Dear Authors,

I found the integration of Transformer-based forecasting and physical constraints in the proposed PCBT framework interesting. I would appreciate clarification on several methodological details that seem important for reproducibility and for interpreting the physics-constrained component of the model.

1. Implementation of the flow-continuity constraint

Equation (10) introduces $h(x,t)$, the spatial coordinate x, and the derivatives $\partial h/\partial t$ and $\partial Q/\partial x$, which are stated to be evaluated using automatic differentiation.

Does x denote the spatial location, for example the along-channel position, of the hydrological gauging stations? If so, could the authors clarify how these spatial coordinates are incorporated into the PCBT input and computational graph? The model description and Table 1 appear to list antecedent precipitation and discharge, $P(t-1:t-7)$ and $Q(t-1:t-7)$, as the principal model inputs.

In particular, is $\partial Q/\partial x$ obtained by automatic differentiation with respect to an explicit coordinate input x, or is it approximated from discharge differences between gauging stations? Similarly, could the authors clarify how $h(x,t)$ is obtained: is water level/depth an additional model input or output, or is it derived from another variable?

It would also be helpful to clarify the dimensional form of Eq. (10). As written,

$$\frac{\partial h}{\partial t} + c\frac{\partial Q}{\partial x},$$

if h denotes water depth, Q denotes discharge, and c denotes wave propagation velocity, the two terms do not appear to have identical physical dimensions under their usual definitions. Is an additional transformation or normalization involved, or is a different definition of these variables intended?

2. Definition of the rainfall–runoff constraint

Equation (11) combines Q_t, Q_{t-1}, and basin-average precipitation P_t:

$$Q_t = Q_{t-1} + P_t - \alpha Q_{t-1}$$

Since discharge is expressed in $m^3\,s^{-1}$, could the authors specify the units and preprocessing of P_t before it is combined with discharge? For example, is precipitation converted into an equivalent discharge using basin area and the time interval, or is this constraint evaluated using normalized/non-dimensional variables?

3. Information used during multi-step forecasting

The model is described as using antecedent P and Q, while Eq. (11) requires P_t and Q_{t-1}. For D1–D7 forecasting, could the authors clarify whether P_t during the forecast period represents observed future precipitation, forecast precipitation, or whether Eq. (11) is evaluated only over the antecedent period?

Similarly, after the first forecast step, does Q_{t-1} represent the previous model prediction $\hat{Q}_{t-1}$, or the observed value Q_{t-1}^{obs}? This distinction determines whether the constraint is recursively imposed on the predicted hydrograph or is applied under a teacher-forcing setting, and also affects the interpretation of the experiment as antecedent-only forecasting versus forecasting with information from the forecast period.

Clarification of these points would substantially improve the reproducibility of the proposed framework and help readers better understand how the physical constraints are implemented during multi-step forecasting.

Best regards,

Jiawei