Appendix A: How the EnKF works

The EnKF is essentially a Bayesian inference process that operates via Bayes' theorem:

$$P(\boldsymbol{\psi} \mid \boldsymbol{y}) \propto P(\boldsymbol{y} \mid \boldsymbol{\psi}) P(\boldsymbol{\psi}).$$

The proportionality constant is simply a normalization constant. Here $\boldsymbol{\psi}$ is a vector containing a list of model concentrations and $\boldsymbol{y}$ is a vector containing laboratory measurements. According to the Bayesian interpretation, $P(\boldsymbol{\psi})$ is the probability density function that $\boldsymbol{\psi}$ is true, before looking at the laboratory measurements (prior pdf). $P(\boldsymbol{y} \mid \boldsymbol{\psi})$ is the probability density function of measuring the values in $\boldsymbol{y}$, given that $\boldsymbol{\psi}$ is true. The result of combining the two pdfs through Bayes' theorem is the probability that $\boldsymbol{\psi}$ is true, given the laboratory measurement $\boldsymbol{y}$ (posterior pdf).

Under the assumption that all three pdf's are Gaussian, the three pdfs can be written as:

$$P(\boldsymbol{\psi}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{\psi} - \langle\boldsymbol{\psi}_b\rangle)^{\mathsf{T}} \boldsymbol{B}^{-1}(\boldsymbol{\psi} - \langle\boldsymbol{\psi}_b\rangle)\right)$$

$$P(\boldsymbol{y} \mid \boldsymbol{\psi}) \propto \exp\left(-\frac{1}{2}[\boldsymbol{h}(\boldsymbol{\psi}) - \boldsymbol{y}]^{\mathsf{T}} \boldsymbol{R}^{-1}[\boldsymbol{h}(\boldsymbol{\psi}) - \boldsymbol{y}]\right)$$

$$P(\boldsymbol{\psi} \mid \boldsymbol{y}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{\psi} - \langle\boldsymbol{\psi}_a\rangle)^{\mathsf{T}} \boldsymbol{A}^{-1}(\boldsymbol{\psi} - \langle\boldsymbol{\psi}_a\rangle)\right).$$

In the EnKF, the prior pdf are represented by the outcomes of an ensemble of simulations. $\langle\boldsymbol{\psi}_b\rangle$ is the ensemble-averaged concentrations, and $\boldsymbol{B}$ is the covariance matrix derived from the ensemble of concentrations. $\boldsymbol{h}$ is an operator that ingests the model results and extracts data that corresponds to the laboratory measurement, and $\boldsymbol{R}$ is the error covariance of the laboratory measurements. For most purposes, $\boldsymbol{R}$ is assumed to be a diagonal matrix (i.e., the errors are not correlated between laboratory measurements of different quantities).

When the EnKF is applied, all of the members of the prior ensemble gets updated to generate a new ensemble, which represents the posterior pdf. As in the case of the prior pdf, the vector ψ_a is the ensemble-average of the posterior ensemble and A is the posterior covariance matrix derived from the posterior ensemble.

To get a feeling of how the EnKF estimates the hidden concentrations, we can explicitly solve for $\boldsymbol{\psi}_a$, the posterior estimate of all concentrations. It can be shown that [see Kalnay (2002)]:

$$\langle \boldsymbol{\psi_a} \rangle - \langle \boldsymbol{\psi_b} \rangle = \delta \langle \boldsymbol{\psi_a} \rangle = \boldsymbol{BH}^\top (\boldsymbol{HBH}^\top + \boldsymbol{R})^{-1}[\boldsymbol{y} - \boldsymbol{h}(\langle \boldsymbol{\psi_b} \rangle)]$$

where

$$\boldsymbol{H} \equiv \frac{\partial \boldsymbol{h}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}}\Big|_{\boldsymbol{\psi} = \langle \boldsymbol{\psi_a} \rangle}.$$

For the simplicity of subsequent discussions, we will assume that h is a linear operator. In other words,

$$\boldsymbol{h}(\boldsymbol{\psi}) = \boldsymbol{H}\,\boldsymbol{\psi}$$

$$\frac{\partial \boldsymbol{h}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}} = \boldsymbol{H} \quad \forall \quad \boldsymbol{\psi} \in \mathbb{R}^n$$

In the limit of a single measurement, the estimate reduces to

$$\delta \langle \boldsymbol{\psi_a} \rangle \propto Cov(\boldsymbol{\psi}, \boldsymbol{H}\boldsymbol{\psi})[\boldsymbol{y} - \boldsymbol{H}\langle \boldsymbol{\psi_b} \rangle]$$

In other words, the difference between the prior ensemble's estimate of the concentration (i.e., the ensemble mean) and the measured concentrations are broadcasted to all of the concentrations contained in $\boldsymbol{\psi}$. This broadcasting is done through the prior ensemble covariance between all concentrations and the quantity to be measured. This broadcasting through the prior covariance is the reason why the EnKF can be used to estimate hidden concentrations.

Aside from that, the estimate from applying the EnKF is also more precise than the prior ensemble. This can be deduced by solving for the posterior covariance matrix [see Kalnay (2002)]:

$$\boldsymbol{A} = \boldsymbol{B} - \boldsymbol{BH}^\top (\boldsymbol{HBH}^\top + \boldsymbol{R})^{-1}\boldsymbol{HB}$$

For ease of illustration, we will consider a one-variable model. The posterior covariance matrix reduces to a posterior variance ($\sigma_a{}^2$) and likewise for the prior covariance matrix ($\sigma_b{}^2$). Thus:

$$\sigma_a{}^2 = \sigma_b{}^2 \left(1 - \frac{\sigma_b{}^2}{\sigma_b{}^2 + \sigma_o{}^2}\right) < \sigma_b{}^2$$

where $\sigma_o{}^2$ is the observation error variance. From this equation, it is clear that the estimate from the EnKF is more precise than that of the prior ensemble.

Appendix B: RESCUE modelling equations

Equilibrium curve in phase space

When a system is out of equilibrium, the new equilibrium state can be obtained by considering the chemical equilibrium equation and the conservation of mass:

$$[E]\,[S] = K_{eq}[ES]$$
$$[E] + [ES] = E_T$$
$$[S] + [ES] = S_T$$

During the equilibration process, E_T and S_T are constants (mass conservation). Solving this set of equations yields the following:

$$[ES] = \frac{1}{2}(E_T + S_T + K_{eq}) \pm \frac{1}{2}\sqrt{(E_T + S_T + K_{eq})^2 - 4\,E_T\,S_T}$$

Clearly, there can only be one possible concentration of $[ES]$, if given a value of E_T and S_T. The unrealistic solution is one which would definitely result in a negative concentration if plugged into the conservation equations. From the conservation equations, for concentrations to always be non-negative, the conditions are

$$0 \le [ES] \le E_T \quad \text{and} \quad 0 \le [ES] \le S_T.$$

Since E_T, S_T and K_{eq} are always positive, then it is clear that

$$[ES] = \frac{1}{2}(E_T + S_T + K_{eq}) - \frac{1}{2}\sqrt{(E_T + S_T + K_{eq})^2 - 4\,E_T\,S_T}$$

In our simulations, if we assume equilibrium to always hold, the enzyme-substrate concentrations must always fall on this curve. To recast this into a form suitable for time-integrations, we apply the chain rule. For ease of writing, we define:

$$f(E_T, S_T) \equiv \frac{1}{2}(E_T + S_T + K_{eq}) - \frac{1}{2}\sqrt{(E_T + S_T + K_{eq})^2 - 4\,E_T\,S_T}$$

Then,

$$\partial_t\,[ES] = \partial_{E_T} f(E_T, S_T)\,\partial_t\,E_T + \partial_{S_T} f(E_T, S_T)\,\partial_t\,S_T$$

The ∂_x denotes a partial derivative with respect to variable x.

As such, to determine the evolution of the enzyme-substrate concentration, we need the time-derivatives of the total enzyme and total substrate concentrations.

Evolution of total enzyme due to ES ⇒ P + E

Assuming no creation or destruction of cas13-gRNA, then,

$$\partial_t \, E_T = 0$$

I.e.,

$$\partial_t \, [ES] = \ \partial_{S_T} \, f(E_T, S_T) \, \partial_t \, S_T$$

Evolution of total substrate

The substrate is generated constitutively and decays over time. Furthermore, the ES $\Rightarrow$ P + E reaction removes ES. Hence, the total substrate's rate of change is:

$$\partial_t \, S_T = g_S - d_P[S] - K_P[ES]$$

g_S and d_S are the constitutive production and decay constants. The last term is the loss of substrate from ES $\Rightarrow$ P + E.

Evolution of modified mRNA (P)

$$\partial_t \, [P] = K_P[ES] - d_P[P]$$

d_P is the decay constant for the modified mRNA. It is likely that $d_P = d_S$.

Production and decay of reporter proteins

$$\partial_t \, [GC] = \gamma_{GC}[S] - d_{GC}[GC]$$
$$\partial_t \, [C] = \gamma_C[P] - d_C[C]$$