Representation Errors

The dynamic models used in data assimilation and prediction systems are most likely imperfect. Uncertainties in parameterization and local forcings give rise to model errors, which should be accounted for in data assimilation to achieve optimal state estimate. "Representation" error, or "representativeness" error, is due to model's limited ability to fully represent the physical processes as in nature, causing a mismatch between the climatologies of observations and model states.

- unresolved scales in a coarse-resolution model
- wrong model parameters

1. Methods to account for model errors in background error covariance. Lack of resolved small scales shows up as ensemble spread deficiency $\rightarrow$ Inflate the ensemble with additive noises.

- use simplified stochastic process, such as a Wiener process (Brownian motion), to represent unresolved scales.

$$X_{t+1} = m(X_t) + \overbrace{\sigma \dot{W}}^{\varepsilon^m}, \quad \varepsilon^m \sim N(0, Q)$$

the amount of this additive noise, σ, can be "optimally" tuned $\Rightarrow$ "Linear theory for two-scale Lorenz system" (Berry and Harlim 2014)

$\Rightarrow$ use innovation statistics over time to obtain an estimate for σ either online or offline. (Berry and Sauer 2013)

$$\Gamma_0 = \mathbb{E}\left[d_t^{o-b}\left(d_t^{o-b}\right)^T\right] = HP_t^b H^T + R$$

$$\Gamma_1 = \mathbb{E}\left[d_{t+1}^{o-b}\left(d_t^{o-b}\right)^T\right] = HMP_t^b H^T - HMK\Gamma_0$$

- Assuming a stochastic process to account for a missing dynamical process is sometimes too simplified
  - → Stochastically perturbed physics Tendencies (SPPT; Buizza et al. 1999)
  - → Stochastic Kinetic Energy Back scatter (SKEB; Shutts 2005, Berner et al. 2009)
  - → use multi-physics (Meng and Zhang 2007) or multi-core/model ensemble to characterize background error covariance.
  - → Random Transport in advection terms (Resseguier et al. 2016)

  Idea: introduce random errors in model physics so that the unresolved scales and uncertainties in parameters are accounted for. ⇒ stochastic parameterization

- More sophisticated methods:

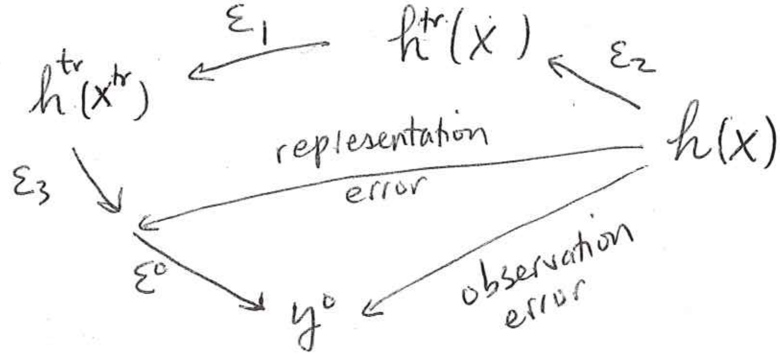
  Actually run the high-resolution model to resolve the small scales, then use the climatology from these runs to obtain model parameters for the reduced-resolution ⇒ "super parameterization"

  Example: - Couple local Large-Eddy Simulations to a mesoscale simulation of hurricane.
  - use cloud resolving simulation on a local periodic domain, and use the cloud statistics in climate model radiative schemes.

- Eventually: enough computational resource to perform cloud-resolving simulations for the whole globe?

2. Methods that adjust observation to fit model climatology:
The mismatch between observation and state climatologies is often handled in terms of observation error:

Janjić et al. 2017: $\quad y^o = h^{tr}(x^{tr}) + \varepsilon_1 + \varepsilon_2 + \varepsilon_3 + \varepsilon^o$



ε^o instrument error

ε_1 errors due to unresolved scales or model physics parameters

ε_2 errors due to observation operator

ε_3 errors due to quality control and pre-processing

$\left.\begin{matrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{matrix}\right\}$ representation errors

$\Rightarrow$ representation error due to unresolved scales (ε_1) introduce spatial, temporal correlation in observation errors; observations appear less smoothed (more details) than background states.

$\rightarrow$ Observation thinning: reduces correlation

Super observation ("superob"): taking average of near-by observations to make observation less noisy.

Hodyss and Nichols 2015 :

Assume distributions of the truth and forecast (background) are separately Gaussian, and there is a linear map between them.

$$x^b = F(x^{tr}) \quad , \quad p(x^b | x^{tr}) = \delta(x^b - F(x^{tr}))$$

The lack of inverse for F leads to errors of representation.

(Liu and Rabier 2002 ; Waller et al. 2013 interpret F as a smoothing operator)

$$p(x^{tr} | x^b) = \frac{p(x^{tr})}{p(x^b)} \delta(x^b - F(x^{tr}))$$

when we know x^b, the probability of x^{tr} is a weighted collection of delta functions located on possible $x^b = F(x^{tr})$ solutions.

observation likelihood $\quad p(y^o | x^{tr}) = \mathcal{N}(Hx^{tr}, R)$

but $\quad p(y^o | x^b) = \mathcal{N}(H\bar{x}_c, R + HP_cH^T)$

where $\quad \bar{x}_c = \int x^{tr} p(x^{tr} | x^b) \, dx^{tr}$

$$P_c = \int (x^{tr} - \bar{x}_c)(x^{tr} - \bar{x}_c)^T p(x^{tr} | x^b) \, dx^{tr}$$

Data assimilation $\quad p(x | y^o, x^b) = p(y^o | x^b) \, p(x^b | x^{tr}) \, p(x^{tr}) / \text{Norm}$

innovation $d = y^o - H^b\bar{x}^b$, there is a bias $\mathbb{E}(d) = H\bar{x}_c - H^b\bar{x}^b$

Modified update eqn :

$$\bar{x}^a = \bar{x}^b + G\Big(\overbrace{d - \mathbb{E}(d)}^{= d'}\Big) \qquad x^{b\prime} = x^b - \bar{x}^b$$

$$G = (P^b H^{bT} + P^{bd})\Big(\underbrace{H^b P^b H^{bT}}_{\mathbb{E}(H^b x^{b\prime} x^{b\prime T} H^{bT})} + \underbrace{H^b P^{bd} + P^{bdT} H^{bT}}_{\mathbb{E}(H^b x^{b\prime} d^{\prime T})} + \underbrace{P^{dd}}_{\mathbb{E}(d' d^{\prime T})} + \bar{R}^b\Big)^{-1}$$

$$\bar{R}^b = R + \int HP_cH^T p(x^b) \, dx^b$$