Forecast probabilities - work in progress

Nicolas Blöchliger, Institute of Medical Microbiology, University of Zurich

24/10/2016

Contents

1 Goal

The main goal of the present work is to estimate the probability that a strain is wild-type (wt) given an observed diameter y.

2 Model

We assume that the distribution of the observed diameter Y of a given antibiotic-species combination is a mixture of three components with weights $w_i = p(C = i)$, where C encodes the component. The observed diameter is 6 mm for the first component and normally distributed for the other two components:

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

where

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

Thus,

$$f_Y(y) = w_1\delta_6 + \sum_{i=2}^{3} w_i\phi(y; \mu_i, \sigma_Y^2).$$

The idea is that the last component corresponds to the wild-type strains while the other two components correspond to strains that are resistant to the given antibiotic.

Note:

- We do not account for the fact that the observed data are rounded to integer values.
- No model for the technical error is needed for the present work.

2.1 Estimation of model parameters

- We estimate w_1 as the fraction of data points in the sample that are equal to 6 mm.
- The parameters of the second and third component of Y, i.e. w_i, μ_i, and σ_Y^2, are estimated by fitting a normal mixture model with two components to the data in the sample with diameters greater than 6 mm. We use the R package `mclust`.

2.2 Limitations of the model

- The model does not account for the fact that $Y \geq 6$ mm. As long as the means of the two 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 two components are sufficiently small (say $40 \text{ mm} - \mu_i > 2\sigma$), this should not cause problems.
- The distributions of the second and the third components are assumed to be normal. This seems fine for the component corresponding to wild-type strains. However, the distribution of the non-wild-type strains with $X > 6$ mm might not be adequately modelled.
- The variances of the two components are assumed to be equal. This assumption is problematic but has the advantage of guaranteeing that there is only one decision boundary if strains are assigned to the most likely component.

3 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

3.1 Ground truth for beta-lactams

The strains for which measurements of zone diameters of beta-lactams are available were assigned to 8 classes based on the zone diameters additional confirmation tests as follows.

1. Confirmation tests were conducted if the FOX diameter was $\leq$ 18 mm (suspicion of AmpC production), if a strain was categorized as I or R for 3rd/4th generation cephalosporins according to EUCAST CBPs or synergy with clavulanic acid or tazobactam was observed (ESBL screening markers), if the MEM diameter was $\leq$ 25 mm (carbapenemase screening marker), and in some other cases [refer to manual bacteriology IMM UZH and Giorgia Valsesia for additional explanations]. In particular, note that the AmpC confirmation test was run in parallel with the ESBL and carbapenemase confirmation assays even if the FOX diameter was $>$ 18 mm.
2. Strains were assigned to the classes AmpC, ESBL, and carba (or combinations thereof) if the corresponding confirmation test was conducted and positive.
3. The remaining strains were assigned to BSBL/OTHERS if the AM10 diameter was $\leq$ 14 mm and to wild-type otherwise.

We used these class assignments to define the antibiotic-specific wild type, which is used as the ground truth to evaluate our model:

antibiotic	definition of antibiotic-specific wild type
amoxicillin - clavulanic acid	wt
piperacillin - tazobactam	wt
cefuroxime	wt and BSBL/OTHERS
cefotaxime	wt and BSBL/OTHERS

antibiotic	definition of antibiotic-specific wild type
ceftazidime	wt and BSBL/OTHERS
ceftriaxone	wt and BSBL/OTHERS
cefepime	wt, BSBL/OTHERS, and AmpC
ertapenem	wt, BSBL/OTHERS, AmpC, and ESBL
imipenem	wt, BSBL/OTHERS, AmpC, and ESBL
meropenem	wt, BSBL/OTHERS, AmpC, and ESBL

The antibiotics ampicillin, cephalothin, cefoxitin, and cefpodoxime were used to define the classes and are thus excluded.

General characteristics of the data associated with the beta-lactams are given in the appendix.

4 Results

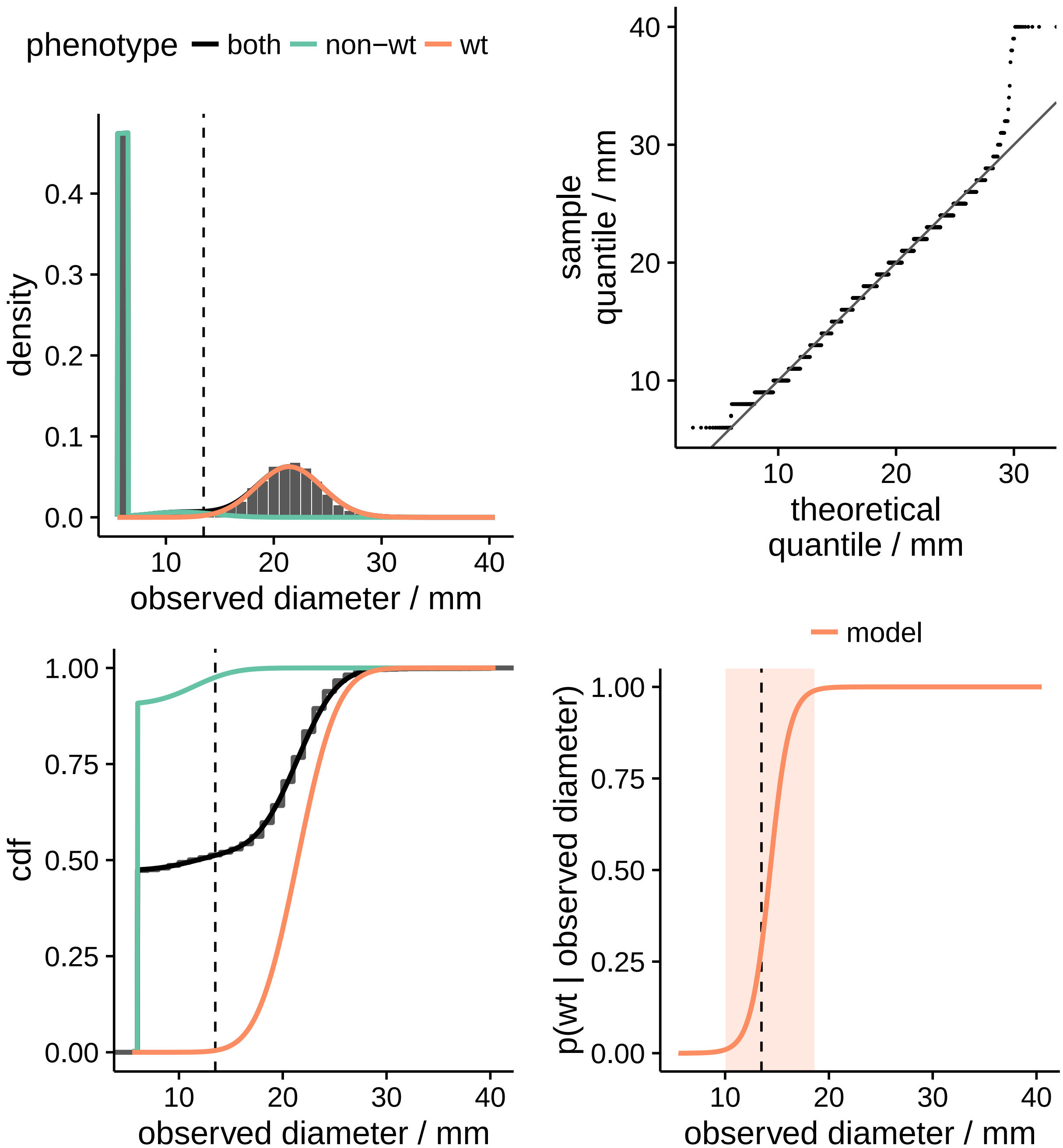
In the figures on the next pages, dashed vertical lines indicate CBPs (if available) according to http://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Breakpoint_tables/v_6.0_Breakpoint_table.xls (13.05.2016, see Appendix). The figures are organized as follows. Note that the the first and the second component are combined for visualisation.

- 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].
- Bottom-left: Empirical cumulative distribution function (cdf) of Y (grey), its estimate (black) and estimated cdfs for the components of Y (coloured).
- Top-right: Q-Q plot. If the estimated density of Y explained the data perfectly, all points would lie on the identity line (grey).
- Bottom-right: $p(\text{wt}|Y = y)$, i.e. the probability that a strain is wt given an observed diameter y for the model (blue) and according to the ground truth (grey, if available). For this calculation, the last component is assumed to describe the wt population. A diameter value is defined to be in a zone of uncertainty if the associated model-based probability for wt is between 0.01 and 0.99, and this zone of uncertainty is indicated with a blue rectangle. The widths of these zones listed towards the end of this document.

4.1 Beta-lactams

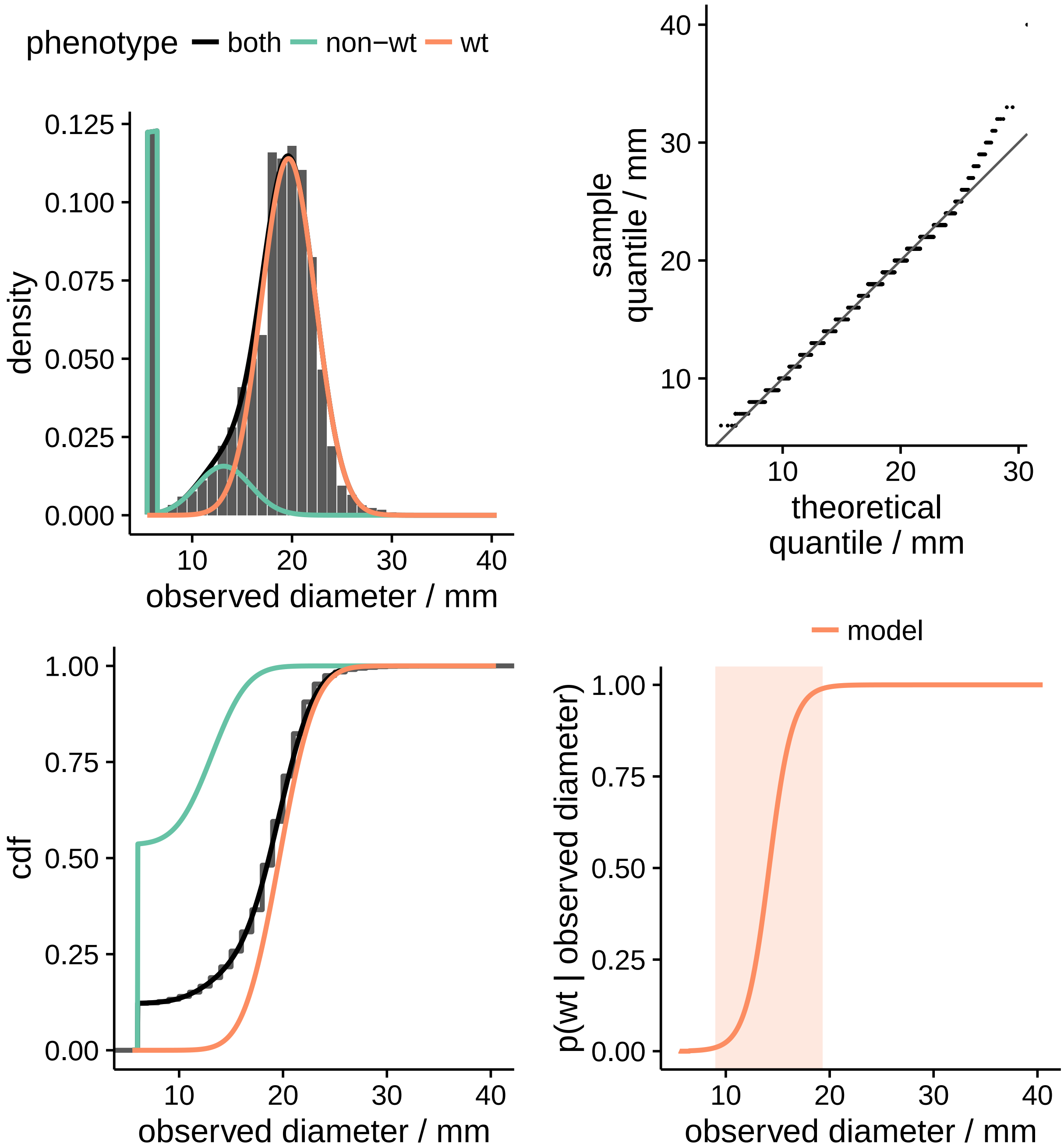
4.1.1 AM10

component	mean/mm	sd/mm	weight
1	6	0	0.47
2	11	3	0.05
3	21	3	0.48



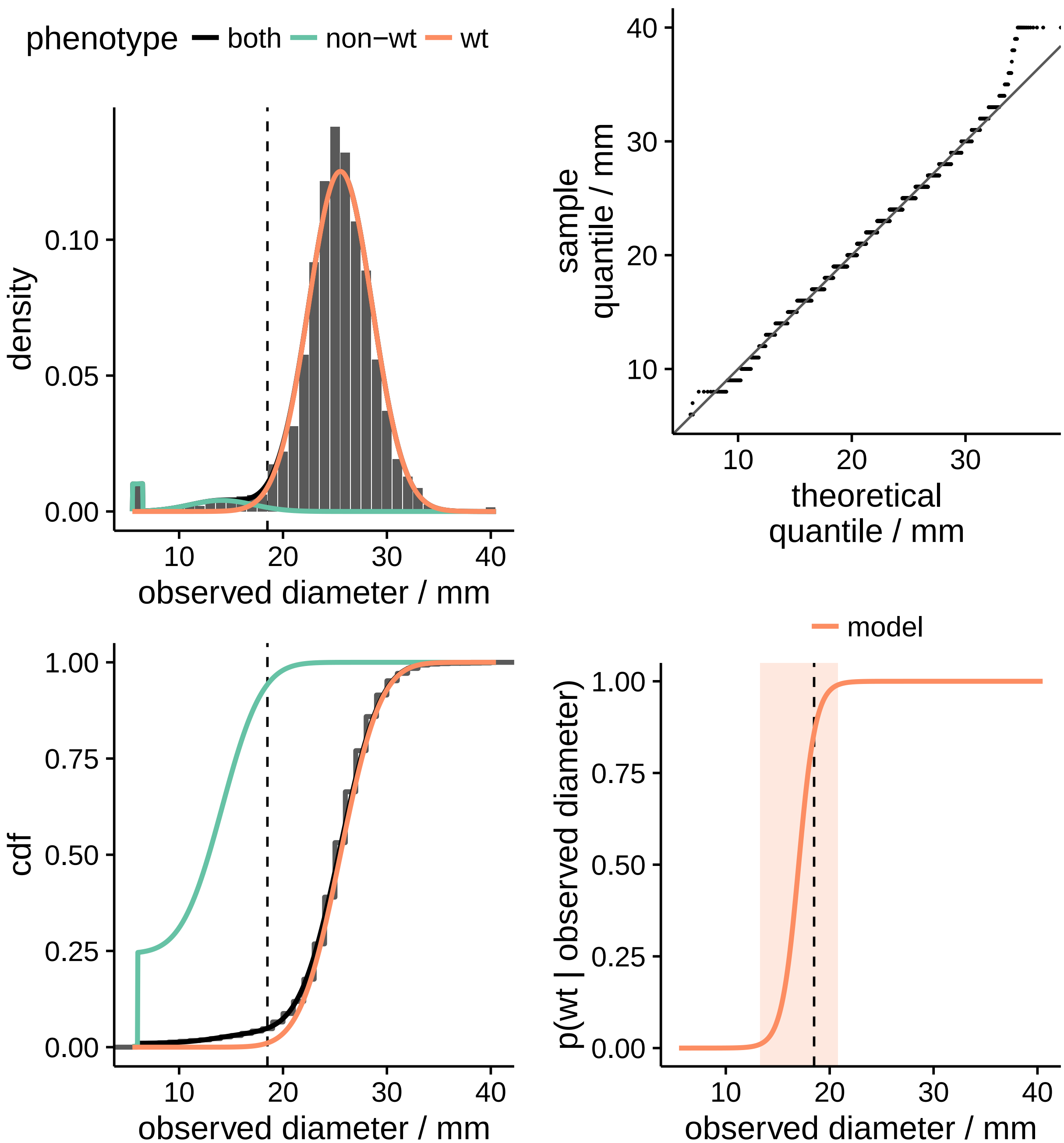
4.1.2 KF

component	mean/mm	sd/mm	weight
1	6	0.0	0.12
2	13	2.7	0.11
3	20	2.7	0.77



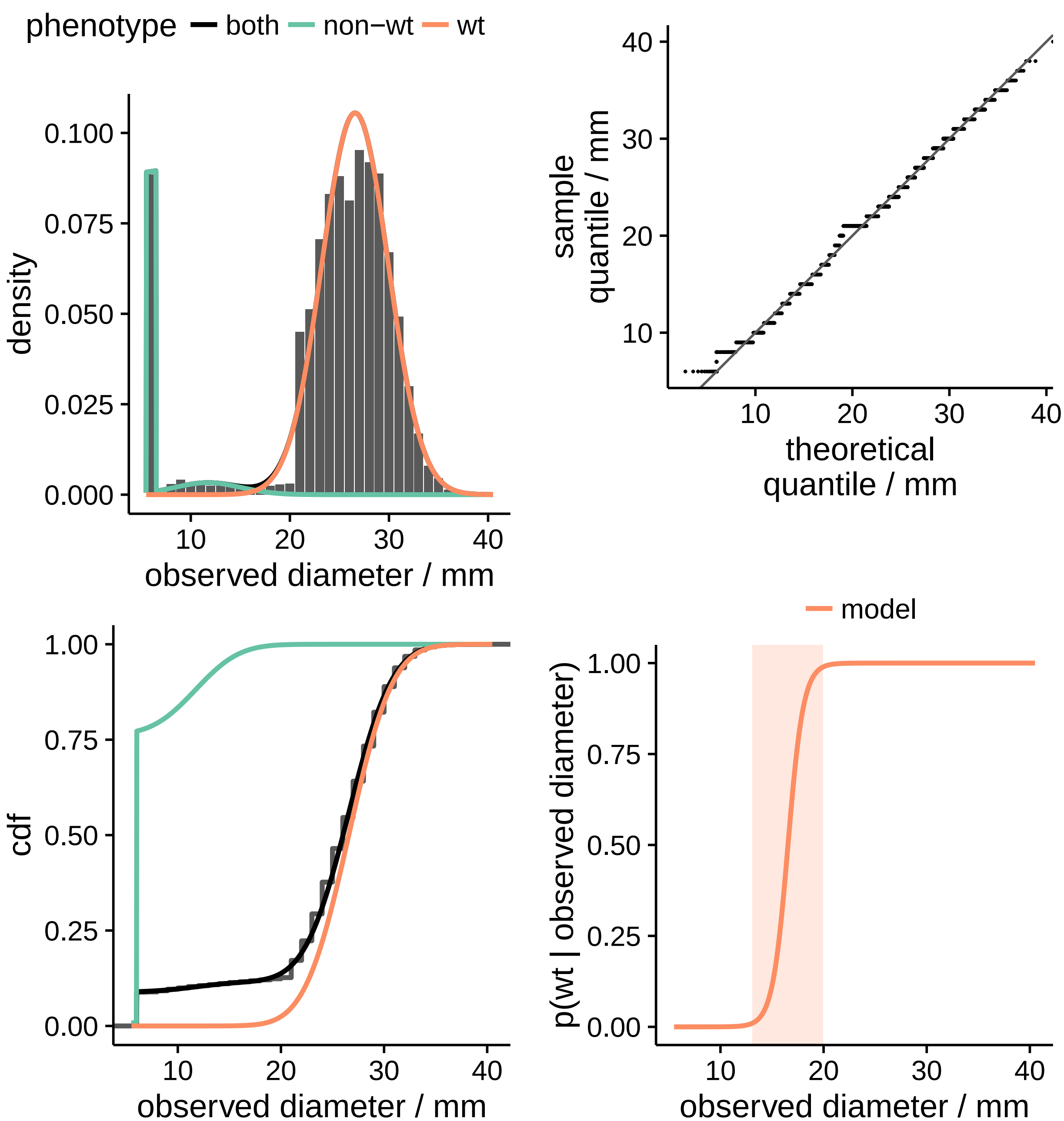
4.1.3 FOX

component	mean/mm	sd/mm	weight
1	6	0.0	0.01
2	14	3.1	0.03
3	26	3.1	0.96



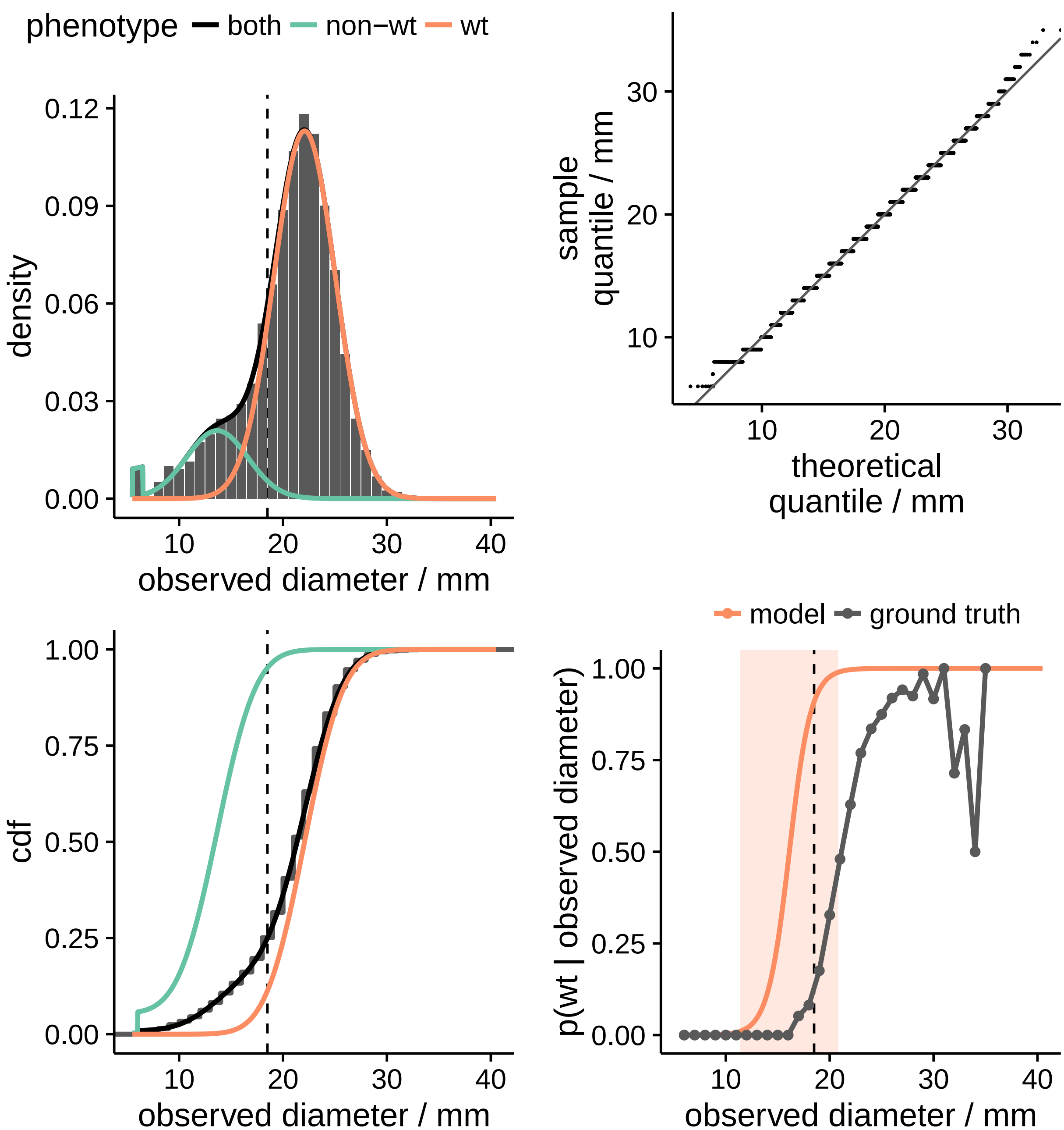
4.1.4 CPD

component	mean/mm	sd/mm	weight
1	6	0.0	0.09
2	12	3.3	0.03
3	27	3.3	0.88



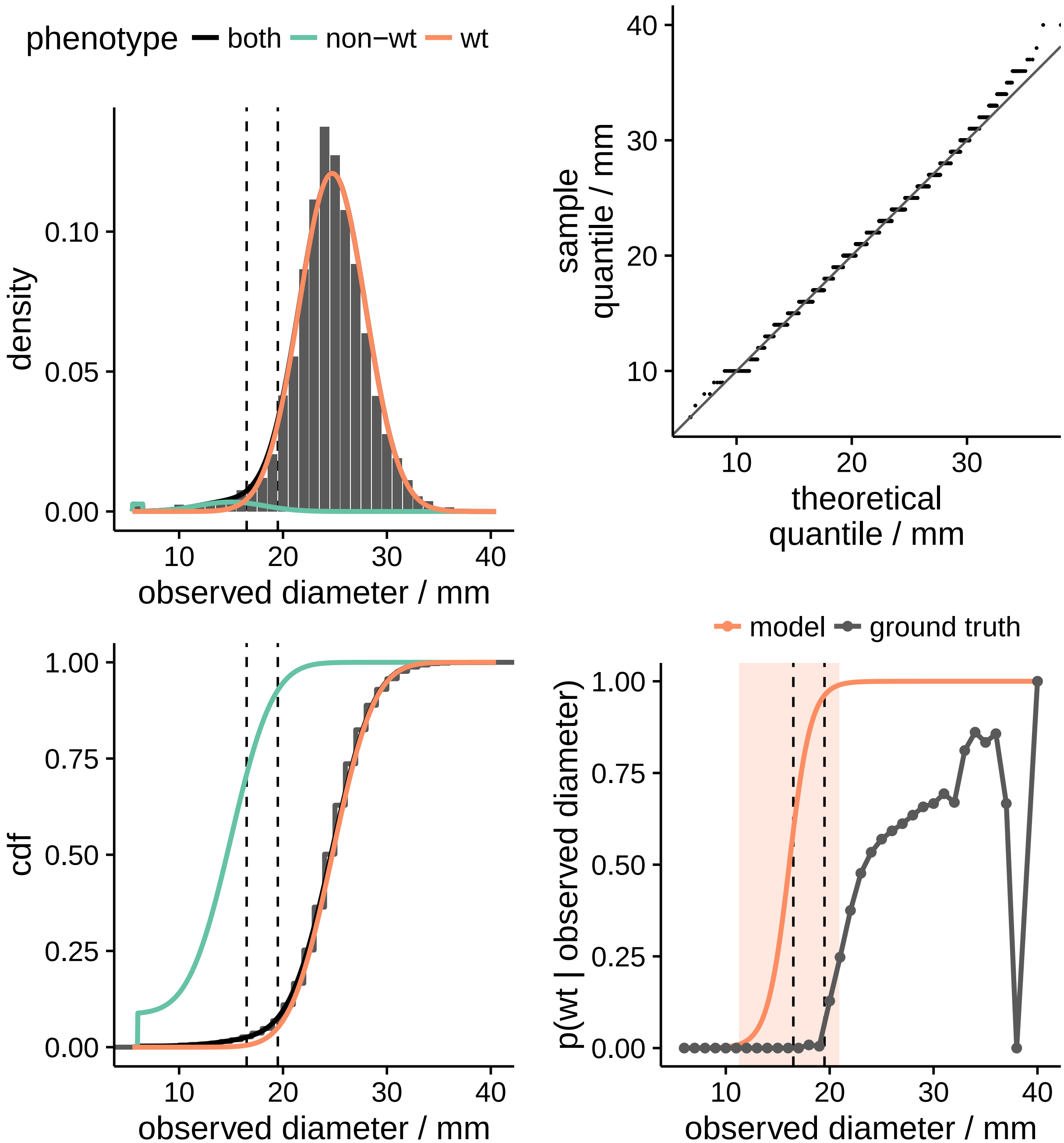
4.1.5 AMC

component	mean/mm	sd/mm	weight
1	6	0	0.01
2	14	3	0.15
3	22	3	0.84



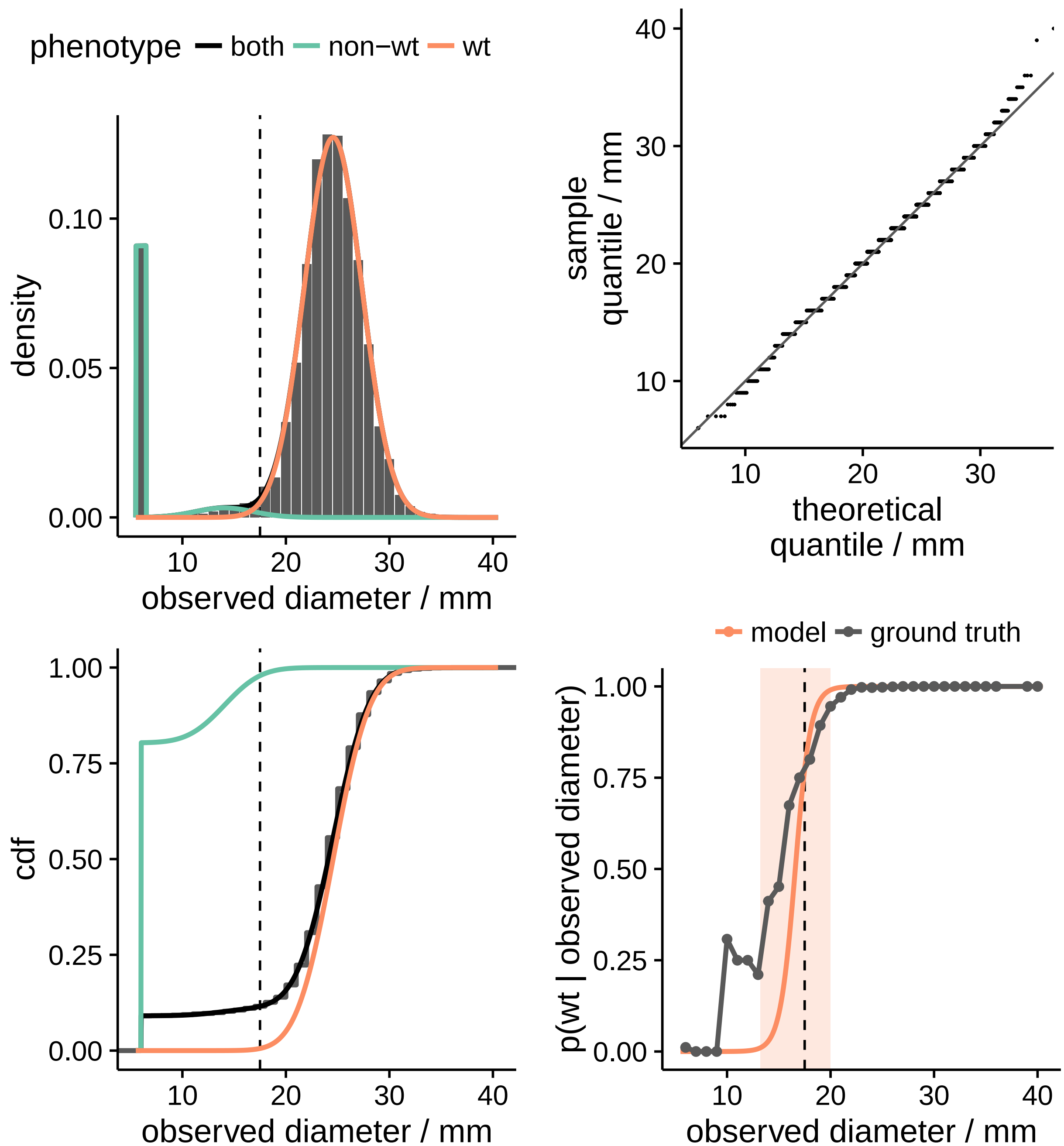
4.1.6 TPZ

component	mean/mm	sd/mm	weight
1	6	0.0	0.00
2	15	3.2	0.03
3	25	3.2	0.97



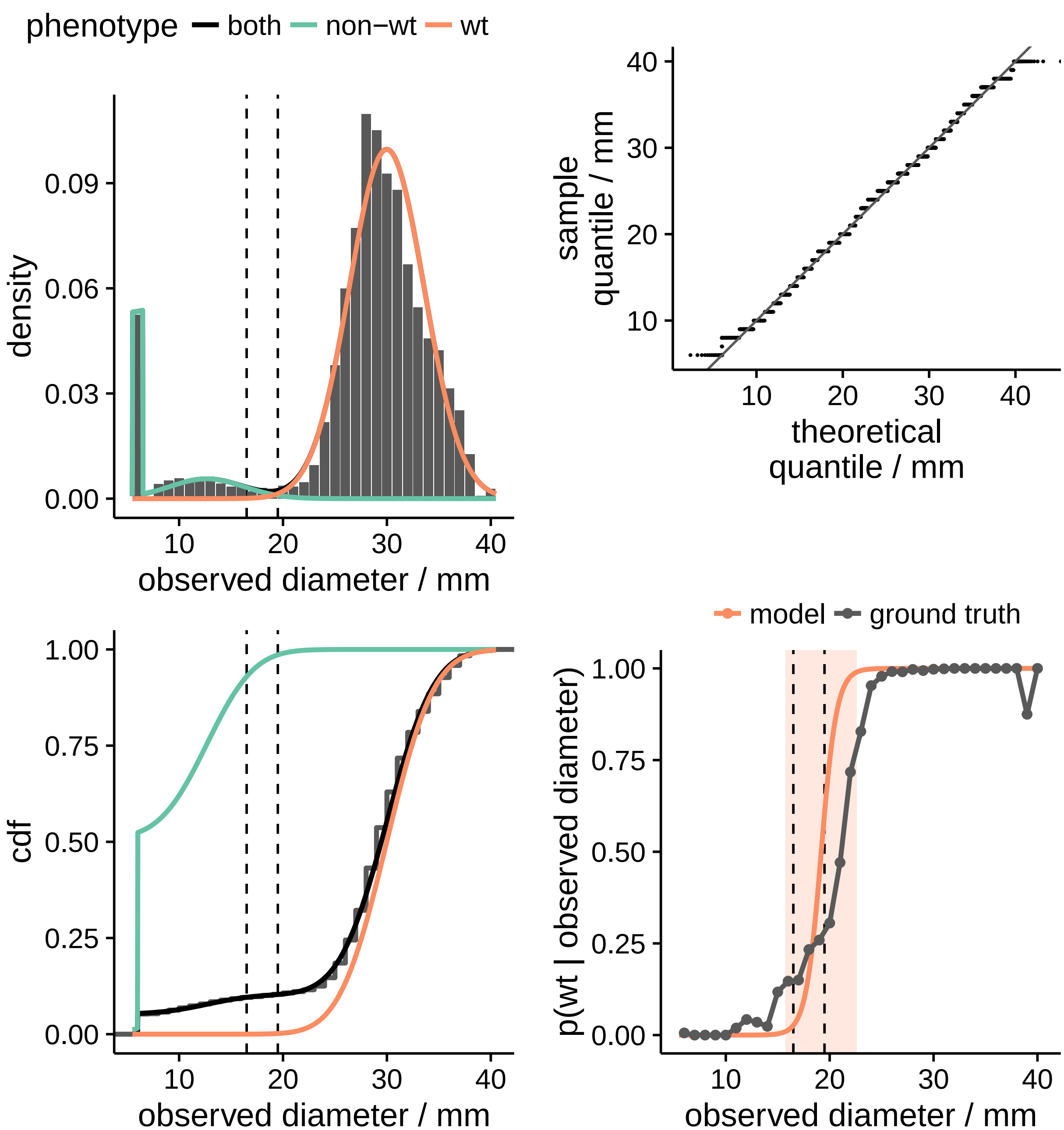
4.1.7 CXM

component	mean/mm	sd/mm	weight
1	6	0.0	0.09
2	14	2.8	0.02
3	25	2.8	0.89



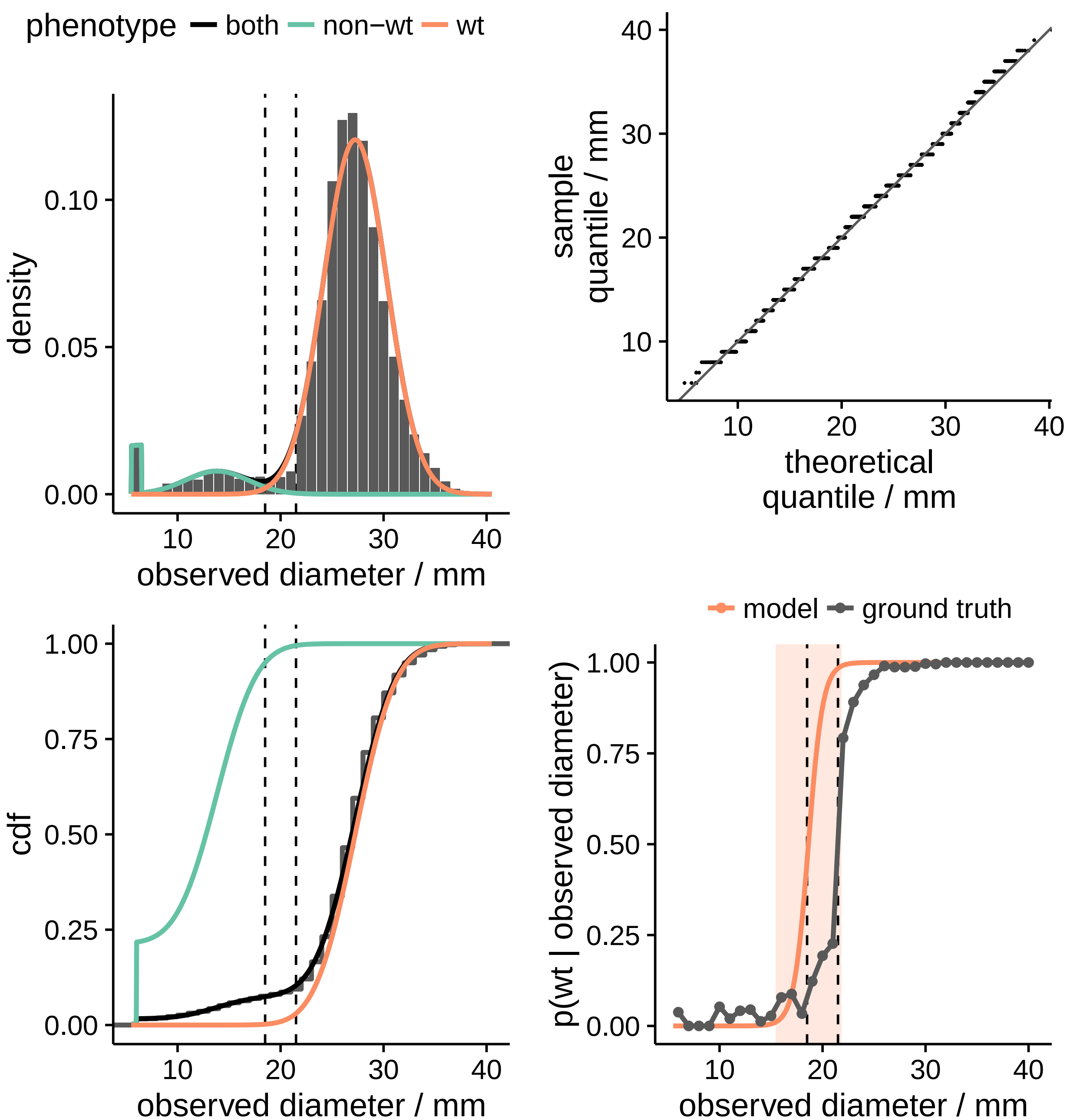
4.1.8 CTX

component	mean/mm	sd/mm	weight
1	6	0.0	0.05
2	13	3.6	0.05
3	30	3.6	0.90



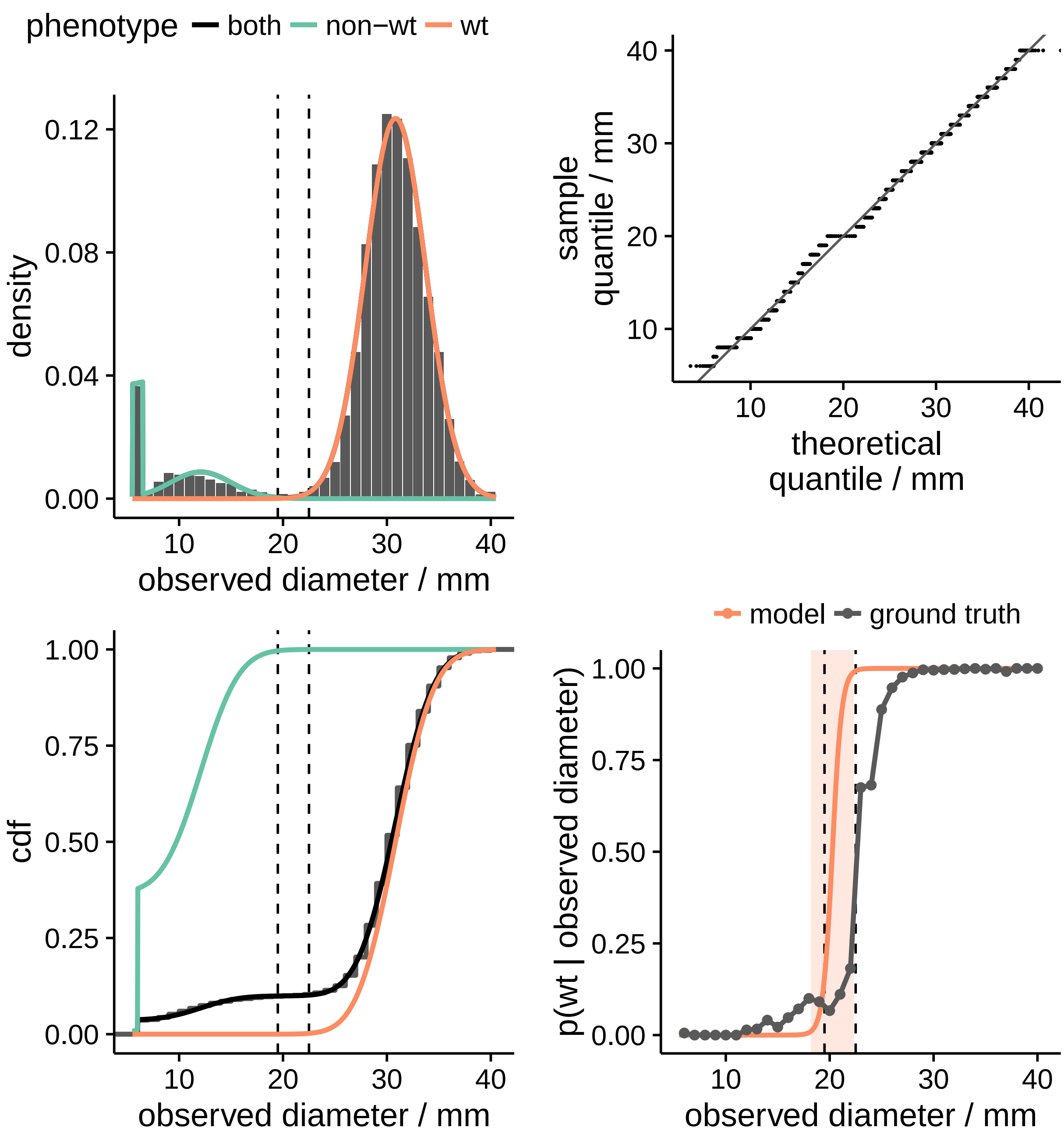
4.1.9 CAZ

component	mean/mm	sd/mm	weight
1	6	0.0	0.02
2	14	3.1	0.06
3	27	3.1	0.92



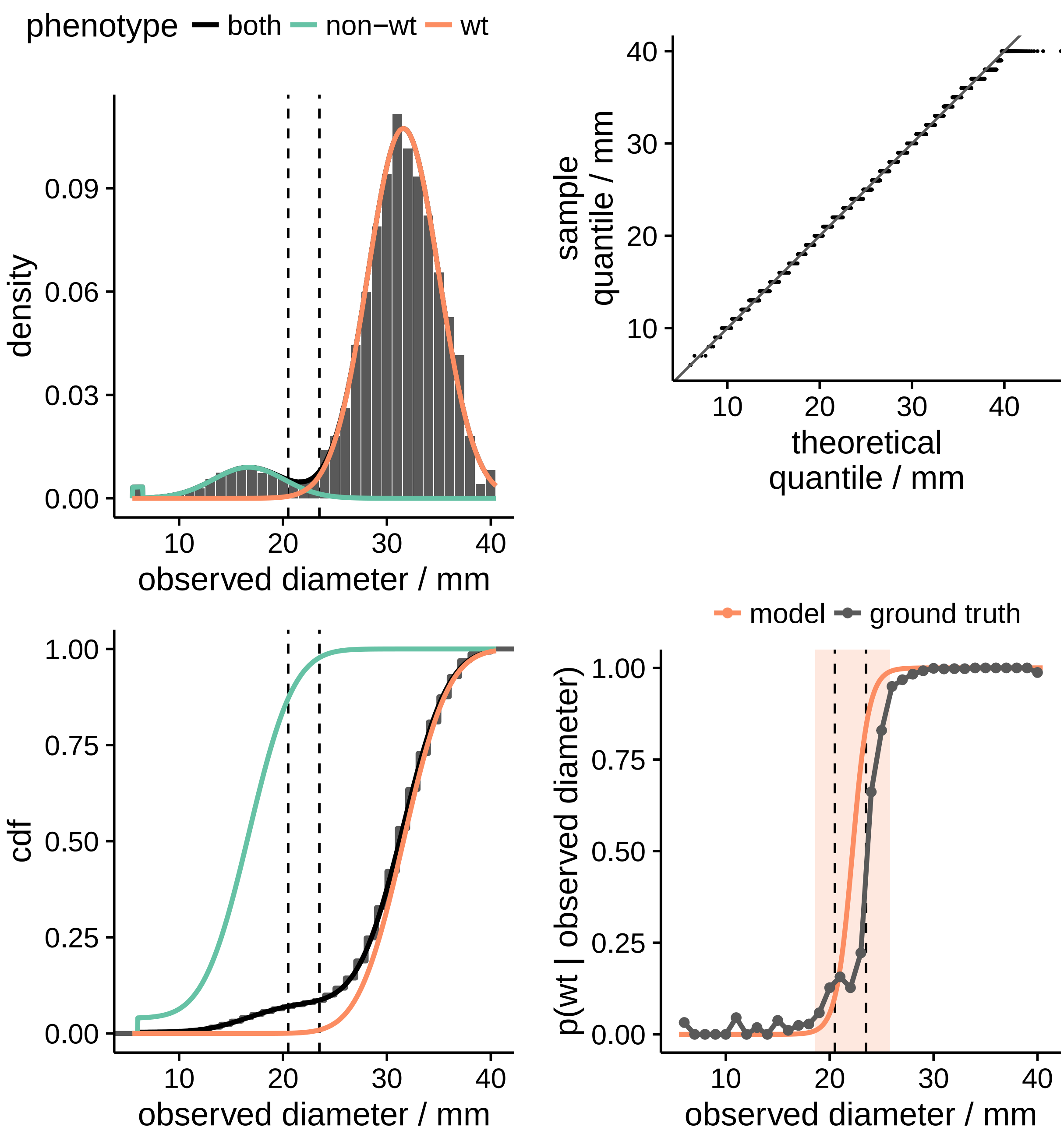
4.1.10 CRO

component	mean/mm	sd/mm	weight
1	6	0.0	0.04
2	12	2.9	0.06
3	31	2.9	0.90



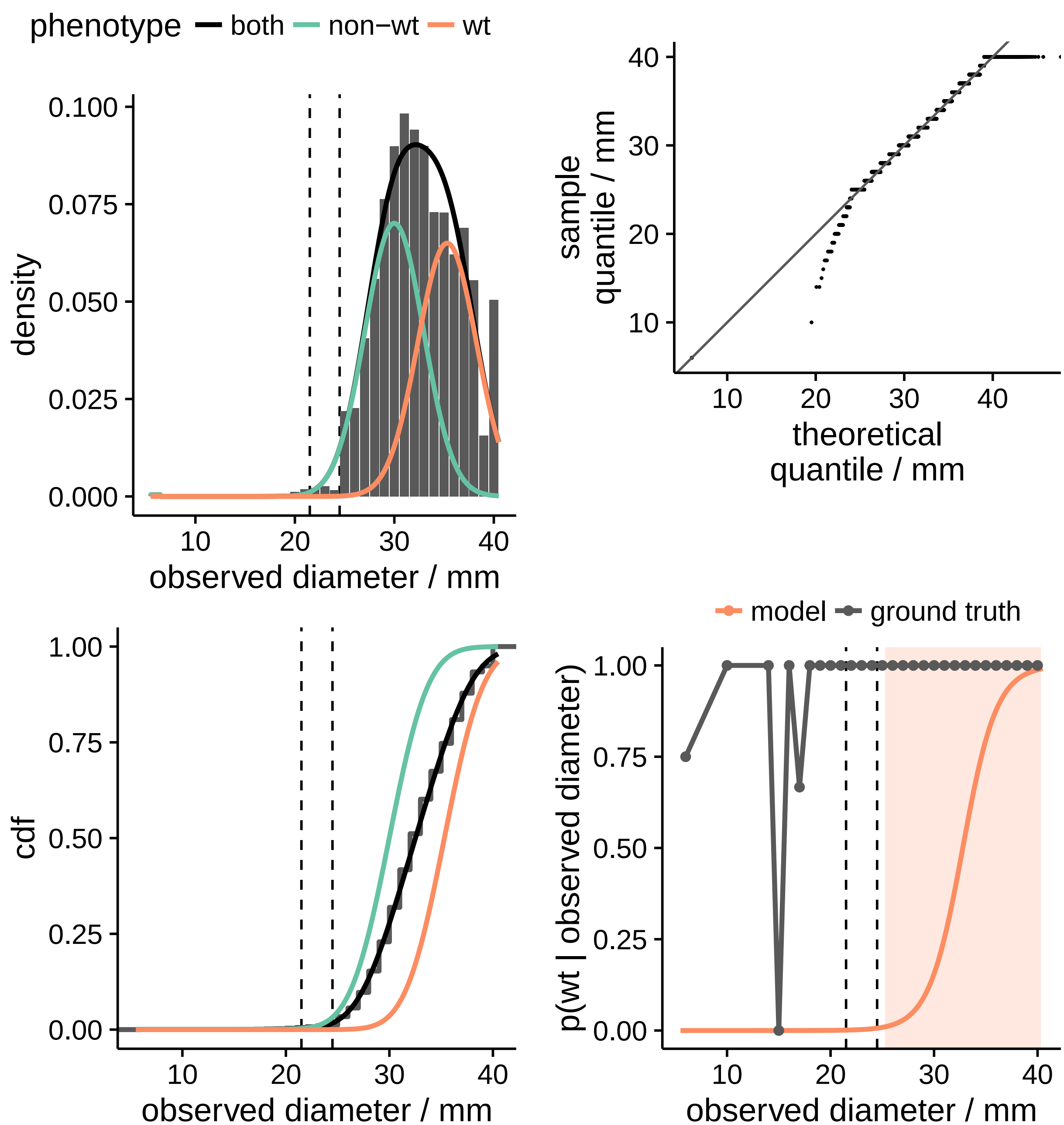
4.1.11 FEP

component	mean/mm	sd/mm	weight
1	6	0.0	0.00
2	17	3.4	0.08
3	32	3.4	0.92



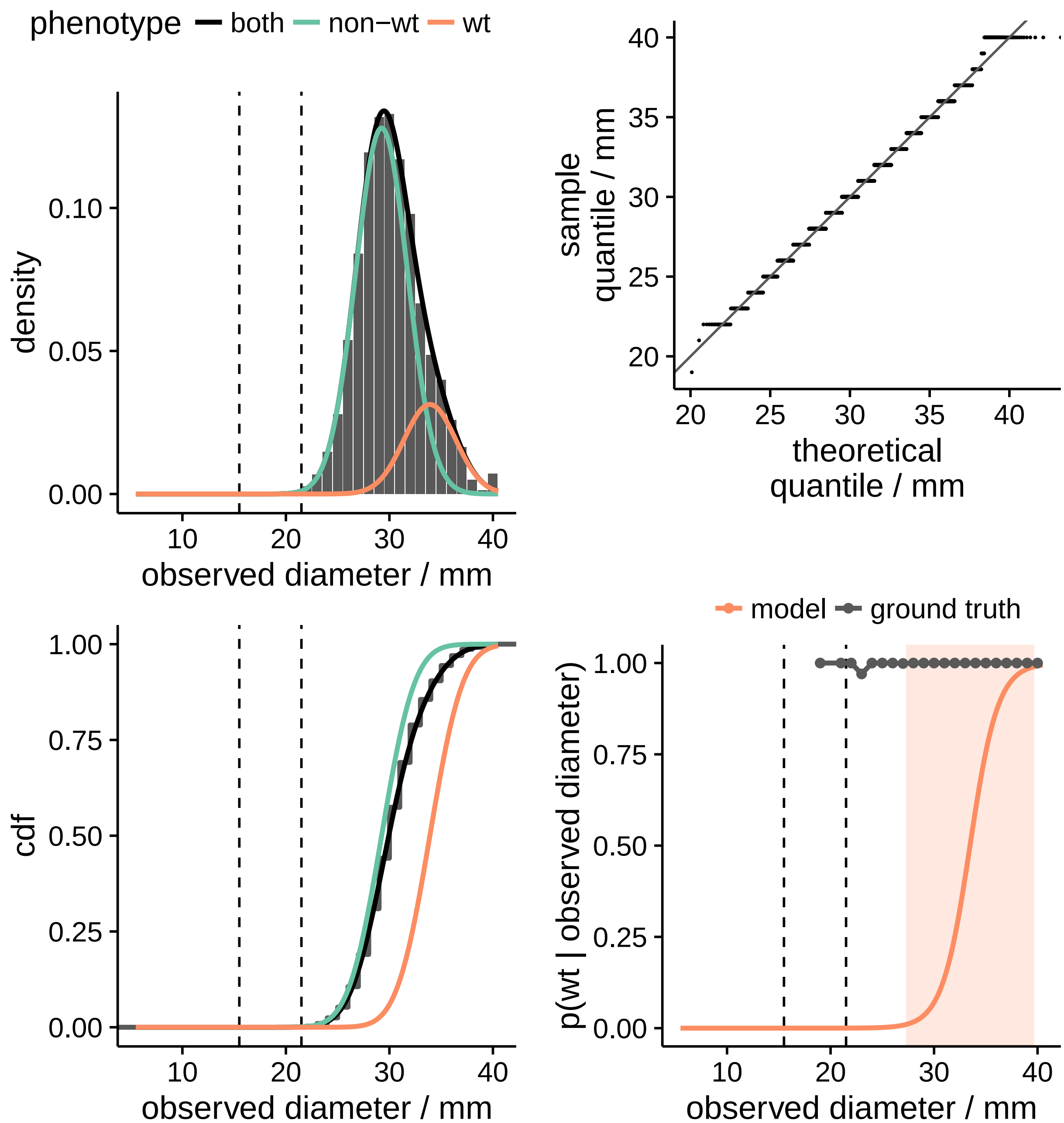
4.1.12 ETP

component	mean/mm	sd/mm	weight
1	6	0	0.00
2	30	3	0.52
3	35	3	0.48



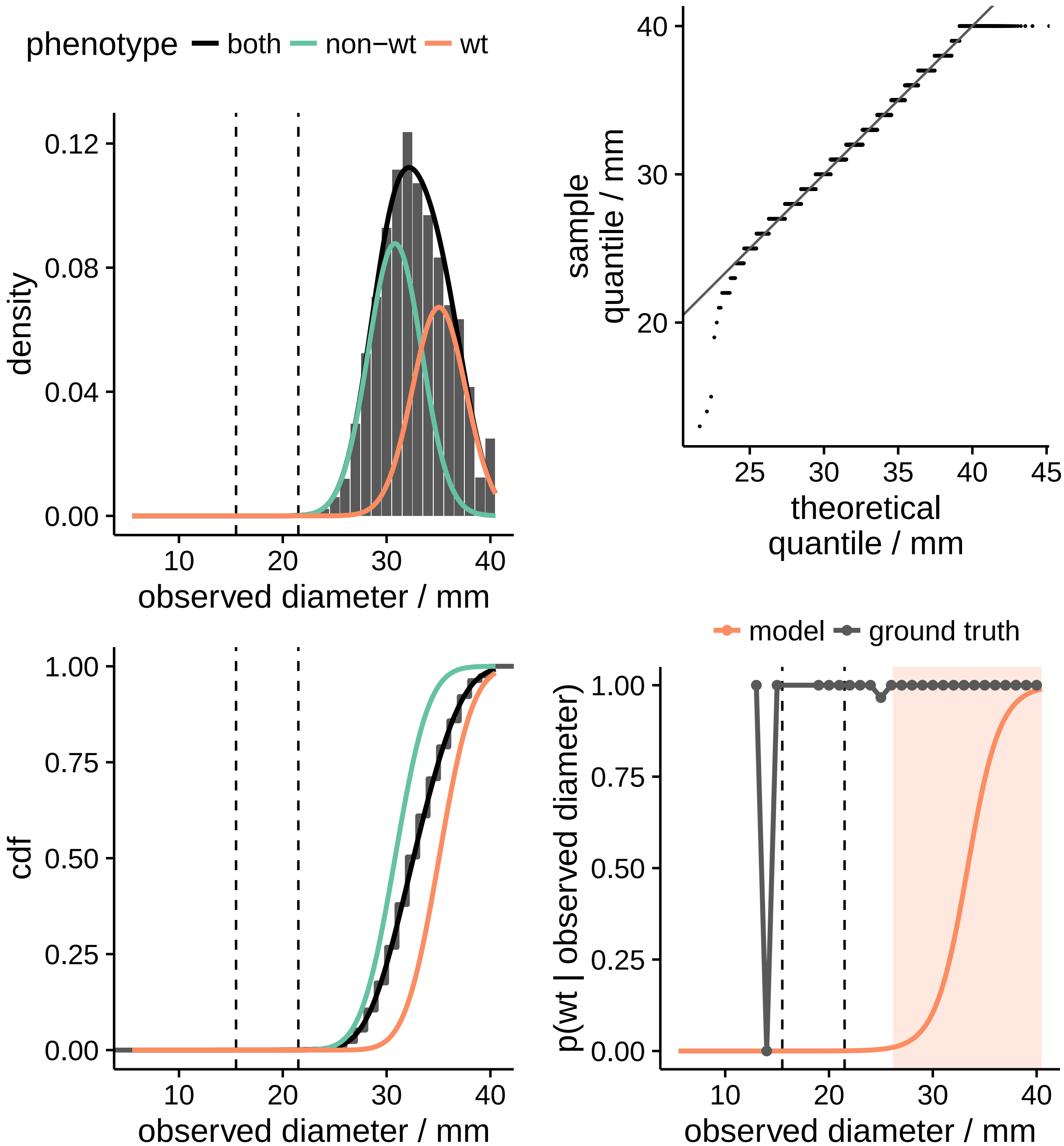
4.1.13 IPM

component	mean/mm	sd/mm	weight
1	29	2.5	0.8
2	34	2.5	0.2



4.1.14 MEM

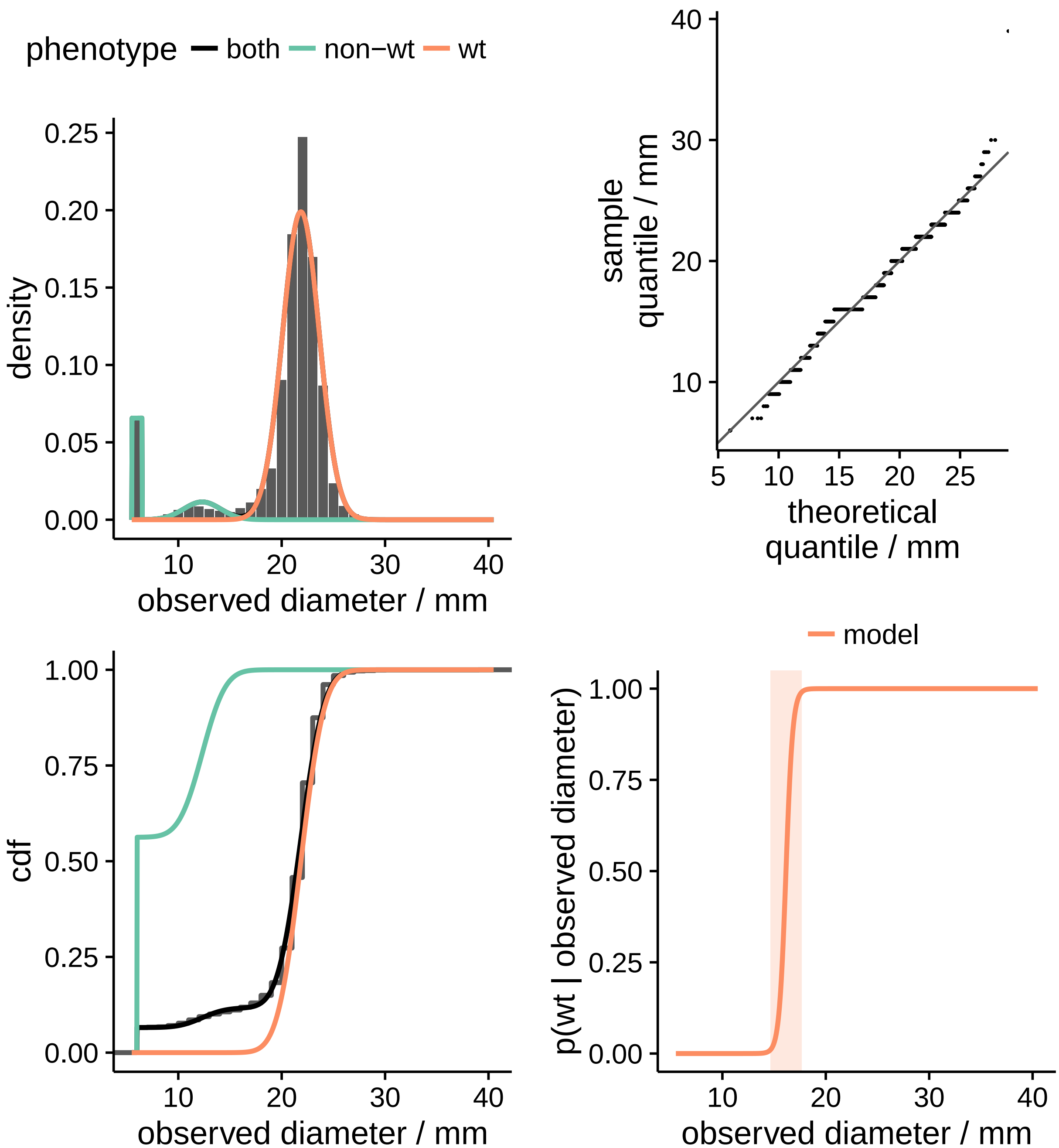
component	mean/mm	sd/mm	weight
1	31	2.6	0.57
2	35	2.6	0.43



4.2 Aminoglycosides

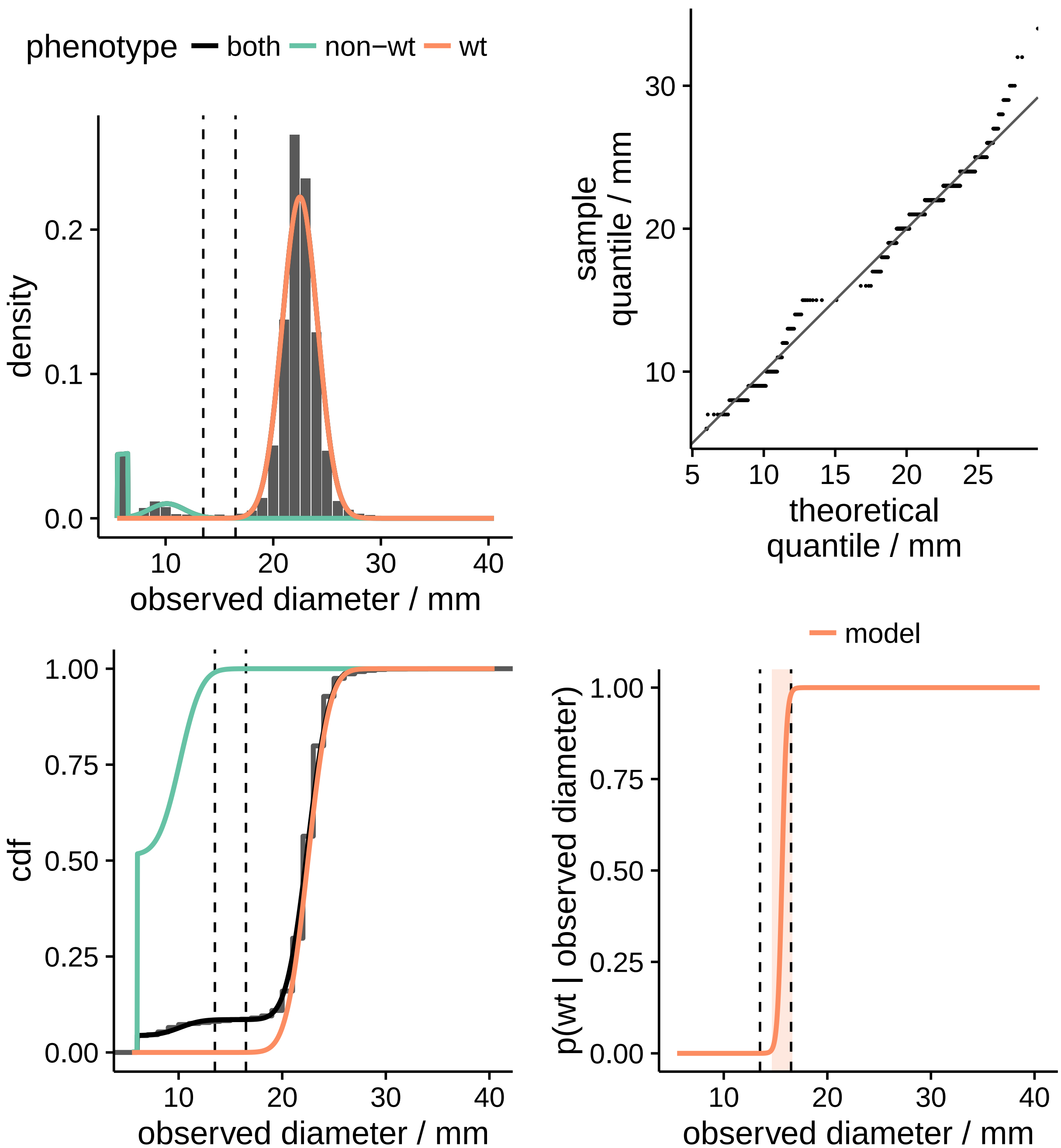
4.2.1 KAN

component	mean/mm	sd/mm	weight
1	6	0.0	0.07
2	12	1.8	0.05
3	22	1.8	0.88



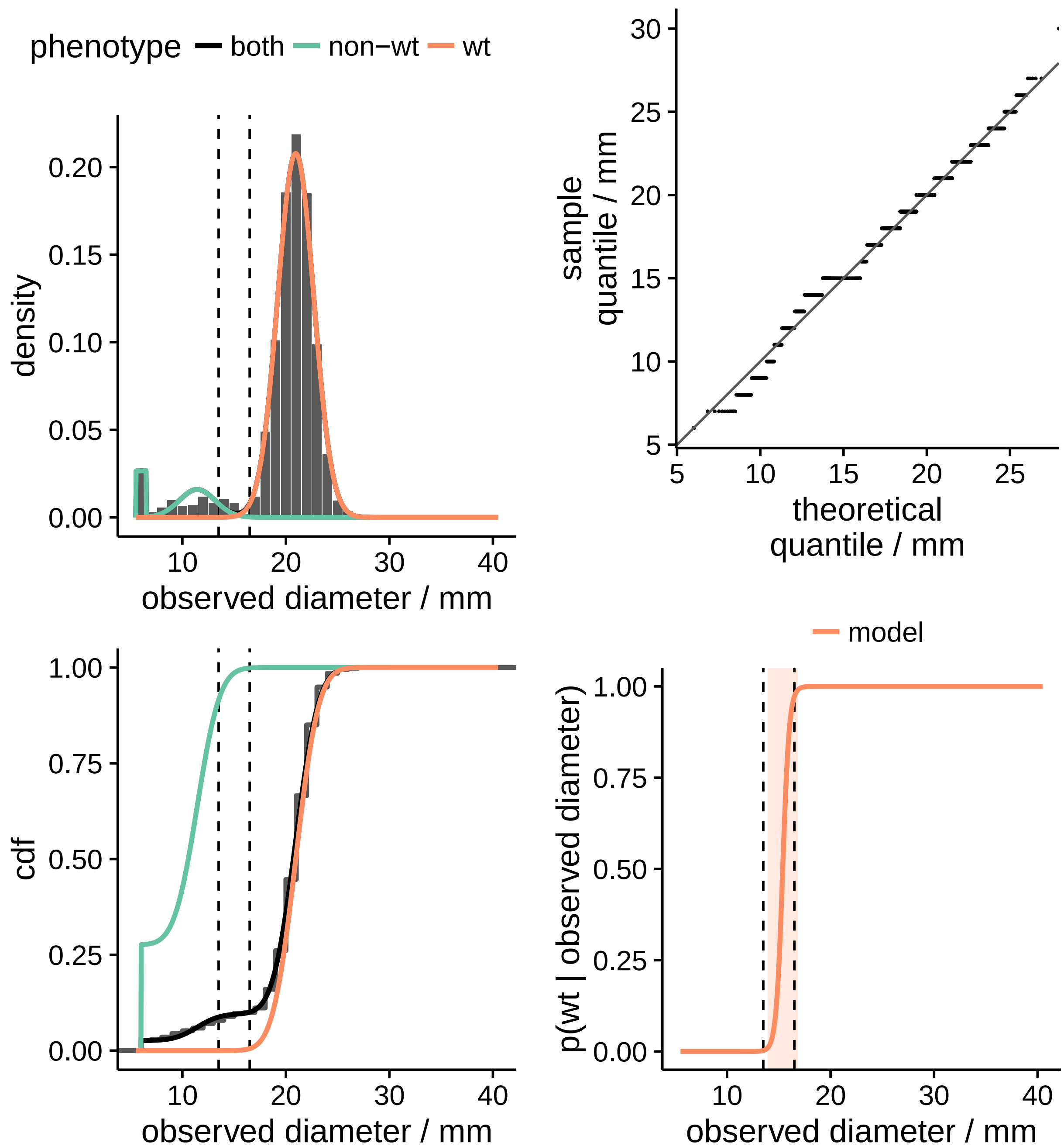
4.2.2 GEN

component	mean/mm	sd/mm	weight
1	6	0.0	0.04
2	10	1.6	0.04
3	22	1.6	0.91



4.2.3 TOB

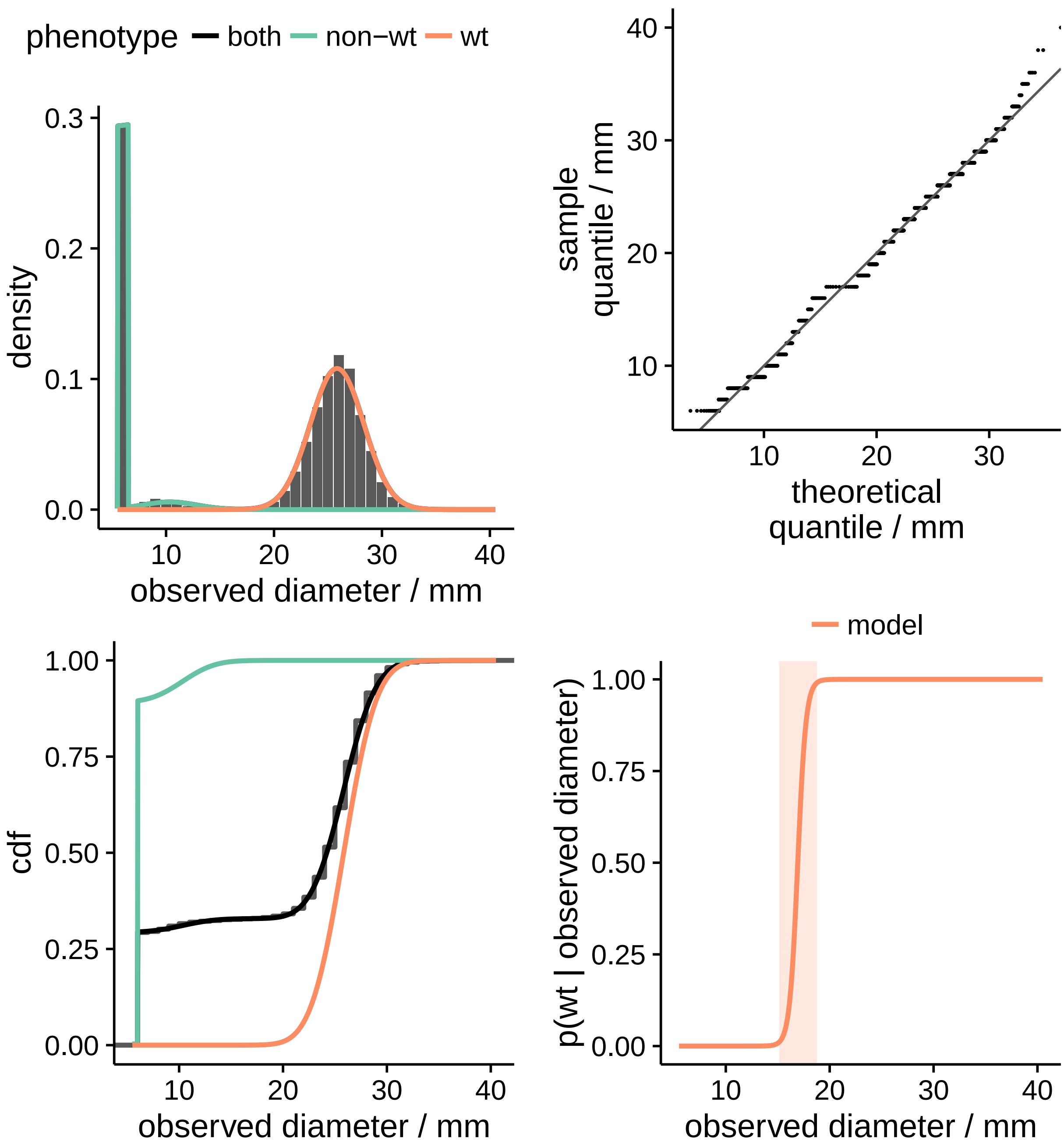
component	mean/mm	sd/mm	weight
1	6	0.0	0.03
2	11	1.7	0.07
3	21	1.7	0.90



4.3 Quinolones

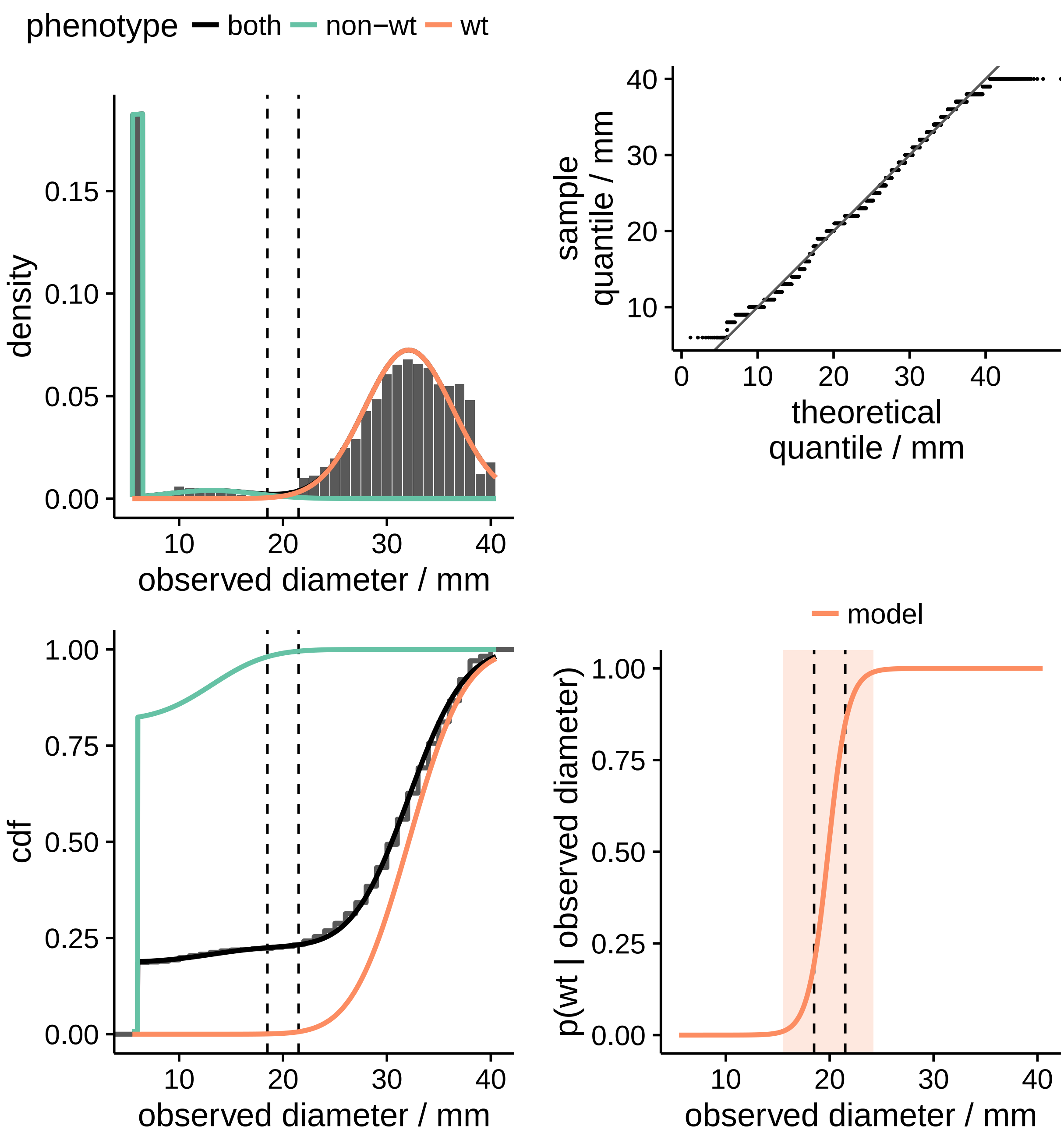
4.3.1 NAL

component	mean/mm	sd/mm	weight
1	6	0.0	0.29
2	10	2.5	0.04
3	26	2.5	0.67



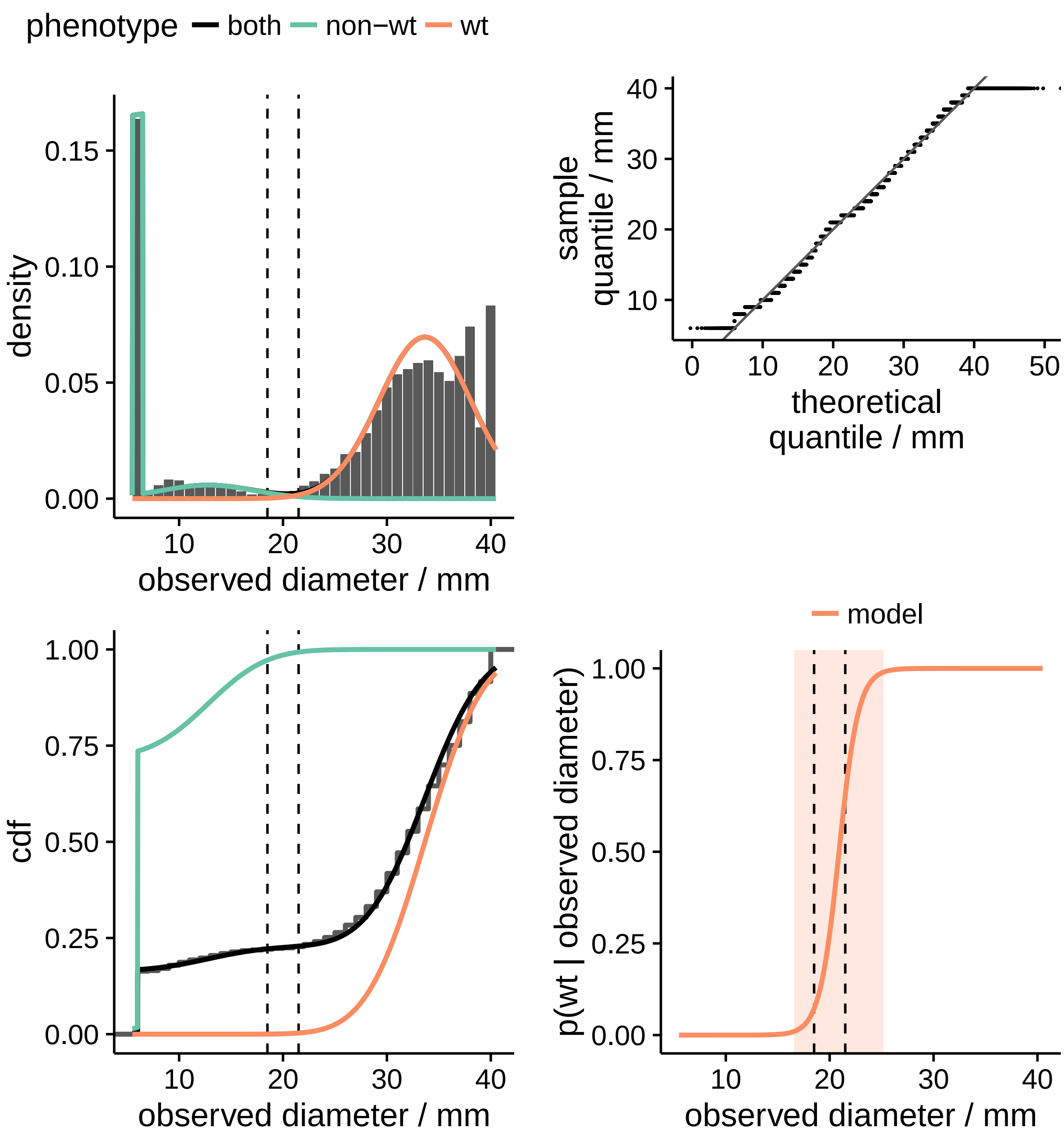
4.3.2 NOR

component	mean/mm	sd/mm	weight
1	6	0.0	0.19
2	13	4.2	0.04
3	32	4.2	0.77



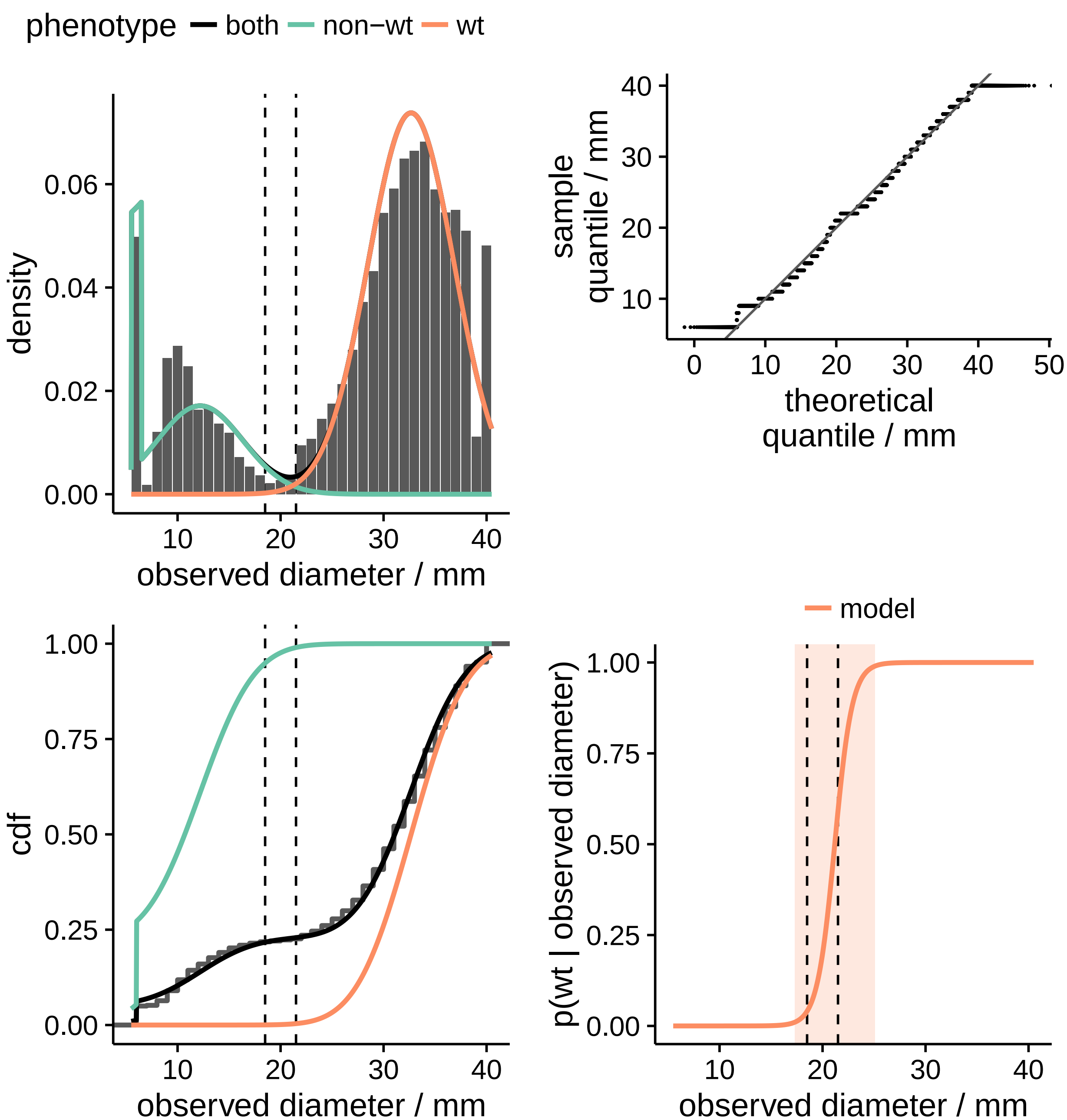
4.3.3 CIP

component	mean/mm	sd/mm	weight
1	6	0.0	0.16
2	13	4.4	0.06
3	34	4.4	0.77



4.3.4 LEV

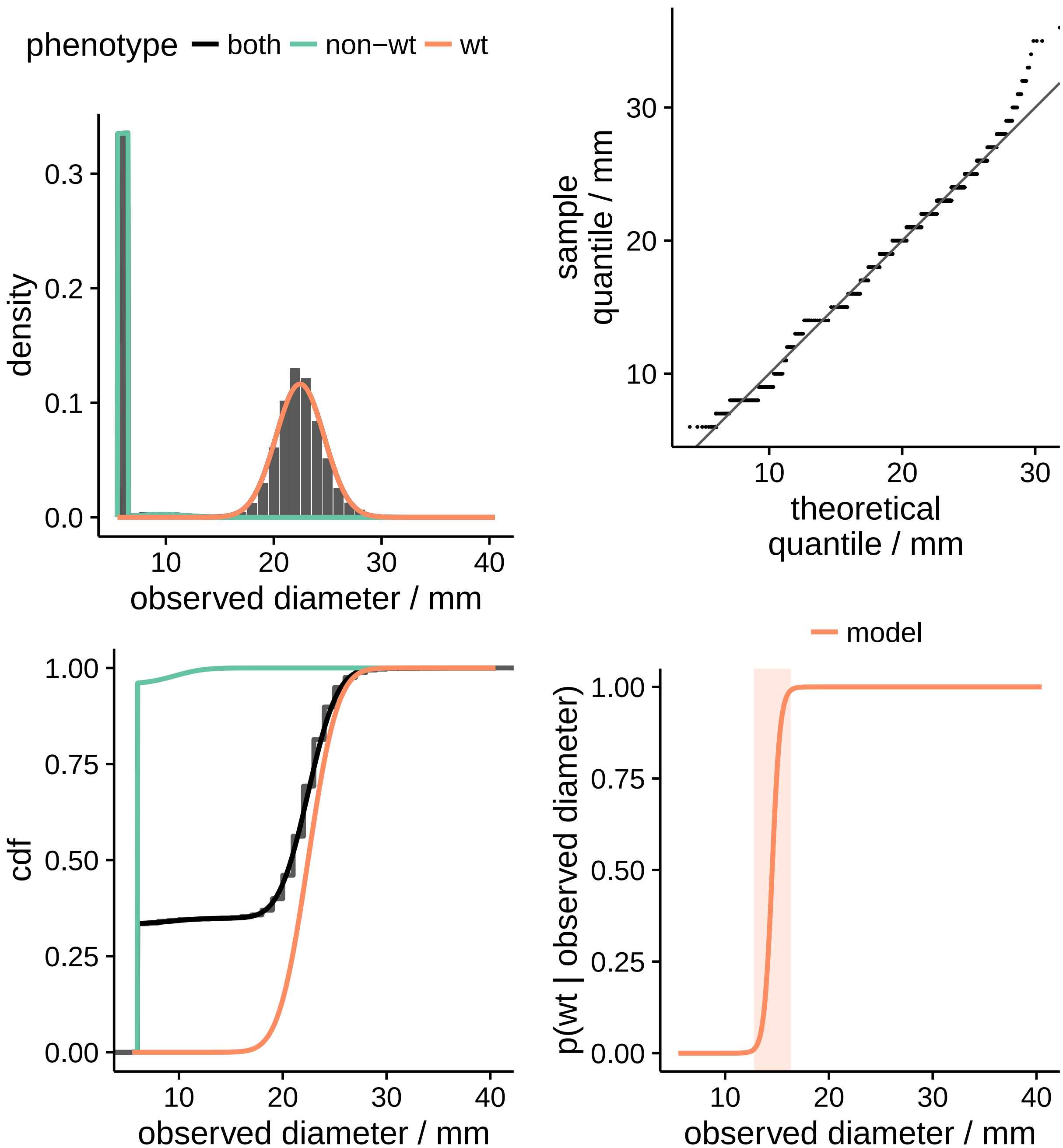
component	mean/mm	sd/mm	weight
1	6	0.0	0.05
2	12	4.2	0.18
3	33	4.2	0.77



4.4 Tetracyclines

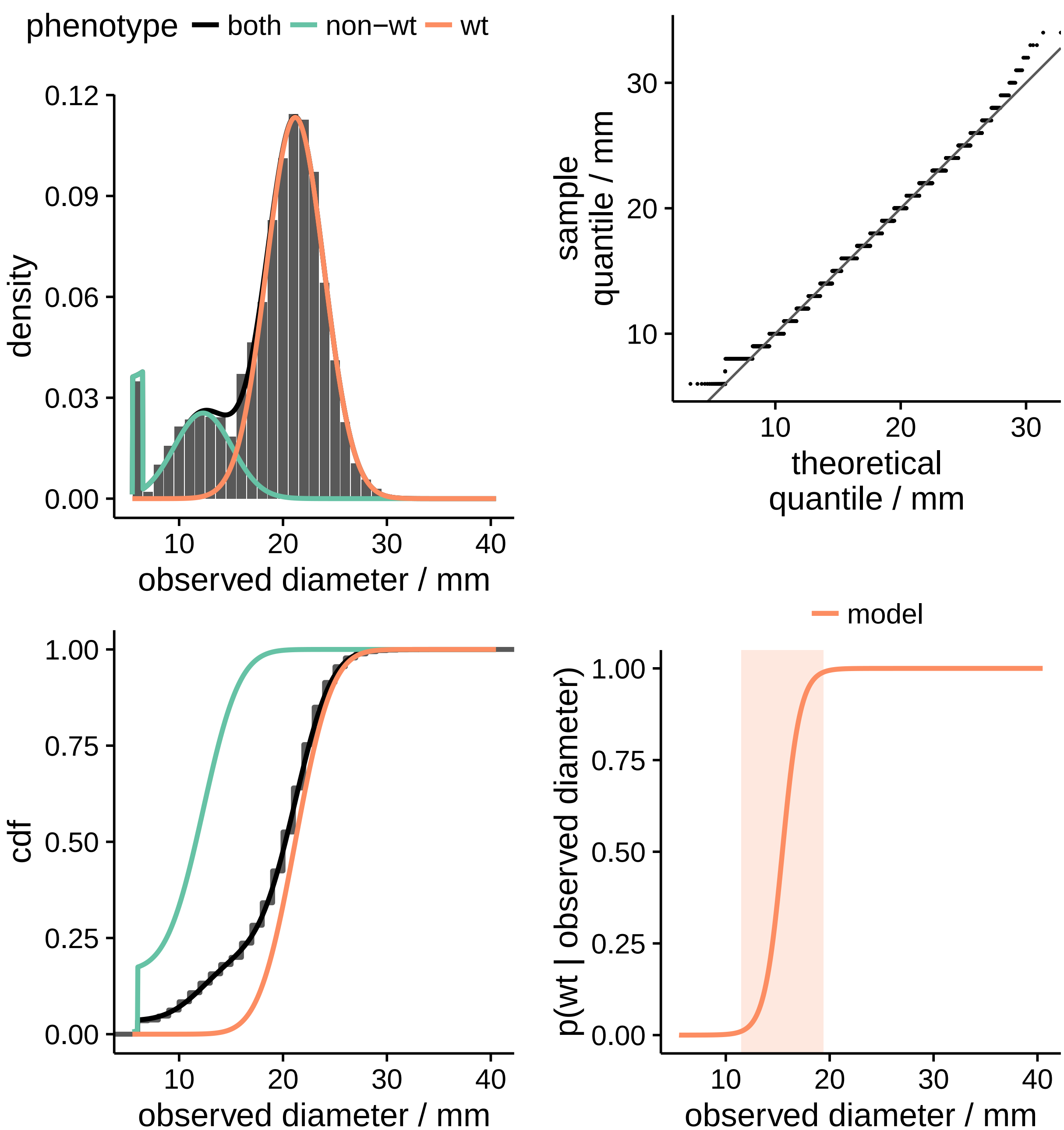
4.4.1 TE

component	mean/mm	sd/mm	weight
1	6.0	0.0	0.33
2	9.6	2.2	0.01
3	22.4	2.2	0.65



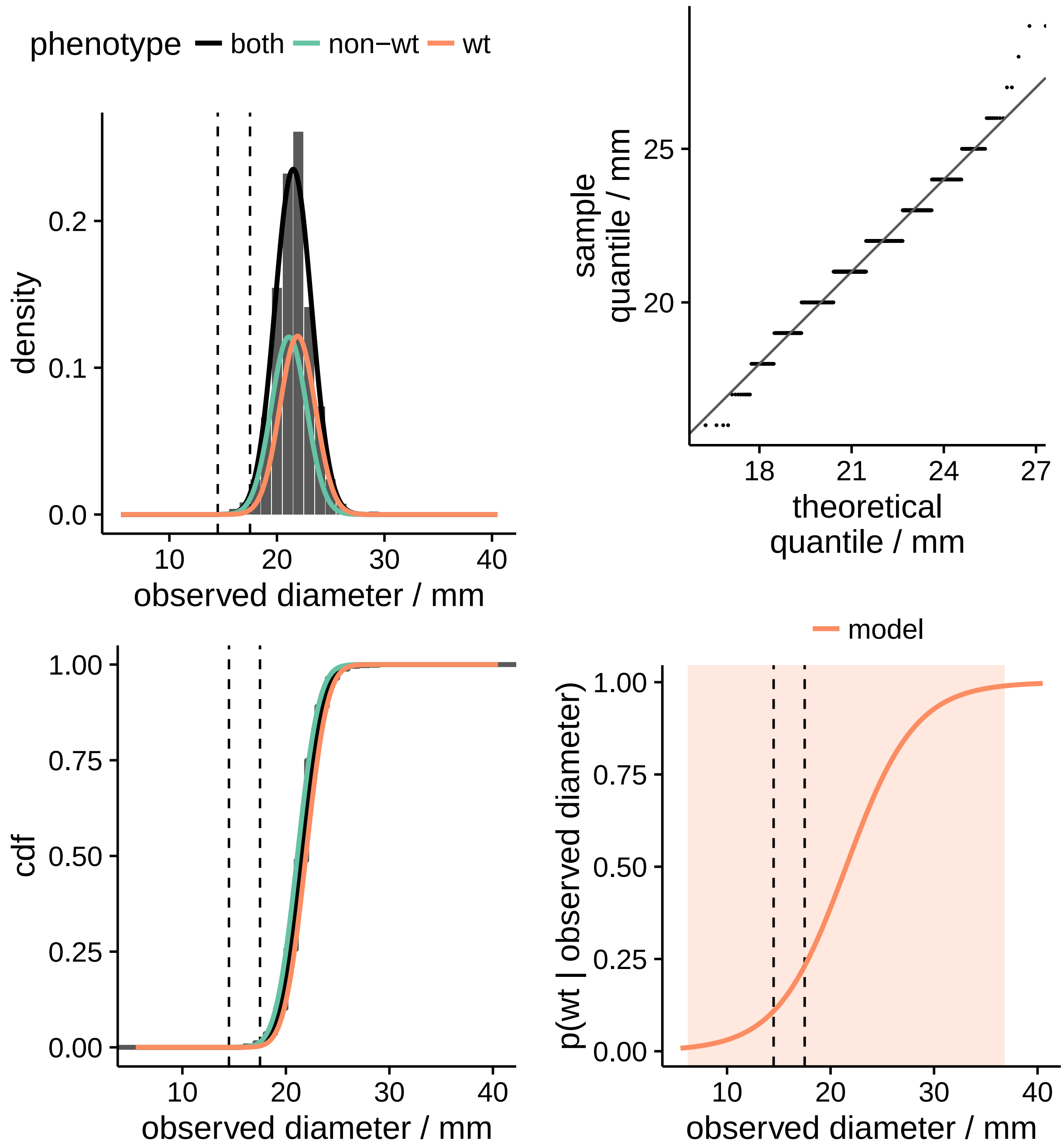
4.4.2 MI

component	mean/mm	sd/mm	weight
1	6	0.0	0.03
2	12	2.8	0.18
3	21	2.8	0.79



4.4.3 TGC

component	mean/mm	sd/mm	weight
1	21	1.6	0.5
2	22	1.6	0.5



antibiotic	width of zone of uncertainty / mm
AM10	8.6
KF	10.3
FOX	7.5
CPD	6.9
AMC	9.5
TPZ	9.6
CXM	6.8
CTX	6.8
CAZ	6.4
CRO	4.1
FEP	7.2
ETP	15.0
IPM	12.4
MEM	14.3
KAN	3.0
GEN	2.0
TOB	2.9
NAL	3.6
NOR	8.7
CIP	8.6
LEV	7.8
TE	3.5
MI	7.9
TGC	30.6

5 Discussion of agreement with ground truth for beta-lactams

- **AMC, TPZ:** We observe no agreement at all between model and ground truth due to the fact that the strains assigned to BSBL/OTHERS are defined to be non-wt for AMC and TPZ. If BSBL/OTHERS is defined to be wt for AMC and TPZ the results are qualitatively similar to the ones for CXM (data not shown).
- **CXM:** For diameters above 15 mm the model agrees with the ground truth. For diameters in the range of 10 to 15 mm, the probabilities for wt derived from the ground truth are higher than the ones derived from the model. This is due to the fact that the strains assigned to BSBL/OTHERS are defined to be wt for CXM. This might be problematic given the differences between the distribution of diameters for BSBL/OTHERS and wt (see Appendix).
- **CTX, CAZ, CRO, and FEP:** The curves of the probabilities for wt derived from the model are shifted towards lower diameters by about 2 to 3 mm with respect to the curves based on the ground truth.
- **ETP, MEM, and ITP:** The two normally distributed components obtained from fitting the observed distribution do not correspond to resistant and susceptible components. This is most likely due to the fact that we have only three carba strains.

6 Appendix

6.1 Clinical breakpoints

Table 28: 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

6.2 Ground truth for beta-lactams

6.2.1 Classes

The 9766 strains are distributed among the 8 classes as follows.

class	n
AmpC	173
BSBL/OTHERS	3933
carba	1
ESBL	896
ESBL and carba	1
ESBL and AmpC	19
ESBL, AmpC, and carba	1
wild-type	4742

6.2.2 Distributions

```
## Warning: `origin` is deprecated. Please use `boundary` instead.
```

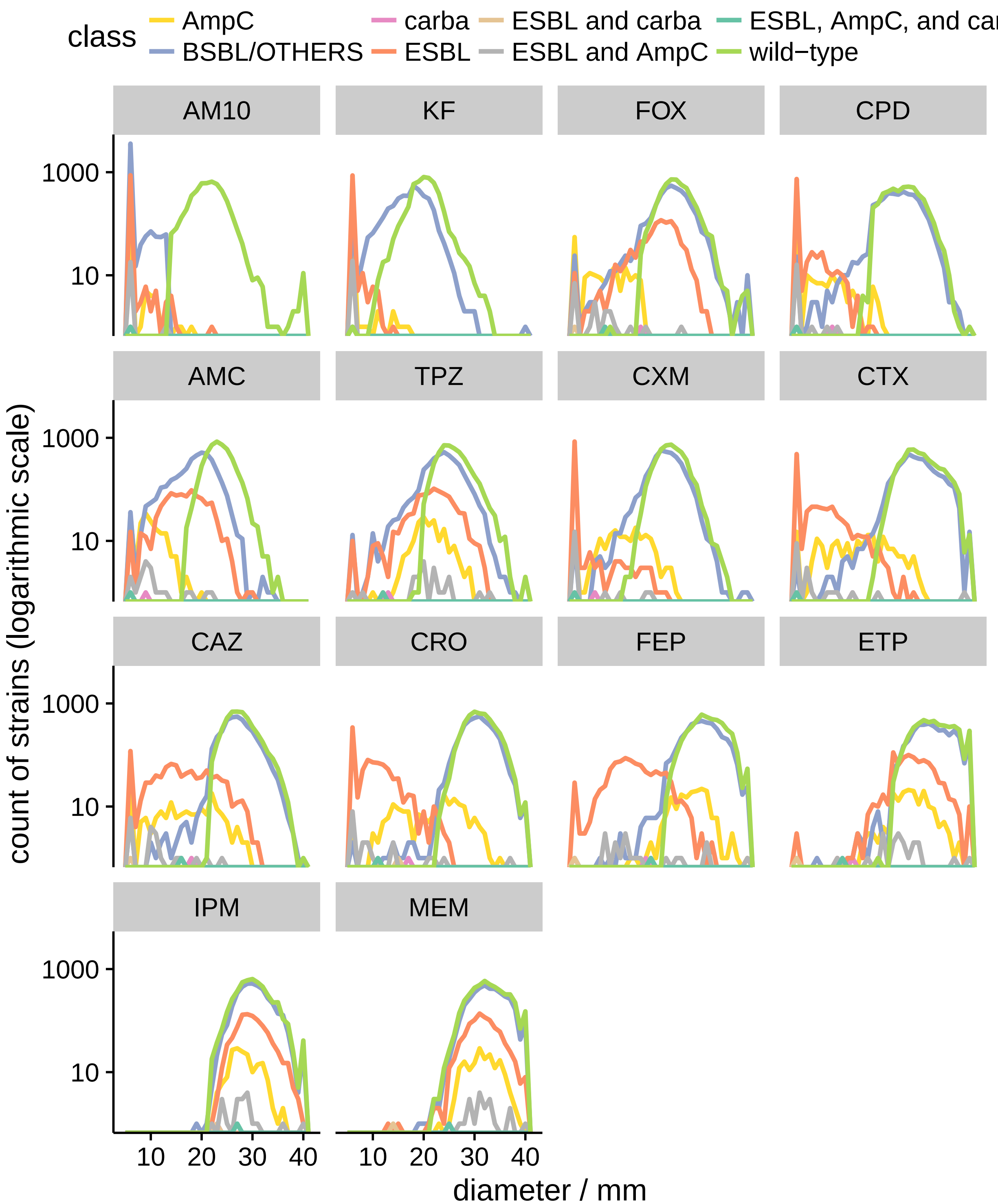


Figure 1: Distributions of diameters for beta-lactams.