https://doi.org/10.5194/egusphere-2026-3652
Preprint. Discussion started: 8 July 2026

High-resolution rain-on-grid hydrodynamic modelling can replace hydrological models for catchment-scale flood simulation: the case of the 2021 Ahr catchment flood

Shahin Khosh Bin Ghomash[1] and Daniel Caviedes-Voullième[1,2]

[1]Chair for Environmental Fluid Dynamics and Modeling, TUD Dresden University of Technology, Dresden, Germany
[2]Simulation and Data Lab Terrestrial Systems, Institute of Bio- and Geosciences Agrosphere and Jülich Supercomputing Centre, Forschunszentrum Jülich, Jülich, Germany

Correspondence: Shahin Khosh Bin Ghomash (shahin.khoshbin@tu-dresden.de)

Abstract.

Flood simulation and forecasting in mid- to large-sized catchments has long relied on coupling hydrological and hydrodynamic models, a two-tiered modelling strategy driven mainly by the assumption that physically-based hydrodynamic simulations at the resolutions needed to resolve channels and floodplains are computationally prohibitive at catchment scale. Recent advances in multi-GPU high-performance computing, together with increasing availability of high-resolution geospatial data are challenging this assumption. In this study, we test whether a fully hydrodynamic, rain-on-grid approach can be a standalone alternative to the traditional coupled chain, using the July 2021 flood in the Ahr catchment, Germany. We apply the performance-portable multi-GPU shallow-water solver SERGHEI to the entire ∼900 km^2 Ahr catchment at dx = 2, 5, and 10 m, forcing it directly with the 5-minute RADOLAN precipitation product. Assessed against the observed event, the rain-on-grid setup reproduces the flood characteristics remarkably well. Model skill metrics are comparable to –and in some metrics improving–, those reported by previous inundation studies relying on reconstructed hydrographs or hydrological-hydrodynamic two-tiered modelling chains for the same event. Remarkably the present setup using no hydrological model, no reconstructed inflow hydrograph, and no parameter calibration, and covering a domain roughly an order of magnitude larger, achieves very good results. We further show that the approach scales efficiently across two generations of GPU supercomputers, completing the 96-hour event between roughly 18 and 420 times faster than real time, well within operational early-warning requirements. The results suggest that, for catchments of this scale and event types of this nature, catchment-scale rain-on-grid hydrodynamic modelling has reached a level of maturity that justifies its use as a standalone alternative to the more traditional coupled hydrological–hydrodynamic chain.

1 Introduction

Flood simulation and forecasting in mid- to large-sized catchments has long relied on the coupling between hydrological and hydrodynamic models (Teng et al., 2017; Falter et al., 2016; Grimaldi et al., 2019). In such setups, a distributed or semi-distributed hydrological model translates rainfall into runoff and channel discharge across the catchment, which is

https://doi.org/10.5194/egusphere-2026-3652
Preprint. Discussion started: 8 July 2026

then provided as boundary forcing to a hydrodynamic model that simulates inundation along the river network (Falter et al., 2016; Najafi et al., 2024). This separation-of-concerns reflects several long-standing constraints. The most prominent has been computational: running a fully physically based 2D hydrodynamic model forced directly by spatially distributed rainfall, over catchments spanning hundreds of square kilometres and at resolutions sufficient to resolve channels, floodplains, and urban features, has historically been considered prohibitively expensive (Hill et al., 2023). Beyond cost, the division of work also reflects other constraints such as fundamental differences in process representation between the two model types, the need for accurate initial and boundary conditions, and the practical overhead of maintaining, calibrating, and coupling two separate modelling components in an operational forecasting chain. Recent advances in multi-GPU high-performance computing (HPC) and the development of performance-portable shallow-water solvers (Caviedes-Vouillième et al., 2023; Morales-Hernández et al., 2021; Khosh Bin Ghomash et al., 2026) are now challenging this assumption, raising the question of whether a hydrological model is still strictly necessary for catchment-scale flood simulation, or whether modern hydrodynamic models alone, empowered by HPC, can resolve the full hydrodynamic/hydrological process chain from rainfall to inundation, an idea which has long been present and progressively establishing in the community (e.g., Caviedes-Voullième et al., 2012; Costabile and Costanzo, 2021; David and Schmalz, 2021; Ennouini et al., 2024; Godara et al., 2024; Liang et al., 2016; Xing et al., 2018; García-Alén et al., 2022).

Two-dimensional hydrodynamic models, typically based on the shallow-water equations (SWE), have a long-standing history in flood modelling and have been applied across a wide range of scenarios (e.g. Bates et al., 2010; Dullo et al., 2021; Pasculli et al., 2021; Khosh Bin Ghomash et al., 2022, 2025b, 2026; Xu et al., 2025). They offer the most robust and physical correct framework for simulating flood propagation, inundation extent, water depths, and flow velocities, all of which are essential for impact-based forecasting and risk assessment (Costabile et al., 2019, 2023; Khosh Bin Ghomash et al., 2025b; Bachmann et al., 2021). However, the practical use of such models over large catchments and at high spatial resolutions has been historically limited by their substantial computational demands (de Almeida and Bates, 2013; Caviedes-Vouillième et al., 2020; Morales-Hernández et al., 2020). This is particularly relevant for floods in mid- to large-sized catchments, where the modelling domain can span thousands to tens of thousands of square kilometres and where capturing the geometry of narrow river channels, urban areas, and other small-scale features requires resolutions of a few meters or less (Chakraborty et al., 2025; Schuermans et al., 2022; Khosh Bin Ghomash et al., 2024).

In recent years, this picture has changed substantially. Advances in high-performance parallel computing, and in particular the adaptation of hydrodynamic solvers to graphical processing units (GPUs), have led to dramatic reductions in simulation runtimes (Morales-Hernández et al., 2020; Caviedes-Vouillième et al., 2023). Multi-GPU implementations now make it possible to run high-resolution shallow-water simulations across very large domains within operationally relevant timeframes (Morales-Hernández et al., 2021; Khosh Bin Ghomash et al., 2026). Recent studies have demonstrated that fully dynamic 2D SWE solvers can deliver simulations several times faster than real time even at meter-scale resolutions over urban or catchment-scale domains (Khosh Bin Ghomash et al., 2024, 2025b, 2026). This shift has fundamentally expanded the range of problems that can be tackled with hydrodynamic modelling and opens the door to applications that were previously considered computationally infeasible.

https://doi.org/10.5194/egusphere-2026-3652
Preprint. Discussion started: 8 July 2026

A particularly compelling motivation for advancing catchment-scale, high-resolution flood modelling is the catastrophic flood event that struck western Germany in mid-July 2021. Triggered by the slow-moving low-pressure system "Bernd", extreme rainfall over the Eifel region produced devastating flash flooding, severely affecting Belgium, the Netherlands, and Germany (Chakraborty et al., 2025; Schäfer et al., 2021; Mohr et al., 2023). The Ahr Valley was among the hardest-hit regions, with 134 fatalities along the Ahr River alone, accounting for the majority of the more than 180 flood-related deaths in Germany (Truedinger et al., 2023; Thieken et al., 2023; Schüttrumpf et al., 2026). The narrow, gorge-like morphology of the Ahr basin severely limited the river's capacity to convey the unprecedented inflows, leading to rapid and destructive inundation across densely populated valley settlements (Khosh Bin Ghomash et al., 2024; Roggenkamp et al., 2024). The disaster also exposed important shortcomings in the existing flood warning chain, including limited spatial detail in warnings, the absence of impact-based information, and the difficulty of communicating the severity of the event to the affected population (Busker et al., 2026; Rocha Silva et al., 2024; Thieken et al., 2023). Together, these aspects highlighted the need for high-resolution, catchment-scale flood simulations capable of resolving both flood-generating hydrological processes and detailed inundation dynamics in the affected valleys.

For catchments of this size, flood simulation has traditionally relied on the coupled hydrological–hydrodynamic chain, in which a hydrological model translates rainfall into runoff and channel discharge that then serve as boundary forcing for a 2D hydrodynamic model simulating inundation along the river network (Falter et al., 2016; Najafi et al., 2024). Prominent recent examples of this approach applied to the 2021 Ahr Valley flood include the coupling of the mesoscale Hydrologic Model (mHM) with the 2D hydrodynamic model RIM2D, both to introduce the concept of spatial counterfactuals to flood hazard assessment (Merz et al., 2024) and to provide an analysis of alternative realisations of the 2021 Ahr event (Vorogushyn et al., 2025). Moreover, this setup was used for presenting an impact-based early warning system for riverine flooding by Najafi et al. (2024), for several catchments, including the Ahr. In all of these studies, mHM handles the rainfall–runoff transformation and channel routing across the catchment, while RIM2D is restricted to the lower, inundation-prone reaches of the river network. Coupled approaches have proven valuable and have enabled large-scale flood risk assessments and forecasting applications (Falter et al., 2016; Najafi et al., 2024). However, this approach also introduces additional sources of uncertainty and complexity. The hydrological and hydrodynamic models need to be set up, parameterised, and calibrated separately, with hydrological models in particular requiring the calibration of a large number of internal parameters governing infiltration, evapotranspiration, soil moisture and routing, typically constrained only by gauge discharge time series whose availability is often limited and which constrain the integrated catchment response without uniquely identifying the spatial pattern of runoff generation. Moreover, errors in the hydrological component can propagate into the inundation simulation, while the coupling itself is typically one-way and cannot accommodate feedbacks such as backwater effects, channel-floodplain exchange, or surface ponding influencing local runoff generation. A further consequence is a substantial resolution gap between the two stages: hydrological models commonly operate at sub-catchment or even kilometer-scale resolutions, whereas the hydrodynamic component requires metre-scale grids, so the fine-scale spatial structure of the rainfall forcing which is particularly important for the highly localised, convective precipitation patterns that drive flash floods is partially averaged out before reaching the inundation stage. Overall, the spatial structure of runoff generation, overland flow, and channel routing is fragmented

https://doi.org/10.5194/egusphere-2026-3652
Preprint. Discussion started: 8 July 2026

across two distinct modelling frameworks with different assumptions, resolutions, and process representations. It must be acknowledged that fully physically-based approaches using the integrated surface-subsurface hydrological model ParFlow have also been explored for this catchment and event. Saadi et al. (2023); Goergen et al. (2025). Our approach differs strongly in the target resolutions, which are at the scale of flood impact, with coarsest grids at 10 m resolution, whereas Goergen et al. (2025) and Saadi et al. (2023) computed at 611 m resolution, albeit with a larger domain encompassing the entire Eifel region and computing also subsurface fluxes.

Recent advances in multi-GPU hydrodynamic solvers raise the question of whether such a coupling is still strictly necessary for catchment-scale flood simulation. In particular, fully dynamic 2D SWE models such as SERGHEI (Caviedes-Vouillième et al., 2023), when deployed on multi-GPU HPC systems, are now capable of simulating both rainfall-runoff and inundation processes within a single, physically consistent framework, even across large catchments and at very high spatial resolutions (Caviedes-Vouillième et al., 2020; Khosh Bin Ghomash et al., 2026). A direct rain-on-grid hydrodynamic approach offers several conceptual and practical advantages over the traditional coupled chain. First, runoff generation, overland flow, channel concentration, streamflow, and floodplain inundation are simulated by the same set of physical equations on the same computational grid, removing the need to transfer fluxes between two separate models with potentially inconsistent assumptions and resolutions. Second, the spatial structure of the rainfall forcing is preserved throughout the modelling chain, which can be critical in flash flood events driven by highly localised, convective precipitation such as the 2021 Ahr event (Mohr et al., 2023). Third, the modelling chain itself is simplified, reducing the number of parameters that need to be calibrated and the number of model interfaces at which errors can accumulate. Fourth, by avoiding the lumped or semi-distributed conceptualisations typically employed in hydrological models, rain-on-grid hydrodynamic simulations can better capture the highly nonlinear and rapid response of small headwater catchments to extreme rainfall. Finally, from an operational perspective, a single integrated model is easier to drive directly with radar-based precipitation products and numerical weather prediction outputs, which supports a tighter integration into early warning systems.

In this study, we explore whether such a direct, catchment-scale rain-on-grid hydrodynamic approach is feasible and advantageous for fluvial flood simulation in mid-sized catchments, using the 2021 Ahr Valley flood as a case study. We employ the multi-GPU shallow-water solver SERGHEI (Caviedes-Vouillième et al., 2023), forced by 5-minutely, 1km resolution RADOLAN radar precipitation data from the German Weather Service, to simulate the entire Ahr catchment (approximately 900 km^2) at spatial resolutions ranging from 10 down to 2 m. We assess the model's ability to reproduce the observed flood extent, water depths, and channel hydrographs without resorting to an intermediate hydrological model. The objective is to evaluate whether high-resolution, catchment-scale, fully hydrodynamic flood simulation has reached a level of maturity that justifies its use as a standalone alternative to the traditional coupled hydrological-hydrodynamic chain for catchment-scale flood simulation, hazard mapping, and early warning.

https://doi.org/10.5194/egusphere-2026-3652
Preprint. Discussion started: 8 July 2026

2 Methods

2.1 Hydrodynamic model: SERGHEI

SERGHEI (Simulation EnviRonment for Geomorphology, Hydrodynamics, and Ecohydrology in Integrated form) is a performance-portable, multi-GPU shallow-water solver designed for hydrology and environmental hydraulics applications across a wide range of scales (Caviedes-Vouillième et al., 2023). SERGHEI solves the fully dynamic two-dimensional shallow-water equations, retaining all advective and inertial terms allowing a full representation of flow dynamics under transient and rapidly varying conditions, including the steep wavefronts, supercritical regions, and complex flow interactions that are characteristic of flash flood events such as the 2021 Ahr Valley flood (Khosh Bin Ghomash et al., 2024).

The fully dynamic shallow-water equations solved by SERGHEI can be written in conservative form as:

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} + \frac{\partial \mathbf{G}}{\partial y} = \mathbf{S}_b + \mathbf{S}_f, \tag{1}$$

with the conserved variables, fluxes, and source terms defined as:

$$\mathbf{U} = \begin{bmatrix} h \\ q_x \\ q_y \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} q_x \\ \frac{q_x^2}{h} + \frac{1}{2}gh^2 \\ \frac{q_x q_y}{h} \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} q_y \\ \frac{q_x q_y}{h} \\ \frac{q_y^2}{h} + \frac{1}{2}gh^2 \end{bmatrix}, \tag{2}$$

$$\mathbf{S}_b = \begin{bmatrix} 0 \\ -gh\frac{\partial z}{\partial x} \\ -gh\frac{\partial z}{\partial y} \end{bmatrix}, \quad \mathbf{S}_f = \begin{bmatrix} 0 \\ -\sigma_x \\ -\sigma_y \end{bmatrix}. \tag{3}$$

Here, h denotes the water depth [L], q_x and q_y are the unit-width discharge components [L^2/T] in the Cartesian directions, z is the bed elevation [L], and g is the gravitational acceleration [L/T^2]. The friction slopes σ_x and σ_y are evaluated using Manning's equation.

SERGHEI discretises the shallow-water equations using a finite-volume scheme on a structured Cartesian grid, with fluxes computed via a Riemann solver, ensuring conservation of mass and momentum, robust shock-capturing, and a well-balanced treatment of wetting and drying over irregular topography (Caviedes-Vouillième et al., 2023). Time integration is performed explicitly under the CFL stability condition. The model is implemented in C++ with a hybrid MPI–Kokkos parallelisation strategy (Trott et al., 2022) that enables performance-portable deployment across multi-core CPUs and CUDA-enabled NVIDIA GPUs without code modifications, allowing the model to scale across hundreds of GPUs to handle the hundreds of millions of computational cells that arise in catchment-scale, meter-resolution flood simulations (Caviedes-Vouillième et al., 2023).

For rainfall-driven catchment-scale flood simulation, SERGHEI supports spatially distributed precipitation forcing applied directly to the computational grid as a source term in the continuity equation (the so-called rain-on-grid approach). Thus,

https://doi.org/10.5194/egusphere-2026-3652
Preprint. Discussion started: 8 July 2026

SERGHEI directly simulates runoff generation, overland flow, streamflow concentration, and floodplain inundation within a single physically consistent hydrodynamic framework. Infiltration losses can be incorporated through infiltration models (currently user-defined infiltration rates or the Horton infiltration model), and surface roughness is parameterised through spatially distributed Manning coefficients, which can be in turn informed by land cover. SERGHEI is distributed as open-source software under a three-clause BSD license.

While this study employs only the SERGHEI-SWE module, SERGHEI is developed as a modular ecohydrological framework in which additional process modules can be coupled within the same multi-GPU infrastructure. A physically based, variably-saturated subsurface flow solver (SERGHEI-RE; Li et al. 2025) and a coupled surface–subsurface configuration (SERGHEI-SWE-RE; Zheng et al. 2026) are already available, providing a pathway to more complete representations of infiltration, soil moisture dynamics, and shallow groundwater that go beyond the simple infiltration-capacity formulation used in this work. The present study is deliberately restricted to the SWE module to isolate the question of whether catchment-scale hydrodynamic surface flow simulation (with an infiltration model), on its own, is sufficient for fluvial flash flood reproduction; the broader ecohydrological modules of SERGHEI are the natural pathway to extend the approach to events in which subsurface and vegetation processes are more influential.

2.2 Case Study

This study focuses on the Ahr catchment, located in the federal states of Rhineland-Palatinate and North Rhine-Westphalia in west Germany. The Ahr River is an 86 km long river, draining a catchment area of approximately 900 km^2 in the Eifel low-mountain region before joining the Rhine near the town of Sinzig. The catchment is characterised by a complex and high relief topography, with steep, forested hillslopes, narrow tributary valleys, and a confined main valley that gradually widens toward its downstream end. Elevations range from approximately 50 m a.s.l. near the confluence with the Rhine to over 700 m a.s.l. in the upper headwaters. Land cover is dominated by forest and agriculture in the upper and mid-catchment, while the valley hosts a series of towns and villages, including Altenahr, Mayschoss, Rech, Dernau, Bad Neuenahr-Ahrweiler, and Sinzig. The catchment has a temperate oceanic climate with an average annual precipitation of around 675 mm (Truedinger et al., 2023), but the combination of steep slopes, narrow valleys, and a high degree of stream-network confinement makes it particularly susceptible to rapid runoff concentration and flash flooding during extreme rainfall events (Roggenkamp and Herget, 2022).

In mid-July 2021, the slow-moving low-pressure system “Bernd” caused exceptionally heavy rainfall over large parts of western and central Europe, with the Eifel region among the most severely affected areas (Schäfer et al., 2021; Mohr et al., 2023). Within approximately 24 hours, several rainfall stations in and around the Ahr catchment recorded rainfall accumulations exceeding 100 mm, triggering a catastrophic flash flood that propagated rapidly through the Ahr Valley. Water levels in the Ahr River reached extreme values, with peak water depths estimated at around 10 m at the Altenahr gauge (Mohr et al., 2023), although the exact peak water levels remain uncertain as most gauging stations along the Ahr were damaged or destroyed during the event. The flood caused widespread devastation throughout the valley, with extensive damage to buildings, transportation infrastructure, and critical services, and accounted for the majority of the more than 180 flood-related fatalities recorded in Germany during the event (Truedinger et al., 2023; Thieken et al., 2023). The 2021 Ahr Valley flood represents

https://doi.org/10.5194/egusphere-2026-3652
Preprint. Discussion started: 8 July 2026

an ideal case study for evaluating catchment-scale, high-resolution rain-on-grid hydrodynamic flood simulation the catchment is sufficiently large to challenge the computational capabilities of fully dynamic 2D shallow-water solvers while remaining well-defined and bounded. Moreover, given the magnitude and societal relevance, the event is exceptionally well documented, with reconstructed inflow hydrographs, an officially mapped flood extent and post-event high-water measurements available for model validation. Figure 1 shows the extent of the Ahr catchment, its topography, the observed flood extent between Altenahr and Sinzig, location of the measured high water marks, and the location of the Altenahr and Bad Bodendorf gauges.

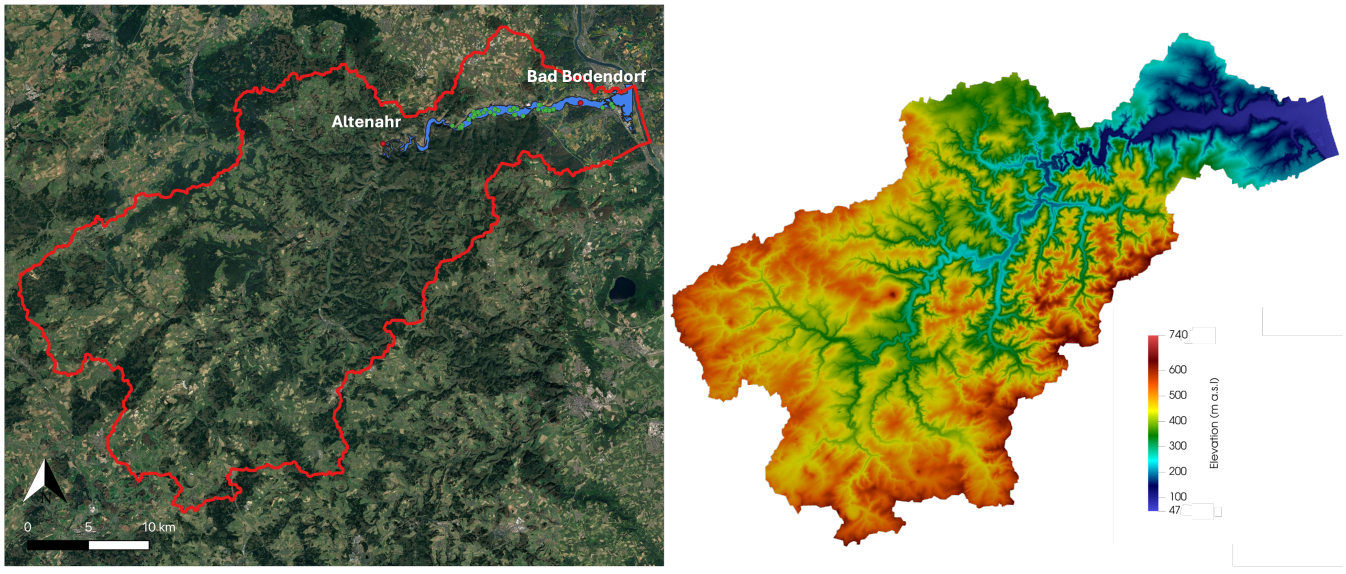


Figure 1. Right panel: topography of the Ahr catchment. Left panel: the red line marks the boundary of the Ahr catchment and the simulation domain, the red dots indicate the locations of the Altenahr and Bad Bodendorf gauges, the green dots show the 45 observed water marks used for model validation, and the blue area represents the observed flood extent during the July 2021 event. Satellite imagery: © Google Earth 2026.

2.3 Data and model set-up

Elevation in the domain is represented by the DGM1 digital elevation model (DEM), available with a resolution $dx = 1$m and provided by the German Federal Agency for Cartography and Geodesy (BKG). In order to examine the behaviour of SERGHEI across a range of spatial resolutions and to manage computational demand, the original 1 m DEM was aggregated through averaging to grid resolutions of 2, 5, and 10 m. During the simulations, only the cells within the catchment boundary (red line in Figure 1 were valid computational cells. This yielded approximately 225 million cells at 2 m, 36 million cells at 5 m, and 9 million cells at 10 m resolution. Building footprints obtained from OpenStreetMap were removed from each of the four DEMs, so that building outlines act as internal reflective barriers to the flow during the simulations. This approach has been shown to provide the best flood simulation results in terms of flood extent and water depths in the Ahr Valley (Khosh Bin Ghomash et al., 2025c). An initial water depth of 1 m was assigned to all cells along the main Ahr channel at the start of each simulation,

https://doi.org/10.5194/egusphere-2026-3652
Preprint. Discussion started: 8 July 2026

consistent with the typical low-flow water depths of the Ahr River (Apel et al., 2022), ensuring a physically plausible initial state for the model. This ensures that the channel is not initialised in a dry state, so that the flood wave propagates over an already wetted bed. The channel bathymetry is likely not well captured by the DTM (due to limitations of LiDAR technology over wet surfaces), and thus there are errors associated to both the channel geometry and the initial state. However, since the flood wave is extreme, the potential buffering effect of filling the real channel bathymetry and the initial state likely have only a very small impact, and mostly in the initial times of the simulation. Thus, the key aspect is only to guarantee an initially wet channel.

Surface roughness was parameterised through spatially distributed Manning roughness coefficients, assigned according to the 2020 Sentinel-2 based land cover classification of Germany (Riembauer et al., 2021). The classification draws on atmospherically corrected Sentinel-2 imagery processed via the MAJA algorithm, distributed by the EOC Geoservice of the German Aerospace Centre (DLR), and complemented by reference training data from sources such as OpenStreetMap. This product was selected primarily for its 10 m native spatial resolution, which is well-suited to catchment-scale hydrodynamic modelling. Following this classification, Manning roughness coefficients of 0.04, 0.03, 0.1, 0.025, 0.035, and 0.035 $\mathrm{m}^{-1/3}\mathrm{s}$ were applied to the classes vegetation, water bodies, forest, built-up areas, bare soil, and agricultural land, respectively.

Infiltration was represented by setting heterogeneous, but constant (in time) infiltration capacities. This is the most parameter-lean approach, requiring a single parameter per infiltration class. Infiltration capacities were derived by combining two datasets. The soil texture distribution was obtained from the Soil Map of Germany 1:200,000 (BÜK200) at a 150 m grid resolution, published by the Federal Institute for Geosciences and Natural Resources (Federal Institute for Geosciences and Natural Resources (BGR), 2020). The BÜK200 organises soil information through mapping units, each characterised by a vertical sequence of horizons (discrete texture layers defined by depth intervals), with sand fraction, clay fraction, and bulk density recorded for each horizon. For each horizon, the sand and clay fractions were classified into a USDA soil texture class (Soil Survey Staff, 1999) and a literature-based saturated hydraulic conductivity (K_s) value was assigned following the pedotransfer relationships of Saxton and Rawls (2006) and Schaap et al. (2001). Horizon-specific K_s values were then aggregated over the soil profile using a depth-weighted harmonic mean, which is the appropriate operator for vertical flow through a series of layers under Darcy's law, as the flux is controlled by the least-permeable horizon and the harmonic mean weights low-K_s layers more strongly than the arithmetic mean would (Zhu and Mohanty, 2002; Sobieraj et al., 2002). To represent the influence of urban sealing on infiltration, the Imperviousness Density raster from the Copernicus Land Monitoring Service (European Environment Agency, 2020), available at 10 m resolution, was used to estimate the fraction of pervious surface in each grid cell. The final effective infiltration capacity for each grid cell was obtained following

$$I_{cap} = K_s\,(1 - f_i) \tag{4}$$

where $f_i \in [0,1]$ is the imperviousness density of the cell retrieved from the Copernicus raster, such that fully sealed cells ($f_i = 1$) are represented by zero infiltration capacity and fully pervious cells ($f_i = 0$) retain the unmodified soil K_s.

https://doi.org/10.5194/egusphere-2026-3652
Preprint. Discussion started: 8 July 2026

Precipitation forcing was provided by the 5-minute RADOLAN radar product of the German Weather Service (DWD), with national coverage at 1000 × 1000 m spatial resolution. The radar fields for 12–15 July 2021 (96 hours = 1152 timesteps) were used to force the model, with the precipitation intensity applied directly on the computational grid as a spatially and temporally varying source term in the continuity equation (i.e., rain-on-grid approach). This RADOLAN product was selected as the precipitation forcing because it represents the highest spatial (1 km) and temporal (5 min) resolution radar-based precipitation product operationally available for Germany (Winterrath et al., 2018). The fine temporal resolution is particularly relevant for the present application, as sub-hourly intensity peaks associated with convective rainfall are substantially smoothed out when aggregated to hourly intervals, leading to significant underestimation of peak rainfall intensities driving flash floods (Lengfeld et al., 2021; Mohr et al., 2023; Khosh Bin Ghomash et al., 2023).

While SERGHEI has previously been applied to and validated against a number of flood scenarios (Caviedes-Vouillième et al., 2023), including the 2021 Ahr Valley event itself (Khosh Bin Ghomash et al., 2024), the simulations presented here are deliberately configured using a blind approach, i.e., without any tuning or calibration of roughness coefficients nor infiltration capacities. This decision is the consequence of the broader aim of this study to evaluate whether catchment-scale, fully hydrodynamic rain-on-grid simulation is feasible as a standalone alternative to the traditional coupled hydrological–hydrodynamic chain, under conditions that are realistic for operational forecasting deployment. In operational contexts, calibration data are rarely available in sufficient quantity or quality especially for comparatively smaller catchments, which are the ones typically affected by flash floods. Moreover, detailed event-specific tuning is not an option when models need to be deployed rapidly, or across many catchments, and evidently simply unfeasible for a future (perhaps unprecedented) extreme event. Relying on standard parameterisations derived from spatial datasets is, in practice, the most realistic (and demanding) baseline. While calibration would undoubtedly improve quantitative agreement with observations, we deliberately refrain from such tuning here in order to assess the predictive capability of the uncalibrated, rain-on-grid SERGHEI setup on its own terms.

2.4 Model assessment

The simulated flood is evaluated against three observational datasets covering complementary aspects of the event: the officially mapped flood extent provided by the State Agency for the Environment of Rhineland-Palatinate (LfU), 45 post-event high-water marks reported by residents and surveyed across the inundated lower valley, and the reconstructed stage hydrograph at the Altenahr gauge (Mohr et al., 2023).

To assess model skill in terms of flood extents, a set of binary classification metrics widely used in hydrodynamic model evaluation was applied (Wing et al., 2017; Khosh Bin Ghomash et al., 2025a). Each cell within the assessment area was classified into one of four categories according to Table 1, with SERGHEI outputs treated as the simulated state and the LfU flood extent treated as the observation. The resulting counts of true positives, false negatives, false positives, and true negatives were then aggregated into compute five domain-wide performance indicators (Table 2): the Critical Success Index (CSI), Hit Rate (HR), False Alarm ratio (FA), Error Bias (EB), and Bias Percentage Indicator (BPI).

An important consideration in this context is that SERGHEI simulates both pluvial and fluvial flooding processes simultaneously, whereas the LfU flood extent represents the fluvial flooding observed along the Ahr river. Including pluvially-flooded cells away

https://doi.org/10.5194/egusphere-2026-3652
Preprint. Discussion started: 8 July 2026

Table 1. Inundation confusion matrix. Each cell within the assessment area is compared between the SERGHEI simulation and the LfU flood extent and classified accordingly.

		Simulated	
		Wet	Dry
Observed	Wet	True Positive (TP)	False Negative (FN)
	Dry	False Positive (FP)	True Negative (TN)

Table 2. Flood inundation performance metrics used for evaluating the SERGHEI simulations against the LfU observed flood extent.

Metric	Equation	Poor	Perfect	Description
Critical Success Index (CSI)	$\frac{TP}{TP+FP+FN}$	0	1	Ratio of correctly predicted wet cells to the total number of wet and missed wet cells.
Hit Rate (HR)	$\frac{TP}{TP+FN}$	0	1	Proportion of observed wet cells correctly identified by the simulation.
False Alarm (FA)	$\frac{FP}{TP+FP}$	1	0	Proportion of predicted wet cells that were not actually wet.
Error Bias (EB)	$\frac{FP}{FN}$	0 or ∞	1	Ratio of overpredicted wet cells to underpredicted wet cells.
Bias Percentage Indicator (BPI)	$100\left(\frac{TP+FP}{TP+FN}-1\right)$	-100 or 100	0	Relative percentage error in the total extent of the predicted flood area.

from the valley and its floodplains (i.e., disconnected from the stream) in the comparison would unfairly penalise the model and obscure its actual performance in reproducing the fluvial) flood extent that the LfU map captures. To address this, skill metric computation was restricted to an area encompassing a 1 km buffer on either side of the main Ahr river channel. This buffer comfortably encloses the entire observed inundation extent and includes a substantial margin of additional floodplain area beyond it, ensuring that the assessment is not biased by truncation of the predicted flood footprint, while still excluding pluvial wet (flooded) cells on the upland hillslopes of the catchment.

In addition to the comparison of flood extents, the simulated maximum water depths were evaluated against the 45 high-water marks documented across the lower valley after the event. These water marks were reported by residents and subsequently surveyed and quality-controlled using reference objects such as building facades, lampposts, curb stones, and bridge structures visible in post-event photographs and field inspections (Khosh Bin Ghomash et al., 2025a). For each water mark, the simulated maximum water depth was extracted from the SERGHEI output at the corresponding grid cell, and the resulting set of paired simulated-observed depths was summarised using the root mean square error (RMSE) and the mean bias:

https://doi.org/10.5194/egusphere-2026-3652
Preprint. Discussion started: 8 July 2026

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}, \qquad \mathrm{Bias} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i), \tag{5}$$

where n is the number of water marks, $\hat{y}_i$ is the simulated water depth at the location of water mark i, and y_i is the corresponding observed water depth.

Finally, the in-channel hydrodynamics were evaluated against the reconstructed stage hydrograph at the Altenahr gauge. This gauge was destroyed during the event, and the available hydrograph is a reconstruction by the LfU using a combination of post-event evidence and hydraulic modelling (Mohr et al., 2023). Despite the reconstructed nature, this stage hydrograph remains the only available approximations of the temporal evolution of the flood wave at a fixed point along the river and therefore provides a valuable reference for evaluating the simulated hydrograph in terms of timing, peak magnitude, and overall shape.

3 Results and discussion

3.1 Simulated flood extent and resolution sensitivity

The ability of the rain-on-grid approach to reproduce the observed inundation pattern was evaluated by comparing simulated maximum water depths against the official post-event flood extent map provided by the State Agency for the Environment of Rhineland-Palatinate (LfU). Figure 2 shows the maximum water depths obtained with SERGHEI at dx = 2 m, 5 m, and 10 m. Only cells with a maximum water depth of at least 0.01 m are displayed to facilitate visualization and interpretation. The overall inundation pattern is highly consistent across the three resolutions, with extensive flooding along the main Ahr Valley and its major tributaries, including the populated settlements around Altenahr, Mayschoss, Dernau, and Bad Neuenahr-Ahrweiler, reproduced in similar detail. While the 2 m simulation resolves finer overland flow paths, narrow tributary channels, and small-scale topographic effects such as embankments, road cuttings, and individual valley constrictions, the broad-scale flood extent and the general distribution of inundated zones remain robust even at 10 m resolution.

The quantitative performance assessment was carried out within the 1 km buffer along the main Ahr river channel, considering only cells with simulated or observed water depths of at least 0.10 m (the rationale is explained in subsection 2.4). The results are summarised in Figure 3. The Critical Success Index (CSI) ranges from 0.875 at 2 m to 0.861 at 5 m and 0.846 at 10 m resolution, indicating very good agreement with the official LfU map across all three grids. The Hit Rate (HR) is essentially insensitive to resolution and remains above 0.955 throughout, while the complement of the False Alarm ratio (1−FA) decreases moderately from 0.912 at 2 m to 0.878 at 10 m. The Error Bias (EB) and Bias Percentage Indicator (BPI) consistently show a moderate overprediction of the inundated area, which grows with coarsening resolution: EB increases from 2.085 at 2 m to 3.245 at 10 m, and BPI from 4.79 % to 9.23 %.

https://doi.org/10.5194/egusphere-2026-3652
Preprint. Discussion started: 8 July 2026

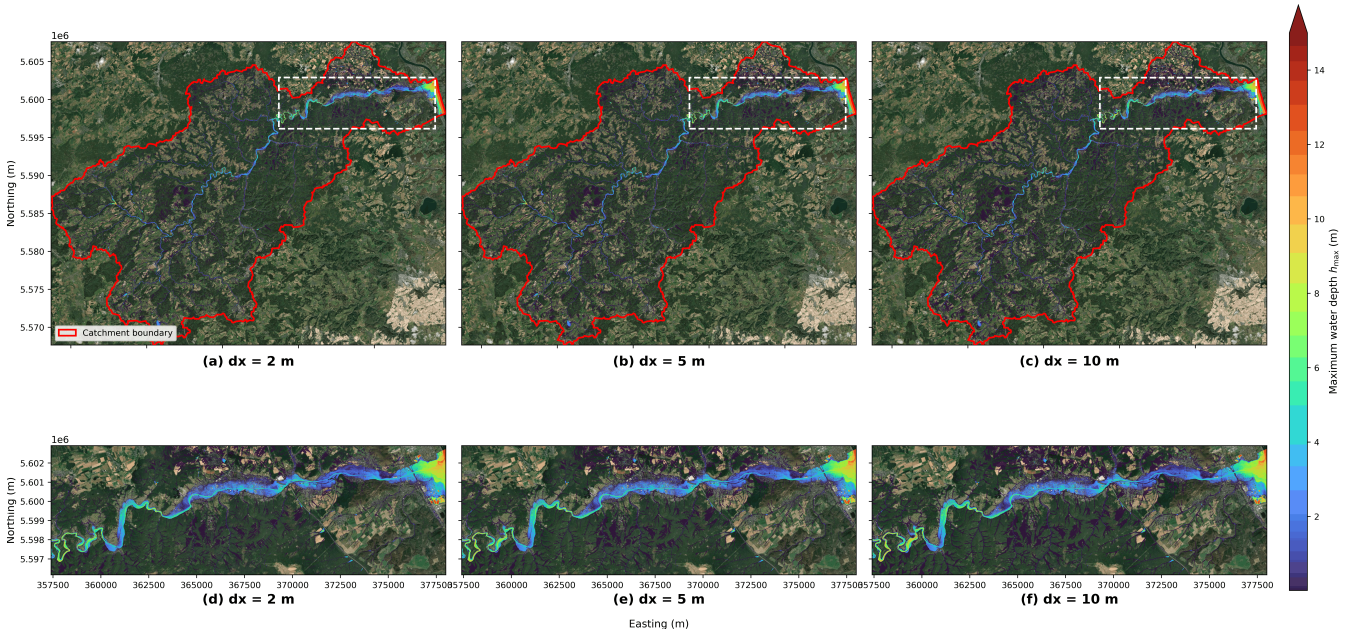


Figure 2. Maximum water depth ($h_{\max}$) simulated with SERGHEI at dx = 2 m (a, d), 5 m (b, e), and 10 m (c, f) for the 2021 Ahr flood. Top row (a–c): full catchment overview; the red line shows the catchment boundary and the white dashed rectangle indicates the extent of the zoomed panels. Bottom row (d–f): zoomed view of the main Ahr Valley inundation corridor. Only cells with $h_{\max} \geq 0.01$ m are shown. Basemap: World Imagery, Source: Esri, Maxar, Earthstar Geographics, and the GIS User Community | Powered by Esri.

The pattern revealed by these metrics is internally consistent and physically interpretable. The near-constant HR across resolutions indicates that almost all of the observed inundated cells are correctly captured by all simulations. The progressive decline in CSI and $1-\mathrm{FA}$ with coarsening resolution, combined with the increase in EB and BPI, is driven by a moderate but growing tendency to over-extend the predicted flood beyond the observed limits as the grid is coarsened. This behaviour is well documented in 2D hydrodynamic flood modelling (Horritt and Bates, 2001; Fewtrell et al., 2008) and reflects two compounding effects of grid coarsening: (i) the smoothing of fine-scale topographic barriers such as levees, embankments, road embankments, and narrow valley constrictions (Caviedes-Voullième et al., 2012), which on the natural terrain confine the flow but become partially averaged out at coarser resolutions, and (ii) the increased numerical diffusion of the flood wave on coarser grids, which spreads water over a slightly larger footprint than the physical solution would suggest (Khosh Bin Ghomash et al., 2024). Importantly, this overprediction is of moderate magnitude (BPI below 10 % even at 10 m) and is biased towards false positives rather than false negatives, which from an operational and impact-based forecasting perspective is the less problematic of the two error types: it favours warnings on the side of caution rather than missing exposed areas.

The skill scores achieved here compare well with previously reported performance for high-resolution 2D hydrodynamic simulations of the 2021 Ahr flood, all of which relied on either reconstructed discharge boundaries or the classical coupled hydrological–hydrodynamic chain. Apel et al. (2022) reported a CSI of 0.845 for the RIM2D simulation of the lower Altenahr–Sinzig reach (approximately 30 km) driven by the reconstructed stage hydrograph at Altenahr. Vorogushyn et al. (2025) coupled

https://doi.org/10.5194/egusphere-2026-3652
Preprint. Discussion started: 8 July 2026

the mesoscale Hydrologic Model (mHM) with RIM2D for the spatial counterfactual analysis of the same event, adopted the same calibrated and validated RIM2D setup of Apel et al. (2022) and inherited its CSI of 0.845 as the reference baseline for the unperturbed (factual) 2021 simulation. Najafi et al. (2024) further coupled the RIM2D configuration downstream of an mHM driven by ICON-D2-EPS precipitation (weather) forecasts to generate an impact-based early warning system emphasizing forecast lead time and exceedance probabilities rather than reporting a single CSI value for the inundated area. In this context, the CSI values of 0.846–0.875 obtained here directly with a catchment-scale rainfall-runoff SERGHEI-SWE simulation (without any intermediate hydrological model, reconstructed inflow hydrograph, or parameter calibration) are essentially indistinguishable from the CSI of 0.845 reported by Apel et al. (2022) and Vorogushyn et al. (2025) for the coupled mHM–RIM2D chain. This is a key result of the present study: at catchment scale and at meter-scale resolutions, a fully hydrodynamic, rain-on-grid simulation is not only feasible, but delivers a flood-extent skill on par with the established hydrological-hydrodynamic chain approaches that have been considered the standard for the 2021 Ahr event so far.

It is also useful to compare the present results with those of Khosh Bin Ghomash et al. (2024), who applied SERGHEI-SWE to the lower Altenahr–Sinzig reach –identical to that of Apel et al. (2022), using the same Manning roughness parameterisation used in this study, but with the model driven by the reconstructed stage hydrograph at Altenahr rather than by rainfall. In that configuration, SERGHEI achieved CSI values of 0.849, 0.854, and 0.853 at 10, 5, and 2 m resolution, respectively, over the approximately 30 km reach. The CSI values of 0.846–0.875 obtained here are thus essentially equivalent, and at 2 m resolution even exceed the hydrograph-forced baseline, despite the present simulations resolving the entire rainfall–runoff–inundation chain across the full $\sim$900 km^2 catchment. This close agreement further reinforces the central finding of this study: a standalone rain-on-grid hydrodynamic approach can deliver flood-extent skill that is fully comparable to that of the traditional coupled hydrological–hydrodynamic modelling chain, even when applied across a domain roughly an order of magnitude larger.

From an operational standpoint, the relatively small degradation in skill when coarsening the grid from 2 m to 10 m (CSI drop of less than 0.03 and HR essentially unchanged) is particularly relevant. The 10 m configuration captures the fluvial flood extent of the 2021 Ahr event with a level of agreement that, in most operational contexts, would be considered fully acceptable, with a significantly reduced computational cost relative to the 2 m setup (for details see subsection 3.4). This supports the use of the rainfall-forced hydrodynamic modelling approach at computationally efficient resolutions for operational forecasting deployment, while finer resolutions are able to resolve flow paths through individual streets and around building footprints, providing the spatial detail needed for impact-based assessment where identifying which specific structures and infrastructure are affected matters most. Crucially, such detailed simulations are now computationally feasible at the catchment scale, as demonstrated by the 2 m configuration completing the full 96-hour event in under four hours (see subsection 3.4).

https://doi.org/10.5194/egusphere-2026-3652
Preprint. Discussion started: 8 July 2026

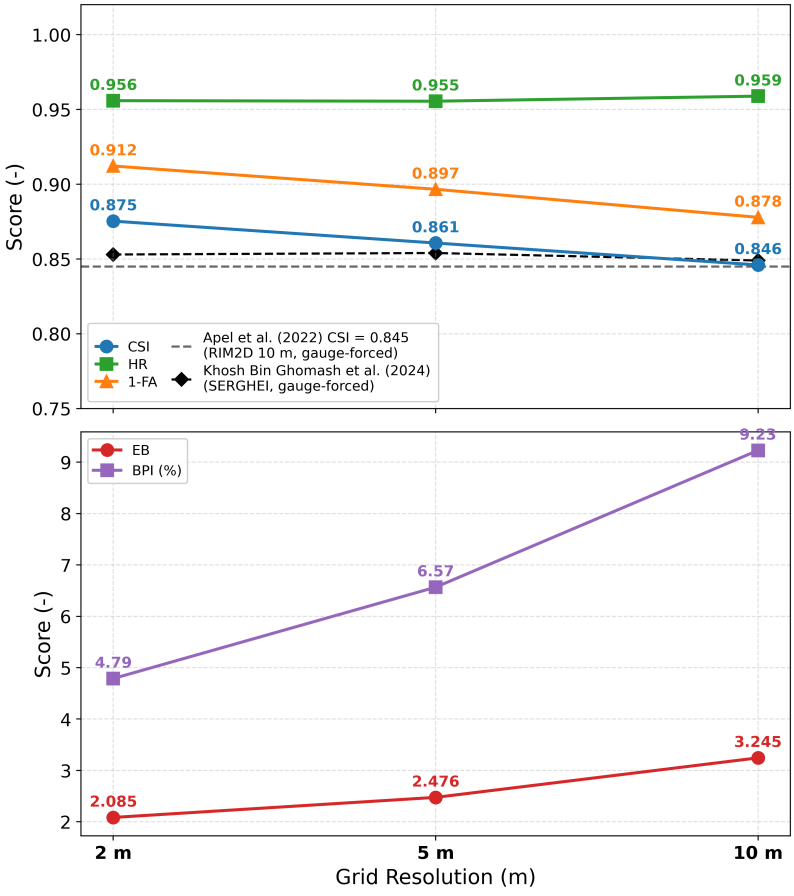


Figure 3. Flood extent performance metrics against the observed flood map, shown as a function of grid resolution. Skill scores of CSI = Critical Success Index, HR = Hit Rate, and 1−FA = 1 – False Alarm ratio, EB = Error Bias and BPI = Bias Percentage Indicator are presented. Metrics were calculated within a 1 km buffer along the main Ahr river channel using a minimum water depth threshold of 0.10 m. The horizontal dashed line marks the CSI of 0.845 reported by Apel et al. (2022) for the gauge-forced RIM2D setup of the lower Ahr Valley, and the dashed line with diamond markers shows the resolution-specific CSI values of Khosh Bin Ghomash et al. (2024) for the gauge-forced SERGHEI setup of the lower Ahr Valley.

3.2 Water depths at high-water marks

In addition to the comparison of inundation extents, the simulated peak water depths were evaluated at the 45 high-water marks documented across the Ahr Valley after the event. For each watermark, the simulated maximum water depth ($h_{\max}$) was extracted from the corresponding grid cell of the SERGHEI output, and the resulting set of paired simulated and observed depths was summarised through the BIAS, the root mean square error (RMSE), and the Pearson correlation coefficient r. Figure 4 presents the resulting scatter plots of simulated versus observed water depths at dx = 2, 5, and 10 m. All three simulations track the 1:1 line closely across the full range of observed depths (approximately 0.3–4 m), with the vast majority

https://doi.org/10.5194/egusphere-2026-3652
Preprint. Discussion started: 8 July 2026

of points falling within the ±0.5 m envelope and Pearson correlation coefficients of r = 0.88, 0.88, and 0.86 at dx = 2, 5, and 10 m, respectively. The BIAS values indicate a slight overall over-prediction of water depths at all three resolutions, ranging from +0.05 m at 2 m to +0.18 m at 5 m and +0.10 m at 10 m, while RMSE values increase modestly from 0.46 m at 2 m to 0.52 m at 5 m and 0.57 m at 10 m. The progressive increase in RMSE with coarsening resolution is expected, as coarser grids progressively smooth out the fine-scale topographic features (channel banks, road embankments, narrow constrictions) that locally control peak water depths in the valley, while the BIAS remains close to zero throughout, indicating no systematic over- or underestimation of flood severity at the catchment scale.

These results compare very well with previous high-resolution hydrodynamic simulations of the 2021 Ahr flood, all of which relied on either a reconstructed discharge boundary or a coupled hydrological–hydrodynamic chain rather than direct rainfall forcing. Apel et al. (2022) validated the original RIM2D setup of the Altenahr–Sinzig reach against the surveyed high-water marks and reported a BIAS of +0.09 m and an RMSE of 0.30 m, with the model driven by the reconstructed water depth hydrograph at Altenahr. The BIAS values obtained here (+0.05 to +0.18 m) are essentially indistinguishable from the +0.09 m of Apel et al. (2022), indicating that the rain-on-grid SERGHEI setup reproduces the maximum inundation marks response at the lower valley as accurately as the gauge-forced RIM2D benchmark. The somewhat higher RMSE of the present simulations relative to the 0.30 m of Apel et al. (2022) is expected, since (i) the present analysis covers the entire 900 km^2 catchment rather than the lower 30 km reach, which introduces greater complexity in terms of the different processes involved and the need to resolve runoff generation and routing across the full drainage network, (ii) the simulations are driven by spatially distributed precipitation rather than a gauge-reconstructed hydrograph that effectively absorbs all upstream (catchment and workflow-wise)y uncertainties before the simulation begins, and Given these compounding factors, an RMSE that is only 0.16–0.27 m higher than the gauge-forced benchmark is a good result for a fully standalone catchment-scale rain-on-grid simulation, and reinforces the conclusion drawn from the flood-extent analysis: at the spatial scales and event types considered here, a multi-GPU shallow-water solver driven directly by rainfall delivers water-depth skill that is competitive with the traditional coupled-chain approach.

https://doi.org/10.5194/egusphere-2026-3652
Preprint. Discussion started: 8 July 2026

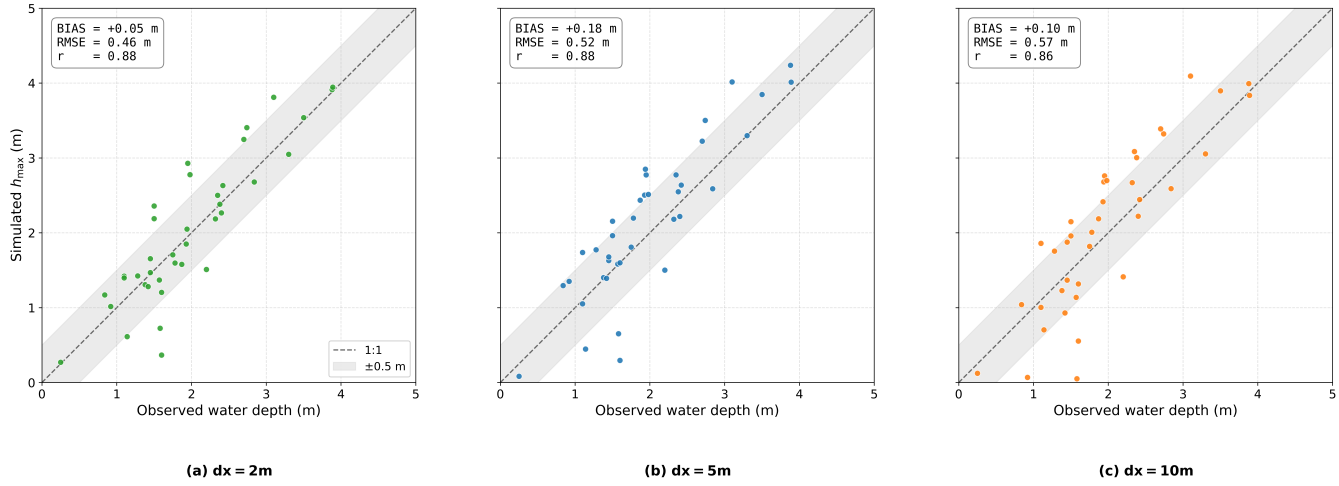


Figure 4. Simulated versus observed peak water depths at the documented high-water marks across the Ahr Valley for the SERGHEI rain-on-grid simulations at (a) 2 m, (b) 5 m, and (c) 10 m resolution. The dashed line shows the 1:1 reference, the shaded band the ±0.5 m envelope. Inset values report the overall bias (BIAS), root-mean-square error (RMSE), and Pearson correlation coefficient (r) between simulated and observed depths for each resolution.

The simulated and observed stage hydrograph at the Altenahr gauge are compared in Figure 5. All three resolutions reproduce the overall shape of the reconstructed hydrograph reasonably well, including the rapid rising limb associated with the main rainfall pulse on 14 July and the sharply peaked maximum. The recession behavior, however, varies across resolutions: the 5 m simulation captures the falling limb most faithfully, while both the 2 and 10 m simulations deviate more noticeably. This non-monotonic resolution dependence suggests that the recession is governed by how the grid resolves the drainage pathways controlling post-peak floodplain emptying, rather than by grid refinement alone. Before the arrival of the flood wave, the simulated water levels at the Altenahr cell remain essentially constant at the level associated with the prescribed 1 m initial channel depth, reflecting the absence of any antecedent flow in the rain-on-grid setup. This initial stage differs between resolutions purely because the bed elevation extracted from each DEM differs by up to about 1 m at the gauge cell, which was used to convert water depth to water level (in turn showing also the relevance of resolution to resolve the stream itself). A small precursor pulse, approximately 1 m above the baseline channel level, appears in the 5 m simulation around $t \approx 165{,}000$ s in response to the first moderate rainfall pulse on 13 July, and is less pronounced at the other resolutions. This precursor is well below the operationally relevant threshold for the event and does not really affect the simulation of the main flood wave. At the peak, the 10 m simulation reaches 169.0 m a.s.l., just under the observed peak of 170.7 m a.s.l. while the 2 m simulation slightly under-predicts (169.9 m a.s.l., +0.9 m above the 10 m value but still below the observation) and the 5 m simulation produces the largest peak (172.5 m a.s.l., +1.8 m above the observation). All three simulations also exhibit timing leads of the simulated peak relative to the observed peak between two to three hours. This offset is can plausibly be attributable to a combination of factors, including (i) the absence of antecedent baseflow and soil moisture conditioning in the uncalibrated rain-on-grid setup, which can shift the timing of the rising limb and (ii) backwater effects from bridge clogging observed downstream of

https://doi.org/10.5194/egusphere-2026-3652
Preprint. Discussion started: 8 July 2026

the Altenahr gauge during the event (Poppema et al., 2025), which might have affected the actual rise at the gauge and are not represented in the model. Despite this timing offset, the reproduction of the peak magnitude and the overall hydrograph shape, without any hydrological model and without calibration, is a strong demonstration that the rain-on-grid approach can also capture the dominant in-channel flood dynamics at the catchment scale.

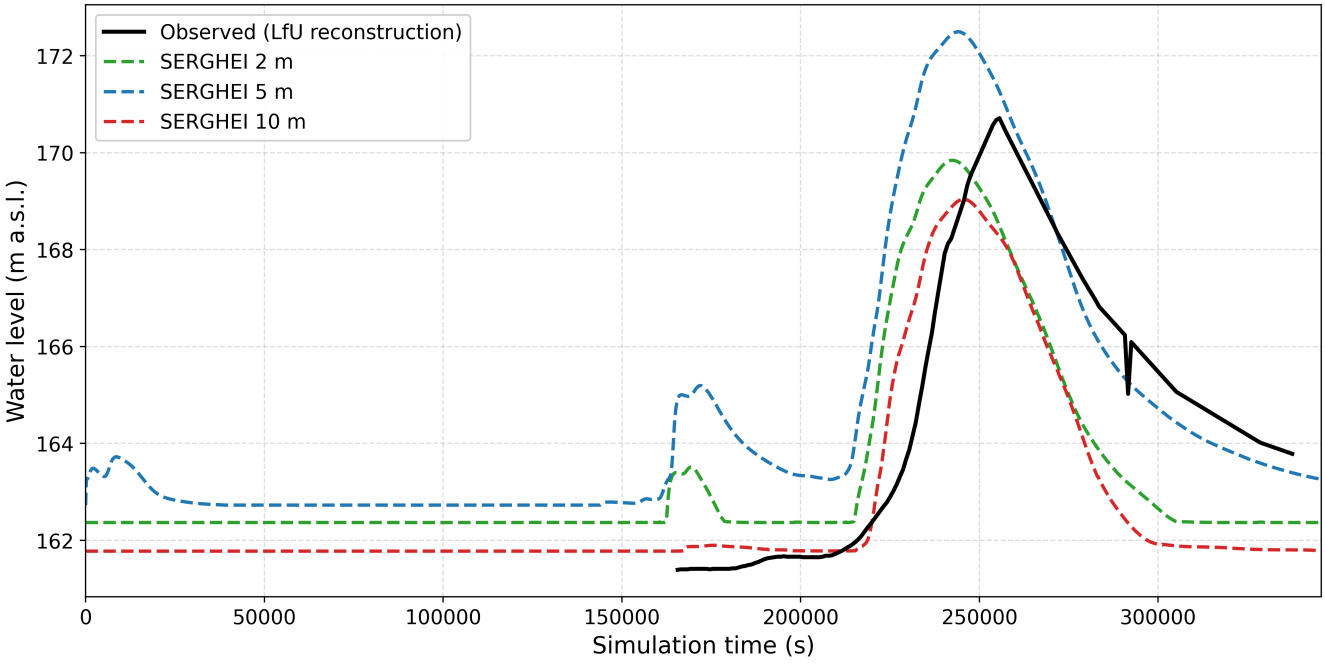


Figure 5. Simulated water levels at the Altenahr gauge from the SERGHEI rain-on-grid simulations at 2 m (green), 5 m (blue), and 10 m (red) resolution, compared against the LfU-reconstructed water level time series (black). Simulated water depths are converted to water levels by adding the bed elevation at the gauge cell of each grid (162.367, 162.727, and 161.777 m a.s.l. at 2, 5, and 10 m, respectively).The start of the simulation ($t = 0$), corresponds to 12 July 2021, 00:00 CEST.

3.3 Water balance and runoff partitioning

One of the central advantages of the rain-on-grid approach is that runoff generation, infiltration, channel routing, and floodplain storage are all simulated within a single physically consistent framework, governed by the same governing equations, numerical scheme and computational grid. To assess whether SERGHEI can reproduce a plausible catchment-scale water balance without an external hydrological model, we analysed the domain-integrated time series of rainfall input, infiltration loss, and surface storage extracted from the SERGHEI diagnostic output at each resolution.

Figure 6 presents the cumulative catchment-scale water balance for the $dx =$ 2, 5, and 10 m simulations. The total rainfall depth integrated over the ∼900 km^2 Ahr catchment is 107.7 mm in all three cases, as expected since the same RADOLAN forcing is applied at every resolution. The temporal evolution of the rainfall reproduces the well-documented multi-pulse structure of the 12–15 July 2021 event (Mohr et al., 2023), with a first, moderate pulse during the morning of 13 July (around

https://doi.org/10.5194/egusphere-2026-3652
Preprint. Discussion started: 8 July 2026

simulation hour 25–30), a second pulse later that day (around hour 45), and the main convective burst on 14 July (between hours 55 and 70) that delivered approximately 93 mm (around 86% of the total event rainfall) in approximately 15 h and produced the catastrophic flood wave. By the end of the 96-hour simulation, infiltration losses had removed 48.2, 49.1, and 50.0 mm at 2, 5, and 10 m resolution, respectively, corresponding to a near-identical infiltration fraction of 45–46 % of the total rainfall input across all three grids. Surface storage, which aggregates water held in the channel network, on the floodplain, and on wet hillslope cells across the catchment, increases over the course of the event to a peak of approximately 63 mm shortly after the end of the main rainfall pulse, then slowly recedes to 59.5, 58.5, and 57.7 mm at 2, 5, and 10 m, respectively, as water continues to infiltrate and drain through the stream network. Numerical mass closure, evaluated as the change in surface storage relative to the prescribed initial condition minus the net of accumulated rainfall and infiltration, remained below 10^{-3} mm in all three cases, as expected from a fundamentally mass-conservative finite-volume solver.

Two aspects of these results are particularly noteworthy. First, the agreement between resolutions is worth highlighting: the cumulative curves for rainfall, infiltration, and surface storage at 2, 5, and 10 m are very similar, with the maximum spread between resolutions never exceeding 1.8 mm for any of the three quantities at any point during the simulation. This indicates that the runoff-infiltration partitioning (here governed by a constant effective infiltration capacity set by soil texture and imperviousness fraction) is already well captured at operationally feasible resolutions, with the domain-integrated infiltration (~45–46 % of rainfall) and runoff coefficient (~54–55 %) remaining essentially consistent across the 2, 5, and 10 m grids. Finer resolutions primarily refine the spatial pattern of overland flow, as reflected in the improving extent metrics (CSI, EB, and BPI), rather than altering the integrated catchment response. Second, the resulting catchment runoff coefficient (the fraction of rainfall not lost to infiltration) of approximately 54–55 % is high but physically consistent with both the antecedent wet conditions preceding the event (Mohr et al., 2023) and the extreme rainfall intensities, which substantially exceeded the typical infiltration capacity of the catchment soils. As an independent check, we integrated the LfU-reconstructed Altenahr inflow hydrograph (Apel et al., 2022) over the event window using a rating curve calibrated to the documented peak discharge of $\sim$1000 m^3 s^{-1}, which yields a total flood volume of approximately 43 million m^3. Distributed over the catchment area upstream of Altenahr ($\sim$750 km^2) and compared to the event-integrated rainfall, this corresponds to an observed runoff coefficient of approximately 52–54 %, in close agreement with the simulated value. This consistency is noteworthy given that Saadi et al. (2023) demonstrated substantial spread in both precipitation totals and simulated peak flows for the same July 2021 event depending on the radar QPE product used, underscoring that the precipitation input remains the single largest source of uncertainty in the water balance.

Taken together, these results demonstrate that a fully dynamic shallow-water model, when forced directly by precipitation and parameterised with standard, openly available geospatial datasets is able to simulate the complete rainfall–runoff–inundation chain at the catchment scale with a water balance that is conservative, internally consistent across resolutions, and physically plausible. This is precisely the role that a coupled hydrological model fulfills in the traditional modelling chain. The fact that SERGHEI delivers it within a single integrated framework is a strong indication that the rain-on-grid approach is sufficiently mature to serve as a standalone alternative for catchment-scale flood simulation of flash-flood events of this type.

https://doi.org/10.5194/egusphere-2026-3652
Preprint. Discussion started: 8 July 2026

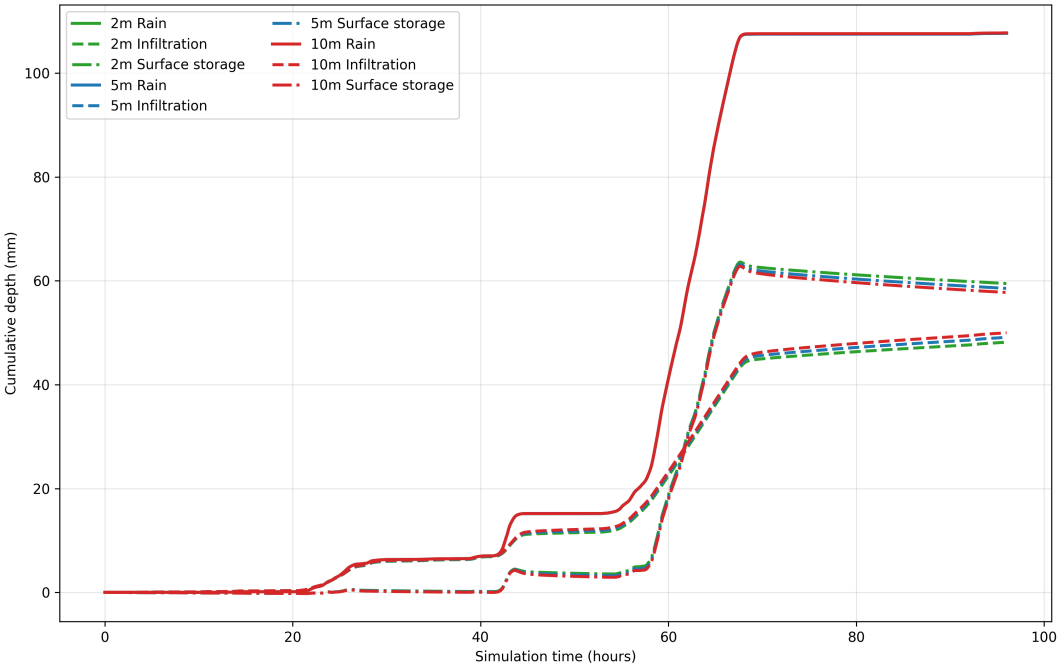


Figure 6. Cumulative catchment-scale water balance for the SERGHEI rain-on-grid simulations of the 2021 Ahr flood at dx = 2, 5, and 10 m. Solid lines show the cumulative rainfall input, dashed lines the cumulative infiltration loss, and dash-dotted lines the integrated surface storage (water held in the channel, on the floodplain, and on all wet cells across the ∼900 km^2 catchment). All quantities are expressed as catchment-averaged depths. The start of the simulation ($t = 0$), corresponds to 12 July 2021, 00:00 CEST.

3.4 Computational performance

All simulations were carried out on HPC systems at the Jülich Supercomputing Centre. We ran simulations in the JUWELS Booster module, equipped with four NVIDIA A100-SXM4-40GB GPUs per node, and the recent JUPITER Booster system equipped with four NVIDIA GH200 Superchips per node. The full ∼900 km^2 Ahr catchment was distributed across multiple nodes and GPUs using SERGHEI's hybrid MPI–Kokkos parallelisation strategy, i.e., each MPI rank manages one GPU, and the computational domain is partitioned into rectangular subdomains exchanging halo data at each time step. At the time of writing of this paper, the JUWELS Booster system (number 48 in the November 2025 TOP500 list (TOP500.org, 2025)) was over 6 years old, and soon to be decommissioned. In contrast the JUPITER Booster system (number 4 in the November 2025 TOP500 list) was in early operational deployment.

Figure 7 summarises the runtime characteristics of the three fastest simulations in the JUWELS Booster system. The dx = 10 m simulation, employing 9 million computational cells distributed across 8 GPUs (~ 1.125 million cells per GPU), completed the 96-hour event in 22.6 minutes of wall-clock time, corresponding to a simulation speed-up of approximately 255× real time. The 5 m simulation, with 36 million cells distributed across 32 GPUs (~ 1.125 million cells per GPU), required 61.2 minutes (94× real time). The finest 2 m simulation, with 225 million cells distributed across 256 GPUs (~ 0.88

https://doi.org/10.5194/egusphere-2026-3652
Preprint. Discussion started: 8 July 2026

460 million cells per GPU), completed the full 96-hour event in approximately 3.75 hours, still 26× faster than real time. The decrease in speed-up with finer resolution reflects two combined effects: the well-known reduction in the explicit time step imposed by the Courant–Friedrichs–Lewy condition as the grid spacing is halved, and the increased relative weight of inter-GPU communication overhead as the domain is partitioned across a much larger number of GPUs (256 versus 8). Even so, the 2 m configuration delivers the entire catchment-scale, 2m-resolution simulation in less than four hours on a HPC system at 465 the end of its life, demonstrating that fully hydrodynamic flood modelling at this scale and resolution has moved well into the regime of practical feasibility, thus solving what is perceived as the greatest impediment to deploy these models at such scale (Hill et al., 2023).

These runtimes have direct operational implications. At 10 m resolution, the simulation runs more than two orders of magnitude faster than real time, comfortably within the lead-time requirements of operational flood forecasting: a 24-hour 470 rainfall forecast can be translated into the corresponding inundation forecast in under three minutes on 8 GPUs. The 5 m configuration also remains operationally feasible, requiring approximately one hour of compute time on 32 GPUs to process a four-day forcing window (effectively keeping the lead time of the precipitation forecast which would drive the model). The 2 m setup, requiring almost four hours and a substantially larger GPU allocation, remains applicable to time-sensitive contexts such as ensemble flood forecasting where sufficient lead time is available, while also being well suited to offline applications 475 such as detailed post-event reconstruction, impact assessment, and scenario analysis. Crucially, these results demonstrate that the entire chain, from spatially distributed rainfall input to high-resolution inundation output, simulated over the full catchment by a single fully dynamic 2D shallow-water solver, can be executed within operationally relevant timeframes using a modest GPU allocation on HPC systems.

https://doi.org/10.5194/egusphere-2026-3652
Preprint. Discussion started: 8 July 2026

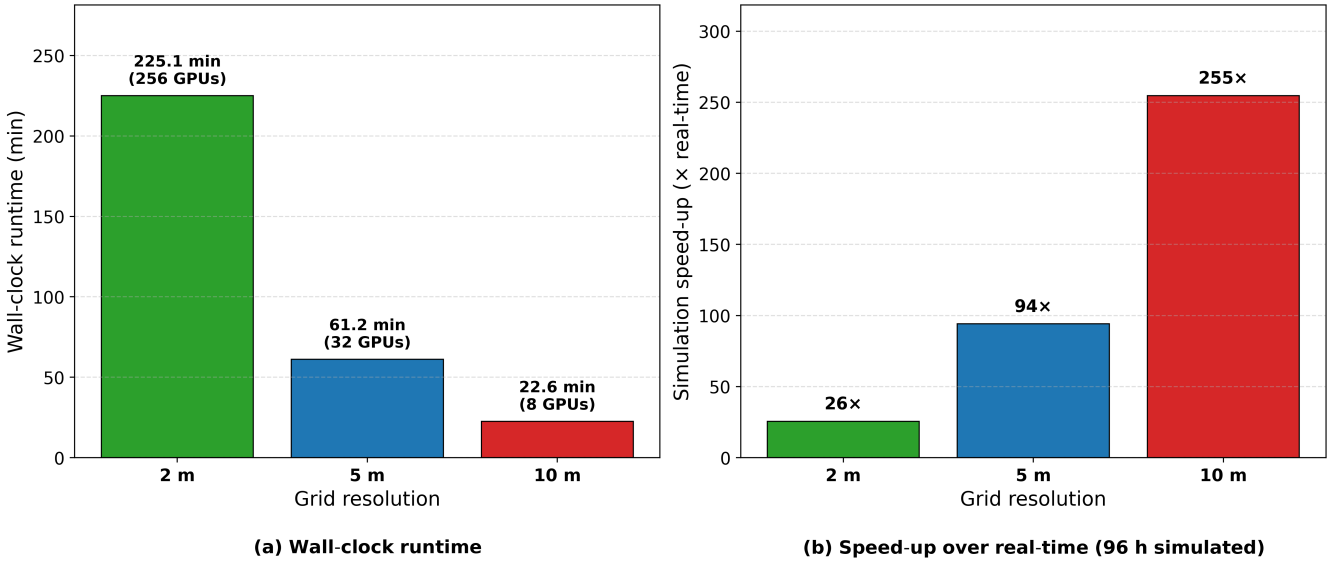


Figure 7. Computational performance of the SERGHEI simulations at dx = 2, 5, and 10 m resolutions on the JUWELS Booster module (NVIDIA A100-SXM4-40GB GPUs). (a) Wall-clock runtime in minutes, and number of GPUs used for each configuration. (b) Simulation speed-up over real time, defined as the ratio of the simulated event duration (96 h) to the wall-clock runtime.

The results reported above and in Figure 7 are the fastest results achieved in JUWELS Booster. Although the fastest runtimes are of critical interest for operational forecasting purposes, it is necessary to evaluate the dependency of this runtime to the number of computational resources used. That is, the strong-scaling behaviour is relevant to assess both the relative costs of operation and the possibility to use different sets of resources depending on urgency and availabilty required to meet a given early-warning deadline. Figure 8 shows, for all three resolutions and both HPC systems, the wall-clock time (panel a) and the corresponding speed-up over real time (panel b) as a function of GPU count. Two features stand out. First, all tested configurations run far faster than real time, on both systems, confirming that even the finest catchment-scale setups remain comfortably within early-warning-relevant runtimes, despite running on a system close to decommisioning. Second, for all resolutions the runtime decreases as GPUs are added. Increasing resources pays off for all cases in the range of resources tested here. Only the 2m case with 256 GPUs shows a distinctive saturation in scalability. This flexibility means that the allocation can be tuned to the operational context, trading additional GPUs for faster turnaround when forecast lead time is critical, or using fewer devices when turnaround is less urgent.

A further point evident from panel (a) is that the runtime cannot be reduced indefinitely by adding GPUs, as is well known for strong scaling behaviour. As the number of devices increases, the gains progressively diminish, until adding more GPUs yields almost no further speed-up. This is most pronounced for the 2 m grid on JUWELS, where doubling the allocation from 128 to 256 GPUs reduces the runtime only marginally, from 230 to 225 min despite the massive increase in resources (and in required energy to run it). This is the result of the combined workload reduction per GPU and the latency introduced by inter-GPU communications. In practice, this means there is a limit beyond which allocating additional GPUs is simply not

https://doi.org/10.5194/egusphere-2026-3652
Preprint. Discussion started: 8 July 2026

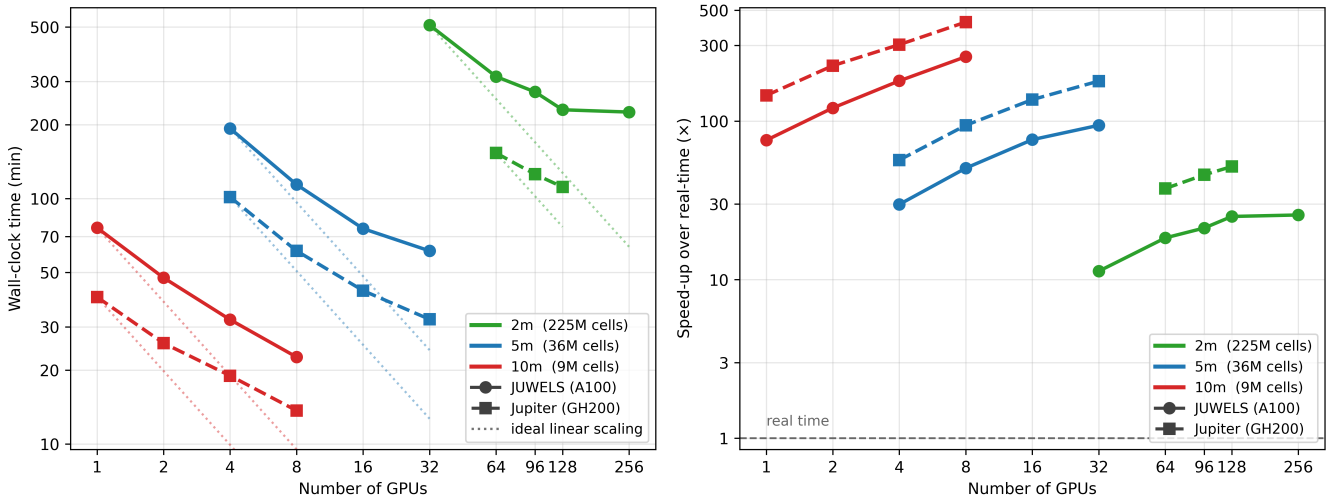
worthwhile, and the most effective configuration is the one that keeps each device sufficiently loaded rather than simply the one with the largest GPU count. Although for early warning applications time-to-solution is likely the critical performance metric, we highlight that energy-to-solution (or energy efficiency) is a growing concern in HPC, including operational HPC services for weather forecasting (Bauer et al., 2021), for which a good understanding of scalability is required.

The comparison between the two systems further illustrates how directly the approach benefits from advances in GPU hardware. For the same spatial resolution, the GH200-based JUPITER system is between roughly 1.6 and 2.2 times faster than the A100-based JUWELS Booster, with the largest gains realised for the most computationally demanding, finest-resolution configurations. For instance, the 2 m simulation on 128 GPUs, for example, completes in under two hours on JUPITER compared with almost four hours on JUWELS. But also for the coarser 10 m simulation, running in JUPITER on 8 GH200 results in runtimes 420× faster than real time, compared to the 255× achieved in JUWELS. Perhaps the most important feature of the speed-up obtained when moving from JUWELS to JUPITER is that, for a given resolution, a specific lead time target –which translates to specific speed-up-over-real-time (e.g, 100×)–, can be achieved with JUPITER with significantly fewer resources, e.g., 4 times fewer GPUs in the case of the 5m resolution case. For coarser problems the gain is not as large. But for finer problems, JUPITER enables speed-up-over-real-time simply not achievable with JUWELS regardless of the number of resources. This is clearly conveyed in Figure 8 for the 2 m resolution case, which is in the order of 35× with 64 GPUs, whereas in JUWELS for 256 GPUs it is in the order of 26×, i.e, around 1.4× faster in JUPITER with 4 times fewer resources. We further highlight that the JUWELS system was close to decommissioning at the time of writing this paper, whereas JUPITER was in its early operational stage. Thus our results illustrate the gains for one generation HPC systems for these purposes. Crucially, this acceleration is obtained with the identical source code and no resolution- or architecture-specific re-engineering, a direct consequence of the performance-portable design of SERGHEI (Caviedes-Vouillième et al., 2023), enabled by Kokkos. This suggests that the computational feasibility of catchment-scale rain-on-grid modelling will continue to improve with newer hardware generations at essentially no additional development cost, so that the runtimes reported here represent a conservative upper bound (with current technology) rather than a fixed ceiling (for future technologies).

Taken together, these results carry concrete implications for agencies and practitioners considering catchment-scale hydrodynamic modelling for operational flood forecasting. The core message is that fully hydrodynamic, meter-scale simulation of a $\sim$900 km^2 catchment is no longer a privilege reachable only via the largest supercomputers: a 10 m simulation of the entire event completes in little more than an hour on a single GPU and in minutes on a handful, while even the finest 2 m configuration is delivered within a few hours on a moderate allocation. For an operational deployment, the relevant decision is therefore a three-way trade-off between spatial resolution (and hence the level of detail and local accuracy required), the number of GPUs that can be allocated (and their associated hardware and energy costs, and their availability), and the forecast lead time that must be met. The strong-scaling curves in Figure 8 make this trade-off explicit and quantitative, allowing a forecasting centre to select the coarsest resolution that satisfies its accuracy requirements and then size the GPU allocation to keep each device adequately loaded, rather than over-provisioning hardware for diminishing returns. Because the model is performance-portable, the same configuration can be run on whatever GPU system an institution has access to, from a small in-house cluster to a national or Tier-0 supercomputer. Combined with the high flood-extent and water-depth skill demonstrated above, these performance

https://doi.org/10.5194/egusphere-2026-3652
Preprint. Discussion started: 8 July 2026

characteristics indicate that the principal historical barrier to standalone catchment-scale hydrodynamic forecasting, namely computational cost, has been substantially lowered.



(a) Strong scaling: runtime vs GPU count

(b) Speed-up over real-time (96 h simulated)

Figure 8. Strong-scaling performance of SERGHEI-SWE for the 2021 Ahr flood simulation at 2, 5, and 10 m resolutions, on two GPU-based supercomputers at the Jülich Supercomputing Centre: JUWELS Booster (NVIDIA A100 GPUs, solid lines and circles) and Jupiter (NVIDIA GH200 GPUs, dashed lines and squares). For each resolution the problem size (number of computational cells) is held fixed while the number of GPUs is varied. (a) Wall-clock time as a function of GPU count, with dotted lines indicating ideal linear (strong) scaling anchored at the smallest GPU count of each resolution and system; (b) corresponding speed-up over real time, defined as the 96 h simulated event duration divided by the wall-clock runtime, with the dashed grey line marking real-time performance ($1\times$). Colors denote resolution and marker/line styles denote the computing system. Both axes are logarithmic.

3.5 Sources of uncertainty, implications, and limitations

Here we discuss the main sources of uncertainty in the simulations, the broader implications of the results for catchment-scale flood modelling, and the limitations that bound the conclusions drawn here. The most influential source of uncertainty can be the precipitation forcing itself. The RADOLAN product is well established for operational and research applications in Germany (Mohr et al., 2023), however, remains subject to the well-known limitations of radar-based quantitative precipitation estimation, including beam blockage, attenuation by intense convective cells, and uncertainties in the Z–R relationship (Villarini and Krajewski, 2010). For the 2021 Ahr event, post-event analyses suggest local deviations from gauge-measured totals, particularly in the Eifel headwaters where radar coverage is poorest and rainfall was most intense (Mohr et al., 2023). Since the precipitation enters the SERGHEI mass balance directly, any bias in the input propagates into the simulated runoff volume and the resulting water depths and flood extents (Costabile et al., 2023). This, together with the absence of a calibration procedure, implies that the model simply propagates the precipitation errors with no means of correction. This is characteristically different

https://doi.org/10.5194/egusphere-2026-3652
Preprint. Discussion started: 8 July 2026

from classical hydrological-hydrodynamic modelling chains, where these errors are partially absorbed by the calibrated hydrological model. Additional uncertainty stems from the parameterisation of infiltration and surface roughness and from the validation data themselves. The infiltration field relies on the BÜK200 soil map, the Copernicus imperviousness raster, and literature-based pedotransfer relationships (Saxton and Rawls, 2006; Schaap et al., 2001), each step introducing its own error and neglecting sub-grid heterogeneity of soil hydraulic properties; in a calibrated configuration these uncertainties are partly compensated through parameter tuning (Khosh Bin Ghomash et al., 2025a), but in the uncalibrated setup adopted here they translate directly into output uncertainty. Despite the unavoidable uncertainties, the good performance metrics reported here can be interpreted as a realistic lower bound on the achievable skill of a catchment-scale, rain-on-grid SERGHEI setup in quasi-operational conditions, rather than as the upper limit of what the model can deliver.

Beyond these uncertainties, the results of this work carry broader implications for how catchment-scale flash-flood modelling may be structured. For catchments of the order of 10^3 km^2 and for flash-flood events of the type considered here, the long-standing technological limitations that have justified coupling a hydrological model with a hydrodynamic model are no longer binding. Current multi-GPU shallow-water solvers can simulate the full process chain from rainfall to inundation within a single physically consistent framework, at meter-scale resolutions, and within early-warning relevant runtimes. Some benefits of this integration are as follows: 1. Eliminating the explicit coupling interface between a hydrological and a hydrodynamic model removes the artifacts introduced when fluxes are transferred between two models that differ in their process assumptions, spatial grids, and time-stepping. 2. It preserves the full spatial structure of the rainfall forcing throughout the modelling chain. Conventional lumped or semi-distributed hydrological models aggregate precipitation over subcatchments, smoothing out the fine-scale gradients of localised convective cells, even though the position of the rainfall maximum relative to the drainage network strongly controls the magnitude and timing of the flood peak in fast-responding catchments (Sangati et al., 2009; Saharia et al., 2021). By applying each radar pixel directly to the computational grid, the rain-on-grid approach retains this information down to the resolution of the precipitation product, which is particularly advantageous for convective flash floods such as the 2021 Ahr event. 3. The parameters governing the present setup are physically interpretable and can be assigned in advance from independently mappable properties, surface roughness from land cover, infiltration capacity from soil texture and imperviousness via established pedotransfer relationships, rather than being effective quantities identifiable only through calibration against gauged discharge and regionalisation to ungauged basins, a procedure that is inherently prone to non-uniqueness. (Clark et al., 2016; Samaniego et al., 2010). Calibration can still refine these parameters, but the approach does not depend on it for a physically based first estimate. This is consistent with other fully physically-based efforts, showing that uncalibrated physically-based models can deliver very good results, and thus strongly questioning the need for calibration (Goergen et al., 2025). 4. The relatively small loss of skill when coarsening the grid from 2 to 10 m, combined with the resulting reduction in computational cost, suggests that the rain-on-grid approach is already practical at operationally feasible resolutions, while the finest grids lend themselves to offline applications such as detailed impact assessment, post-event reconstruction, and scenario analysis. The present results do not, however, rule out that ultra-high resolutions, e.g., 1 m or even higher, which may also add value in a warning context. This value may come, for instance, by resolving impact-relevant flow paths through streets and around buildings, which in turn is highly relevant if simulations also include sediment and debris

https://doi.org/10.5194/egusphere-2026-3652
Preprint. Discussion started: 8 July 2026

transport – processes which have been noted to have been particularly relevant in this event (Dietze et al., 2022; Lehmkuhl et al., 2022). Establishing where this added detail justifies its computational cost remains an open question, which is partially constrained both by the availability of computational resources, topography data, and detailed flood observations to compare against.

Structurally, the present single-module configuration does not explicitly represent evapotranspiration, subsurface flow, soil moisture dynamics, snowmelt, or vegetation interception, processes that are of secondary importance for the short, high-intensity, partially saturated conditions (Mohr et al., 2023) of the 2021 Ahr event, but whose omission would introduce systematic biases for events of longer duration or for catchments where antecedent moisture and groundwater interactions are more decisive. It should be pointed out that SERGHEI is designed as a modular ecohydrological framework, and a three-dimensional, variably-saturated subsurface flow solver based on the Richards equation (SERGHEI-RE) is already part of the framework (Li et al., 2025), together with a coupled surface–subsurface flow configuration (SERGHEI-SWE-RE) that combines the SWE and Richards modules within the same performance-portable, multi-GPU infrastructure (Zheng et al., 2026). These modules provide a natural pathway for replacing the simple infiltration formulation used here with a physically-based representation of soil moisture dynamics and shallow groundwater –in a fashion similar to the study by Goergen et al. (2025), but at higher resolution–, and SERGHEI also allows further ecohydrological processes such as canopy interception and evapotranspiration to be included. This, of course, comes with significant added computational cost.

Finally, the results presented here pertain to a single catchment and a single event, and the conclusions drawn regarding the feasibility and skill of the rain-on-grid approach should not be directly generalised to other catchments or events without further testing. The Ahr basin is a particular case in several respects: a relatively medium-sized, steep and narrow, gorge-like valley experiencing a black-swan extreme, predominantly convective flash flood on partially saturated soils, in which the flood response is dominated by fast surface runoff. The performance of the same approach on, for example, a larger lowland catchment with slow-responding hillslope and groundwater contributions or on a snowmelt-driven event remains an open question. Systematic testing across different catchment and event types is therefore a necessary step before the rain-on-grid approach can be more broadly recommended as a standalone alternative to the well-established coupled hydrological–hydrodynamic chain. Moreover, the performance reported here is that of an uncalibrated model, and the published results of Khosh Bin Ghomash et al. (2025a) suggest that targeted calibration of roughness and infiltration parameters would further improve water-depth and in-channel flow skill. Coupling the present setup with ensemble numerical weather prediction inputs, as in Najafi et al. (2024), would in addition allow the full forecast lead-time and probabilistic dimension of early warning to be addressed. It is very likely that calibration, ensemble NWP forcing, and the progressive activation of the additional ecohydrological modules of SERGHEI will constitute concrete pathways to improve model skill and applicability in subsequent studies.

https://doi.org/10.5194/egusphere-2026-3652
Preprint. Discussion started: 8 July 2026

4 Conclusions

This study set out to address a single question: do we still need a hydrological model for catchment-scale flash-flood simulation, or can a fully hydrodynamic, rain-on-grid approach now stand on its own? We addressed it by applying the uncalibrated SERGHEI-SWE solver to the entire ∼900 km^2 Ahr catchment at 2, 5, and 10 m resolution, forcing it directly with 5-minute RADOLAN radar precipitation and validating against three datasets: the observed flood-extent of the event, 45 post-event high-water marks, and the reconstructed Altenahr hydrograph.

The results consistently support a positive answer to the research question. The simulated flood extent reproduces the observed extent map with Critical Success Index values of 0.846–0.875 across the three resolutions, Hit Rates above 0.955, and only moderate overprediction of the inundated area (BPI below 10 % even at 10 m resolution), with all metrics degrading only slightly as the grid is coarsened from 2 to 10 m. The simulated peak water depths at the surveyed high-water marks approximate the observations well, with Pearson correlations of r = 0.86–0.88, near-zero overall BIAS (+0.05 to +0.18 m), and RMSE values of 0.46–0.57 m across the entire 900 km^2 catchment. The domain-integrated water balance is fully conservative, consistent across resolutions, and produces a catchment runoff coefficient (54–55 %) consistent with the reported values from the 2021 event. Crucially, these scores are achieved without an intermediate hydrological model, without a reconstructed inflow hydrograph at an arbitrary inflow point leading into an arbitrarily defined reach, and without any prior calibration of roughness or infiltration parameters. The skill metrics are essentially on par with the CSI of 0.845 and the BIAS of +0.09 m / RMSE of 0.30 m reported by Apel et al. (2022) for the gauge-forced RIM2D setup of the lower 30 km of the valley, and with the subsequent coupled mHM–RIM2D applications of Najafi et al. (2024) and Vorogushyn et al. (2025) that have been considered the state of the art for the 2021 Ahr flood so far, but extend over a domain approximately one order of magnitude larger. They are likewise comparable to the earlier SERGHEI application to the 2021 Ahr flood (Khosh Bin Ghomash et al., 2024), which achieved CSI values of 0.849–0.854 using the same hydrodynamic solver and roughness parameterisation, but driven by the reconstructed Altenahr stage hydrograph and confined to the lower 30 km of the valley.

Moreover, the time-to-solution for all simulations is fast enough to allow for operational forecasting and sufficient lead time for early warning, ranging from 25 to 420× faster than the duration of the event depending on resolution and hardware. A specific time-to-solution –likely a key operational constraint– is modulated by model resolution and computational resources in a three-way-trade-off. Nevertheless, the capacity of the SERGHEI-SWE solver to efficiently use large resource sets allows for a variety of operational choices, limited mainly by operational requirements and available computational resources.

Taken together, these results show that catchment-scale, rain-on-grid hydrodynamic modelling can deliver flood-extent and water-depth skill comparable to the established coupled hydrological–hydrodynamic chain for an event of this type. Where computational resources allow, the hydrological and hydrodynamic components of flash-flood modelling need then not always be treated as separate sequential stages, but can be combined within a single, physically consistent framework.

https://doi.org/10.5194/egusphere-2026-3652
Preprint. Discussion started: 8 July 2026

Code availability. SERGHEI is open-source software distributed under a three-clause BSD license. The source code is available at https://gitlab.com/serghei-model/serghei. The simulations presented in this study were performed with the SERGHEI-SWE module only (Caviedes-Vouillième et al., 2023).

Data availability. The DGM1 digital elevation model is provided by the German Federal Agency for Cartography and Geodesy (BKG) and is available through the Geoportal of the State Agency for Surveying and Geobasis Information of Rhineland-Palatinate (https://lvermgeo.rlp.de). The Soil Map of Germany 1:200,000 (BÜK200), used to derive soil texture and horizon-level pedological data for the infiltration parameterisation, is provided by the Federal Institute for Geosciences and Natural Resources (BGR) and is available from the BGR Geoportal (https://www.bgr.bund.de/EN/Themen/Boden/Projekte/Flaechen_Rauminformationen_Boden/BUEK200/BUEK200.html) (Federal Institute for Geosciences and Natural Resources (BGR), 2020). The OSM building footprints used in this study are freely available at https://download.geofabrik.de/europe/germany.html. The 2020 Sentinel-2 based land cover classification of Germany, employed to assign surface roughness values within the simulation domain, can be accessed at https://www.mundialis.de/en/germany-2020-land-cover-based-on-sentinel-2-data/ (Riembauer et al., 2021). The Imperviousness Density raster from the Copernicus Land Monitoring Service (CLMS), used in the infiltration parameterisation, is available from the European Environment Agency (European Environment Agency, 2020). The 5-minute RADOLAN-YW radar precipitation product is provided by the German Weather Service (DWD) through the open data portal at https://opendata.dwd.de/climate_environment/CDC/grids_germany/5_minutes/radolan/.

Author contributions. SKBG: Conceptualization, Methodology, Investigation, Simulation, Software, Analysis, Visualization, Model validation, Writing; DCV: Conceptualization, Methodology, Software, Writing, Review, Supervision

Competing interests. The authors declare that they don't have any competing interest within the frame of the content of the manuscript.

Acknowledgements. The authors gratefully acknowledge the Earth System Modelling (ESM) project of the Helmholtz Association for supporting this work by providing computing time on the ESM partition of the JUWELS supercomputer (Jülich Supercomputing Centre, 2021) at the Jülich Supercomputing Centre (JSC) through the compute time project Environmental Hydrodynamics, runoff and transport across scales with the SERGHEI model (EHRTAS), as well as the ESM Test compute project.

The authors thank the Gauss Centre for Supercomputing e.V. (GCS) for providing computing time on the Supercomputer JUPITER at the Jülich Supercomputing Centre via the John von Neumann Institute for Computing. JUPITER was supported by the EuroHPC JU and GCS through funding by the European Commission, the German Federal Ministry of Research, Technology and Space, and the Ministry of Culture and Science of the State of North Rhine-Westphalia. The authors would also like to acknowledge the JUPITER Research and Early Access Program (JUREAP) for providing early access to JEDI (JUPITER Exascale Development Instrument) and JUPITER enabling all the preparatory work for deploying SERGHEI on JUPITER.

https://doi.org/10.5194/egusphere-2026-3652
Preprint. Discussion started: 8 July 2026

References

Apel, H., Vorogushyn, S., and Merz, B.: Brief communication: Impact forecasting could substantially improve the emergency management of deadly floods: case study July 2021 floods in Germany, Natural Hazards and Earth System Sciences, 22, 3005–3014, https://doi.org/10.5194/nhess-22-3005-2022, 2022.

Bachmann, D., Khosh Bin Ghomash, S., and Schotten, R.: Neue entwicklungen in der hochwasserrisikoanalyse: Niederschlagsgeneratoren und kritische infrastrukturen, WasserWirtschaft, 111, 32–38, 2021.

Bates, P. D., Horritt, M. S., and Fewtrell, T. J.: A simple inertial formulation of the shallow water equations for efficient two-dimensional flood inundation modelling, Journal of Hydrology, 387, 33–45, https://doi.org/10.1016/j.jhydrol.2010.03.027, 2010.

Bauer, P., Dueben, P. D., Hoefler, T., Quintino, T., Schulthess, T. C., and Wedi, N. P.: The digital revolution of Earth-system science, Nature Computational Science, 1, 104–113, https://doi.org/10.1038/s43588-021-00023-0, 2021.

Busker, T., Rodriguez Castro, D., Vorogushyn, S., Kwadijk, J., Zoccatelli, D., Oliveira, R. G. L., Murdock, H. J., Pfister, L., Dewals, B., Slager, K., Thieken, A. H., Verkade, J., Willems, P., and Aerts, J. C. J. H.: Comparing flood forecasting and early warning systems in northwestern Europe, Natural Hazards and Earth System Sciences, 26, 1457–1478, https://doi.org/10.5194/nhess-26-1457-2026, 2026.

Caviedes-Vouillième, D., Fernández-Pato, J., and Hinz, C.: Performance assessment of 2D zero-inertia and shallow water models for simulating rainfall-runoff processes, Journal of Hydrology, 584, 124 663, https://doi.org/10.1016/j.jhydrol.2020.124663, 2020.

Caviedes-Vouillième, D., Morales-Hernández, M., Norman, M. R., and Özgen-Xian, I.: SERGHEI (SERGHEI-SWE) v1.0: a performance-portable high-performance parallel-computing shallow-water solver for hydrology and environmental hydraulics, Geoscientific Model Development, 16, 977–1008, https://doi.org/10.5194/gmd-16-977-2023, 2023.

Caviedes-Voullième, D., García-Navarro, P., and Murillo, J.: Influence of mesh structure on 2D full shallow water equations and SCS Curve Number simulation of rainfall/runoff events, Journal of Hydrology, 448-449, 39 – 59, https://doi.org/https://doi.org/10.1016/j.jhydrol.2012.04.006, 2012.

Chakraborty, P., Dessers, C., Archambeau, P., Pirotton, M., Erpicum, S., and Dewals, B.: The mirage of the silver bullet: Exploring the limitations of high-resolution data in flood model validation, Journal of Hydrology, p. 134578, https://doi.org/10.1016/j.jhydrol.2025.134578, 2025.

Clark, M. P., Schaefli, B., Schymanski, S. J., Samaniego, L., Luce, C. H., Jackson, B. M., Freer, J. E., Arnold, J. R., Moore, R. D., Istanbulluoglu, E., and Ceola, S.: Improving the theoretical underpinnings of process-based hydrologic models, Water Resources Research, 52, 2350–2365, https://doi.org/10.1002/2015WR017910, 2016.

Costabile, P. and Costanzo, C.: A 2D SWEs framework for efficient catchment-scale simulations: hydrodynamic scaling properties of river networks and implications for non-uniform grids generation, Journal of Hydrology, p. 126306, https://doi.org/10.1016/j.jhydrol.2021.126306, 2021.

Costabile, P., Costanzo, C., Lorenzo, G. D., and Macchione, F.: Is local flood hazard assessment in urban areas significantly influenced by the physical complexity of the hydrodynamic inundation model?, Journal of Hydrology, 580, 124 231, https://doi.org/10.1016/j.jhydrol.2019.124231, 2019.

Costabile, P., Costanzo, C., Kalogiros, J., and Bellos, V.: Toward street-level nowcasting of flash floods impacts based on HPC hydrodynamic modeling at the watershed scale and high-resolution weather radar data, Water Resources Research, 59, e2023WR034 599, https://doi.org/10.1029/2023WR034599, 2023.

https://doi.org/10.5194/egusphere-2026-3652
Preprint. Discussion started: 8 July 2026

David, A. and Schmalz, B.: A Systematic Analysis of the Interaction between Rain-on-Grid-Simulations and Spatial Resolution in 2D Hydrodynamic Modeling, Water, 13, 2346, https://doi.org/10.3390/w13172346, 2021.

de Almeida, G. A. M. and Bates, P.: Applicability of the local inertial approximation of the shallow water equations to flood modeling, Water Resources Research, 49, 4833–4844, https://doi.org/10.1002/wrcr.20366, 2013.

Dietze, M., Bell, R., Ozturk, U., Cook, K. L., Andermann, C., Beer, A. R., Damm, B., Lucia, A., Fauer, F. S., Nissen, K. M., Sieg, T., and Thieken, A. H.: More than heavy rain turning into fast-flowing water – a landscape perspective on the 2021 Eifel floods, Natural Hazards and Earth System Sciences, 22, 1845–1856, https://doi.org/10.5194/nhess-22-1845-2022, 2022.

Dullo, T. T., Gangrade, S., Morales-Hernández, M., Sharif, M. B., Kao, S.-C., Kalyanapu, A. J., Ghafoor, S., and Evans, K. J.: Simulation of Hurricane Harvey flood event through coupled hydrologic-hydraulic models: Challenges and next steps, Journal of Flood Risk Management, https://doi.org/10.1111/jfr3.12716, 2021.

Ennouini, W., Fenocchi, A., Petaccia, G., Persi, E., and Sibilla, S.: A complete methodology to assess hydraulic risk in small ungauged catchments based on HEC-RAS 2D Rain-On-Grid simulations, Natural Hazards, https://doi.org/10.1007/s11069-024-06515-2, 2024.

European Environment Agency: Imperviousness Density 2018 (raster 10 m), Europe, 3-yearly, https://doi.org/10.2909/3bf542bd-eebd-4d73-b53c-a0243f2ed862, 2020.

Falter, D., Dung, N. V., Vorogushyn, S., Schröter, K., Hundecha, Y., Kreibich, H., Apel, H., Theisselmann, F., and Merz, B.: Continuous, large-scale simulation model for flood risk assessments: proof-of-concept, Journal of Flood Risk Management, 9, 3–21, https://doi.org/10.1111/jfr3.12105, 2016.

Federal Institute for Geosciences and Natural Resources (BGR): Soil Map of Germany 1:200,000 (BÜK200), https://www.bgr.bund.de/EN/Themen/Boden/Projekte/Flaechen_Rauminformationen_Boden/BUEK200/BUEK200.html, 2020.

Fewtrell, T. J., Bates, P. D., Horritt, M., and Hunter, N. M.: Evaluating the effect of scale in flood inundation modelling in urban environments, Hydrological Processes, 22, 5107–5118, https://doi.org/10.1002/hyp.7148, 2008.

García-Alén, G., González-Cao, J., Fernández-Nóvoa, D., Gómez-Gesteira, M., Cea, L., and Puertas, J.: Analysis of two sources of variability of basin outflow hydrographs computed with the 2D shallow water model Iber: Digital Terrain Model and unstructured mesh size, Journal of Hydrology, 612, 128 182, https://doi.org/10.1016/j.jhydrol.2022.128182, 2022.

Godara, N., Bruland, O., and Alfredsen, K.: Comparison of two hydrodynamic models for their rain-on-grid technique to simulate flash floods in steep catchment, Frontiers in Water, 6, https://doi.org/10.3389/frwa.2024.1384205, 2024.

Goergen, K., Belleflamme, A., Hammoudeh, S., Vanderborght, J., and Kollet, S.: The July 2021 flood event in the Eifel-Ardennes mountains as simulated by the high-resolution integrated hydrologic model ParFlow, Frontiers in Water, 7, https://doi.org/10.3389/frwa.2025.1571704, 2025.

Grimaldi, S., Schumann, G. J.-P., Shokri, A., Walker, J. P., and Pauwels, V. R. N.: Challenges, opportunities, and pitfalls for global coupled hydrologic-hydraulic modeling of floods, Water Resources Research, 55, 5277–5300, https://doi.org/10.1029/2018WR024289, 2019.

Hill, B., Liang, Q., Bosher, L., Chen, H., and Nicholson, A.: A systematic review of natural flood management modelling: Approaches, limitations, and potential solutions, Journal of Flood Risk Management, https://doi.org/10.1111/jfr3.12899, 2023.

Horritt, M. S. and Bates, P. D.: Effects of spatial resolution on a raster based model of flood flow, Journal of Hydrology, 253, 239–249, https://doi.org/10.1016/S0022-1694(01)00490-5, 2001.

Jülich Supercomputing Centre: JUWELS Cluster and Booster: Exascale Pathfinder with Modular Supercomputing Architecture at Jülich Supercomputing Centre, Journal of large-scale research facilities, 7, A138, https://doi.org/10.17815/jlsrf-7-183, 2021.

https://doi.org/10.5194/egusphere-2026-3652
Preprint. Discussion started: 8 July 2026

Khosh Bin Ghomash, S., Bachmann, D., Caviedes-Vouillième, D., and Hinz, C.: Impact of rainfall movement on flash flood response: a synthetic study of a semi-arid mountainous catchment, Water, 14, 1844, https://doi.org/10.3390/w14121844, 2022.

Khosh Bin Ghomash, S., Bachmann, D., Caviedes-Voullième, D., and Hinz, C.: Effects of Within-Storm Variability on Allochthonous Flash Flooding: A Synthetic Study, Water, 15, 645, 2023.

Khosh Bin Ghomash, S., Apel, H., and Caviedes-Vouillième, D.: Are 2D shallow-water solvers fast enough for early flood warning? A comparative assessment on the 2021 Ahr valley flood event, Natural Hazards and Earth System Sciences, 24, 2857–2874, https://doi.org/10.5194/nhess-24-2857-2024, 2024.

Khosh Bin Ghomash, S., Yeste, P., Apel, H., and Nguyen, V. D.: Monte Carlo-based sensitivity analysis of the RIM2D hydrodynamic model for the 2021 flood event in western Germany, Natural Hazards and Earth System Sciences, 25, 975–990, https://doi.org/10.5194/nhess-25-975-2025, 2025a.

Khosh Bin Ghomash, S., Apel, H., Schröter, K., and Steinhausen, M.: Rapid high-resolution impact-based flood early warning is possible with RIM2D: a showcase for the 2023 pluvial flood in Braunschweig, Natural Hazards and Earth System Sciences, 25, 1737–1749, https://doi.org/10.5194/nhess-25-1737-2025, 2025b.

Khosh Bin Ghomash, S., Nallasamy, N. D., and Apel, H.: How can building representation influence flood hazard and impact modelling: insights from the 2021 Ahr Valley flood, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2025-2304, 2025c.

Khosh Bin Ghomash, S., Deng, S., and Apel, H.: Enabling real-time high-resolution flood forecasting for the entire state of Berlin through multi-GPU accelerated physics-based modeling, Natural Hazards and Earth System Sciences, 26, 85–101, 2026.

Lehmkuhl, F., Schüttrumpf, H., Schwarzbauer, J., Brüll, C., Dietze, M., Letmathe, P., Völker, C., and Hollert, H.: Assessment of the 2021 summer flood in Central Europe, Environmental Sciences Europe, 34, https://doi.org/10.1186/s12302-022-00685-1, 2022.

Lengfeld, K., Walawender, E., Winterrath, T., and Becker, A.: CatRaRE: a catalogue of radar-based heavy rainfall events in Germany derived from 20 years of data, Meteorologische Zeitschrift, 30, 469–487, https://doi.org/10.1127/metz/2021/1083, 2021.

Li, Z., Caviedes-Vouillième, D., Özgen-Xian, I., Norman, M. R., Morales-Hernández, M., and Zheng, N.: SERGHEI v2.0: introducing a performance-portable, high-performance, three-dimensional variably saturated subsurface flow solver (SERGHEI-RE), Geoscientific Model Development, 18, 547–569, https://doi.org/10.5194/gmd-18-547-2025, 2025.

Liang, Q., Xia, X., and Hou, J.: Catchment-scale High-resolution Flash Flood Simulation Using the GPU-based Technology, Procedia Engineering, 154, 975 – 981, https://doi.org/http://dx.doi.org/10.1016/j.proeng.2016.07.585, 12th International Conference on Hydroinformatics (HIC 2016) - Smart Water for the Future, 2016.

Merz, B., Nguyen, V. D., Guse, B., Han, L., Guan, X., Rakovec, O., Samaniego, L., Ahrens, B., and Vorogushyn, S.: Spatial counterfactuals to explore disastrous flooding, Environmental Research Letters, 19, 044 022, https://doi.org/10.1088/1748-9326/ad22b9, 2024.

Mohr, S., Ehret, U., Kunz, M., Ludwig, P., Caldas-Alvarez, A., Daniell, J. E., Ehmele, F., Feldmann, H., Franca, M. J., Gattke, C., Hundhausen, M., Knippertz, P., Küpfer, K., Mühr, B., Pinto, J. G., Quinting, J., Schäfer, A. M., Scheibel, M., Seidel, F., and Wisotzky, C.: A multi-disciplinary analysis of the exceptional flood event of July 2021 in central Europe – Part 1: Event description and analysis, Natural Hazards and Earth System Sciences, 23, 525–551, https://doi.org/10.5194/nhess-23-525-2023, 2023.

Morales-Hernández, M., Sharif, M. B., Gangrade, S., Dullo, T. T., Kao, S.-C., Kalyanapu, A., Ghafoor, S. K., Evans, K. J., Madadi-Kandjani, E., and Hodges, B. R.: High-performance computing in water resources hydrodynamics, Journal of Hydroinformatics, 22, 1217–1235, https://doi.org/10.2166/hydro.2020.163, 2020.

https://doi.org/10.5194/egusphere-2026-3652
Preprint. Discussion started: 8 July 2026

Morales-Hernández, M., Sharif, M. B., Kalyanapu, A., Ghafoor, S., Dullo, T., Gangrade, S., Kao, S.-C., Norman, M., and Evans, K.: TRITON: a multi-GPU open source 2D hydrodynamic flood model, Environmental Modelling & Software, 141, 105 034, https://doi.org/10.1016/j.envsoft.2021.105034, 2021.

Najafi, H., Shrestha, P. K., Rakovec, O., Apel, H., Vorogushyn, S., Kumar, R., Thober, S., Merz, B., and Samaniego, L.: High-resolution impact-based early warning system for riverine flooding, Nature Communications, 15, 3726, https://doi.org/10.1038/s41467-024-48065-y, 2024.

Pasculli, A., Cinosi, J., Turconi, L., and Sciarra, N.: Learning case study of a shallow-water model to assess an early-warning system for fast alpine muddy-debris-flow, Water, 13, 750, https://doi.org/10.3390/w13060750, 2021.

Poppema, D. W., Burghardt, L., Benet, L., Wüthrich, D., Klopries, E.-M., Dewals, B., and Erpicum, S.: Bridge Clogging in Belgium and Germany During the 2021 Floods, Water Resources Research, 61, e2024WR039 218, https://doi.org/10.1029/2024WR039218, 2025.

Riembauer, G., Weinmann, A., Xu, S., Eichfuss, S., Eberz, C., and Neteler, M.: Germany-wide Sentinel-2 based land cover classification and change detection for settlement and infrastructure monitoring, pp. 53–56, https://doi.org/10.2760/125905, 2021.

Rocha Silva, A. F., Eleutério, J. C., Apel, H., and Kreibich, H.: Assessing the impact of early warning and evacuation on human losses during the 2021 Ahr Valley flood in Germany using agent-based modelling, https://doi.org/10.5194/nhess-2024-183, 2024.

Roggenkamp, T. and Herget, J.: Assessment of the 2021 summer flood in Central Europe, Environmental Sciences Europe, 34, 107, https://doi.org/10.1186/s12302-022-00685-1, 2022.

Roggenkamp, T., Herget, J., and Roggenkamp, T.: Flood reconstruction – The unexpected rather frequent event at River Ahr in July 2021, Global and Planetary Change, 240, 104 541, https://doi.org/10.1016/j.gloplacha.2024.104541, 2024.

Saadi, M., Furusho-Percot, C., Belleflamme, A., Chen, J.-Y., Trömel, S., and Kollet, S.: How uncertain are precipitation and peak flow estimates for the July 2021 flooding event?, Natural Hazards and Earth System Sciences, 23, 159–177, https://doi.org/10.5194/nhess-23-159-2023, 2023.

Saharia, M., Nikolopoulos, E. I., Bhuiyan, M. A. E., Wang, F., and Schwaller, M.: On the Impact of Rainfall Spatial Variability, Geomorphology, and Climatology on Flash Floods, Water Resources Research, 57, e2020WR029 124, https://doi.org/10.1029/2020WR029124, 2021.

Samaniego, L., Kumar, R., and Attinger, S.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, Water Resources Research, 46, W05 523, https://doi.org/10.1029/2008WR007327, 2010.

Sangati, M., Borga, M., Rabuffetti, D., and Bechini, R.: Influence of rainfall and soil properties spatial aggregation on extreme flash flood response modelling: An evaluation based on the Sesia river basin, North Western Italy, Advances in Water Resources, 32, 1090–1106, https://doi.org/10.1016/j.advwatres.2009.03.004, 2009.

Saxton, K. E. and Rawls, W. J.: Soil water characteristic estimates by texture and organic matter for hydrologic solutions, Soil Science Society of America Journal, 70, 1569–1578, https://doi.org/10.2136/sssaj2005.0117, 2006.

Schaap, M. G., Leij, F. J., and van Genuchten, M. T.: ROSETTA: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions, Journal of Hydrology, 251, 163–176, https://doi.org/10.1016/S0022-1694(01)00466-8, 2001.

Schäfer, A., Mühr, B., Daniell, J., Ehret, U., Ehmele, F., Küpfer, K., Brand, J., Wisotzky, C., Skapski, J., Rentz, L., Mohr, S., and Kunz, M.: Hochwasser Mitteleuropa, Juli 2021 (Deutschland), Tech. rep., CEDIM Forensic Disaster Analysis Group, https://doi.org/10.5445/IR/1000135730, 2021.

Schuermans, N., Knoben, W. J. M., Bertrand, S., De Troch, R., and Willems, P.: Locally relevant high-resolution hydrodynamic modeling of river floods at the regional scale, Water Resources Research, 58, e2021WR030 820, https://doi.org/10.1029/2021WR030820, 2022.

https://doi.org/10.5194/egusphere-2026-3652
Preprint. Discussion started: 8 July 2026

Schüttrumpf, H., Birkmann, J., Stenger-Wolf, S., Klopries, E., and Thieken, A. H.: The Flood Event in July 2021 in Europe—Observations, Conclusions, Challenges, Journal of Flood Risk Management, 19, https://doi.org/10.1111/jfr3.70175, 2026.

Sobieraj, J. A., Elsenbeer, H., Coelho, R. M., and Newton, B.: Spatial variability of soil hydraulic conductivity along a tropical rainforest catena, Geoderma, 108, 79–90, https://doi.org/10.1016/S0016-7061(01)00153-5, 2002.

Soil Survey Staff: Soil Taxonomy: A Basic System of Soil Classification for Making and Interpreting Soil Surveys, United States Department of Agriculture, Natural Resources Conservation Service, Washington, D.C., 2 edn., 1999.

Teng, J., Jakeman, A. J., Vaze, J., Croke, B. F. W., Dutta, D., and Kim, S.: Flood inundation modelling: a review of methods, recent advances and uncertainty analysis, Environmental Modelling & Software, 90, 201–216, https://doi.org/10.1016/j.envsoft.2017.01.006, 2017.

Thieken, A. H., Bubeck, P., Heidenreich, A., von Keyserlingk, J., Dillenardt, L., and Otto, A.: Performance of the flood warning system in Germany in July 2021 – insights from affected residents, Natural Hazards and Earth System Sciences, 23, 973–990, https://doi.org/10.5194/nhess-23-973-2023, 2023.

TOP500.org: TOP500 List – November 2025, https://top500.org/lists/top500/2025/11/, 66th edition; accessed 23 June 2026, 2025.

Trott, C. R., Lebrun-Grandié, D., Arndt, D., Ciesko, J., Dang, V., Ellingwood, N., Gayatri, R., Harvey, E., Hollman, D. S., Ibanez, D., Liber, N., Madsen, J., Miles, J., Poliakoff, D., Powell, A., Rajamanickam, S., Simberg, M., Sunderland, D., Turcksin, B., and Wilke, J.: Kokkos 3: Programming model extensions for the exascale era, IEEE Transactions on Parallel and Distributed Systems, 33, 805–817, https://doi.org/10.1109/TPDS.2021.3097283, 2022.

Truedinger, A. J., Jamshed, A., Sauter, H., and Birkmann, J.: Adaptation after extreme flooding events: moving or staying? The case of the Ahr Valley in Germany, Sustainability, 15, 1407, https://doi.org/10.3390/su15021407, 2023.

Villarini, G. and Krajewski, W. F.: Review of the different sources of uncertainty in single polarization radar-based estimates of rainfall, Surveys in Geophysics, 31, 107–129, https://doi.org/10.1007/s10712-009-9079-x, 2010.

Vorogushyn, S., Han, L., Apel, H., Nguyen, V. D., Guse, B., Guan, X., Rakovec, O., Najafi, H., Samaniego, L., and Merz, B.: It could have been much worse: spatial counterfactuals of the July 2021 flood in the Ahr Valley, Germany, Natural Hazards and Earth System Sciences, 25, 2007–2029, https://doi.org/10.5194/nhess-25-2007-2025, 2025.

Wing, O. E., Bates, P. D., Sampson, C. C., Smith, A. M., Johnson, K. A., and Erickson, T. A.: Validation of a 30 m resolution flood hazard model of the conterminous U nited S tates, Water Resources Research, 53, 7968–7986, 2017.

Winterrath, T., Brendel, C., Hafer, M., Junghänel, T., Klameth, A., Lengfeld, K., Walawender, E., Weigl, E., and Becker, A.: RADOLAN-RW: a product of the operationally gauge-adjusted radar-based precipitation analysis for Germany, Earth System Science Data, 10, 1609–1624, https://doi.org/10.5194/essd-10-1609-2018, 2018.

Xing, Y., Liang, Q., Wang, G., Ming, X., and Xia, X.: City-scale hydrodynamic modelling of urban flash floods: the issues of scale and resolution, Natural Hazards, 96, 473–496, https://doi.org/10.1007/s11069-018-3553-z, 2018.

Xu, D., Bisht, G., Engwirda, D., Feng, D., Tan, Z., and Ivanov, V. Y.: Uncertainties in Simulating Flooding During Hurricane Harvey Using 2D Shallow Water Equations, Water Resources Research, 61, https://doi.org/10.1029/2024wr038032, 2025.

Zheng, N., Li, Z., Rickert, G., Morales-Hernández, M., Özgen-Xian, I., and Caviedes-Vouillième, D.: High-performance coupled surface–subsurface flow simulation with SERGHEI-SWE-RE, Geoscientific Model Development, in press, 2026.

Zhu, J. and Mohanty, B. P.: Effective hydraulic parameters for steady state vertical flow in heterogeneous soils, Water Resources Research, 38, 1153, https://doi.org/10.1029/2001WR000831, 2002.