Stochastic assimilation methods and validation : a full probabilistic approach.

Guillem Candille[1], *J-M. Brankart, P. Brasseur, E. Cosme, S. Metref and J. Verron.*

November 27 2012

SANGOMA

[1]MEOM team, LEGI, CNRS / Grenoble-University

Outline

1. Motivation & context
2. Probabilistic validation
3. Probabilistic scores (univariate)
4. Model benchmarks & metrics

Outline

Ensemble context

SANGOMA purposes :

- Development of advanced stochastic assimilation methods dealing with strongly non-linear and non-gaussian phenomena.
- Provide an uncertainty estimation associated with the analysis process.

Full ensemble analysis schemes :

- Evolution in time of the covariance errors.
- Consider the ensemble (PDF) as a whole
$\longrightarrow$ probabilistic validation.

Beyond the deterministic validation ...

RMS Error :

- $\text{RMSE}^2 = E[(o-m)^2]$
- Deterministic score -negatively oriented- using the ensemble mean as the ensemble estimator.

1st approximation of the ensemble quality ...

Spread Reduction Factor (SRF, Sakov *et al* 2012) :

- $\text{SRF}=\left(\frac{tr(HP^f H^T R^{-1})}{tr(HP^a H^T R^{-1})}\right)^{\frac{1}{2}} - 1$
- SRF=0 $\rightarrow$ no change, SRF=1 $\rightarrow$ uncertainty reduction by 2.

1st approximation of the uncertainty reduction ...
but no information about the consistency with the real errors.

Beyond the deterministic validation ...

RMS Error :

- $\text{RMSE}^2 = E[(o-m)^2]$
- Deterministic score -negatively oriented- using the ensemble mean as the ensemble estimator.

1st approximation of the ensemble quality ...

Spread Reduction Factor (SRF, Sakov *et al* 2012) :

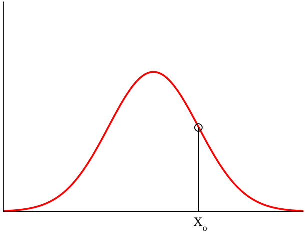
- SRF$=\left(\frac{tr(HP^f H^T R^{-1})}{tr(HP^a H^T R^{-1})}\right)^{\frac{1}{2}} - 1$
- SRF=0 $\rightarrow$ no change, SRF=1 $\rightarrow$ uncertainty reduction by 2.

1st approximation of the uncertainty reduction ...
but no information about the consistency with the real errors.

Outline

How to evaluate an ensemble ?

- 'Forget' the deterministic concepts of validation.

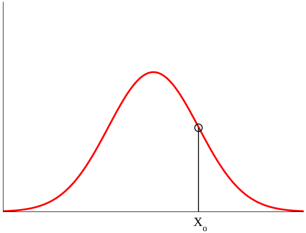


- Ensemble validation by **statistical accumulation**.
  ($\longrightarrow$ the ensemble system is highly reproducible)

- Probabilistic criteria :
  - **reliability**, statistical consistency.
  - **resolution** or **sharpness**, statistical variability.

How to evaluate an ensemble ?

- ‘Forget’ the deterministic concepts of validation.

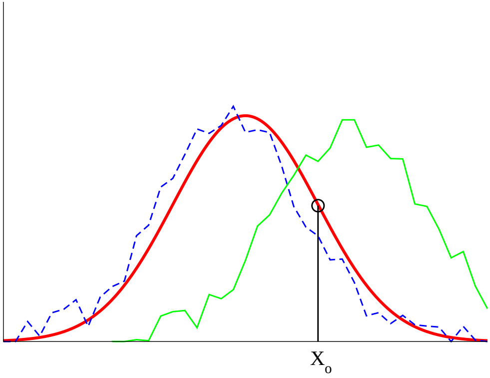


- Ensemble validation by **statistical accumulation**.
  (→ the ensemble system is highly reproducible)

- Probabilistic criteria :
  - **reliability**, statistical consistency.
  - **resolution** or **sharpness**, statistical variability.

Reliability

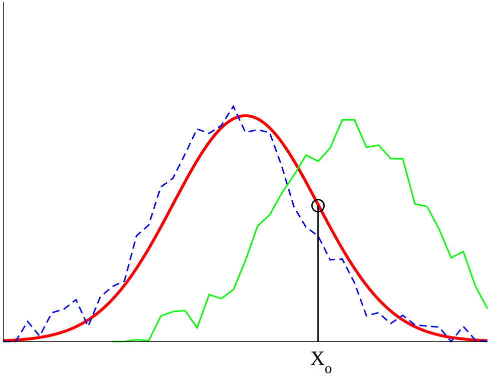
- Statistical consistency between the produced ensembles and the corresponding verifications.



- Produced PDF f.
- f_1' and f_2' : 2 distributions of x_o when f is produced.
- A system is perfectly reliable if and only if $f = f'$ for all f.

Reliability

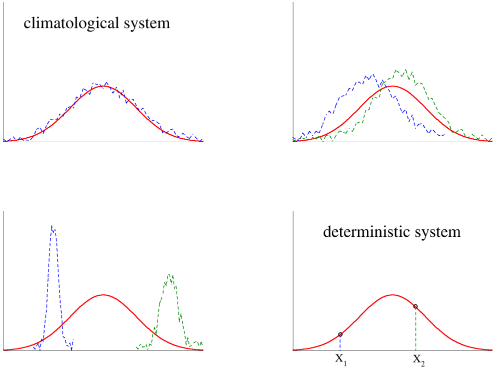
- Statistical consistency between the produced ensembles and the corresponding verifications.



- Produced PDF f.
- f_1' and f_2' : 2 distributions of x_o when f is produced.
- A system is perfectly reliable if and only if $f = f'$ for all f.

Resolution

- Ability of the ensemble system to separate the produced PDF leading to sufficiently distinct corresponding observed distributions (COD).



- COD examples from a minimal resolution (null) for a climatological system to a maximal resolution for a perfect deterministic system.

(*nb* : the curve represents the climatological distribution)

Probabilistic criteria : summary

- *Reliability* and *resolution* are 2 independent properties, necessary and sufficient in order to evaluate the intrinsic quality and the usefulness of an ensemble system.

- First, an ensemble system must be reliable, but also must be able to *a priori* separate the produced PDF into sufficiently various classes so the corresponding observations represent sufficiently distinct situations.

Outline

Reliability scores

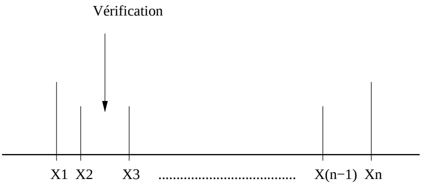
- Ensemble $\longrightarrow$ N independent realizations from a PDF.
- Reliability : statistical consistency between the produced ensembles and the observed verifications.
  $\longrightarrow$ Is the verification a N+1-st realizations of the PDF defined by the N members of the ensemble ?
- Scores :
  - Rank histogram.
  - Reduced Centered Random Variable (RCRV).

Reliability scores

- Ensemble $\longrightarrow$ N independent realizations from a PDF.
- Reliability : statistical consistency between the produced ensembles and the observed verifications.
  $\longrightarrow$ Is the verification a N+1-st realizations of the PDF defined by the N members of the ensemble ?
- Scores :
  - Rank histogram.
  - Reduced Centered Random Variable (RCRV).

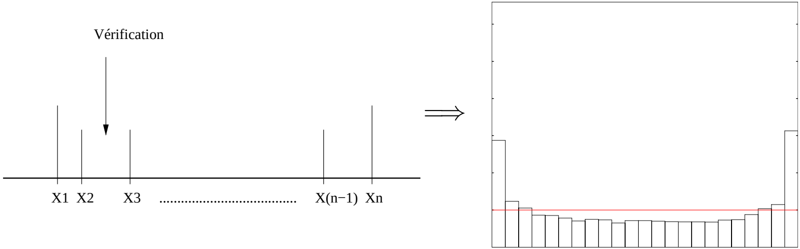
Rank histogram

- Partial order between the N members of the ensemble and the verification.



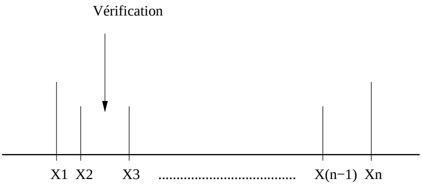
Rank histogram

- Partial order between the N members of the ensemble and the verification.

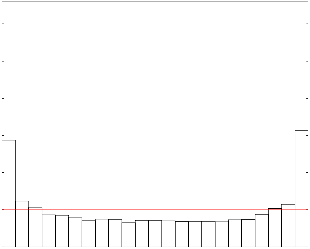


Rank histogram

- Partial order between the N members of the ensemble and the verification.



$\Longrightarrow$



- The verification is statistically *indistinguishable* from the N ensemble values $\longrightarrow$ equally distributed over the N+1 intervals.
- The rank histogram *flatness* is a measure of the ensemble reliability.
- Deviation from the flatness : $\delta = \frac{N+1}{MN} \sum_{k=1}^{M} \left(s_k - \frac{M}{N+1} \right)^2$.
  Reliable system : $\delta = 1$.

Reduced Centered Random Variable (RCRV)

Are the ensemble members and the verification indistinguishable ?

- Decompose the reliability into bias (b) and dispersion (d).
- RCRV :

$$y = \frac{o - m}{\sigma}$$

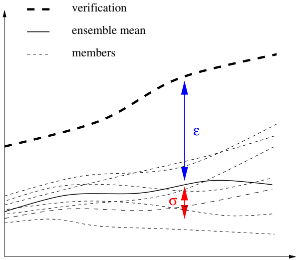
  - $b = E[y]$ measures the weighted bias of the system.
  - $d^2 = E[y^2] - b^2$ measures the agreement between the ensemble spread and the analysis error of the ensemble mean.
  - Reliable system : $b = 0$ and $d = 1$.

- Remarks :
  - Observational error can be intruduced : $\sigma = \sqrt{\sigma_e^2 + \sigma_o^2}$.
  - $\mathrm{RMSE}^2 \approx E[\sigma^2](d^2 + b^2)$

How to improve the ensemble system reliability ?

- Non-reliable ensemble system ...

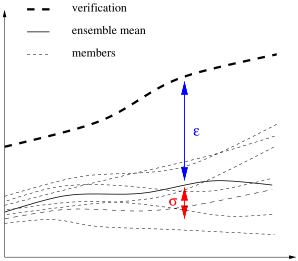
$$d \equiv Var\left(\frac{\varepsilon}{\sigma}\right) > 1$$



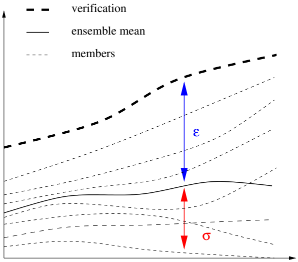
How to improve the ensemble system reliability ?

- Non-reliable ensemble system ... and 2 conceivable corrections.

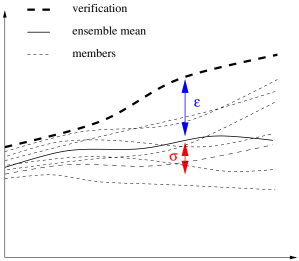
$d \equiv Var\left(\frac{\varepsilon}{\sigma}\right) > 1$



increase the spread



reduce mean the error



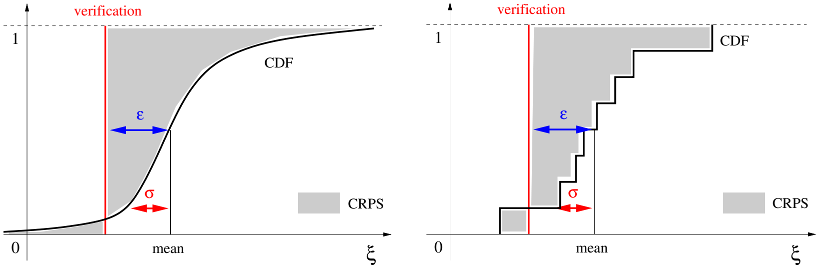
Continuous Ranked Probability Score (CRPS)

- CRPS measures the global quality of an ensemble system :

$$CRPS = E\left[\int_{\Omega} (F_p(\xi) - H(\xi - x_o))^2 \, d\xi\right]$$

F_p is the cumulative density function (CDF) associated with the produced ensemble.

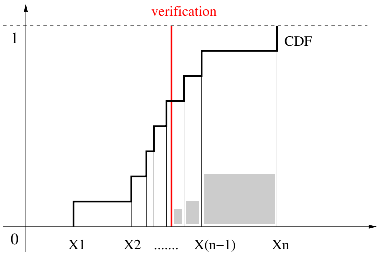
continuous case (theory) discrete case (reality)



- Decomposition (Hersbach 2000) : CRPS = $Reli$ + CRPS_{pot}.

Reli

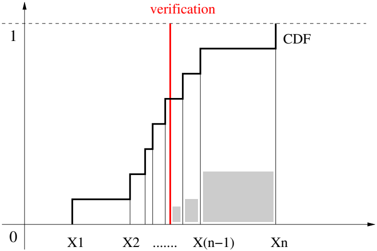
- Coefficients are defined for each $[x_i, x_{i+1}]$ depending on the verification position and the interval size.



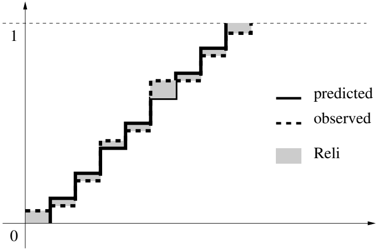
Reli

- Coefficients are defined for each $[x_i, x_{i+1}]$ depending on the verification position and the interval size.

- Build the COD and compare to the mean of the CDF produced by the system.



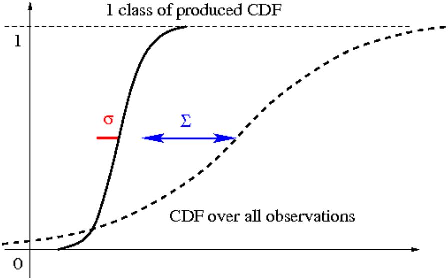
$\Longrightarrow$



CRPS_{pot}

- CRPS_{pot} is the potential value of the CRPS when the ensemble system is reliable, *i.e.* $Reli = 0$.

- CRPS_{pot} = uncertainty - resolution



- $CRPS_{pot} \propto \mathcal{F}(\sigma)$
- uncertainty $\propto \mathcal{F}(\Sigma)$
- resolution $\propto \mathcal{F}(\Sigma - \sigma)$

- The more $\sigma << \Sigma$, better the resolution is.

Summary of the probabilistic tools

bias : $b = E[y]$

Reli (reliability) $\equiv y = \frac{o-m}{\sigma} \approx$ rank histogram

dispersion : $d^2 = E[y^2] - b^2$

$$\text{CRPS} = E\left[\int_\Omega \left(F_p(\xi) - F_o(\xi)\right)^2 d\xi\right]$$

(global score)

$CRPS_{pot}$ (resolution)

- Remark : resampling methods (bootstrap) can be applied in order to assess the statistical uncertainty on the diagnoses due to the limited size of the verification dataset (Candille *et al* 2010).

Outline

Benchmarks

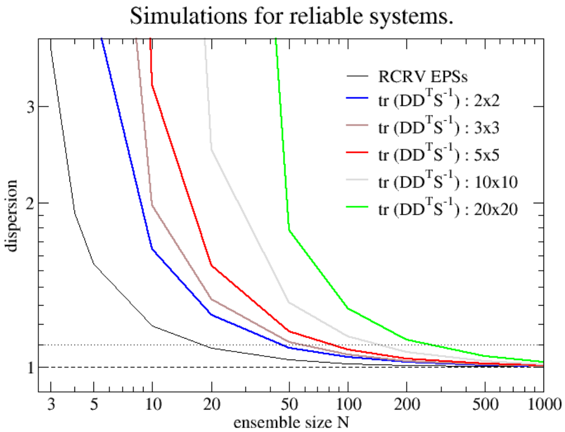
- Small benchmark (L96) : small size model, idealized assimilation problem with no model error, relaxation of the observations.
  $\longrightarrow$ highly reproducible system : metrics with no approximation considering full mathematical generality (multivariate), no restriction on the numerical cost.

- Medium benchmark (SQB) : same as L96 but bigger size model, not all state variable are observed (SSH + some vertical profiles for temperature), relaxation of the observations by simulating satellite traces.
  $\longrightarrow$ reproducible system : approximation on the metrics, numerical efficiency starts to become an issue.

- Large benchmark (NATL025) : much larger size model, real-world observation data, various sources of model errors.
  $\longrightarrow$ hardly reproducible system : restrictions on the validation (univariate), need an independent observation dataset, assumptions on the model errors.

Multivariate issue

RCRV multivariate extension : $M = DD^T S^{-1}$
Reliable system : $E[M] = \mathbf{I}_L$ and $\frac{1}{L} tr(E[M]) = 1$.



Prospective issue

Main questions :

- Consistency between the prior PDF and the PDF estimated by the models ? Investigate the ensemble sample size effect on the full PDF (L96, SQB ?) or on the marginal distributions (NATL025).
  $\longrightarrow$ reliability.
- Are the full (L96) or marginal (SQB, NATL025) posterior distributions consistent with the real errors (L96, SQB) or *independent* observations (NATL025) ? Is there a difference between observed and non-observed data (SQB) ?
  $\longrightarrow$ reliability (+ resolution).
- What is the uncertainty related to the posterior distribution ?
  $\longrightarrow$ resolution.

Bibliography

- Candille G., S. Beauregard, and N. Gagnon. 2010. Bias correction and multiensemble in the NAEFS Context or How to get a 'free calibration' through a multiensemble approach. *Mon. Wea. Rev.*, **138**, pp 4268–4281.

- Hersbach H. 2000. Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weatther and Forecasting*, **15**, pp 559–570.

- Sakov P., F. Counillon, L. Bertino, K.A. Lisæter, P.R. Oke, and A. Korablev. 2012. TOPAZ4 : an ocean-sea ice data assimilation system for the North Atlantic and Artic. *Ocean Sci.*, **8**, pp 633–656.