Using the SAL technique for ensemble forecast verification

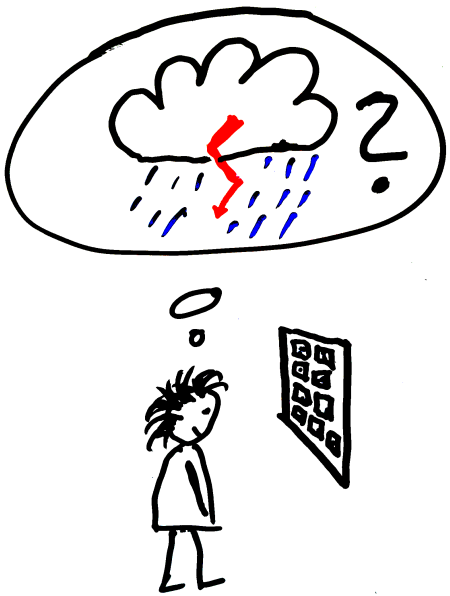
Sabine Radanovics

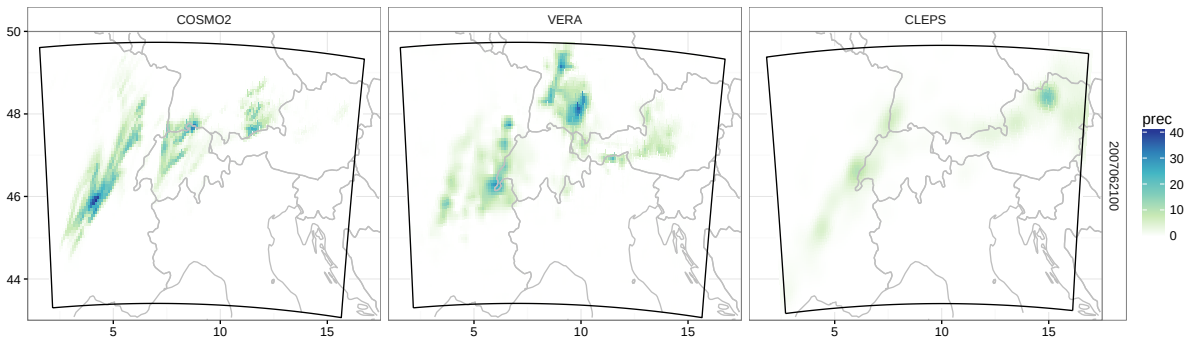
Until March 2017: LSCE/ESTIMR, now: CNRM/GMME/MICADO

May 10, 2017

Why was spatial verification invented?

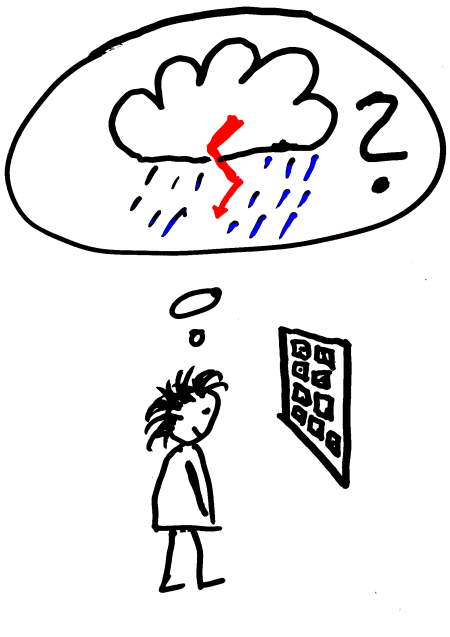
Why was spatial verification invented?

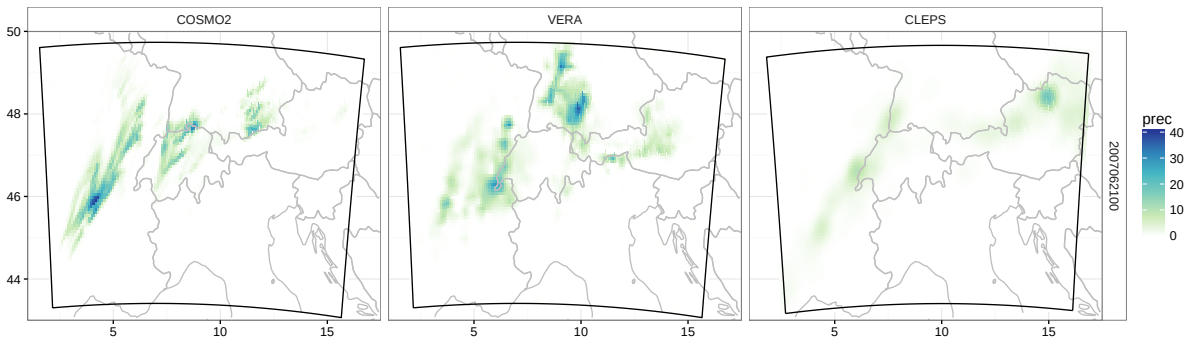




Forecaster Paul

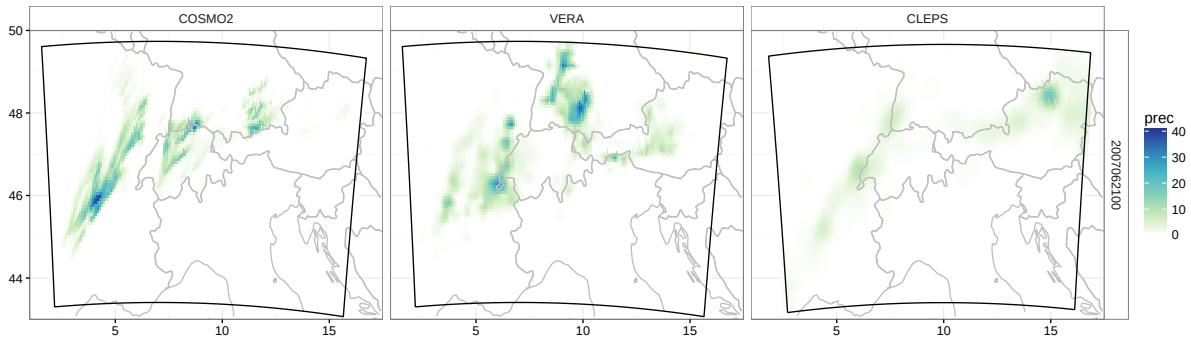
Why was spatial verification invented?





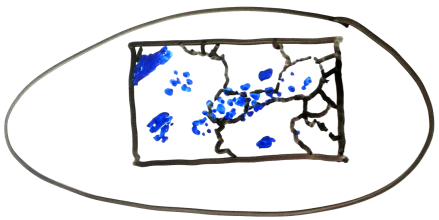
Forecaster Paul

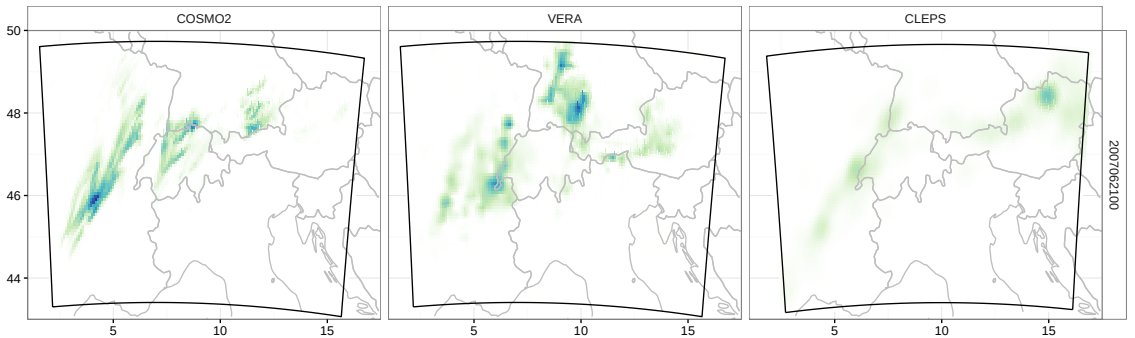
Why was spatial verification invented?



Model developer
Cate

Why was spatial verification invented?



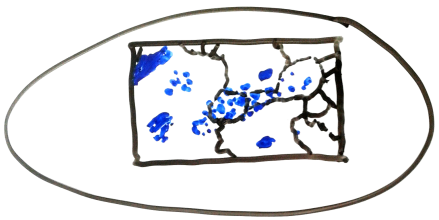


RMSE = 452.02

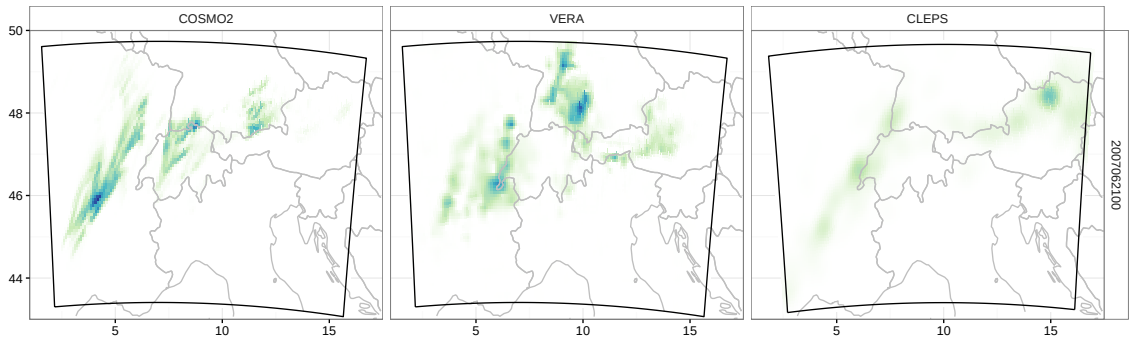
RMSE = 352.03

Model developer
Cate

Why was spatial verification invented?



Model developer
Cate



RMSE = 452.02

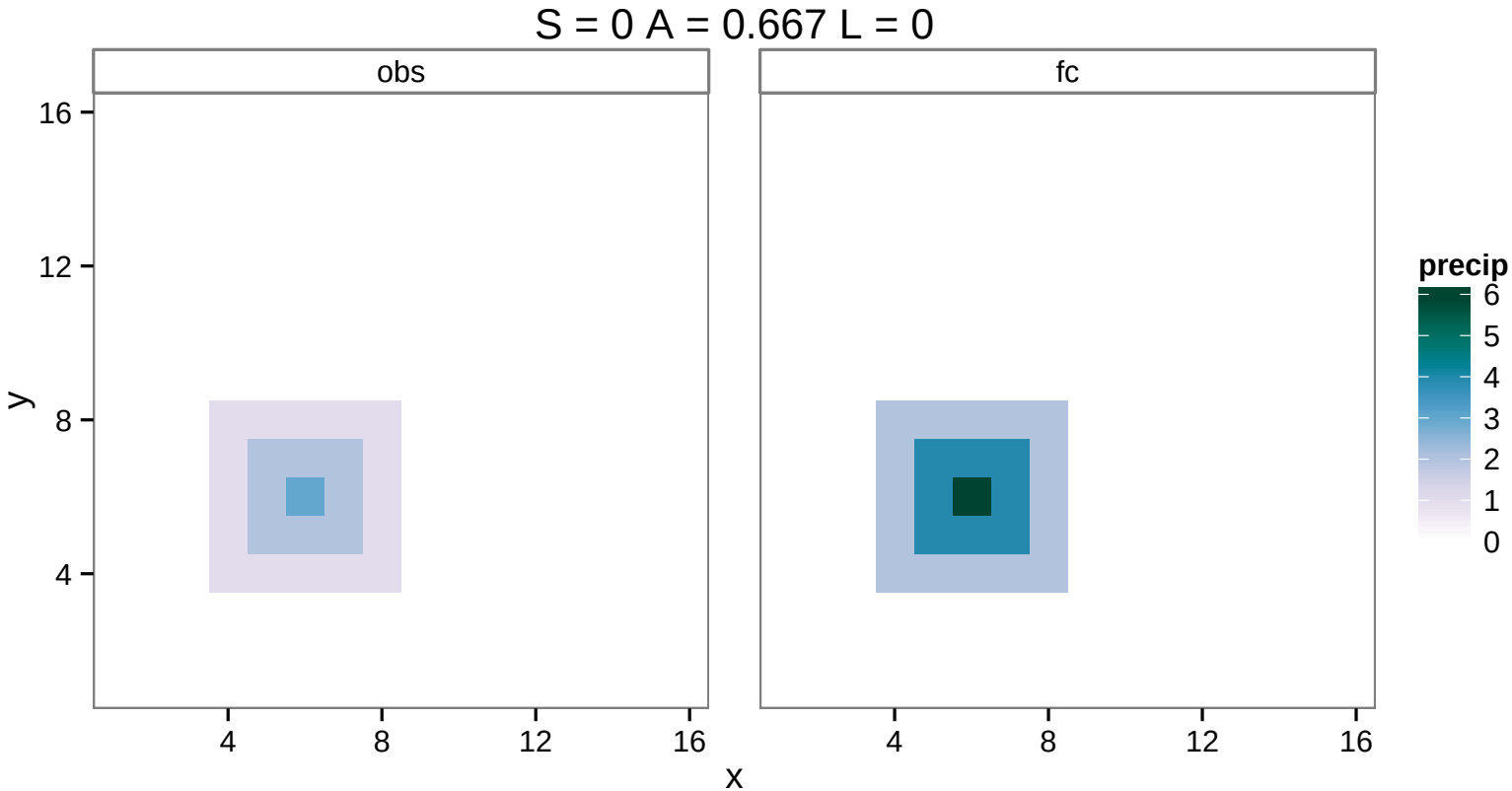
RMSE = 352.03

→ Need to develop verification that gives credit to the attributes Paul appreciates

SAL - Structure, Amplitude, Location

(Wernli et al., 2008)

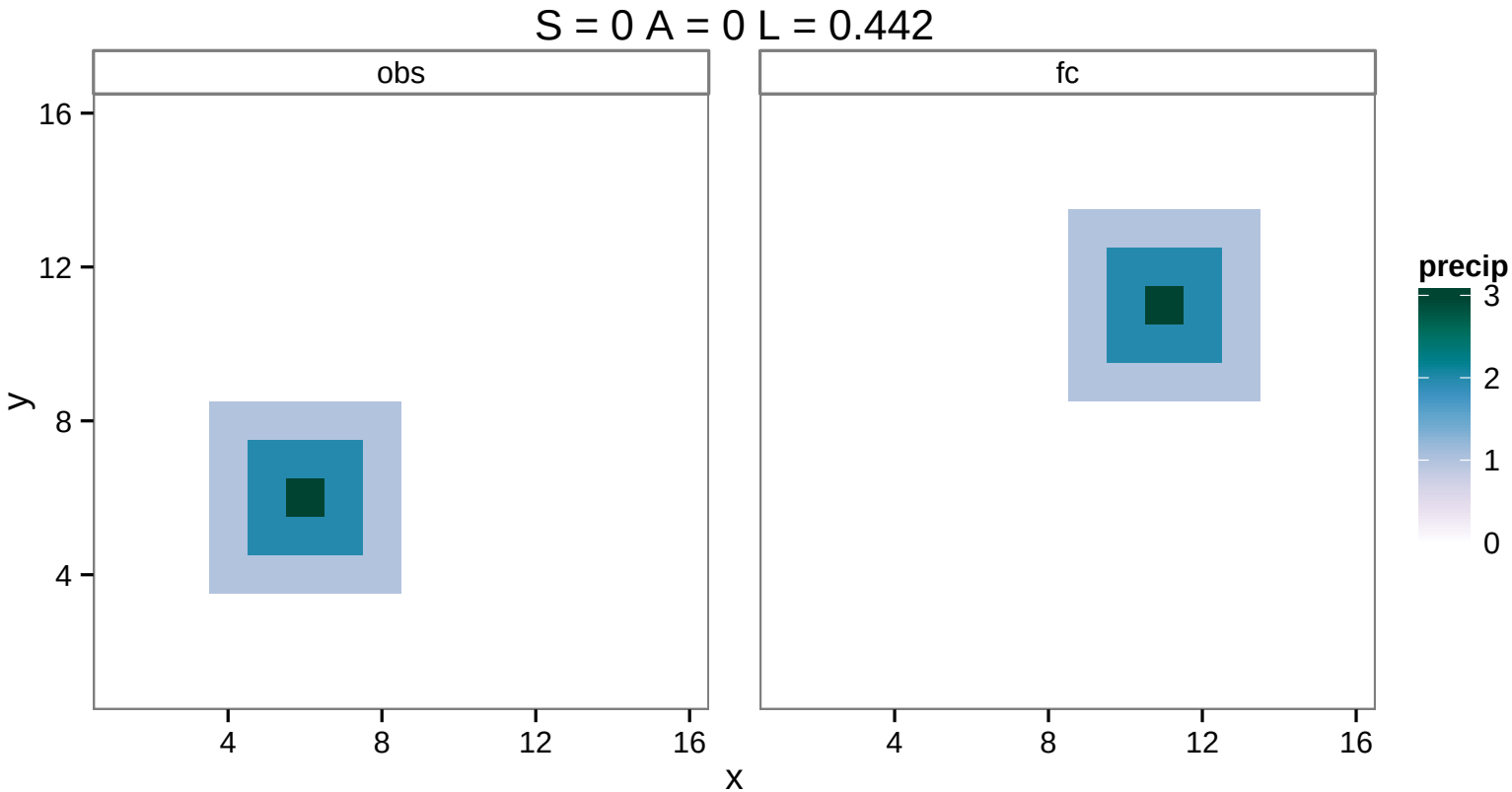
Amplitude component A



- Doubled precipitation
  $\rightarrow$ relative Amplitude error A

SAL - Structure, Amplitude, Location

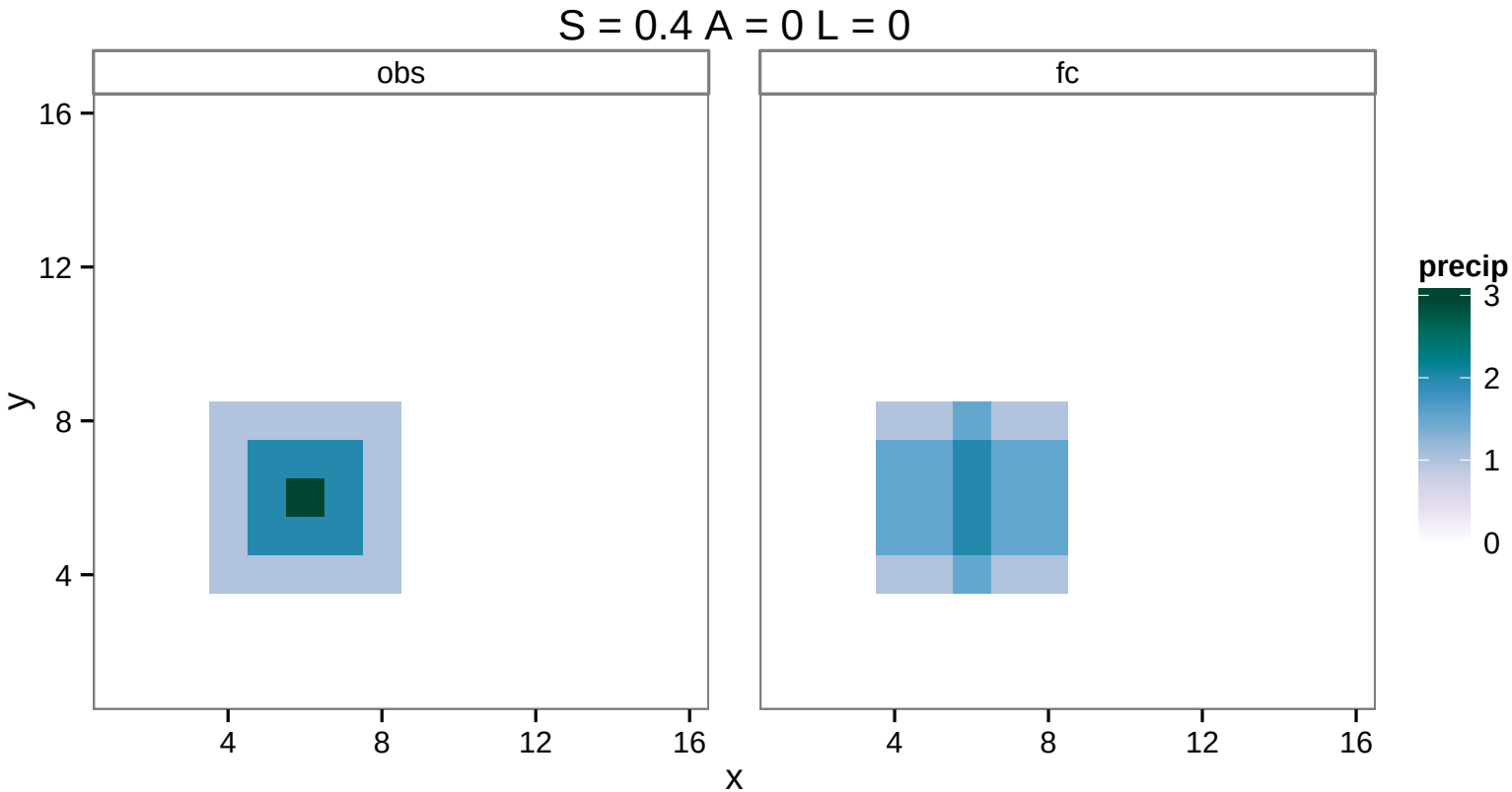
Location component L



- Shifted feature
  $\rightarrow$ location error of center of mass relative to domain size L

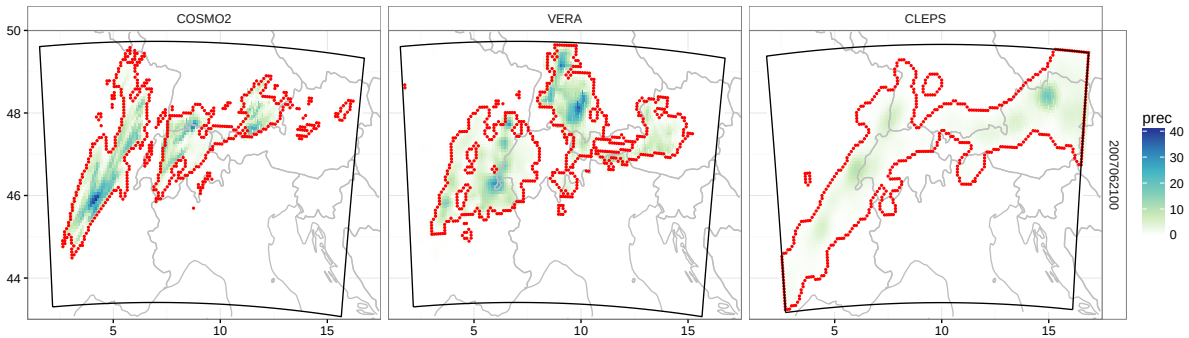
SAL - Structure, Amplitude, Location

Structure component S



- Flatten feature
  $\rightarrow$ structure error in terms of scaled feature volumes S

SAL verification



RMSE = 452.02
S = -0.32
A = -0.22
L = 0.11

RMSE = 352.03
S = 0.55
A = -0.69
L = 0.38

Questions

1. Ensemble verification with SAL?
2. What is the effect of observation uncertainty?

Ensemble verification with SAL?

Ensemble verification with SAL?

Option 1

- Calculate the scores for every ensemble member combination
  $\rightarrow$ study spread of scores

Ensemble verification with SAL?

Option 1

- Calculate the scores for every ensemble member combination
  $\rightarrow$ study spread of scores

Option 2

- Calculate one score for the whole ensemble
  $\rightarrow$ allows comparison with deterministic model

Ensemble SAL

Standard SAL	Ensemble SAL
$A = \frac{\overline{rr_{mod}} - \overline{rr_{obs}}}{0.5[\overline{rr_{mod}} + \overline{rr_{obs}}]}$	$A = \frac{\langle\overline{rr_{mod}}\rangle - \langle\overline{rr_{obs}}\rangle}{0.5[\langle\overline{rr_{mod}}\rangle + \langle\overline{rr_{obs}}\rangle]}$
$L_1 = \frac{\lvert\mathbf{x}(rr_{mod}) - \mathbf{x}(rr_{obs})\rvert}{d}$	$L_1 = \frac{\lvert\langle\mathbf{x}(rr_{mod})\rangle - \langle\mathbf{x}(rr_{obs})\rangle\rvert}{d}$
$L_2 = 2\left[\lvert\frac{\mathbf{r}(mod)}{d} - \frac{\mathbf{r}(obs)}{d}\rvert\right]$	$L_2 = 2 \times CRPS\left(P\left(\frac{\mathbf{r}(mod)}{d}\right), P\left(\frac{\mathbf{r}(obs)}{d}\right)\right)$
$\mathbf{r} = \frac{\sum_i rr_i \lvert\mathbf{x}_i - \mathbf{x}\rvert}{\sum_i rr_i}$	
$S = \frac{V(mod) - V(obs)}{0.5[V(mod) + V(obs)]}$	$S = \frac{\langle V(mod)\rangle - \langle V(obs)\rangle}{0.5[\langle V(mod)\rangle + \langle V(obs)\rangle]}$
$V = \frac{\sum_i rr_i \frac{rr_i}{rr_i^{max}}}{\sum_i rr_i}$	

$\overline{rr}$ domain average precipitation.

$\mathbf{x}$ center of mass of the precipitation field

d largest distance between two domain borders.

$\mathbf{x}_i$ the center of mass of the i^{th} feature

rr_i is the sum of precipitation over all grid cells in feature i

rr_i^{max} maximum precipitation of the i^{th} feature.

MesoVICT project core case 20-22 June 2007

(Mesoscale Verification In Complex Terrain)

3-hourly precipitation

Ensembles

- CLEPS ensemble forecast (16 members) (Marsigli et al., 2005)
- VERA analysis ensemble (50 members) [VERAens]
  (Gorgas and Dorninger, 2012)

Deterministic

- COSMO2 forecasts
  (Ament and Arpagaus, 2009)
- VERA analysis
  (Steinacker et al., 2000)

MesoVICT project core case 20-22 June 2007

(Mesoscale Verification In Complex Terrain)

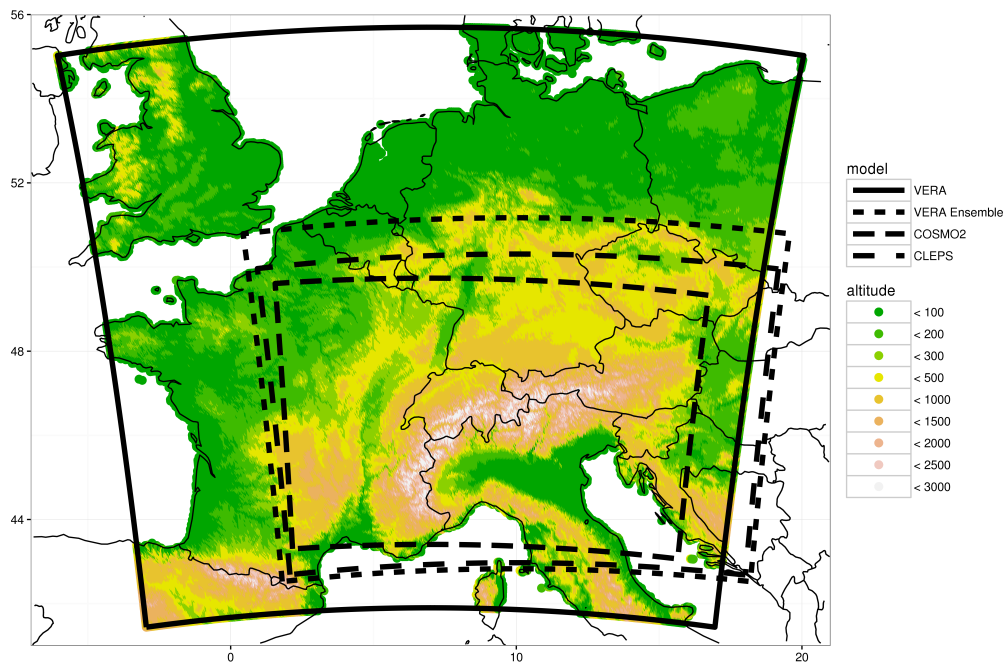
3-hourly precipitation

Ensembles

- CLEPS ensemble forecast (16 members) (Marsigli et al., 2005)
- VERA analysis ensemble (50 members) [VERAens]
  (Gorgas and Dorninger, 2012)

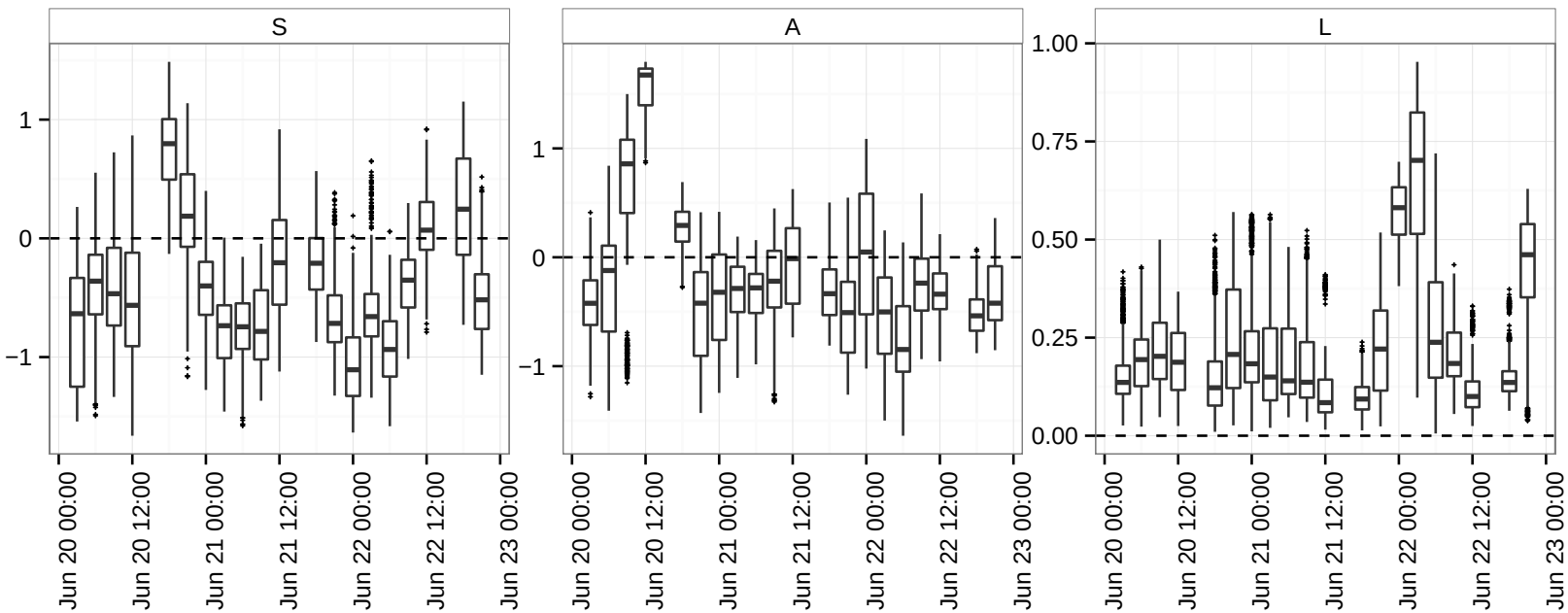
Deterministic

- COSMO2 forecasts
  (Ament and Arpagaus, 2009)
- VERA analysis
  (Steinacker et al., 2000)



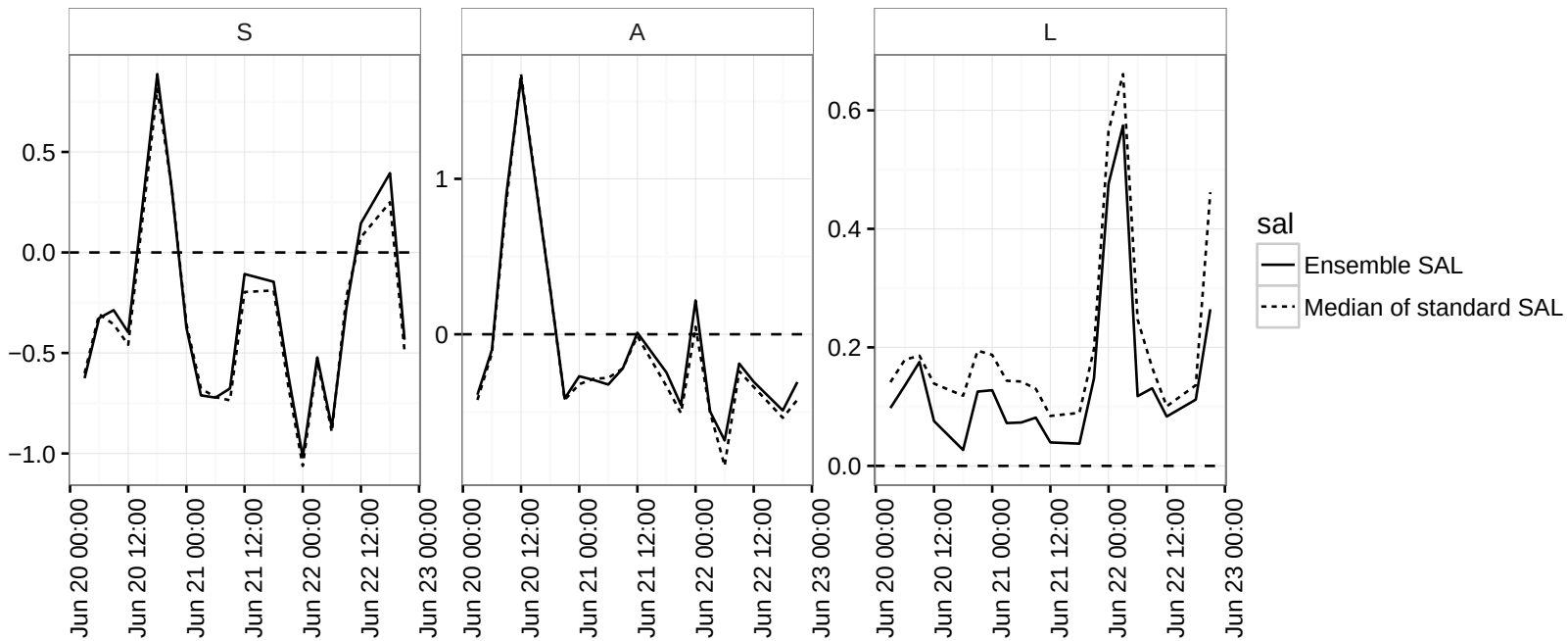
CLEPS-VERAens 16×50 SAL scores

- 3-hourly precipitation, 20-22 June 2007
- 3h-24h lead times (depending on the hour of the day)



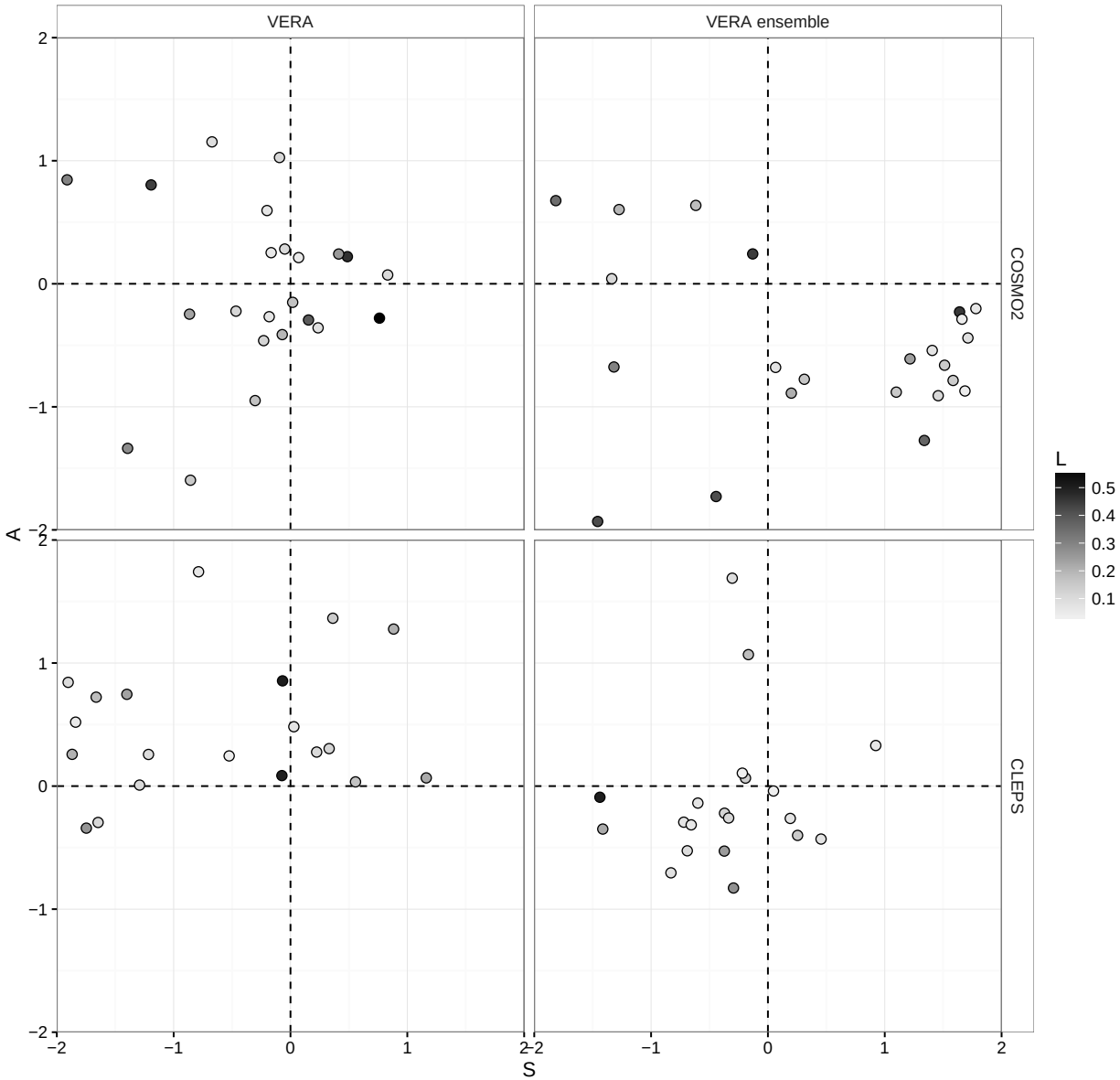
- large spreads in scores, especially S

CLEPS-VERAens Ensemble SAL vs. median of 800 x standard SAL

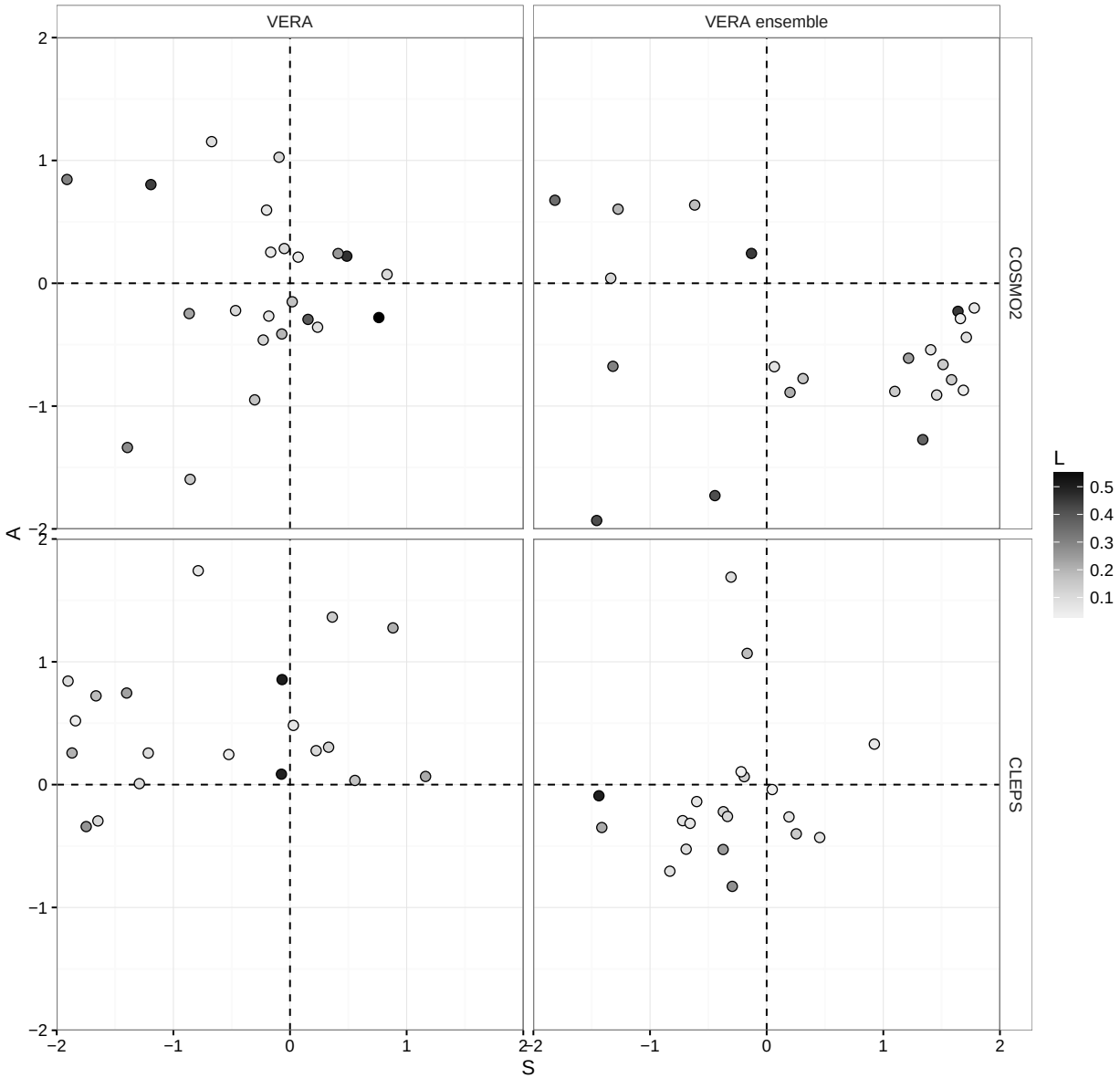


- Similar for 2-sided scores A and S
- Ensemble L tends to be smaller than median of standard L

Compare models - SAL diagrams

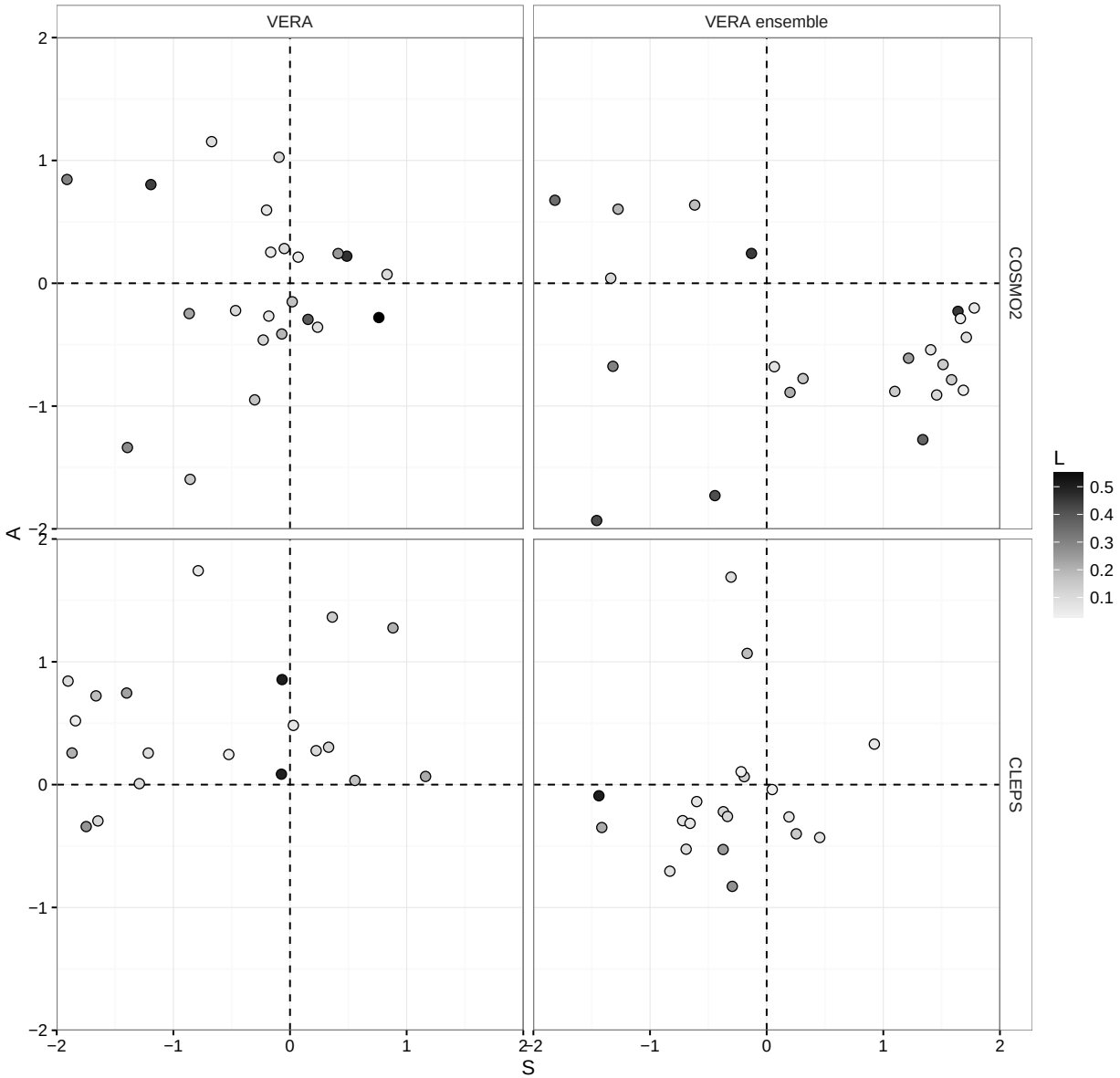


Compare models - SAL diagrams



- difference between VERA and VERAens bigger than difference between forecasts

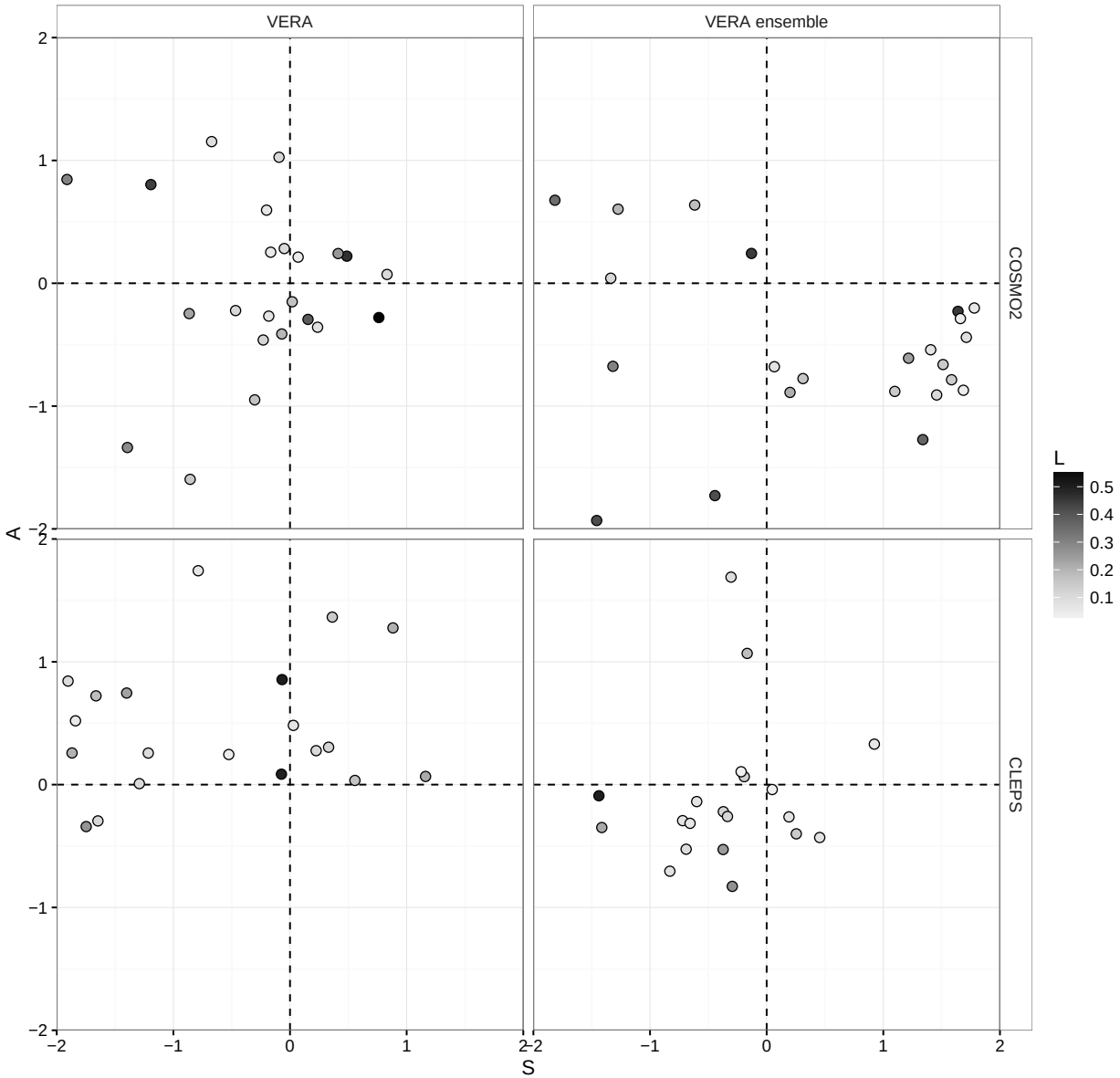
Compare models - SAL diagrams



- difference between VERA and VERAens bigger than difference between forecasts

Why?

Compare models - SAL diagrams

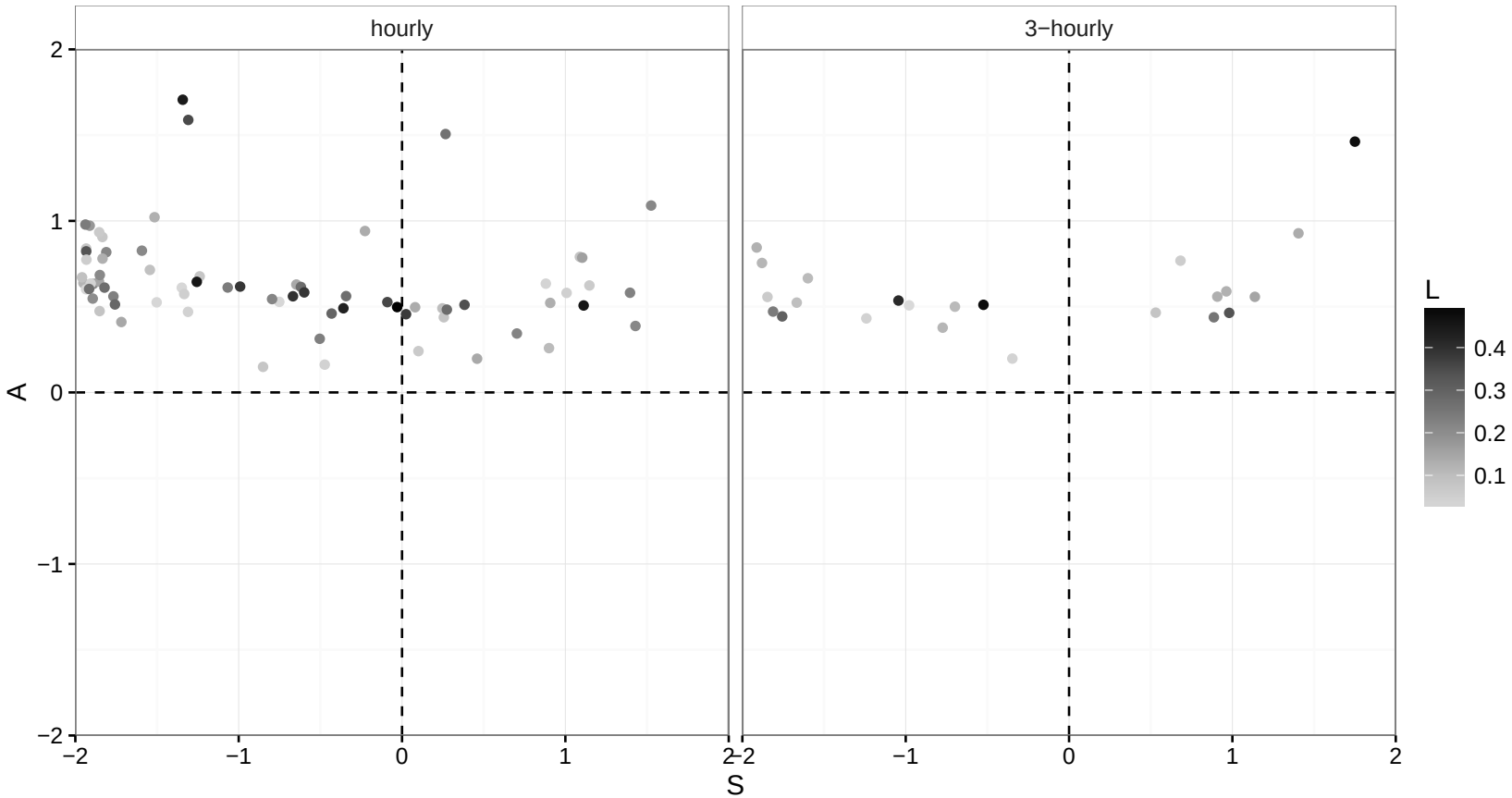


- difference between VERA and VERAens bigger than difference between forecasts

Why?

- Forecasts very similar due to short lead times
- VERAens very different from VERA

VERAens vs. VERA



- Total precipitation A systematically higher in VERAens
- S spans the whole range of values, majority negative S for hourly
- large values of L, same order of magnitude as forecasts

What is known for VERAens

- Gorgas and Dorninger (2012) found that
  1. the quality control procedure to remove unrealistic values like from the perturbed analysis fields tends to remove more of the negative perturbations than from the positive ones $\rightarrow$ high bias in the VERA ensemble.
  2. the standard deviations are on average increased due to the perturbations, but this varies strongly with the ensemble member
- Increased standard deviation $\rightarrow$ smaller/more peaked objects $\rightarrow$ negative S dominates.
- What if the standard deviation due to the perturbations does not only vary with the ensemble member but also over different time steps? This would partly explain the wide range of S values.

Conclusions

- The behaviour of the ensemble SAL seems comparable to the standard SAL
  + less computation than 800 standard SAL
  – loose information on spread
- SAL is very sensitive to small differences in the analysis (or the forecasts) (see also Weniger and Friederichs (2016))
  → may be useful to highlight subtle differences and as a rather qualitative diagnostic in model development
  → probably less useful for ranking models
- There might be an issue with a too large verification domain for L and S.
  → domain size sensitivity experiments needed, in particular with complex terrain

Thank you!

References I

Ament, F. and Arpagaus, M. (2009). dphase_cosmoch2: Cosmo model forecasts (2.2 km) run by meteoswiss for the map d-phase project.

Gorgas, T. and Dorninger, M. (2012). Concepts for a pattern-oriented analysis ensemble based on observational uncertainties. *Quarterly Journal of the Royal Meteorological Society*, 138(664):769–784.

Marsigli, C., Boccanera, F., Montani, A., and Paccagnella, T. (2005). The cosmo-leps mesoscale ensemble system: validation of the methodology and verification. *Nonlinear Processes in Geophysics*, 12(4):527–536.

Steinacker, R., Häberli, C., and Pöttschacher, W. (2000). A transparent method for the analysis and quality evaluation of irregularly distributed and noisy observational data. *Monthly Weather Review*, 128(7):2303–2316.

Weniger, M. and Friederichs, P. (2016). Using the sal technique for spatial verification of cloud processes: A sensitivity analysis. *Journal of Applied Meteorology and Climatology*, 55(9):2091–2108.

Wernli, H., Paulat, M., Hagen, M., and Frei, C. (2008). SAL–a novel quality measure for the verification of quantitative precipitation forecasts. *Monthly Weather Review*, 136(11):4470–4487.