Supplement of Hydrol. Earth Syst. Sci., 28, 5443–5458, 2024
https://doi.org/10.5194/hess-28-5443-2024-supplement

Hydrology and Earth System Sciences
Open Access
EGU

Supplement of

Enhancing generalizability of data-driven urban flood models by incorporating contextual information

Tabea Cache et al.

Correspondence to: Tabea Cache (tabea.cache@unil.ch)

Contents:

- **Figure S1**: Model's architecture.

- **Figure S2**: Augmentation techniques.

- **Figure S3**: Patch locations.

- **Section S1**: Training and evaluating the model without data augmentation.

- **Table S1**: Training, validation and testing metrics of the model trained without patch augmentation.

- **Table S2**: Comparison of training and validation losses for models with various methodological innovations.

- **Table S3**: Residual error in downstream areas, upstream areas, and depressions.

- **Figure S4**: Flood maps for a 100-yr rainfall event in Luzern and Singapore emulated by the single-scale urban flood model presented by Guo et al. (2020) and by the presented multi-scale urban flood model.

- **Figure S5**: Violinplots of the simulation error in Luzern and Singapore for the models without and with transfer learning.

- **Figure S6**: Flood maps and error in Luzern and Singapore, without excluding the results in water bodies.

- **Table S4**: Performance metrics comparison across cities for $P_{19\text{-}1}$, $P_{31\text{-}2}$ and $P_{46\text{-}1}$.

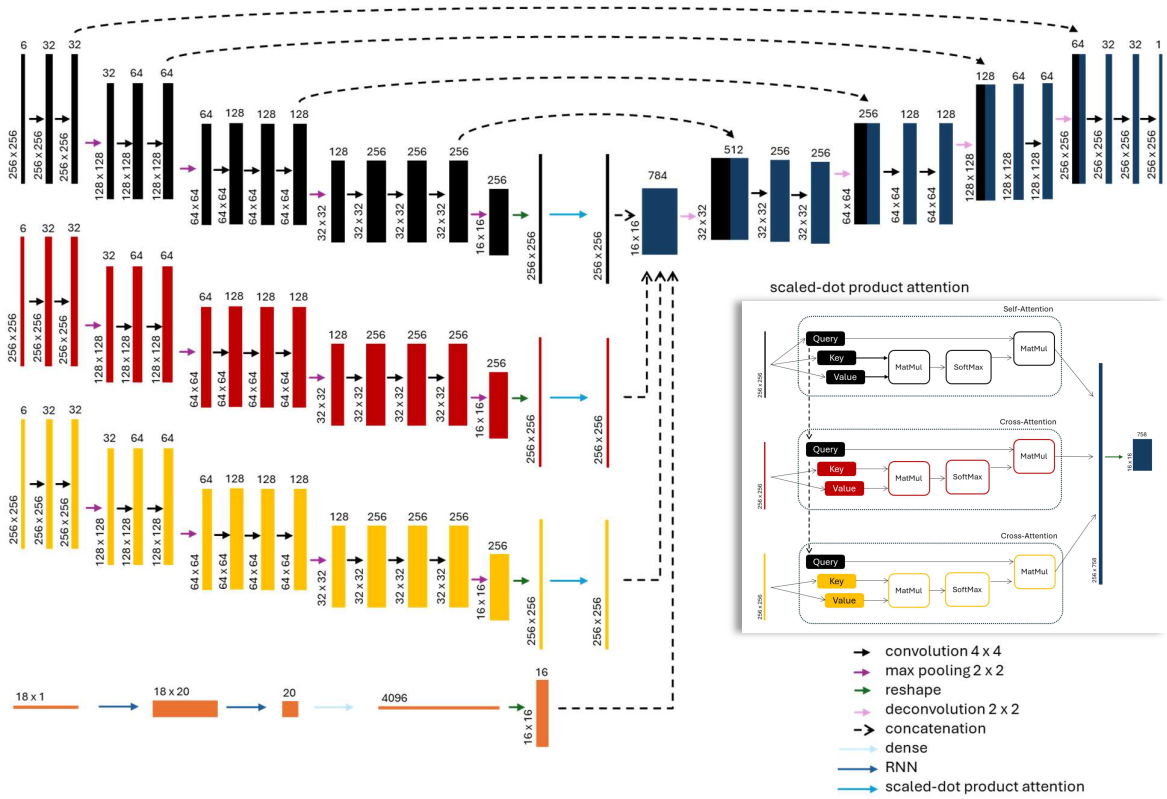

Figure S1. Model's architecture. Each box corresponds to a multi-channel feature map, with the corresponding number of channels denoted on top of the box and its height and width dimensions indicated at the lower left edge of the box. The arrows correspond to different operations and the operations for the scaled-dot product attention are detailed in the figure with shaded contours. The 6 channels of the input images are the terrain features derived from the DEM: DEM, mean curvature, aspect (sine and cosine), depth of the sinks and the slope (in radians).

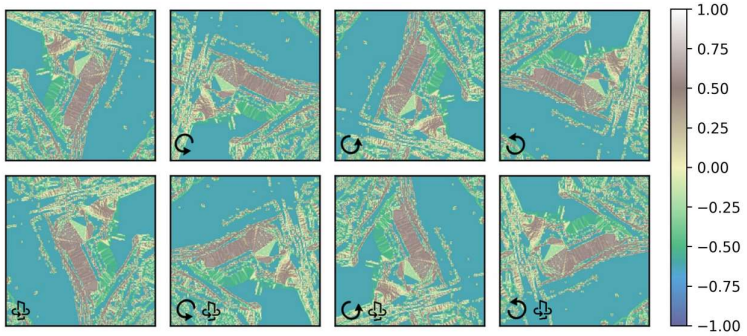

Figure S2. Augmentation techniques that were applied to the patches, consisting of a combination of flips and 90° rotations. The upper left patch is the original patch, without augmentation.

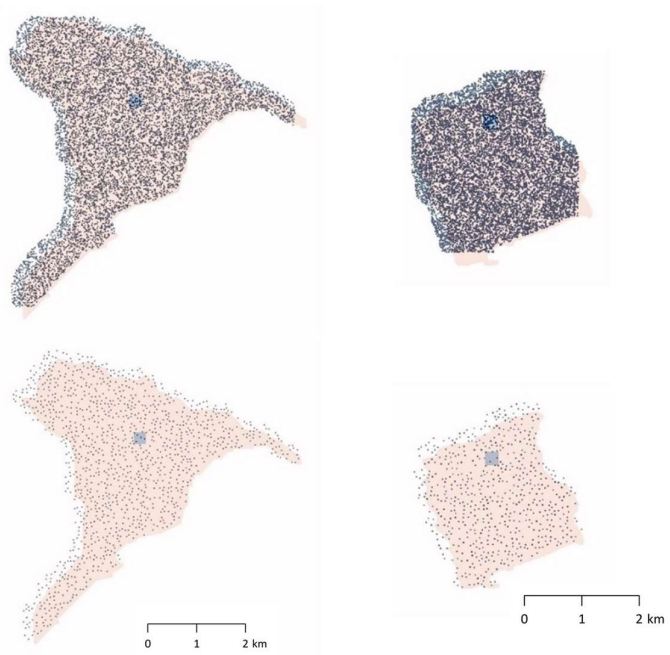

Figure S3. Patch locations in the study areas; the dots indicate the upper left corner of the patches. From left to right: Zürich, Luzern. From top to bottom: number of patches used in a similar study (Guo et al., 2021) and in this study (1250 and 620 respectively for Zürich and Luzern). The light blue square is an example of a patch to help visualize the sampling density.

Section S1. Training and evaluating the model without data augmentation.

We trained a model with identical architecture to the one described in the manuscript but adopting a patch sampling procedure similar to the one from previous studies (Guo et al., 2021; Seleem et al., 2023). We randomly sampled 10'000 patches without patch overlap threshold, with minimum study area coverage of 10% for each individual patch and without data augmentation, i.e. no flip and rotations of the patches. The training and validation losses of the model trained with this alternative patch sampling procedure are reported in Table S1 and suggest a good performance of the model. However, after further analysis of the model performance, it appears that the model is overfitting. In fact, we evaluated the model under two conditions: (1) using the city of Zurich as input, (2) using the city of Zurich with some changes in orientation as input (Table S1). For the latter, we tested the model on two orientation changes (randomly selected; Zurich rotated by 90° and Zurich flipped and rotated by 180°) for the rainfall event $P_{46\text{-}1}$. While the city on which the model is tested is identical to the one used to train and validate the model, the orientation of the city is modified. From the metrics reported in Table S1, we see that the model performs well in Zurich when the city's orientation is not modified (MSE = $0.08 \cdot 10^{-3}$ m^2 and $RMSE_{0.1}$ = $18.2 \cdot 10^{-3}$ m), but that the performance drops when the input city is flipped and rotated, with an MSE in the order of 100 times the MSE for the city without flips and rotations, and an $RMSE_{0.1}$ in the order of 13 times the $RMSE_{0.1}$ for the city without flips and rotations. This suggests that the model is overfitting to the training patches.

Table S1. Training, validation and testing metrics of the model trained without patch augmentation.

		MSE (10^{-3} m^2)	$RMSE_{0.1}$ (10^{-3} m)
Training loss		0.047	-
Validation loss		0.279	-
Model evaluation	Zurich, $P_{46\text{-}1}$	0.080	18.6
	Zurich 90°, $P_{46\text{-}1}$	8.110	231.5
	Zurich flip & 180°, $P_{46\text{-}1}$	8.296	231.3

Table S2. Comparison of training and validation losses for models with various methodological innovations.

	Training MSE (in m^2)	Validation MSE (in m^2)
Model described in the manuscript	0.77e-04	4.84e-04
Model with architecture similar to the one described in the manuscript but without RNN scaling and with the same data pre-processing	1.15e-04	5.04e-04
Model with identical architecture as the one described in the manuscript but with different feature scaling (i.e. scaling at the city-image level instead of at the patch-level)	118e-04	114e-04

Table S3. Residual error in downstream areas (defined as lowest 33% terrain elevation), upstream areas (defined as highest 33% terrain elevation), and depressions.

Residual [10^{-3} m]	Q25	Median	Q75
Downstream	-3.8	0.4	3.4
Downstream (Guo et al., 2021)	-7	5	20
Upstream	-2.7	0.4	5.6
Upstream (Guo et al., 2021)	-2	0	3
Depressions	-16	1.2	10.2
Depressions (Guo et al., 2021)	-60	1	30

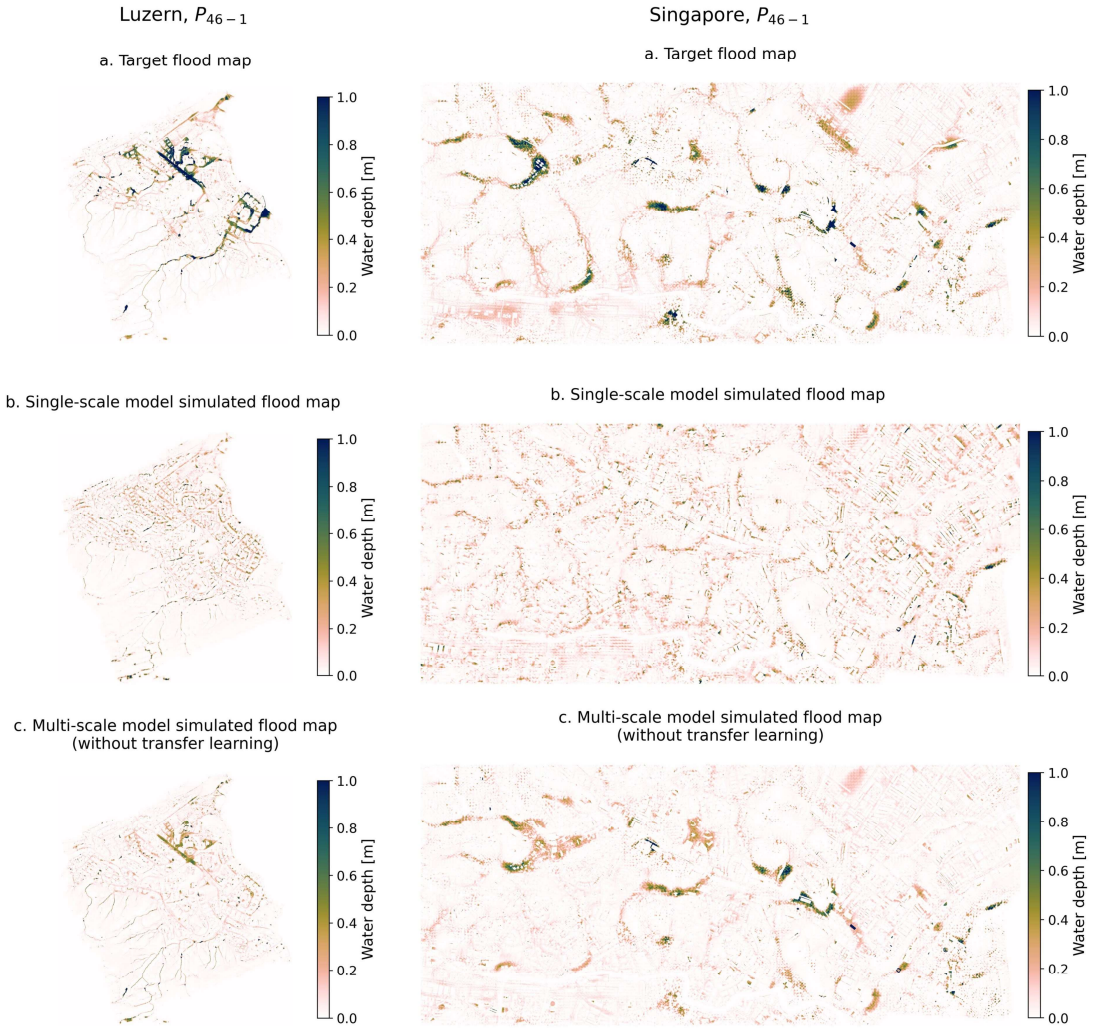

Figure S4. Flood maps for a 100-yr rainfall event in Luzern (left) and Singapore (right) emulated by the single-scale urban flood model presented by Guo et al. (2020; trained model and data available at: https://www.research-collection.ethz.ch/handle/20.500.11850/365484) and by the presented multi-scale urban flood model.

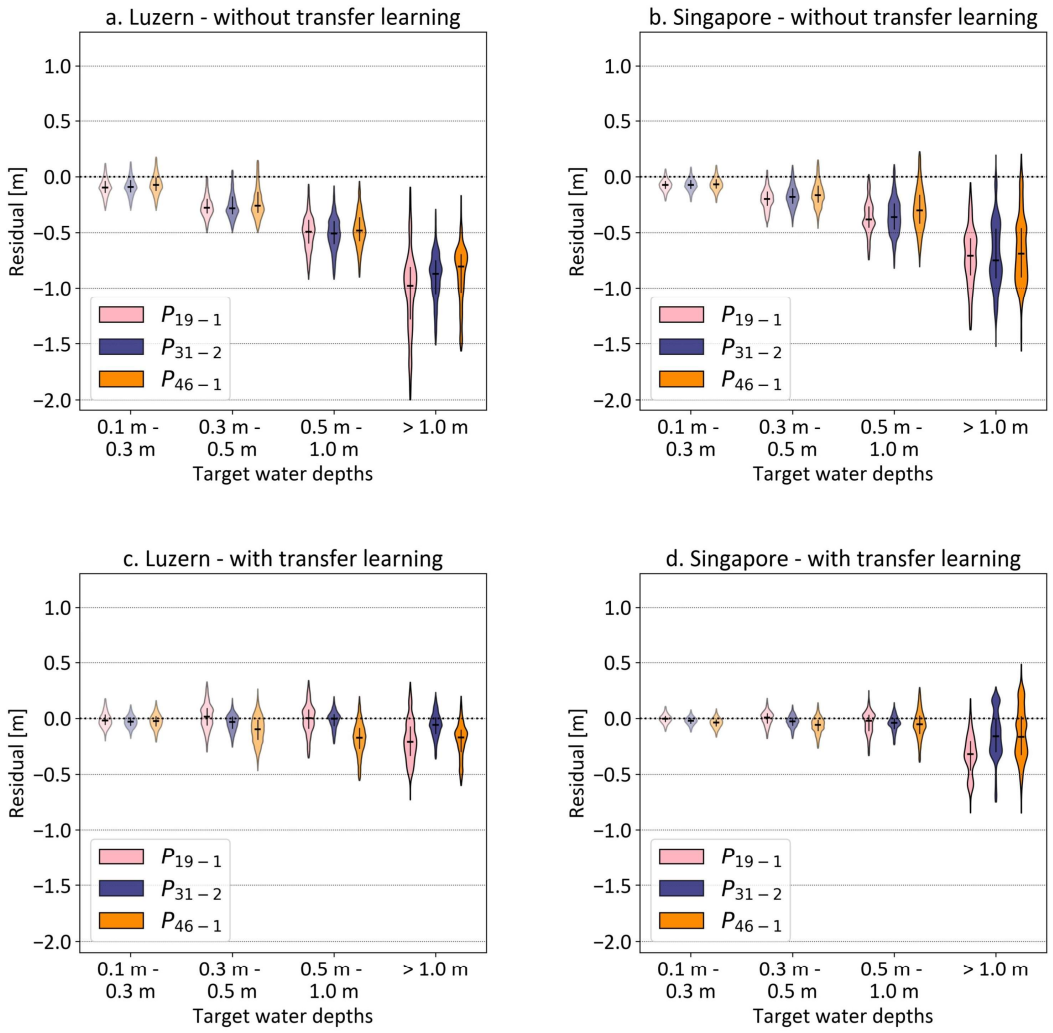

Figure S5. Violinplots of the simulation error in (a. and c.) Luzern and (b. and d.) Singapore, expressed as the difference between the target flood map and simulated flood map, for different target water depth ranges and different rainfall events, using models (a. and b.) without transfer learning, i.e. model presented in Section 5 of the manuscript, and (c. and d.) with transfer learning using event P_{31-2} to retrain the model in the respective cities, i.e. model presented in Section 6 of the manuscript. The vertical black lines range from the 25th to 75th percentiles and the horizontal black line indicates the median value. Negative values correspond to the underprediction of the simulations.

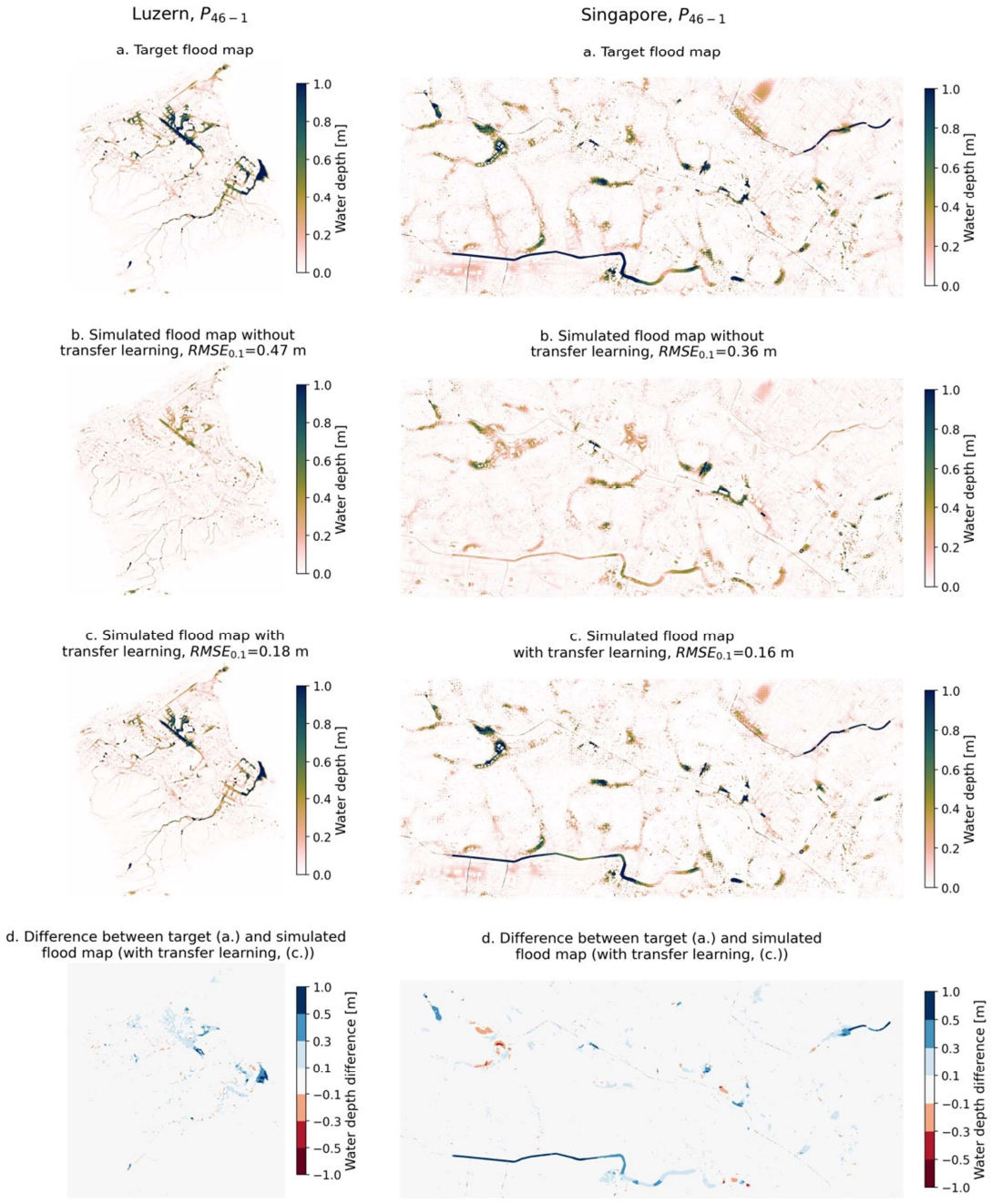

Figure S6. Flood maps and error in Luzern (left) and Singapore (right), without excluding the results in water bodies. Note that the models used to simulate the flood maps 6c were retrained for $P_{31\text{-}2}$.

Table S4. Comparison of the performance metrics $RMSE_{0.1}$, $CSI_{0.1}$ and $CSI_{0.3}$ across cities for $P_{19\text{-}1}$, $P_{31\text{-}2}$ and $P_{46\text{-}1}$.

	Zurich	Luzern	Singapore
Model trained in Zurich for 12 events			
$RMSE_{0.1}$ [m]			
P_{19-1}	0.05	0.27	0.15
P_{31-2}	0.05	0.33	0.17
P_{46-1}	0.04	0.40	0.19
$CSI_{0.1}$			
P_{19-1}	0.75	0.37	0.37
P_{31-2}	0.81	0.43	0.43
P_{46-1}	0.84	0.50	0.48
$CSI_{0.3}$			
P_{19-1}	0.85	0.13	0.17
P_{31-2}	0.87	0.19	0.25
P_{46-1}	0.87	0.32	0.35
Model retrained in the respective cities for P_{31-2}			
$RMSE_{0.1}$ [m]			
P_{19-1}	-	0.11	0.07
P_{31-2}	-	0.08	0.07
P_{46-1}	-	0.17	0.10
$CSI_{0.1}$			
P_{19-1}	-	0.67	0.68
P_{31-2}	-	0.74	0.72
P_{46-1}	-	0.72	0.66
$CSI_{0.3}$			
P_{19-1}	-	0.65	0.66
P_{31-2}	-	0.78	0.75
P_{46-1}	-	0.68	0.67