Continual Model Learning

Donghwa Hong

Compiled on August 14, 2026

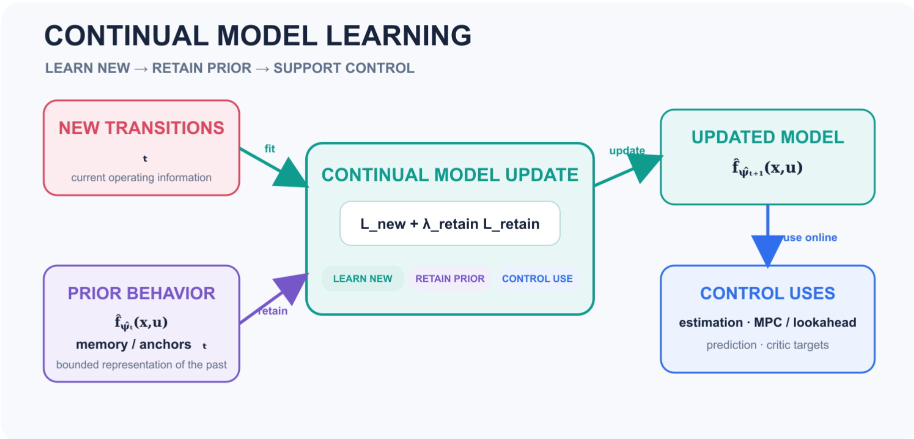

Figure 1: Continual model learning fits new data while retaining prior model behavior.

Physical interaction is indexed by k, whereas learner updates are indexed by t; the two clocks need not advance at the same rate. $\widehat{\psi}_t$ denotes the model parameter available before update t, and $\widehat{\psi}_{t+1}$ denotes the updated parameter.

1 Problem definition

Online Learning-Based Optimal Control requires decision-state and successor information for Bellman-based control. Continual Model Learning addresses the identifier/estimator-side problem: learn a dynamics model from new real measurements without losing earlier model behavior still needed by the estimator or controller.

Consider

$$x_{k+1} = f(x_k, u_k, w_k), \qquad y_k = h(x_k) + v_k, \tag{1}$$

where x_k, u_k, and y_k are the state, control input, and measurement; f and h are the state-transition and measurement maps; and w_k and v_k represent process and measurement uncertainty. Unless a different stochastic predictor is declared, the learned one-step target is the conditional mean

$$\overline{f}(x,u) := \mathbb{E}[x_{k+1} \mid x_k = x, u_k = u] \approx \widehat{f}_{\widehat{\psi}_t}(x,u). \tag{2}$$

The primary continual-learning scope is coverage expansion for a fixed underlying predictor $\overline{f}$: new operating regions are added while selected prior function behavior is retained. Genuine physical drift is a different case. If the same (x,u) acquires a different successor law because of temperature, aging, load, or another condition, that condition should be represented explicitly, for example by $\overline{f}(x,u;\eta)$, or handled by a declared tracking/forgetting rule. Blind retention can otherwise preserve obsolete behavior.

When both state and model are unknown, the basic identification principle is measurement–dynamics consistency:

$$(x^{\star}_{0:k}, \psi^{\star}) \in \arg\min_{x_{0:k},\psi} \left\{\mathcal{L}_{\mathrm{meas}}(x_{0:k}; y_{0:k}) + \mathcal{L}_{\mathrm{dyn}}(x_{0:k}, \psi; u_{0:k-1})\right\}. \tag{3}$$

For example, quadratic measurement and dynamics residuals are

$$\begin{aligned} \mathcal{L}_{\mathrm{meas}}(x_{0:k}; y_{0:k}) &= \sum_{i=0}^{k} \|y_i - h(x_i)\|^2_{W_y}, \\ \mathcal{L}_{\mathrm{dyn}}(x_{0:k}, \psi; u_{0:k-1}) &= \sum_{i=0}^{k-1} \left\|x_{i+1} - \widehat{f}_{\psi}(x_i, u_i)\right\|^2_{W_x}, \end{aligned} \tag{4}$$

where $W_y \succeq 0$ and $W_x \succeq 0$ weight the measurement and dynamics residuals, respectively. An online implementation evaluates these losses over a recent window rather than the entire trajectory.

Real measurements form the stream

$$d_k = (y_k, u_k, y_{k+1}), \qquad \mathcal{D}_{0:k(t)-1} := (d_0, \dots, d_{k(t)-1}), \qquad \mathcal{S}_t \subseteq \mathcal{D}_{0:k(t)-1}. \tag{5}$$

Here $k(t)$ is the latest measurement/state index available at learner update t. The set $\mathcal{S}_t$ may contain only the latest completed transition or a short selected history. If the state is directly measured, replace y_k by x_k and the problem reduces to model learning alone. Otherwise, the latent state trajectory and model parameters must be estimated jointly, or a separate estimator must supply $\widehat{x}_k$.

2 Why this problem matters

A model learned once offline may become incomplete as a mobility system visits new loads, temperatures, road conditions, aging states, or operating regions. Online system identification can reduce the residual on the newest trajectory, but the updated function may no longer represent regions learned earlier.

This distinction matters because a controller does not query the model only at the latest sample. State estimators, MPC or lookahead optimization, and critic targets may evaluate it over many control-relevant states and inputs. Continual learning therefore seeks both **new-region fit** and **prior-function retention**, rather than only a small current residual.

3 Key challenges

- **Limited and policy-dependent data:** real transitions arrive sequentially along the deployed closed-loop trajectory and do not freely cover the full state–input domain.
- **New learning versus prior-function retention (the stability–plasticity trade-off):** the learner must incorporate newly observed behavior without overwriting earlier behavior that remains relevant.
- **State–model coupling:** when x_k is latent, state-estimation error and model error can explain the same residual and must be separated under suitable observability and excitation conditions.
- **Control-relevant online learning:** memory, update time, multistep error, model acceptance, and the effect of each update on downstream control matter in addition to one-step prediction loss.

4 A conventional approach — recent-data residual minimization

One conventional online-identification approach fits the model to the latest transition or a short recent window. For notational simplicity, assume here that the state is available; x_j is replaced by $\widehat{x}_j$ when a separate estimator is used:

$$\begin{aligned} \widehat{\psi}_{t+1} &\approx \arg\min_{\psi} \mathcal{L}_{\text{new}}(\psi; \mathcal{S}_t), \\ \mathcal{L}_{\text{new}}(\psi; \mathcal{S}_t) &:= \sum_{j \in \mathcal{I}_t} \left\| x_{j+1} - \widehat{f}_{\psi}(x_j, u_j) \right\|_{W_x}^2 . \end{aligned} \tag{6}$$

$\mathcal{I}_t$ indexes the recent transitions admitted to update t. This update is useful for reducing current residuals, especially when the physical dynamics truly change. Under limited excitation, however, it becomes a trajectory-local online-identification update: it fits the currently visited region but does not explicitly preserve model behavior learned in earlier regions.

Tracking a genuinely changing model and retaining knowledge over an expanding task domain are therefore different objectives and should be evaluated separately.

5 Research direction — function-retaining continual learning

A continual formulation augments current measurement–dynamics consistency with explicit retention of the previously learned function. Let $k_t := k(t)$ be the newest physical sample available at learner update t, and let M be the estimation-window length. The set $\mathcal{S}_t$ supplies the transitions selected to fit newly observed behavior, whereas $\mathcal{M}_t = \{(\bar{x}_\ell, \bar{u}_\ell)\}_{\ell=1}^{m_t} \subseteq \mathcal{X} \times \mathcal{U}$ is a bounded set of state–input anchors selected to retain prior model behavior. An anchor may be derived from a transition in $\mathcal{S}_t$, but $\mathcal{M}_t$ can additionally contain earlier stored points with pseudo-targets generated from the pre-update model:

$$\begin{aligned}\left(\widehat{x}_{k_t-M:k_t}, \widehat{\psi}_{t+1}\right) \approx \arg \min_{x_{k_t-M:k_t}, \psi} \Big\{ & \mathcal{L}_{\text{meas},t}(x; \mathcal{S}_t) + \mathcal{L}_{\text{dyn},t}(x, \psi; \mathcal{S}_t) \\ & + \lambda_{\text{retain}} \mathcal{L}_{\text{retain}}\left(\widehat{f}_\psi, \widehat{f}_{\widehat{\psi}_t}; \mathcal{M}_t\right) \Big\}.\end{aligned} \tag{7}$$

Here $\mathcal{L}_{\text{meas},t}$ and $\mathcal{L}_{\text{dyn},t}$ are the residual losses defined in Section 1, evaluated on $\mathcal{S}_t$. In a recent-window implementation, $\mathcal{S}_t = \{d_j \mid j = k_t - M, \ldots, k_t - 1\}$ supports the state and measurement window ending at k_t. The retention weight satisfies $\lambda_{\text{retain}} \geq 0$.

For example, functional retention over a bounded anchor set $\mathcal{M}_t$ can be written as

$$\mathcal{L}_{\text{retain}} = \frac{1}{|\mathcal{M}_t|} \sum_{(\bar{x}, \bar{u}) \in \mathcal{M}_t} \left\| \widehat{f}_\psi(\bar{x}, \bar{u}) - \widehat{f}_{\widehat{\psi}_t}(\bar{x}, \bar{u}) \right\|_{W_r}^2 . \tag{8}$$

This expression assumes $|\mathcal{M}_t| > 0$ and a positive-semidefinite weight $W_r \succeq 0$. If no anchors are retained, the retention term is omitted.

The first two losses explain current measurements and dynamics; the last loss retains selected input–output behavior of the model available before the update. If x_k is measured, remove the state-trajectory decision and $\mathcal{L}_{\text{meas},t}$. The retention weight and anchor set define a design trade-off; they do not guarantee global accuracy or eliminate forgetting.

Candidate mechanisms include a small representative replay set, local or RBF activation, pseudo-data generated from the previous model, and functional regularization. These are candidate directions, not achieved MIC Lab results.

Future evidence should compare recent-only fitting, bounded-memory retention, and full replay using new-region error, return-to-prior-region error, multi-step prediction, estimator error, downstream control performance, memory, and update time.

6 Application domains

- **Continual calibration of vehicle, electric-drive, energy-management, and thermal-system models** across operating regions and explicitly represented load, temperature, or aging conditions.
- **Control-oriented model learning for state estimation, MPC, online lookahead, and critic targets**, where the model must remain useful beyond the latest trajectory.
- **Model learning in complex real environments** where complete offline data and repeated full retraining are impractical.

7 References

- D. Hong and K. Choi, "State-Error-Based Online Learning of Control-Oriented Tire Force Models Using Grid Memory," ICROS, accepted, in press, 2026.
- D. Hong and K. Choi, "Integral Error-Based Adaptive Neural Identifier for Nonlinear Lateral Tire Forces," IEEE International Conference on Advanced Motion Control (AMC), pp. 1–6, 2026.