

## Current progress and next steps – meeting 06 Feb 2026

### Current progress

In Phase 1, instability was observed when applying SINDy to the full SIR system using degree-2 polynomial libraries: while restricted libraries enabled reliable recovery, expanding the library led to inconsistent identification of the governing equations.

This instability was hypothesized to arise from manifold-induced collinearity. Specifically, the SIR dynamics evolve on an invariant manifold defined by the conservation law  $S + I + R = 1$  (*after normalization*), which induces strong correlations among degree-2 candidate terms when  $S$ ,  $I$ , and  $R$  are treated as independent coordinates (e.g.,  $SR = S(1 - S - I)$ , which overlaps with  $SI$  and  $S^2$  terms).

To address this, a reduced-coordinate formulation was adopted. Instead of fitting the full SIR system directly, **only the S and I equations were learned** using a degree-2 polynomial library, while **R was reconstructed algebraically** via the conservation law  $R = 1 - S - I$ .

Consistency was verified through an additional diagnostic by integrating

$$\frac{dR}{dt} = -(dS/dt + dI/dt)$$

and comparing the resulting trajectories with the algebraic reconstruction, both within the fitting window and over full-horizon rollouts.

With this reduced-coordinate approach, all three SIR equations **can now be stably and accurately recovered** even when using degree-2 polynomial libraries, whereas recovery was previously unstable under the same library choice when fitting S, I, and R simultaneously.

### Ongoing / Next Steps?

1. Developing symbolic regression methods for the same recovery experiments.
2. Introduce an observation layer to account for reporting delay, reporting rate, and other surveillance effects, and integrate it into the equation discovery framework?