Spatial AR(1) model for the salinity

For the 2nd test in May

1 Model

We assume a spatially discretized domain of n locations; $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$. Times are indicated by $t = 0, 1 \ldots$, discretized in a sampling time interval. We denote the spatio-temporal salinity variable by $\boldsymbol{\xi}_t = (\xi_{t,\boldsymbol{u}_1}, \ldots, \xi_{t,\boldsymbol{u}_n})^T$, $t = 1, \ldots$.

The prior model for salinity is defined by the initial state:

$$\boldsymbol{\xi}_0 \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{1}$$

with mean vector $\boldsymbol{\mu}$ trained by averaging SINMOD data over time at every location. For the covariance, we assume a Matern correlation function so that $\Sigma(i, i') = \sigma^2(1 + \phi h(i, i')) \exp(-\phi h(i, i'))$, for variance σ^2, correlation decay ϕ and Euclidean distance $h(i, i')$ between sites $\boldsymbol{u}_i$ and $\boldsymbol{u}_{i'}$. Parameters σ and ϕ are specified by variogram plots from SINMOD.

The prior model is further defined by the dynamical system for $t = 1, \ldots$:

$$\boldsymbol{\xi}_t = \boldsymbol{\mu} + \rho(\boldsymbol{\xi}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{v}_t, \qquad \boldsymbol{v}_t \sim N(0, (1 - \rho^2)\boldsymbol{\Sigma}), \tag{2}$$

where the scalar autocorrelation parameter is $\rho < 1$. It is trained from correlations over same-location discretized time steps in SINMOD data.

With this formulation, the marginal distribution at any time becomes

$$\boldsymbol{\xi}_t \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \qquad t \geq 0. \tag{3}$$

This is a relatively simple spatio-temporal model, and we should check if it makes sense for the SINMOD data (look at residuals and so on). Then again, it is not straightforward to model currents, winds and tides etc in this setting without making it much more complex. This would cause challenges in the specification of parameters and a non-linear formulation which cannot be run in real-time.

The measurement y_t at each stage or time $t = 1, \ldots$ is modeled by

$$y_t | \boldsymbol{\xi}_t \sim N(\boldsymbol{f}_t^T \boldsymbol{\xi}_t, r^2), \tag{4}$$

where the vector $\boldsymbol{f}_t$ defines the spatial sampling indices at this stage of operation and r is the salinity measurement noise standard deviation.

2 Updating

Starting with $\boldsymbol{m}_{0|0} = \boldsymbol{\mu}$ and $\boldsymbol{S}_{0|0} = \boldsymbol{\Sigma}$, we use the dynamical model to propagate the state variable mean and covariance, and then apply Bayes' rule to achieve data assimilation at times $t = 1, \ldots$. Denoting the predictive mean and covariance by $\boldsymbol{m}_{t|t-1}$ and $\boldsymbol{S}_{t|t-1}$, the recursive formulae for the updated mean and covariance at time $t = 1, \ldots$ given by

$$\begin{aligned}
\boldsymbol{m}_{t|t-1} &= \boldsymbol{\mu} + \rho(\boldsymbol{m}_{t-1|t-1} - \boldsymbol{\mu}) \\
\boldsymbol{S}_{t|t-1} &= \rho^2 \boldsymbol{S}_{t-1|t-1} + (1-\rho^2)\boldsymbol{\Sigma} \\
\boldsymbol{G}_t &= \boldsymbol{S}_{t|t-1}\boldsymbol{f}_t(\boldsymbol{f}_t^T \boldsymbol{S}_{t|t-1}\boldsymbol{f}_t + r^2)^{-1} \\
\boldsymbol{m}_{t|t} &= \boldsymbol{m}_{t|t-1} + \boldsymbol{G}_t(y_t - \boldsymbol{f}_t^T \boldsymbol{m}_{t|t-1}) \\
\boldsymbol{S}_{t|t} &= \boldsymbol{S}_{t|t-1} - \boldsymbol{G}_t \boldsymbol{f}_t^T \boldsymbol{S}_{t|t-1}.
\end{aligned} \tag{5}$$

If the AUV is pausing, the last three steps will not take place, as there is no data updating. In that situation, one will just propagate mean and covariance according to the first two equations. This can also take place over multiple steps, where generally for $s = 0, 1, \ldots$:

$$\begin{aligned}
\boldsymbol{m}_{t+s|t-1} &= \boldsymbol{\mu} + \rho(\boldsymbol{m}_{t+s-1|t-1} - \boldsymbol{\mu}) \\
\boldsymbol{S}_{t+s|t-1} &= \rho^2 \boldsymbol{S}_{t+s-1|t-1} + (1-\rho^2)\boldsymbol{\Sigma}.
\end{aligned} \tag{6}$$

3 AUV sampling design

Like before, the sampling design $\boldsymbol{D}_t$ at this stage t, say directions north, east, west or south, determines the 0 and 1 structure in vector $\boldsymbol{f}_t$ because it directly defines the measurement location. We optimize the design at the current stage from the cost valley calculation based on the current time predicted mean and variance, given all data up to now.

Compute the current cost valley based on $\boldsymbol{m}_{t|t-1}$ and $\boldsymbol{S}_{t|t-1}$, using the same EIBV and VR relations with $\boldsymbol{G}_t$ calculated as in expression (5), and the constraints as before. For a more complex spatio-temporal model, we might want to predict ahead to get $\boldsymbol{m}_{t+s|t-1}$ $S_{t+s|t-1}$, for $s > 0$ in this cost-valley calculation, but lets not go there. (Foss et al. (2022) paper showed that it did not matter so much there.)