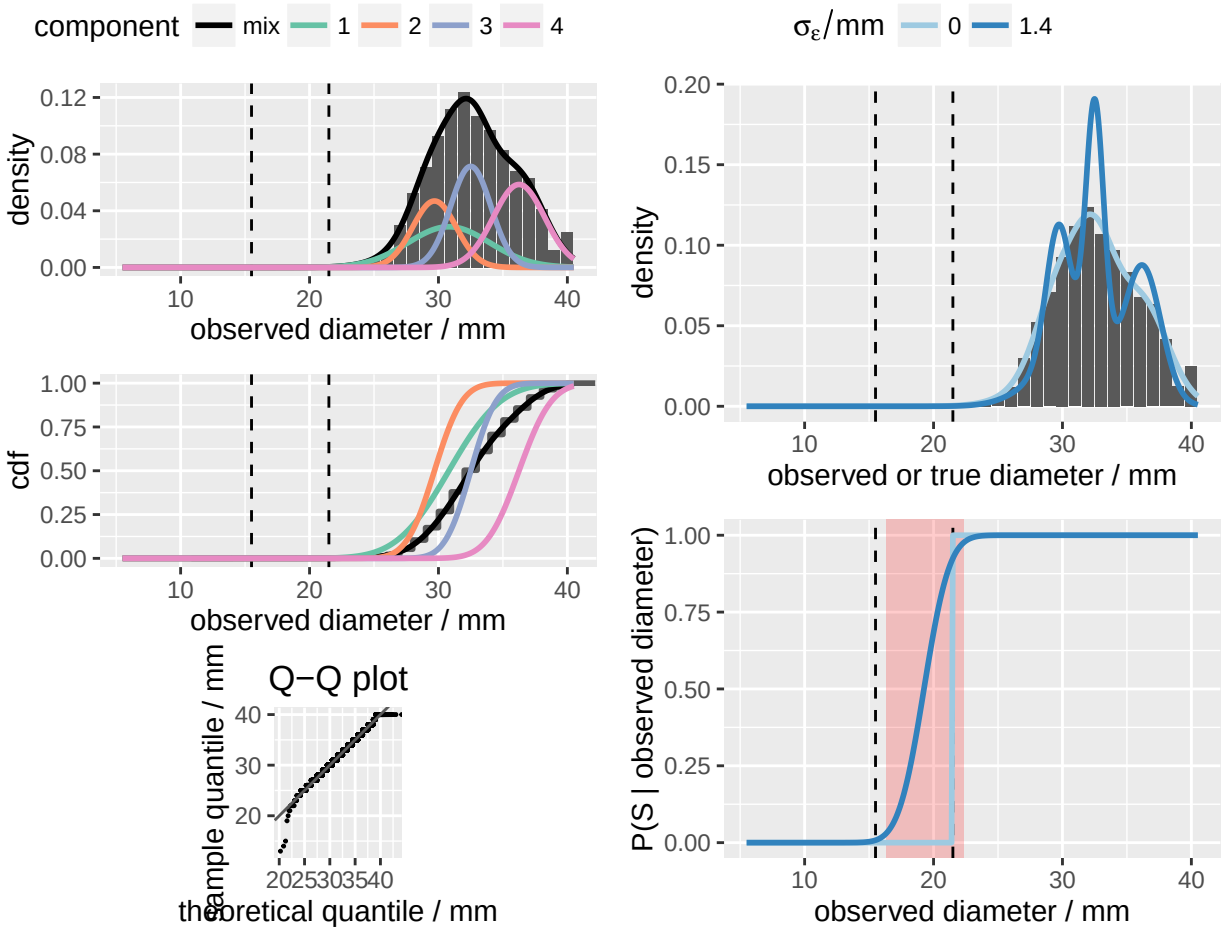
3.1.12 MEM

component	mean/mm	sd/mm	weight
1	30.8	3.18	0.23
2	29.7	1.70	0.20
3	32.5	1.56	0.28
4	36.2	1.99	0.29

Official CBPs: 16 mm and 22 mm. ZTU: [16.3 mm, 22.3 mm].

Table 25: Probabilities of misclassification errors.

basis for prediction	major error	very major error
official CBPs	6.2e-10	3.2e-14
CBP_S increased by 2mm	6.2e-10	1.7e-17
I for ZTU	6.2e-10	2.0e-15



3.2 Aminoglycosides

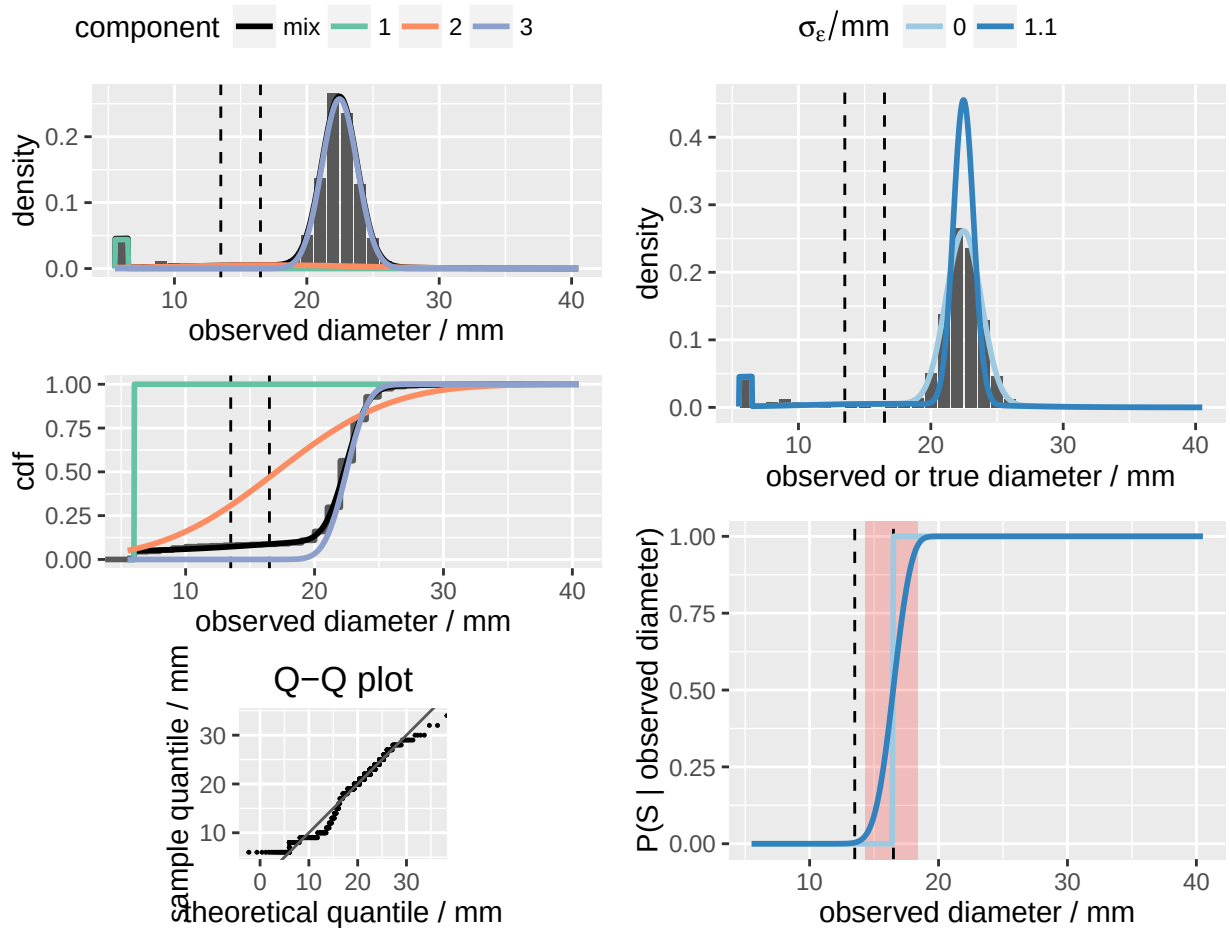
3.2.1 GEN

component	mean/mm	sd/mm	weight
1	6.0	0.00	0.04
2	17.0	7.05	0.09
3	22.5	1.34	0.86

Official CBPs: 14 mm and 17 mm. ZTU: [14.3 mm, 18.4 mm].

Table 27: Probabilities of misclassification errors.

basis for prediction	major error	very major error
official CBPs	5.7e-06	4.9e-06
CBP_S increased by 2mm	5.7e-06	2.8e-09
I for ZTU	5.7e-06	4.4e-09



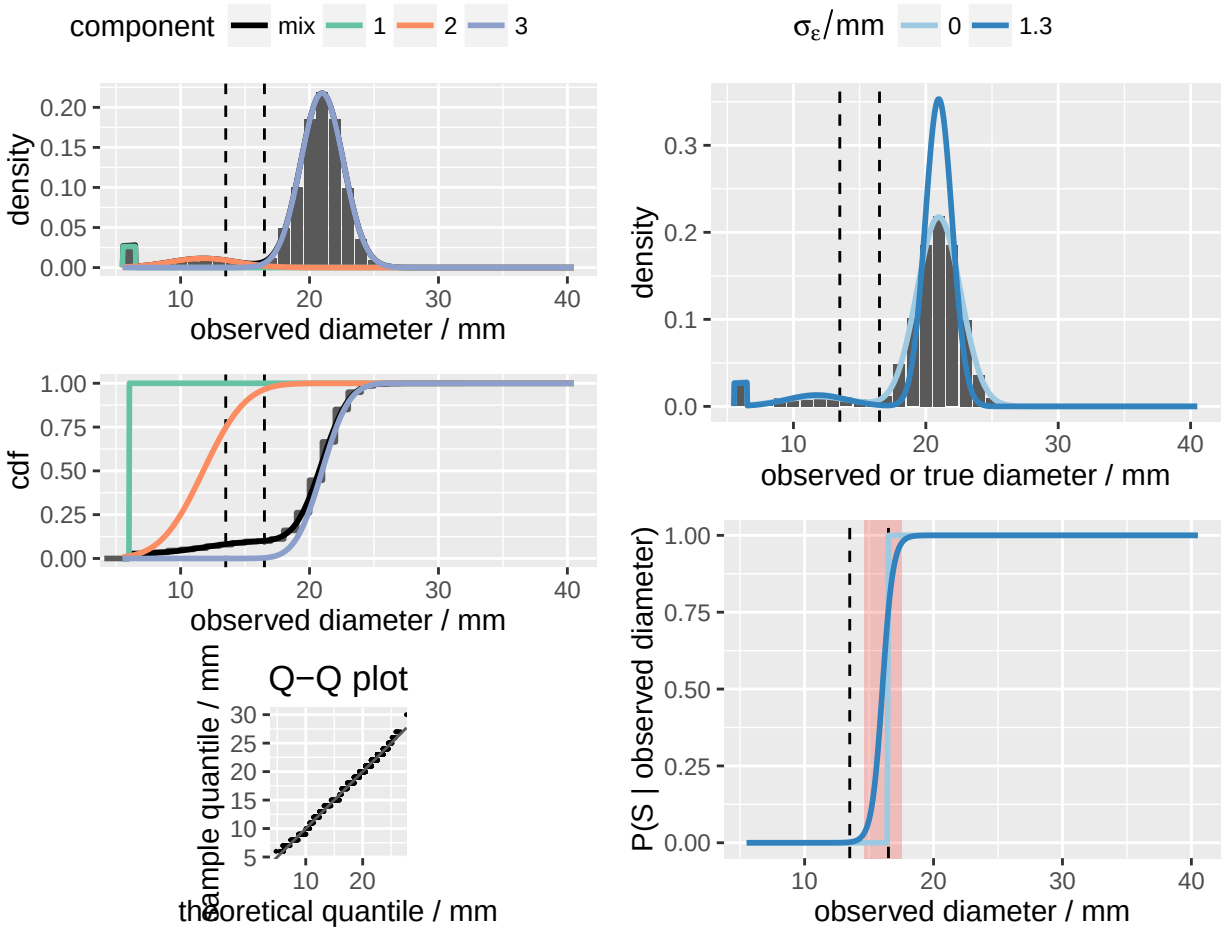
3.2.2 TOB

component	mean/mm	sd/mm	weight
1	6.0	0.00	0.03
2	11.8	2.61	0.07
3	21.0	1.65	0.90

Official CBPs: 14 mm and 17 mm. ZTU: [14.6 mm, 17.5 mm].

Table 29: Probabilities of misclassification errors.

basis for prediction	major error	very major error
official CBPs	7.4e-06	5.1e-05
CBP_S increased by 2mm	7.4e-06	1.9e-07
I for ZTU	7.4e-06	4.0e-06



3.3 Quinolones

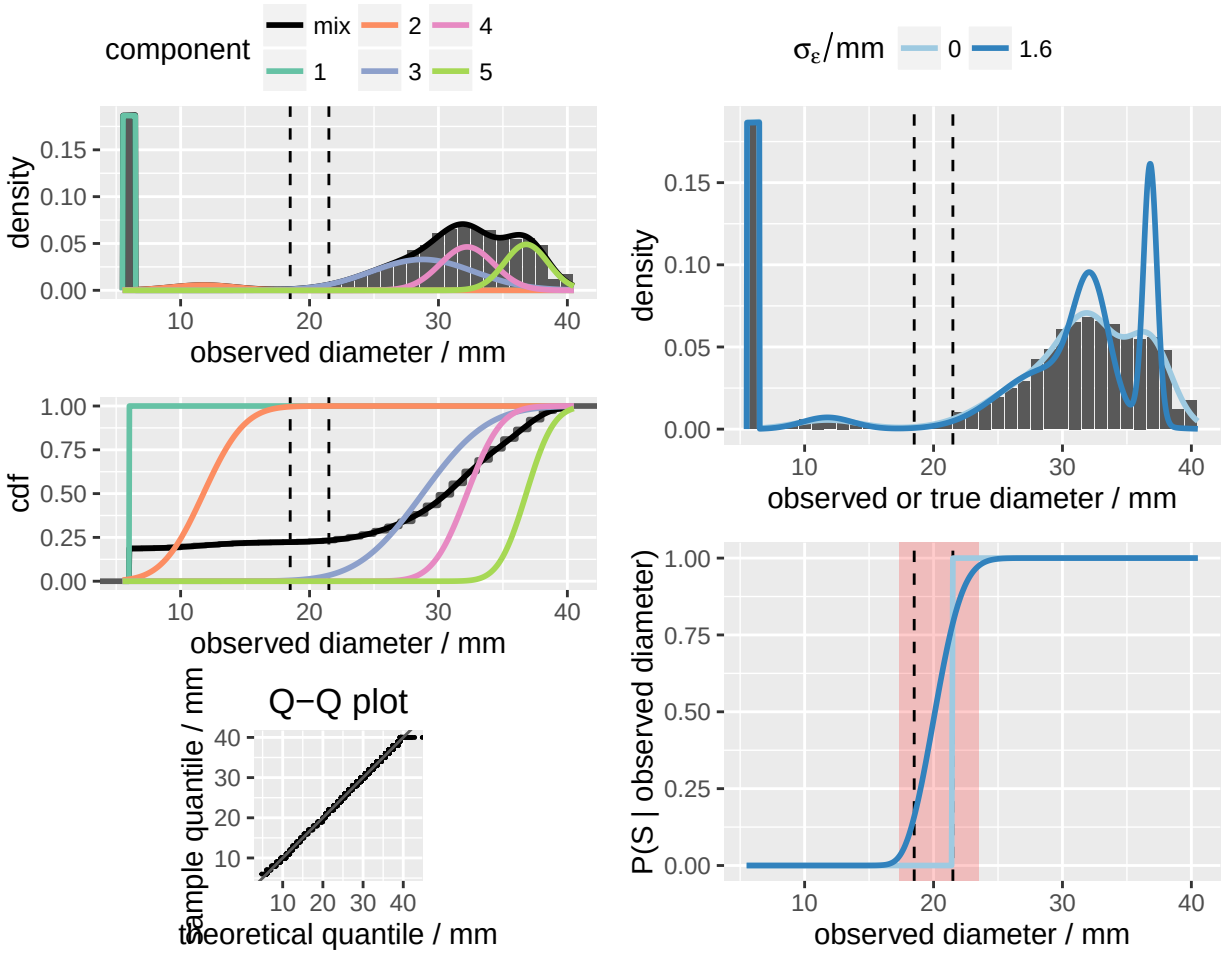
3.3.1 NOR

component	mean/mm	sd/mm	weight
1	6.0	0.00	0.19
2	11.8	2.53	0.04
3	28.8	3.98	0.33
4	32.2	2.09	0.24
5	36.8	1.68	0.21

Official CBPs: 19 mm and 22 mm. ZTU: [17.3 mm, 23.5 mm].

Table 31: Probabilities of misclassification errors.

basis for prediction	major error	very major error
official CBPs	1.3e-04	1.0e-05
CBP_S increased by 2mm	1.3e-04	2.2e-07
I for ZTU	1.4e-05	2.2e-07



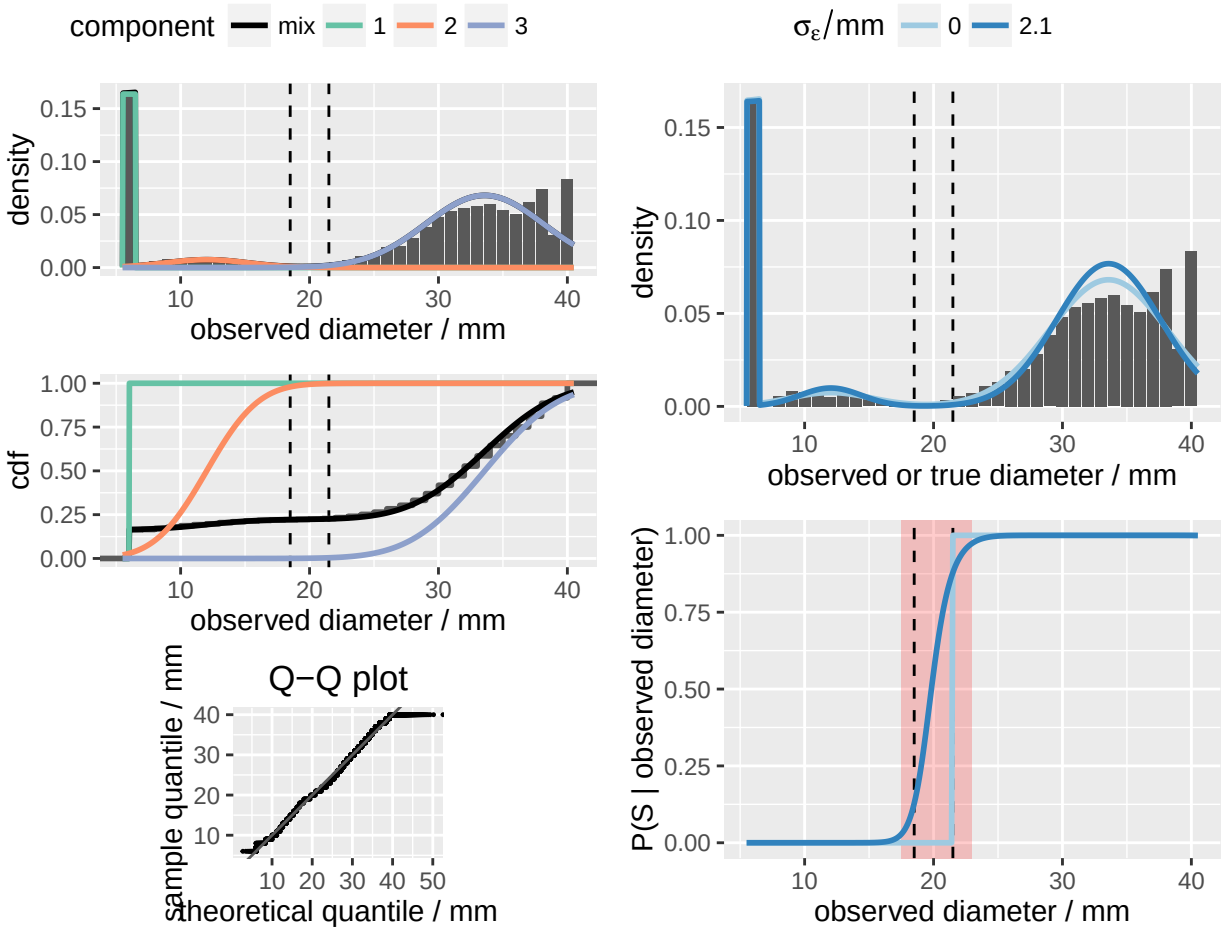
3.3.2 CIP

component	mean/mm	sd/mm	weight
1	6.0	0.00	0.16
2	12.0	3.18	0.06
3	33.6	4.55	0.78

Official CBPs: 19 mm and 22 mm. ZTU: [17.5 mm, 23 mm].

Table 33: Probabilities of misclassification errors.

basis for prediction	major error	very major error
official CBPs	1.3e-04	5.7e-05
CBP_S increased by 2mm	1.3e-04	3.8e-06
I for ZTU	3.8e-05	8.0e-06



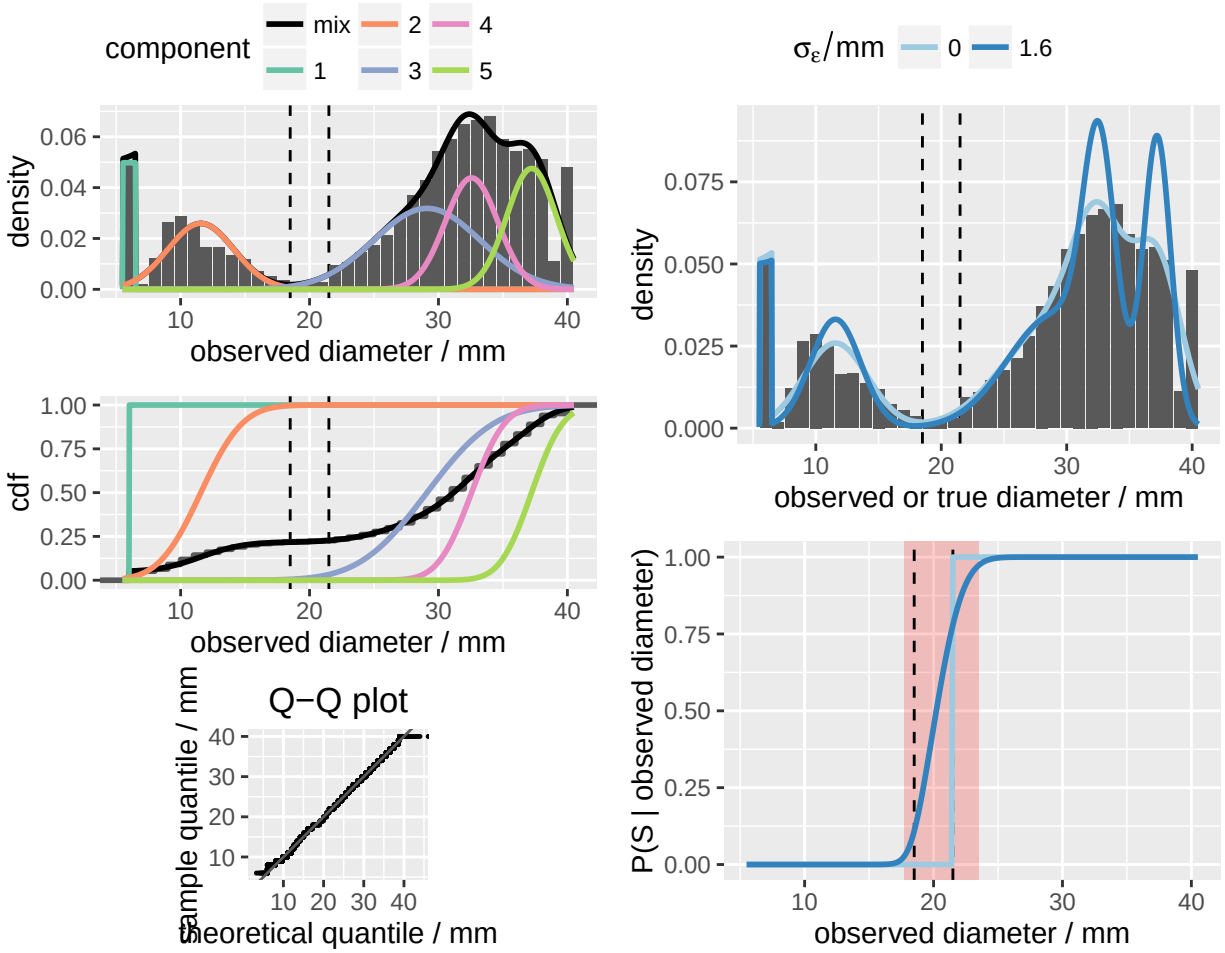
3.3.3 LEV

component	mean/mm	sd/mm	weight
1	6.0	0.00	0.05
2	11.6	2.56	0.17
3	29.1	4.13	0.33
4	32.6	2.04	0.22
5	37.2	1.93	0.23

Official CBPs: 19 mm and 22 mm. ZTU: [17.7 mm, 23.5 mm].

Table 35: Probabilities of misclassification errors.

basis for prediction	major error	very major error
official CBPs	1.2e-04	1.5e-05
CBP_S increased by 2mm	1.2e-04	2.9e-07
I for ZTU	2.9e-05	2.9e-07



3.4 Tetracyclines

3.4.1 TGC

component	mean/mm	sd/mm	weight
1	21.5	1.69	1

Official CBPs: 15 mm and 18 mm. ZTU: [7.8 mm, 15.7 mm].

Table 37: Probabilities of misclassification errors.

basis for prediction	major error	very major error
official CBPs	1.5e-05	4.1e-13
CBP_S increased by 2mm	1.5e-05	2.1e-15
I for ZTU	5.8e-18	4.1e-13

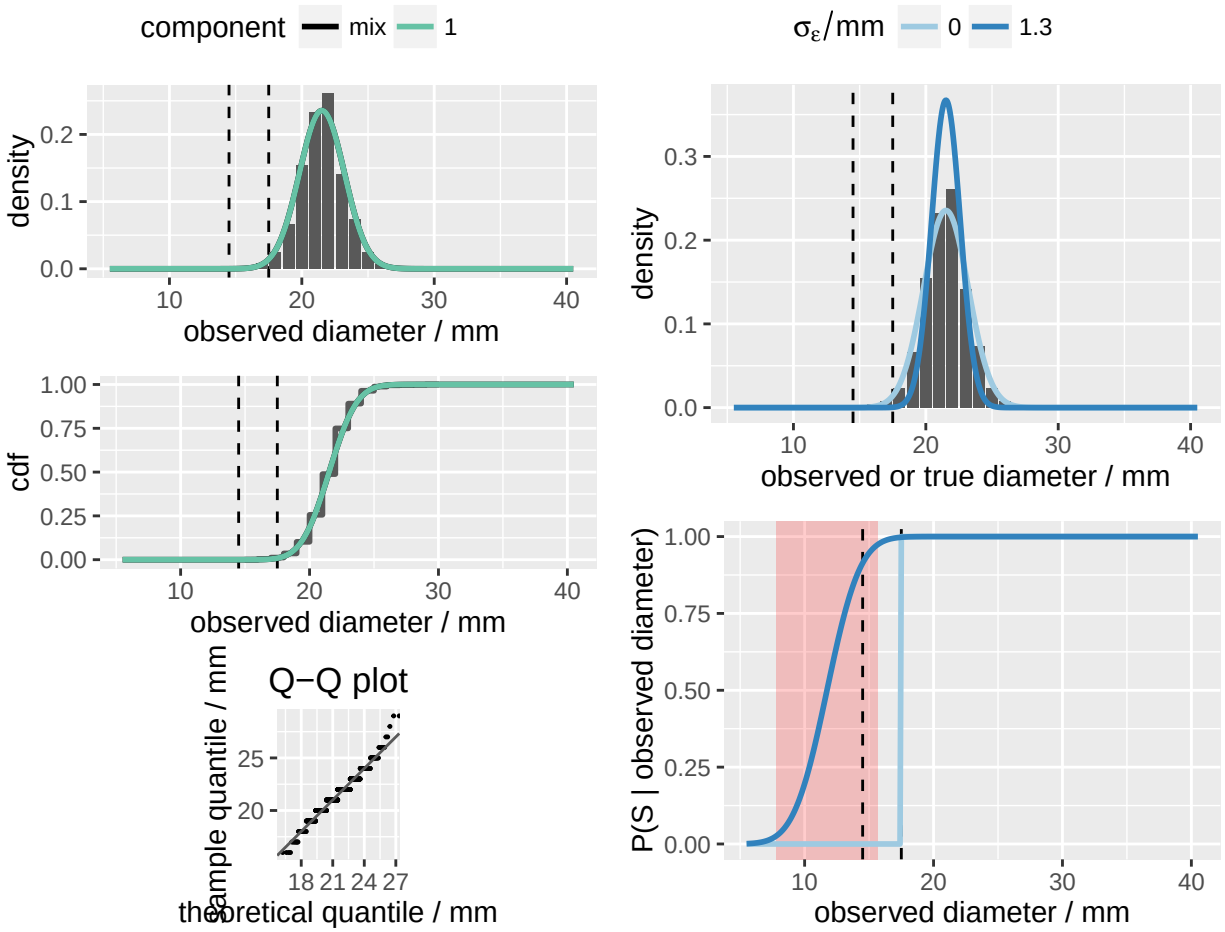
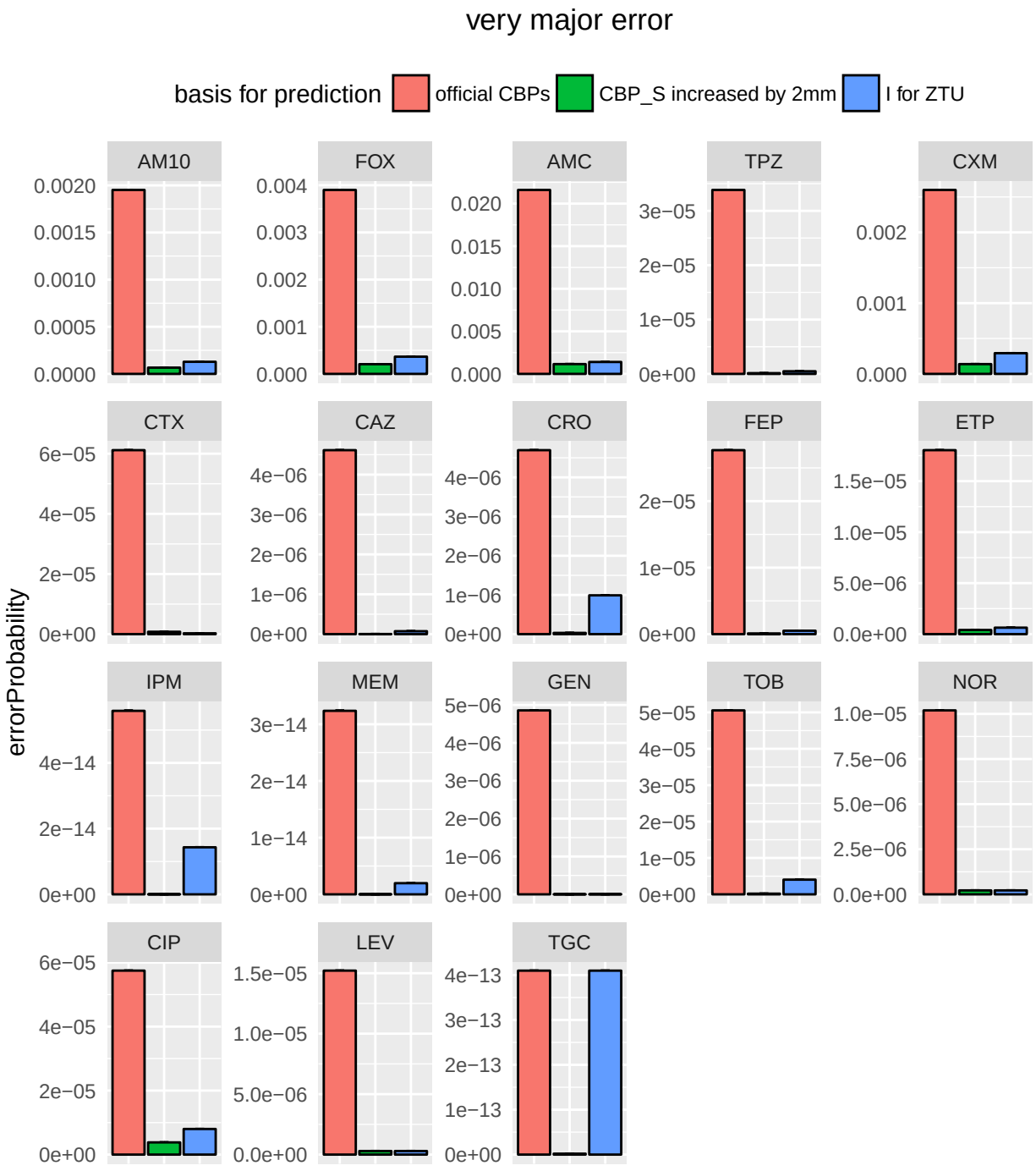
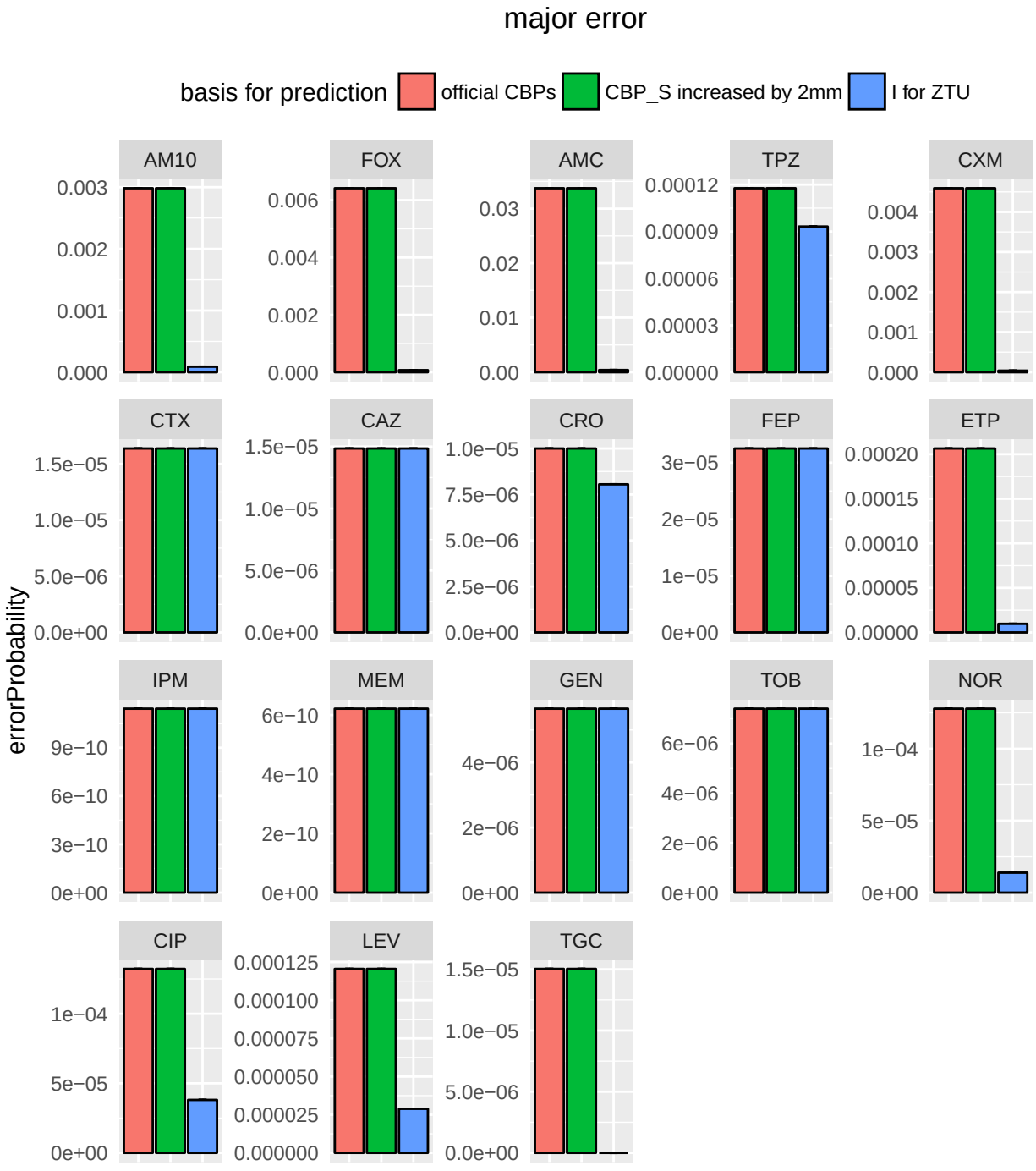


Table 38: Official CBPs (EUCAST) and zones of technical uncertainty (ZTUs) derived from our model. The weight of the ZTU is defined as the fraction of data points that lie in the ZTU.

antibiotic	CBP_S/mm	CBP_R/mm	ZTU: width/mm	ZTU: weight/%
AM10	14	14	3.7	2.82
FOX	19	19	4.3	5.72
AMC	19	19	4.5	27.29
TPZ	20	17	4.7	13.91
CXM	18	18	4.3	3.69
CTX	20	17	4.6	1.30
CAZ	22	19	3.8	4.52
CRO	23	20	3.8	0.88
FEP	24	21	4.2	4.83
ETP	25	22	6.4	5.41
IPM	22	16	5.7	0.02
MEM	22	16	6.0	0.12
GEN	17	14	4.1	1.22
TOB	17	14	2.9	2.22
NOR	22	19	6.2	3.17
CIP	22	19	5.5	2.18
LEV	22	19	5.8	3.17
TGC	18	15	7.9	0.00

AMC and TPZ have ZTUs with exceptionally high weights.


very major error
basis for prediction
official CBPs
CBP_S increased by 2mm
I for ZTU
errorProbability
AM10
FOX
AMC
TPZ
CXM
CTX
CAZ
CRO
FEP
ETP
IPM
MEM
GEN
TOB
NOR
CIP
LEV
TGC


major error
basis for prediction
official CBPs
CBP_S increased by 2mm
I for ZTU
errorProbability
AM10
FOX
AMC
TPZ
CXM
CTX
CAZ
CRO
FEP
ETP
IPM
MEM
GEN
TOB
NOR
CIP
LEV
TGC

4 Appendix

4.1 Clinical breakpoints

Table 39: Clinical breakpoints (CBPs).

antibiotic	CBP_S/mm	CBP_R/mm
AM10	14	14
KF	NA	NA
FOX	19	19
CPD	NA	NA
AMC	19	19
TPZ	20	17
CXM	18	18
CTX	20	17
CAZ	22	19
CRO	23	20
FEP	24	21
ETP	25	22
IPM	22	16
MEM	22	16
KAN	NA	NA
GEN	17	14
TOB	17	14
NAL	NA	NA
NOR	22	19
CIP	22	19
LEV	22	19
TE	NA	NA
MI	NA	NA
TGC	18	15

4.2 Distributions for beta-lactams

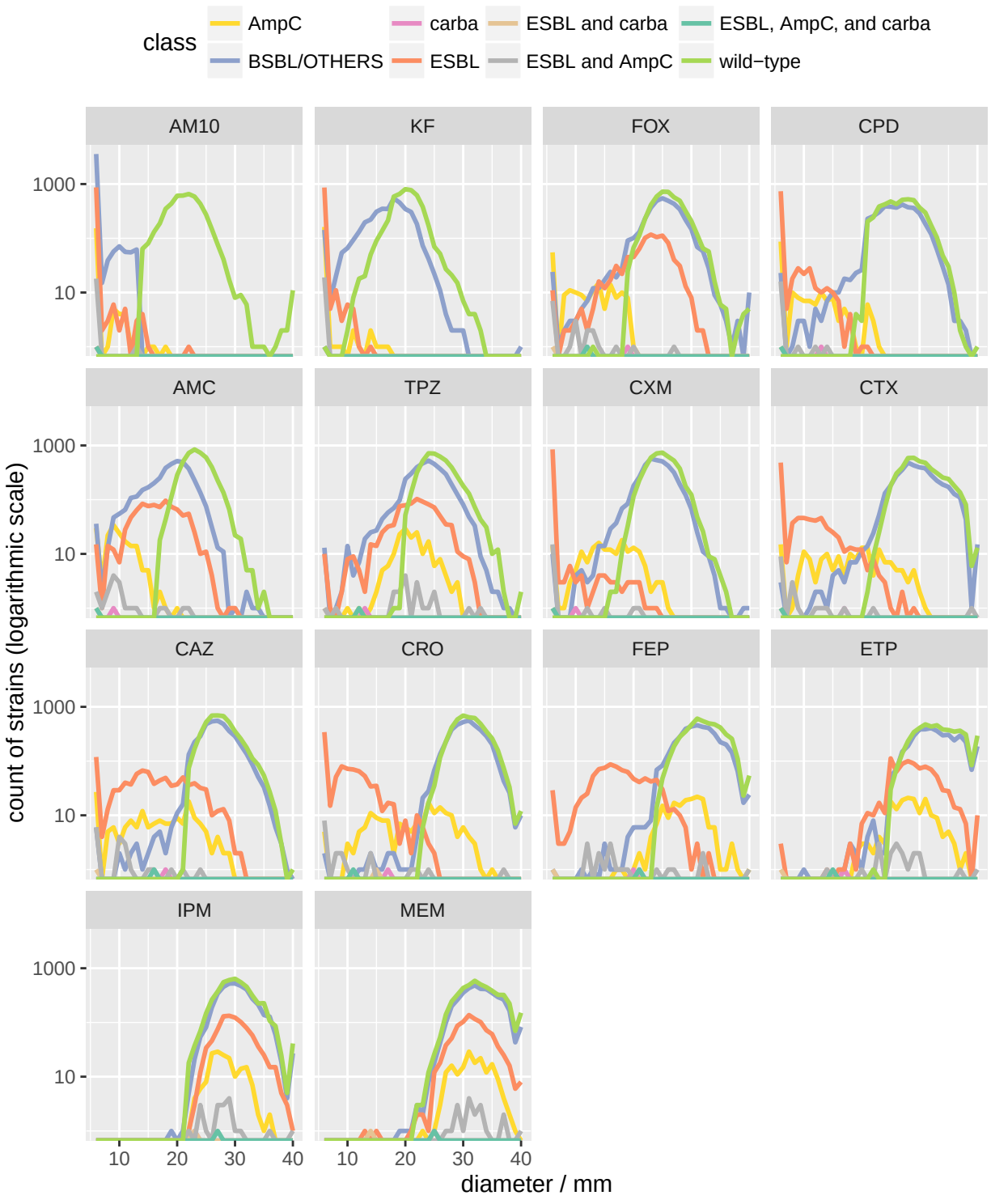


Figure 1: Distributions of diameters for beta-lactams.