Kathrin Baumann-Stanzer, Martin Piringer et al., ZAMG

User experience with model validation exercises

Zentralanstalt für Meteorologie und Geodynamik

Outline

- Methodology
- Validation results for 2 model validation exercises
- “Which model run performed best?”
- Conclusions

Methodology

"… enhance our confidence in a model if it is not rejected according to predefined criteria in a number of tests."

Following the Guidance and Protocol Document of COST 732:

- exploratory data analysis
- residual analysis
- statistical analysis

Quantify the "state of validation" (after Sornette et al. 2007):

$$F \approx \tanh\left(\frac{p}{q}\right)^4$$

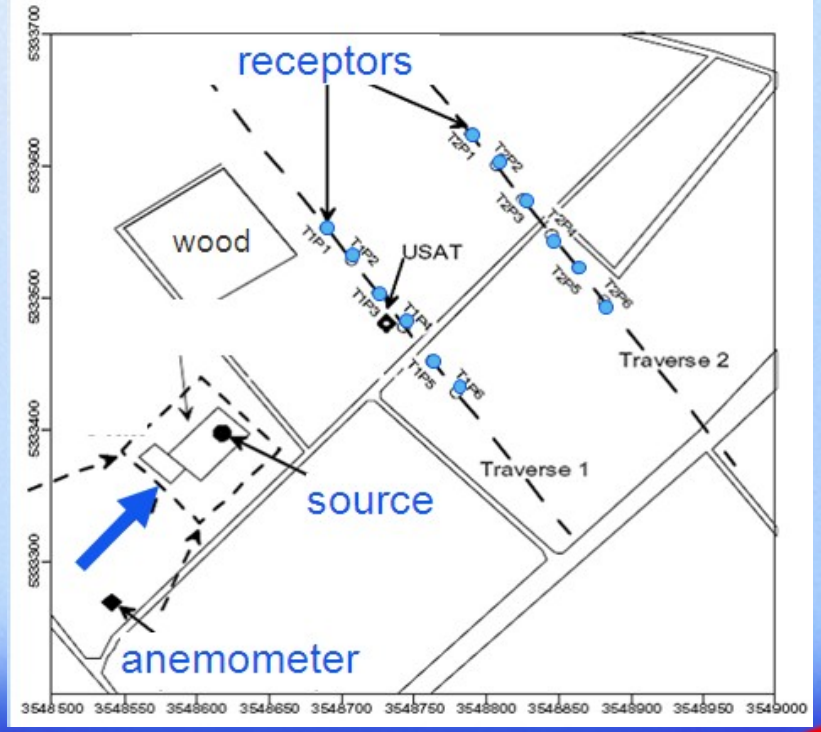
p/q = 0.1 'poor fit'

1.0 'marginally good'

10.0 'good fit'

Odour release and dispersion project

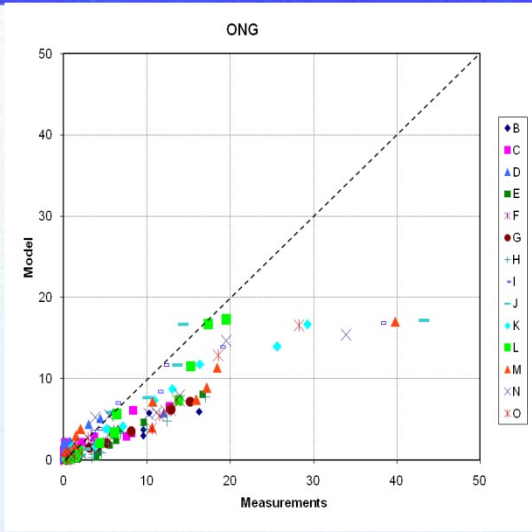
- Data-set available from German environmental program BWPLUS
- Source: pig fattening unit in flat terrain
- 13 (SF6) tracer experiments of 10 minutes duration
- neutral conditions, wind speeds: 2,5 to 7,9 m/s

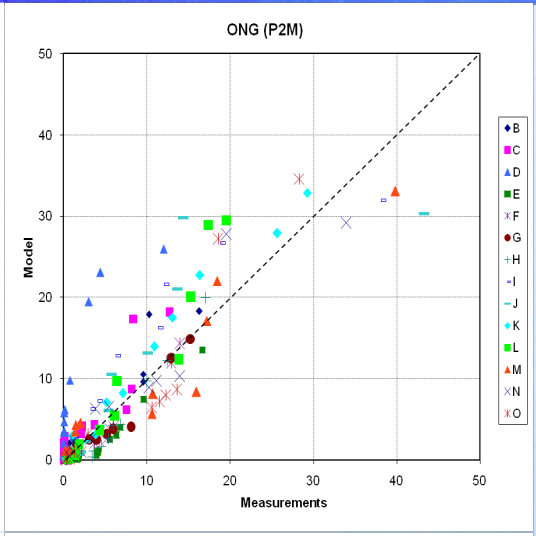


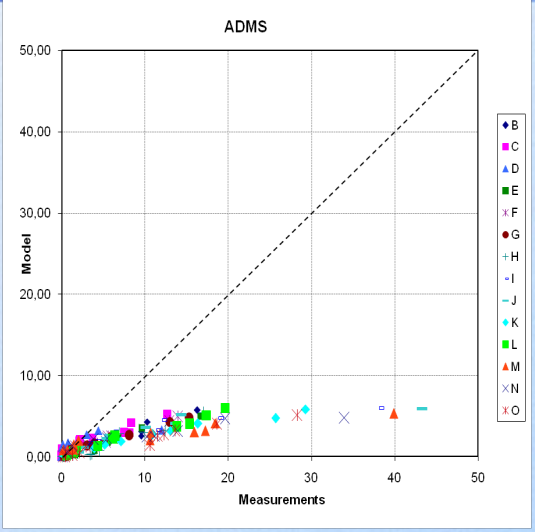
OROD: quantile-quantile plots

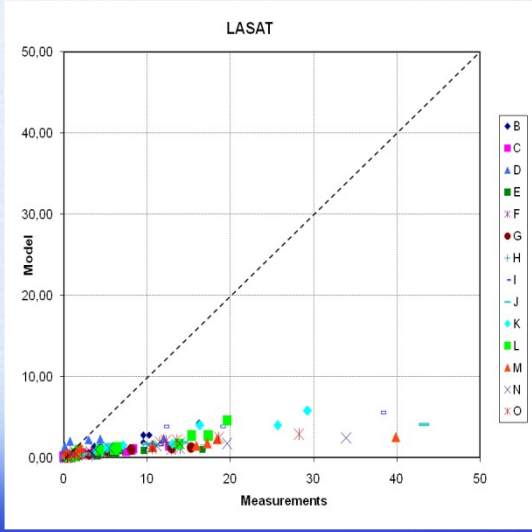
Gaussian

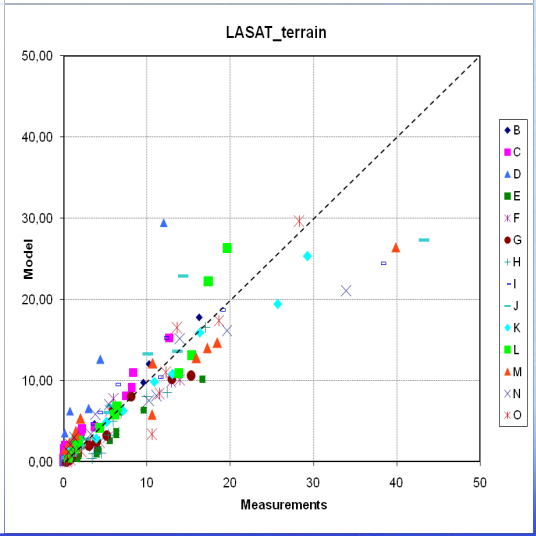
Lagrangian









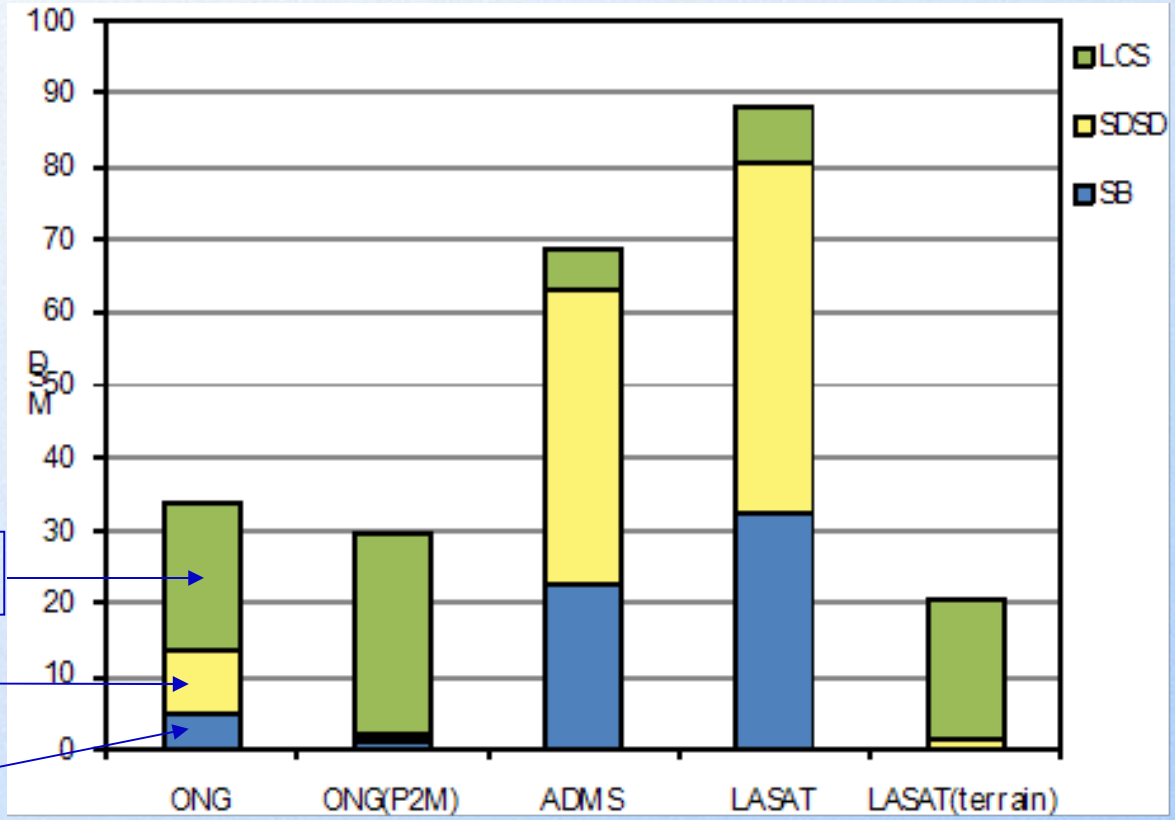


OROD: mean squared deviation

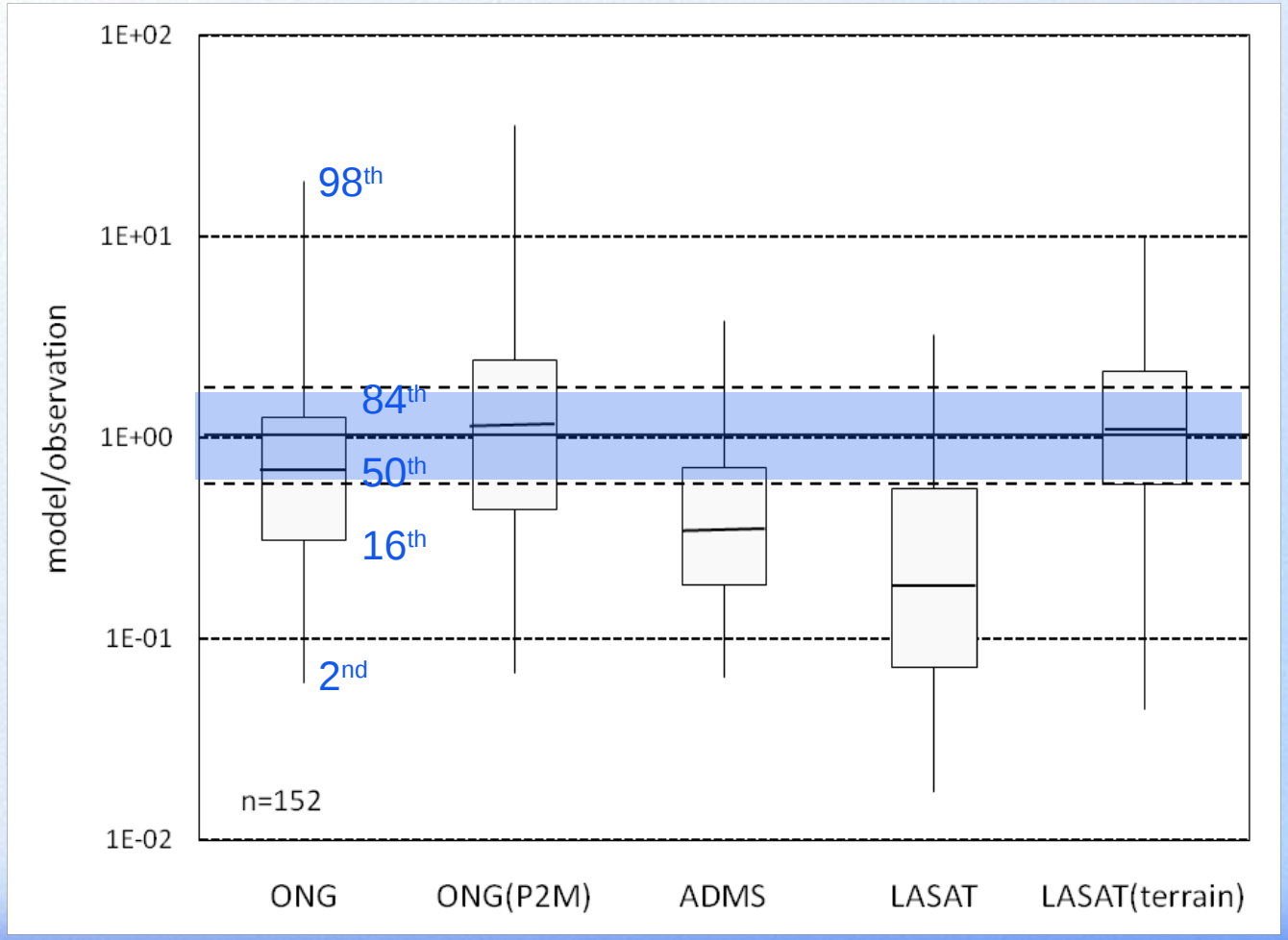
$$LCS = 2SD_{\mathrm{mod}el}SD_{obs}(1-r)$$

$$SDSD = \left(SD_{\mathrm{mod}el} - SD_{obs}\right)^2$$

$$SB = \left(\overline{m} - \overline{o}\right)^2$$



OROD: Residual plot (percentiles)



OROD: metrics

	ONG	ONG(P2M)	ADMS	LASAT	LASAT(terrain)	'acceptable'
MG	1.2	1.0	1.6	2.1	1.0	0,7 < MG < 1,3
VG	1.4	1.4	1.8	2.4	1.4	VG < 1,6
FB	0.3	-0.2	0.9	1.1	0.04	-0,3 < FB < 0,3
NMSE	1.3	0.7	5.7	10.5	0.5	NMSE < 4

MG … geometric mean
VG … geometric variance
FB … fractional bias
NMSE … normalized mean square error

Which model run performed "best"?

$F = F_{MSD}.F_{Res}.F_{MG}.F_{VG}.F_{FB}.F_{NMSE}$

highest score for LASAT(terrain) (F=3)
good - ONG, ONG(P2M) (F=2)
poor - LASAT, ADMS

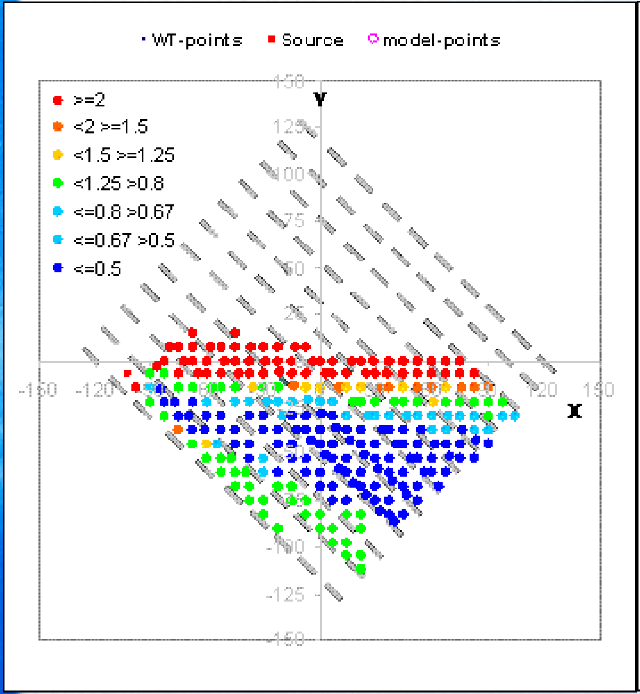
$$F \approx \tanh\left(\frac{p}{q}\right)^4$$

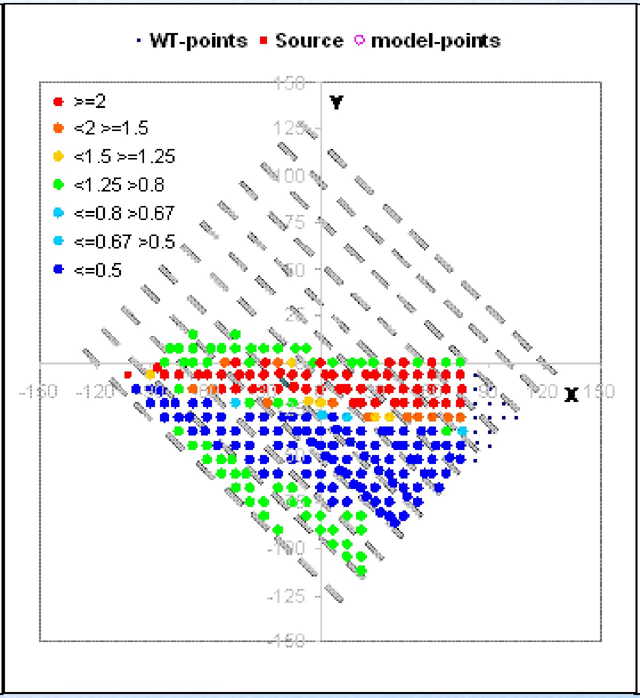
Mock Urban Setting Test (MUST)

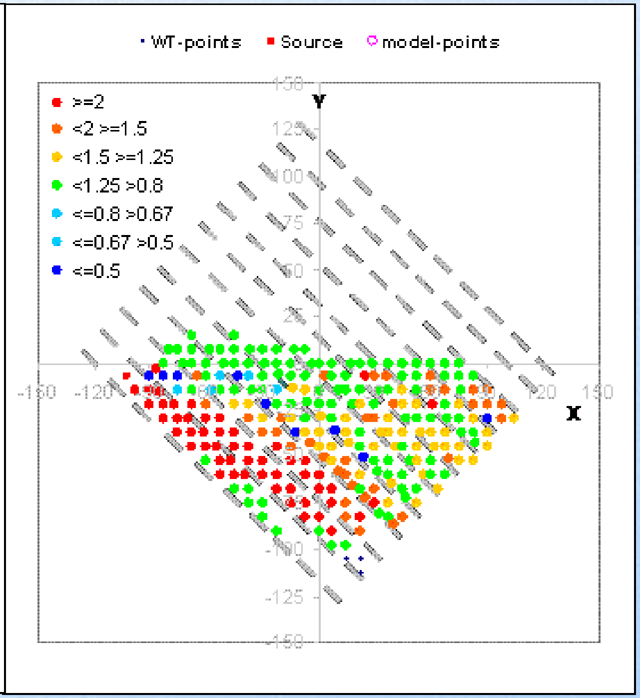
- Wind tunnel experiment (Hamburg University) based on field exp. at Dugway Proving grounds, Utah 2001
- Data provided to COST-Action 732 (Quality Assurance of Micro-scale Meteorological Models)
- Urban area represented by array of 120 shipping containers
- Selected case: NW wind, 7 m/s, neutral conditions

Kathrin Baumann-Stanzer

MUST: model/observation





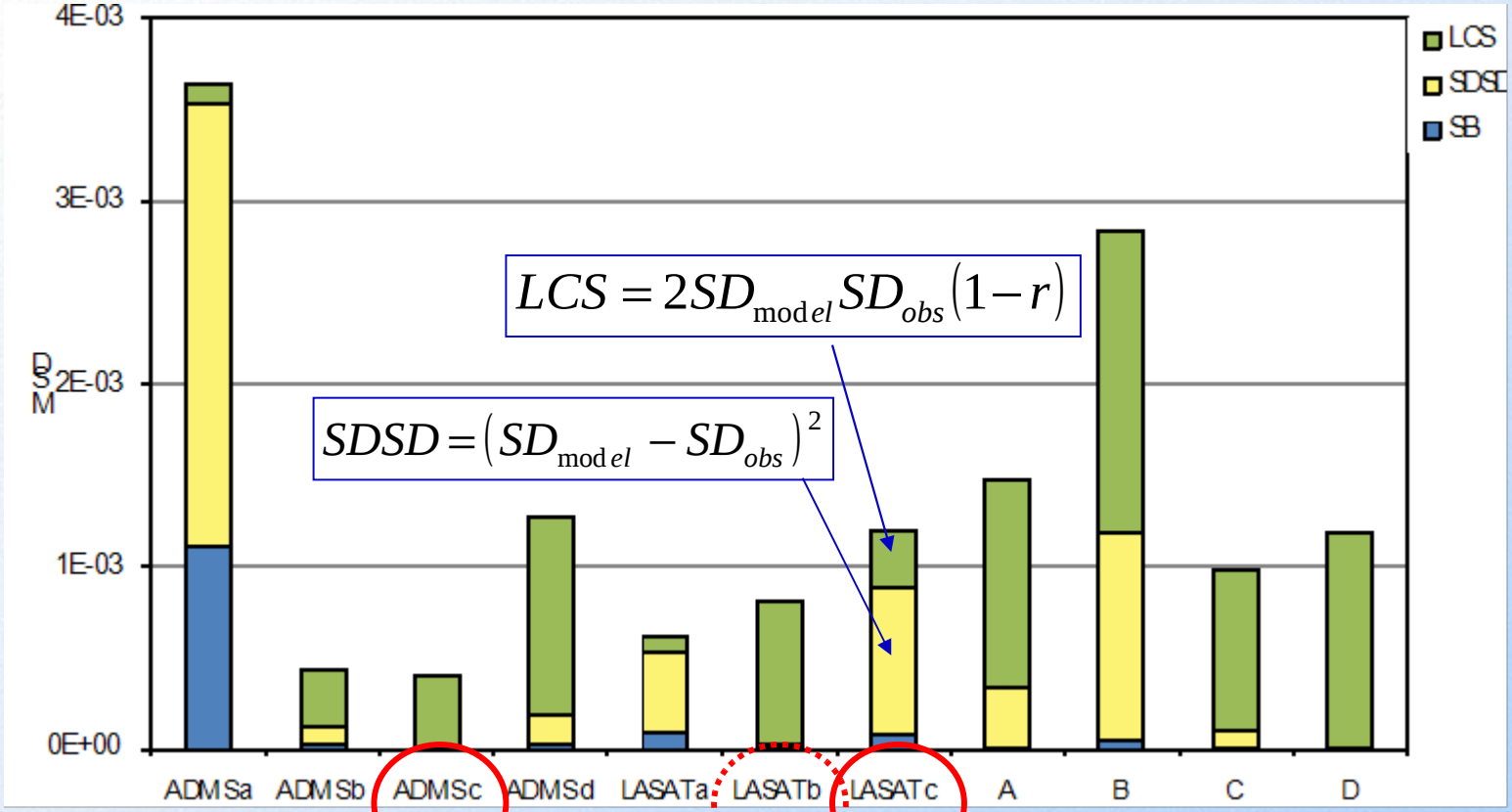


ADMS (Gaussian)

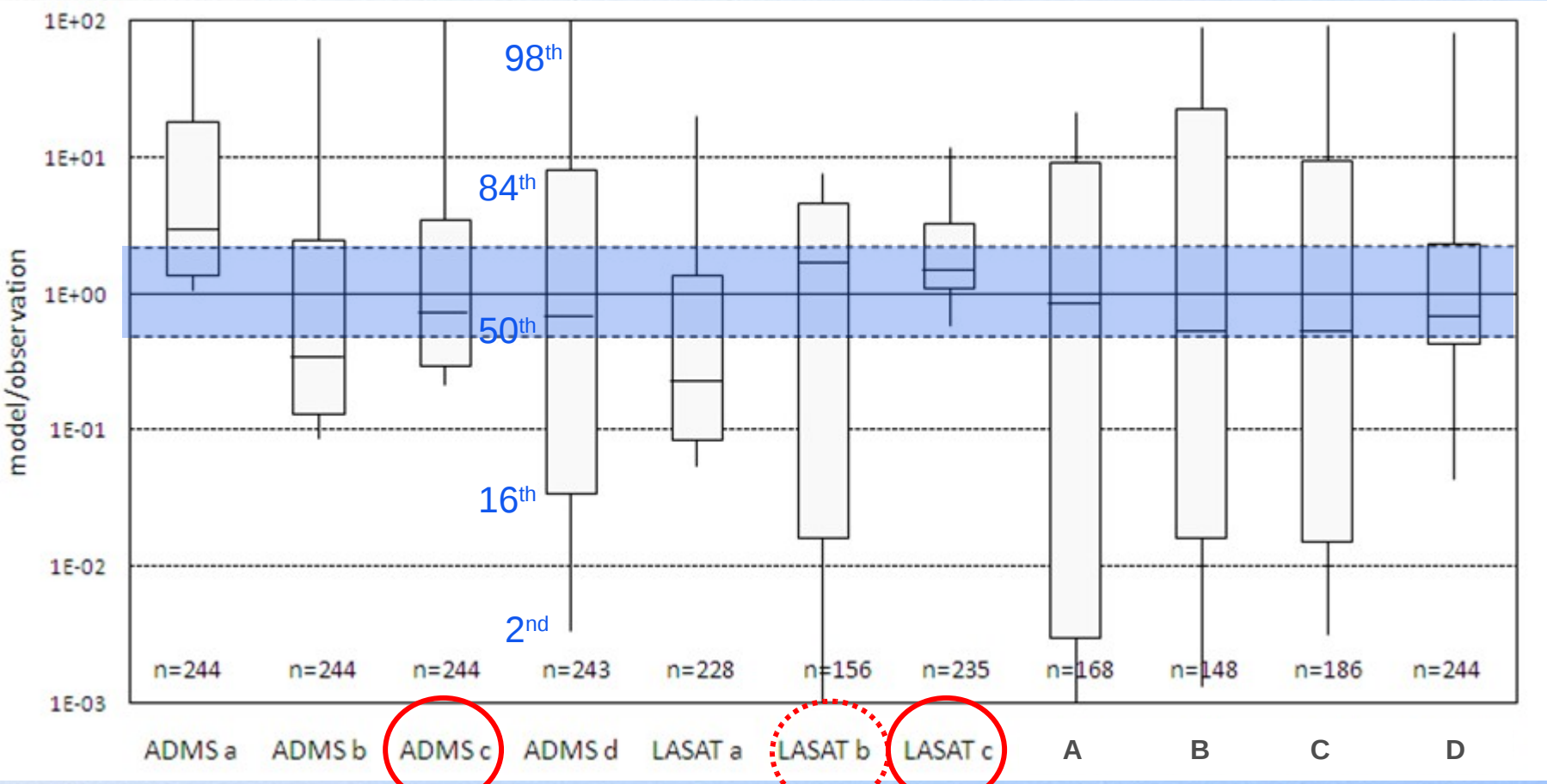
LASAT (Lagrangian)
+ diagnostic wind field

LASAT (Lagrangian)
+ MISKAM wind field

MUST: mean squared deviation



MUST: Residual plot (percentiles)



Gaussian

Lagrangian
+ diagn. w.f.

Lagrangian
+ MISKAM wind field

MUST: metrics

	ADMS a	ADMS b	ADMS c	ADMS d	LASAT a	LASAT b	LASAT c	A	B	C	D
MG	1.3	0.3	1.1	0.9	2.0	1.4	0.7	1.4	1.0	1.1	1.2
VG	5.2	12.0	2.7	6.1	3.0	4.8	1.6	6.6	16.1	6.4	2.9
FB	-1.1	-1.2	0.03	-0.3	1.2	0.2	-0.7	0.2	-0.5	-0.03	0.1
NMSE	100.8	14.0	2.6	5.7	14.3	6.0	17.5	10.7	10.6	5.6	8.0

MG … geometric mean
VG … geometric variance
FB … fractional bias
NMSE … normalized mean square error

Which model run performed “best”?

$F = F_{MSD}.F_{Res}.F_{MG}.F_{VG}.F_{FB}.F_{NMSE}$

highest score - ADMS c (F=2.4)
marginally good - LASAT c (F=0.5)
poor - LASAT b (F<0.2)

$$F \approx \tanh\left(\frac{p}{q}\right)^4$$

Conclusions

- Model validation according to
  “Guidance and Documentation Protocol of COST 732”
- Model performance may vary widely for same model
  with different set-up (parametrisations, input parameters)
- Different validation measures deliver different results
- Combination of results of all validation steps using the approach proposed by Sornette et al. (2007) found to be helpful to assess which model run performed “best”.