Journal of Environmental Radioactivity 298 (2026) 108075

Contents lists available at ScienceDirect

Journal of Environmental Radioactivity

journal homepage: www.elsevier.com/locate/jenr

Multi-model ensemble analysis of JRODOS atmospheric dispersion models for nuclear emergency planning

Ramy-Badr Ahmed [a,*], Thomas Schichtel [a], Dmytro Trybushnyi [a], Wolfgang Raskob [a,b], Sadeeb S. Ottenburger [a]

[a] Karlsruhe Institute of Technology (KIT), Institute for Thermal Energy Technology and Safety (ITES), Resilient and Smart Infrastructure Systems (RESIS) Department, Hermann-von-Helmholtz Platz 1, Eggenstein-Leopoldshafen, 76344, Germany
[b] Walter Bätz Consulting, (NUSAFE), Hermann-von-Helmholtz Platz 1, Eggenstein-Leopoldshafen, 76344, Germany

ARTICLE INFO

Dataset link: JRODOS Multi-Model Atmospheric Dispersion Pipeline for Nuclear Emergency Planning (Original data)

Keywords:
JRODOS
Nuclear emergency management
Atmospheric dispersion models
Ensemble modelling
Emergency planning
Triple modular redundancy
Multi-model decision support

ABSTRACT

The Real-Time On-Line Decision Support System (JRODOS) is operationally used worldwide to support off-site nuclear emergency management. This study presents a new JRODOS module for emergency planning that combines the JRODOS statistical module, which samples meteorological conditions over extended periods for a given site, with a dedicated post-processing workflow to analyze the outputs of the multiple atmospheric dispersion models integrated in JRODOS. Using the full emergency calculation chain, it derives early-phase protective action areas and applies user-defined thresholds to identify higher-probability zones. The approach is demonstrated for three representative nuclear power plant sites with contrasting terrain, and the resulting site- and model-dependent action areas are compared to assess the influence of model choice on planning outcomes.

1. Introduction

The protection of public health and the environment during nuclear emergencies requires rapid and well-informed decision-making supported by reliable computational tools. Following major accidents such as the Chernobyl disaster and the Fukushima Daiichi nuclear disaster, the development and continuous improvement of decision support systems (DSS) for nuclear emergency management has been a central priority for the international nuclear community (UNSCEAR, 2000; Steinhauser et al., 2014; Raskob et al., 2020). These priorities are reflected in international safety standards from the IAEA and the EU, including Council Directive 2013/59/EURATOM, which emphasizes the need for reliable tools and comprehensive emergency planning for nuclear facilities (IAEA, 2017, 2016; EU-Council, 2014).

Atmospheric dispersion modelling underpins such systems by predicting radionuclide transport, deposition and potential public doses, but these predictions remain uncertain because of meteorology, source term specification, terrain complexity and model parameterization (Ulimoen et al., 2022; Hernández-Ceballos et al., 2020; Korsakissok et al., 2020; Tomasi et al., 2019; Sørensen et al., 2020). These uncertainties are particularly important during the early phase of an emergency, when protective actions must be taken under time pressure. Accordingly, ensemble and multi-model approaches, together with structured decision support frameworks, have increasingly been advocated to characterize prediction uncertainty and support planning. In this context, ensemble-style analyses of large simulation sets can identify both typical dispersion patterns and less frequent but potentially critical scenarios (Ulimoen et al., 2022; Korsakissok et al., 2020; Sørensen et al., 2020; Meutter and Delcloo, 2022; Salpadimos et al., 2024; Korsakissok et al., 2019; Vries et al., 2019). Rather than relying on single deterministic runs, such methods help characterize the range of possible outcomes arising from these combined sources of uncertainty, thereby providing probabilistic information that can complement deterministic runs during preparedness and planning (Hernández-Ceballos et al., 2020; Korsakissok et al., 2020; Tomasi et al., 2019; Sørensen et al., 2020; Päsler-Sauer, 2010; Giovannini et al., 2020)

Within the JRODOS DSS (Ievdin et al., 2010; Ehrhardt et al., 1996; Institute for Thermal Energy Technology and Safety (ITES/RESIS), 2019; Raskob et al., 2017), which is used by emergency centres and responsible authorities supporting early off-site countermeasure decisions (Asselt et al., 2020; Ievdin et al., 2010; Ehrhardt et al., 1996;

* Corresponding author.
E-mail address: ramy-badr.ahmed@kit.edu (R.-B. Ahmed).

https://doi.org/10.1016/j.jenvrad.2026.108075
Received 2 April 2026; Received in revised form 31 May 2026; Accepted 7 June 2026
Available online 17 June 2026

Landman et al., 2014; Leung et al., 2018; Institute for Thermal Energy Technology and Safety (ITES/RESIS), 2019), the Local Scale Model Chain (LSMC) is used to assess hazards from nuclear accidents by simulating near-range atmospheric dispersion of released material (Landman et al., 2014). The LSMC integrates three atmospheric dispersion models (ADMs) with different modelling approaches: the Gaussian puff model RIMPUFF, the hybrid Lagrangian–Gaussian model DIPCOT, and the Lagrangian particle model LASAT (Mikkelsen et al., 1984; Andronopoulos et al., 2010; Janicke, 1994). These models have been validated in field experiments and intercomparison studies (Päsler-Sauer, 2010; Davakis et al., 2001, 2003; Dyer and Astrup, 2012), as well as in dedicated comparison analyses, and exhibit different strengths depending on terrain and flow conditions. Puff models are computationally efficient and well suited to flat terrain with homogeneous roughness, whereas Lagrangian particle models are more appropriate for complex terrain and local wind systems, albeit at higher computational cost. Ensemble and uncertainty analyses for nuclear emergency modelling have also been conducted in European research programmes, including CONCERT/CONFIDENCE (Korsakissok et al., 2019; Vries et al., 2019), where RIMPUFF and DIPCOT, or closely related configurations, were included in multi-model studies, whereas LASAT did not participate explicitly.

Beyond real-time support, the JRODOS statistical output module provides functionality to generate ensembles of dispersion and dose simulations by sampling meteorological conditions and, optionally, source term characteristics over extended periods for a given site. It has been used in previous studies to investigate the influence of meteorological variability on accident-consequence assessments and on the design of countermeasure areas, including ensemble-style applications within the CONCERT/CONFIDENCE programme and JRODOS-based scenario studies for German nuclear installations (Walter et al., 2016; Vries et al., 2019).

JRODOS includes an early countermeasure module that proposes action areas, such as evacuation and sheltering zones, based on national intervention criteria consistent with international recommendations and European basic safety standards (IAEA, 2017, 2016; EU-Council, 2014). Consequently, the choice of ADM can directly influence the spatial extent and frequency of proposed countermeasure areas presented to decision-makers. However, despite its operational use for real-time support, JRODOS is not designed, in its standard configuration, to generate the comprehensive emergency response plans required for preparedness activities.

In this study, we address this limitation by combining the existing JRODOS statistical module with a newly developed post-processing module to support emergency planning. Specifically, we investigate how the choice of ADM influences the spatial distribution and frequency of evacuation areas under identical input conditions. Large ensembles of simulations covering the full year 2024 are analyzed to capture realistic meteorological variability across three representative sites: Borssele, Netherlands, representing flat coastal terrain; Cattenom, France, representing a low mountain range; and Leibstadt, Switzerland, representing mountainous terrain.

The resulting statistical outputs are analyzed using frequency plots, spatial difference plots and combined multi-model visualizations. Furthermore, we introduce a multi-model overlay concept based on logical OR, majority and AND rules, inspired by ensemble forecasting (Leutbecher and Palmer, 2008) and triple modular redundancy (TMR) in safety-critical systems (Lyons and Vanderkulk, 1962). This approach enables the identification of regions of model consensus and divergence and provides a basis for deriving robust, decision-relevant action areas for emergency planning.

The contribution of this work is therefore threefold: (i) the post-processing module is a new component integrated with the JRODOS statistical engine — not an external script — and is published as part of the JRODOS toolchain; (ii) the application of triple-modular-redundancy logic to a JRODOS multi-ADM ensemble is, to our knowledge, the first such use in the radiological emergency-planning literature; and (iii) the analysis is planning-oriented, complementing prior JRODOS work that focused on real-time crisis support.

2. Materials and setup

In this study the JRODOS DSS LSMC was used to perform statistical dispersion simulations for three representative nuclear power plant (NPP) sites. Sections Section 2.1–2.8 summarize the LSMC framework, site and scenario specifications, meteorological inputs, computational configuration and statistical sampling. For a given site, all simulations used a common emergency model template that fixed the source term, release height, meteorological provider and background datasets, enabling direct inter-model comparison (Institute for Thermal Energy Technology and Safety (ITES/RESIS), 2019; Raskob et al., 2017; Walter et al., 2016).

2.1. JRODOS software architecture

JRODOS is a modular decision-support platform for nuclear and radiological emergencies that has been developed and maintained by the Karlsruhe Institute of Technology (KIT) over more than two decades (Ievdin et al., 2010; Ehrhardt et al., 1996). It follows a service-oriented architecture (SOA) in which the user interface, workflow control, model execution, data management, and visualization are separated into interoperable components with well-defined interfaces (Fig. 1). This structure supports both operational emergency response and retrospective analyses, while allowing individual models and data-processing components to be extended through a plug-in mechanism.

The Java-based Client provides interactive scenario configuration, simulation control, and visualization of outputs as maps, time series, and tabular reports. The Manager coordinates the overall workflow, including scheduling, plug-in loading, communication with the Client and Engine, and database-mediated data exchange; scheduled and event-driven tasks are handled through dedicated scheduling mechanisms such as Quartz. Computational models are executed by the Engine through model wrappers, while established numerical kernels are commonly preserved as native Fortran/C shared libraries. This allows atmospheric dispersion and dose-assessment models to be integrated into a common operational framework without rewriting their numerical implementations. Model chains can therefore combine dispersion, deposition, dose, food-chain, inhabited-area, hydrological, and data-assimilation components within reproducible JRODOS projects.

Persistent storage is provided by a PostgreSQL/PostGIS backend containing meteorological and radiological monitoring data, numerical weather prediction inputs, geospatial layers, user-defined scenario parameters, and simulation outputs. Database access is mediated through the Manager using a Java persistence layer based on Hibernate ORM and JDBC. Standard preprocessing routines ingest external data such as GRIB1/GRIB2 meteorological files and EURODEP monitoring data and convert them into model-ready formats. Geographic information system (GIS) support, including PostGIS and WMS-compatible layers, enables terrain, land use, population, and infrastructure data to be projected onto computational grids, and simulation results can be exported as GIS-compatible layers for downstream analysis. In this study, the common JRODOS architecture ensures that RIMPUFF, DIPCOT, and LASAT are run with standardized scenario definitions, source terms, meteorological inputs, and output handling; consequently, inter-model differences can be attributed primarily to the atmospheric dispersion models rather than to separate software or data-management workflows.

2.2. JRODOS LSMC

The LSMC simulates atmospheric dispersion, deposition and resulting dose fields for near-range releases. It provides several interchangeable ADMs: the Gaussian puff model RIMPUFF, the hybrid Lagrangian–Gaussian model DIPCOT, and the Lagrangian particle model LASAT. All three models use the common JRODOS meteorological pre-processing chain and therefore operate with identical meteorological input fields,

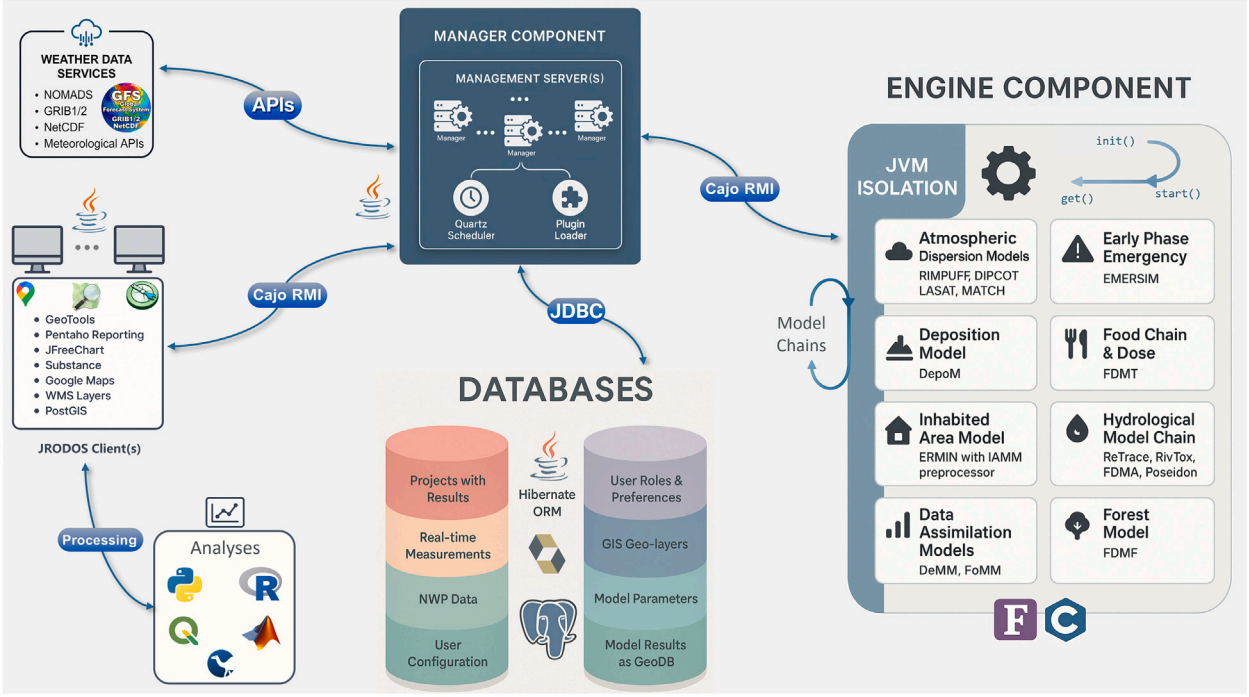


Fig. 1. JRODOS software components and design structure. The Client provides scenario configuration and visualization; the Manager coordinates workflows, scheduling, plug-in loading, and data exchange; the Engine executes computational models through model wrappers and native libraries; and the Database provides persistent storage for meteorological data, measurements, geospatial layers, scenario parameters, and simulation outputs.

allowing inter-model differences to be attributed primarily to the transport and dispersion formulations themselves.

RIMPUFF represents the release as a sequence of Gaussian puffs advected along trajectories driven by the mean wind. Horizontal and vertical puff spreading evolve according to Pasquill–Gifford-type stability classes and atmospheric turbulence, while similarity-scaling methods are used to account for land use and surface roughness effects. Dry and wet deposition are applied at the puff level.

DIPCOT combines Lagrangian particle transport with Gaussian puff reconstruction. In the default JRODOS configuration, computational particles act as centres of three-dimensional Gaussian-distributed puffs, reducing the number of particles required while preserving statistical accuracy. Turbulent fluctuations are derived from similarity theory using the JRODOS meteorological pre-processor, and orographic flow modulation is represented through a diagnostic mass-consistent wind field.

LASAT is a pure Lagrangian particle model in which individual particles follow stochastic trajectories determined by the resolved wind field and turbulent velocity increments generated through a stochastic Markov-chain formulation consistent with VDI 3945. Concentration fields are reconstructed statistically from the particle ensemble, allowing detailed representation of complex terrain effects, local flow structures, and near-source dispersion features.

These methodological differences — particularly the treatment of turbulence, stochastic transport, concentration reconstruction, and terrain interactions — can produce substantially different macroscopic dispersion and dose patterns under realistic meteorological conditions. Because the resulting behaviour emerges from complex non-linear interactions between model formulation, atmospheric dynamics, and local topography, the magnitude and spatial structure of inter-model differences cannot be inferred directly from the mathematical formulations alone and therefore require systematic simulation-based comparison as discussed in Sections 4 and 6.

JRODOS also implements automated decision logic that flags grid cells where projected doses exceed predefined national intervention thresholds — e.g., evacuation, sheltering, relocation — based on dose projections integrated over specified timeframes (Institute for Thermal Energy Technology and Safety (ITES/RESIS), 2019; Raskob et al., 2017).

2.3. Study sites

We considered three European NPPs chosen to represent distinct terrain classes and common planning challenges:

- Borssele, Netherlands — flat coastal terrain at the North Sea coast, representative of simple dispersion conditions.
- Cattenom, France — hilly terrain representative of low-mountain environments.
- Leibstadt, Switzerland — mountainous terrain with complex orography.

This selection permits evaluation of how terrain and local flow regimes modulate ADM behaviour and the resulting action areas.

2.4. Accident scenario and source term

A standardized severe accident source term from the JRODOS library, F6.DWR-FKA, was used for all sites. Although the source term is not known in the early phase of an emergency, using a high-release scenario such as FKA provides a representative basis for assessing worst-case dispersion and countermeasure areas (Walter et al., 2016). F6.DWR-FKA, classified INES 7 (IAEA, 2026), represents a pressurized water reactor event (3733 MW thermal) with an uncovered steam-generator heating pipe leak (6 cm^2) and a failed fresh steam valve; release begins 20.8 h after the initiating event, with an initial ≈7 h short release followed by an extended continuous release yielding a total release duration ≈ 50 h. The source term includes major radionuclides typically associated with large off-site consequences (I-131, Cs-137, Xe-133 and Sr-90) with released activities of I-131 (310 PBq), Cs-137 (29 PBq), Xe-133 (1800 PBq) and Sr-90 (0.88 PBq), respectively as specified in the JRODOS F6.DWR-FKA library entry (Walter et al.,

2016; IAEA, 2026). These radionuclide-specific activities, together with the release timing and dose-conversion assumptions encoded in JRODOS, define the source input used to evaluate the 100 mSv evacuation criterion.

The analysis concentrates on early-phase protective measures. Intervention thresholds applied throughout are the former German values implemented in the JRODOS database: 100 mSv for evacuation and calculated within 7 days exposure in open air, comprising external exposure and committed effective dose from radionuclides inhaled under the assumption of continuous outdoor exposure. Walter et al. (2016).

2.5. Meteorological data

Numerical weather prediction (NWP) data were obtained from the NOAA NOMADS archive in GRIB2 format NOMAD (2024). The data provide:

- 6-hourly initializations at 00, 06, 12, and 18 UTC,
- horizontal grid spacing of $0.25° \times 0.25°$ (approximately 28 km),
- 3-hourly output intervals, and
- 26 vertical pressure levels (1000–10 hPa).

JRODOS meteorological pre-processor ingests these data, applies topographic and land-use corrections, and interpolates the fields to the calculation grid and ADM time steps (Institute for Thermal Energy Technology and Safety (ITES/RESIS), 2019).

2.6. Computational grid and prognosis configuration

Simulations used the JRODOS default Cartesian emergency grid: 8056 square cells arranged in five nested frames (rings) around the release point, covering a horizontal domain of 800 km × 800 km. Cell side length doubles between frames to deliver fine resolution near the source (innermost cell size 1 km) and progressively coarser resolution farther out; the outermost ring extends to approximately 400 km radial distance. Meteorological fields were interpolated to the ADM calculation grid at 30 min intervals. The archived model-output interval was 1 h for all three ADMs, whereas the shorter ADM-internal transport and release steps are model-specific and are reported separately in Table 1.

For production runs, the prognosis duration was set to 72 h to cover the F6.DWR-FKA release (approximately 50 h) plus approximately 22 h for plume passage. Although the German evacuation criterion is formally defined as a 7-day effective dose in open air, the F6.DWR-FKA release terminates essentially within the first 50 h and the dominant inhalation and cloudshine contributions accumulate during plume passage. Integrating the dose over 72 h therefore captures the bulk of the early-phase contribution to the 7-day evacuation dose and provides an operationally relevant approximation of the formal criterion. The remaining contribution from groundshine between days 4 and 7 is comparatively small and mainly affects cells already close to the intervention threshold; consequently, the resulting evacuation fields should be interpreted as early-phase approximations rather than numerically exact reproductions of the full 7-day dose. We return to this approximation in Section 6, where we outline the implications and identify extension to a full 7-day integration as a topic for future work.

The longer 72 h prognosis increased computation times, particularly for LASAT at the mountainous Leibstadt site. Summed over all sites and ADMs, the total CPU time of the ensembles was about 11.5 days. Under ideal parallelization, distributing this workload over N machines would reduce the wall-clock time to approximately $11.5/N$ days; for three workstations, this corresponds to about one-third of the single-machine wall-clock time. In our setup, the production runs on three workstations required about 8–9 days of wall-clock time and produced noticeable differences in LASAT outputs relative to shorter test runs (Section 4).

2.7. ADM-specific numerical configuration

To document ADM-specific configuration details, Table 1 summarizes the numerical controls used by the three ADMs. Because the models use different numerical formulations, the entries are not directly equivalent: RIMPUFF represents the release as sequential Gaussian puffs, DIPCOT is used in a hybrid Lagrangian–Gaussian puff formulation, and LASAT is a Lagrangian particle model with adaptive internal stepping. The table therefore reports the model-control settings relevant to the present simulations rather than implying a common numerical discretization across all ADMs.

2.8. Statistical module application

To capture realistic meteorological variability, the JRODOS statistical module was used with the EMERGENCY template described above. The sampling strategy generated one dispersion simulation per calendar day over the year 2024. For each day the EMERGENCY template parameters (including the F6.DWR-FKA source term and background datasets) were held fixed, while a single release start time t_0 was drawn from a uniform distribution on $[0, 24)$ h within that calendar day. The 72 h prognosis window of each simulation was then launched at t_0, so that for every (site, ADM) combination the procedure produces exactly 365 statistical members covering the full year 2024. The uniform sampling of t_0 was chosen to avoid any preferred phase with respect to the diurnal cycle or to the 6-hourly NWP initialization times; thus the only differences between ensemble members arise from the meteorology (synoptic and diurnal variations) and the relative phasing of the fixed source term with the evolving flow. This setup yields a statistical ensemble of dispersion outcomes that reflects seasonal and synoptic variability for the chosen year while keeping the accident scenario itself fixed.

The workflow applied independently to each ADM (RIMPUFF, DIPCOT, LASAT) and each site produced three statistically consistent ensembles. Outputs were post-processed to construct frequency maps, pairwise difference plots and multi-model overlays as described in Section 3.

3. Statistical analysis methods

3.1. Thresholded frequency analysis

Frequency plots quantify how often a given action criterion, such as evacuation or sheltering, is exceeded across the collection of JRODOS runs. For each site, ADM, action, and grid cell, the number of runs in which the relevant quantity exceeds the intervention level is divided by the total number of available runs. The resulting normalized value approximates the probability that the action criterion is fulfilled in that cell over the year, under the assumed source term and meteorological conditions. Algorithm 1, implemented in our repository (Ahmed, 2026), shows the computation of these exceedance frequencies as the fraction of ensemble members whose simulated values exceed the intervention level.

For each grid cell $j \in G$ and model m, the exceedance frequency is defined as

$$p_j^{(m)} = \frac{1}{N_R^{(m)}} \sum_{i=1}^{N_R^{(m)}} \mathbf{1}\left(Y_{ij}^{(m)} \geq L\right),$$

where $Y_{ij}^{(m)}$ is the JRODOS quantity of interest for run i at grid cell j, $N_R^{(m)}$ is the number of available runs for model m, and L is the intervention level. Here, $\mathbf{1}(\cdot)$ denotes the indicator function. For any logical statement A, it is defined by

$$\mathbf{1}(A) = \begin{cases} 1, & \text{if } A \text{ is true,} \\ 0, & \text{otherwise.} \end{cases}$$

Table 1
ADM-specific numerical settings used in the JRODOS LSMC runs. The entries report the model-control settings relevant to the present simulations; they are not necessarily one-to-one equivalent because the three ADMs use different numerical formulations.

ADM	Vertical discretization	Internal transport/release step	Output interval
RIMPUFF	No fixed Eulerian vertical mesh; vertical spreading is represented through the Gaussian puff dispersion parameter σ_z.	Default puff-release interval $\tau = 300$ s; advection time between successive computations $t_{\mathrm{adv}} = 20$ s.	1 h
DIPCOT	RODOS-DIPCOT concentration cells coincide horizontally with the RODOS grid; for box-counting concentration reconstruction the vertical cell dimension is $\Delta z = 50$ m.	Hybrid Lagrangian–Gaussian puff formulation; in the default setup one puff is released every 10 s.	1 h
LASAT	User-configurable vertical discretization; LASAT allows freely specified vertical grid spacing, with up to 200 arbitrary vertical intervals.	Variable internal particle time step, typically of the order of several seconds and determined internally according to the case.	1 h

Note: The archived output interval was common to all three ADMs. The shorter transport and release steps are ADM-internal numerical controls and are not directly comparable across models.

The quantity $p_j^{(m)}$ represents the empirical frequency, and thus an approximation of the probability, that the action criterion is fulfilled in cell j over the ensemble of runs of model m.

To highlight spatial regions with non-negligible exceedance likelihood, a probability trigger level $T \in [0,1]$ may be applied. In this manuscript the 10% and 15% thresholds are used as visualization and planning filters, not as intervention criteria: they suppress isolated low-frequency activations, retain the spatially persistent cells that matter most for preparedness, and make the map layers readable for decision-makers. For single-ADM maps we use $T = 0.10$, whereas for multi-model overlays we adopt a stricter per-ADM threshold $T = 0.15$ so that, when up to three models can support a cell, the overall false-positive rate remains comparable to the single-model 10% case.

The triggered field

$$p_j^{(\mathrm{rep},m)} = \begin{cases} p_j^{(m)}, & \text{if } p_j^{(m)} > T, \\ 0, & \text{otherwise} \end{cases}$$

is used for map representation of exceedance probabilities. Equivalently, cells with $p_j^{(m)} \le T$ are omitted from the displayed result set. In practice, the computation of $p_j^{(m)}$ is performed efficiently using parallel chunking over runs and grid cells.

3.2. Model difference analysis

Model difference analysis compares the exceedance-frequency fields $p_j^{(m)}$ defined in Section 3.1. These fields are used as input to the pairwise difference procedure summarized in Algorithm 2 and implemented in our repository (Ahmed, 2026). For an ordered ADM pair (m_a, m_b), the signed difference in relative exceedance frequency at grid cell j is defined as

$$d_j^{(m_a,m_b)} = p_j^{(m_a)} - p_j^{(m_b)}.$$

Positive values of $d_j^{(m_a,m_b)}$ indicate that model m_a triggers the action more frequently than model m_b at cell j, while negative values indicate the opposite. Values close to zero correspond to similar behaviour of the two models. We also consider the absolute difference

$$\left| d_j^{(m_a,m_b)} \right|,$$

which measures the magnitude of disagreement irrespective of which model produces the larger exceedance frequency.

To quantify the pairwise ADM differences beyond visual inspection of the difference fields, four scalar inter-model metrics are computed for each ordered model pair: mean bias (MB), root-mean-square error (RMSE), fractional bias (FB), and spatial pattern correlation (r). These metrics are evaluated over the pair-specific union domain

$$S_{ab} = \left\{ j \in G : p_j^{(m_a)} > 0 \ \vee \ p_j^{(m_b)} > 0 \right\}, \qquad n_{ab} = |S_{ab}|.$$

Here, S_{ab} is the set of grid cells where at least one of the two ADMs has a non-zero exceedance frequency, and n_{ab} is the number of cells in that domain. Excluding zero–zero cells avoids inflating apparent agreement by including large numbers of cells where neither model triggers the action.

The mean bias is defined as

$$\mathrm{MB}^{(m_a,m_b)} = \frac{1}{n_{ab}} \sum_{j \in S_{ab}} \left(p_j^{(m_a)} - p_j^{(m_b)} \right).$$

Positive MB values indicate that m_a produces higher exceedance frequencies than m_b on average over the comparison domain, while negative values indicate the opposite.

The root-mean-square error is defined as

$$\mathrm{RMSE}^{(m_a,m_b)} = \sqrt{\frac{1}{n_{ab}} \sum_{j \in S_{ab}} \left(p_j^{(m_a)} - p_j^{(m_b)} \right)^2}.$$

Unlike MB, RMSE is insensitive to the sign of the difference and therefore measures the overall magnitude of pairwise disagreement.

The fractional bias is defined as

$$\mathrm{FB}^{(m_a,m_b)} = \frac{2\left(\bar{p}_{ab}^{(m_a)} - \bar{p}_{ab}^{(m_b)} \right)}{\bar{p}_{ab}^{(m_a)} + \bar{p}_{ab}^{(m_b)}},$$

where

$$\bar{p}_{ab}^{(m)} = \frac{1}{n_{ab}} \sum_{j \in S_{ab}} p_j^{(m)}$$

is the mean exceedance frequency of model m over the pair-specific comparison domain. FB provides a normalized signed measure of the mean difference between the two models. With this convention, positive FB values indicate higher mean exceedance frequencies for m_a, whereas negative values indicate higher mean exceedance frequencies for m_b.

The spatial pattern correlation is defined as

$$r^{(m_a,m_b)} = \frac{\sum_{j \in S_{ab}} \left(p_j^{(m_a)} - \bar{p}_{ab}^{(m_a)} \right) \left(p_j^{(m_b)} - \bar{p}_{ab}^{(m_b)} \right)}{\sqrt{\sum_{j \in S_{ab}} \left(p_j^{(m_a)} - \bar{p}_{ab}^{(m_a)} \right)^2} \sqrt{\sum_{j \in S_{ab}} \left(p_j^{(m_b)} - \bar{p}_{ab}^{(m_b)} \right)^2}}.$$

The correlation coefficient r measures similarity in the spatial pattern of exceedance frequencies independently of the mean magnitude.

Together, MB and FB quantify systematic signed differences, RMSE quantifies the overall magnitude of pairwise disagreement, and r quantifies similarity in the spatial structure of the exceedance-frequency fields.

The pairwise combinations are interpreted separately because they isolate different process contrasts. The RIMPUFF–LASAT comparison mainly probes puff-versus-particle transport and differences in the

Algorithm 1 Parallel computation of (thresholded) grid-cell exceedance frequencies for model m

Require: Grid cells $G = \{1, \dots, N_G\}$;
ensemble runs $R^{(m)} = \{1, \dots, N_R^{(m)}\}$ for model m;
intervention level L;
quantity fields $Y_i^{(m)} = (Y_{i1}^{(m)}, \dots, Y_{iN_G}^{(m)})$ for each run i;
number of workers W;
probability threshold $T \in [0, 1]$.
Ensure: Exceedance frequency $p_j^{(m)}$ and reported exceedance probability $p_j^{(\mathrm{rep},m)}$ for each grid cell $j \in G$.

for all $j \in G$ **do** $c_j \leftarrow 0$ **end for**
Partition $R^{(m)}$ into W disjoint subsets $R_1^{(m)}, \dots, R_W^{(m)}$ such that $R^{(m)} = \bigcup_{w=1}^{W} R_w^{(m)}$. ▷ parallel run distribution
for $w = 1$ **to** W **in parallel do**
  for all $j \in G$ **do** $c_j^{(w)} \leftarrow 0$ **end for**
for all $i \in R_w^{(m)}$ **do**
  $Y_i^{(m)} \leftarrow$ quantity field for run i of model m ▷ read model output for run i
  for all $j \in G$ **do**
    if $Y_{ij}^{(m)} \geq L$ **then**
      $c_j^{(w)} \leftarrow c_j^{(w)} + 1$ ▷ count exceedance at cell j
    end if
  end for
end for
end for ▷ end parallel loop over workers
for all $j \in G$ **do**
  $c_j \leftarrow \sum_{w=1}^{W} c_j^{(w)}$
  $p_j^{(m)} \leftarrow c_j / N_R^{(m)}$ ▷ compute exceedance frequency
  if $p_j^{(m)} > T$ **then**
    $p_j^{(\mathrm{rep},m)} \leftarrow p_j^{(m)}$ ▷ keep cells above probability threshold
  else
    $p_j^{(\mathrm{rep},m)} \leftarrow 0$
  end if
end for
return $\{p_j^{(m)}, p_j^{(\mathrm{rep},m)}\}_{j \in G}$ ▷ exceedance frequencies per grid cell and model

treatment of turbulence and terrain. The RIMPUFF–DIPCOT comparison isolates the effect of replacing the Gaussian puff parameterization with a hybrid Lagrangian–Gaussian scheme. The DIPCOT–LASAT comparison is the most diagnostic comparison for understanding why two nominally Lagrangian models can still diverge when one uses smoother kernel reconstruction and diagnostic wind fields while the other resolves stochastic particle motion more explicitly.

3.3. Multi-model analysis

The multi-model concept extends the analysis from individual models to the joint information from all three ADMs, analogous to multi-model statistical analysis in numerical weather prediction and to redundancy concepts in safety-critical systems. Here, the three ADMs are treated as parallel simulators of the same accident scenario under identical meteorological conditions. Their outputs are combined in an overlay framework that records which models activate a given action in each grid cell. The construction of the thresholded multi-model overlay is summarized in Algorithm 3 and implemented in our repository (Ahmed, 2026).

The multi-model overlay uses the exceedance-frequency fields $p_j^{(m)}$ defined in Section 3.1. For a global probability trigger level $T \in [0, 1]$, model m is considered active in cell j when

$$p_j^{(m)} > T.$$

This condition defines the binary activation signal

$$a_j^{(m)} = \mathbf{1}\left(p_j^{(m)} > T\right), \qquad a_j^{(m)} \in \{0, 1\}.$$

In the three-model analysis, we use the ordered model tuple

$$(m_1, m_2, m_3) = (\mathrm{RIMPUFF}, \mathrm{DIPCOT}, \mathrm{LASAT}).$$

The activation vector

$$\left(a_j^{(m_1)}, a_j^{(m_2)}, a_j^{(m_3)}\right)$$

therefore represents one of the eight possible activation patterns of the three ADMs.

The number of ADMs active in cell j is

$$s_j = \sum_{k=1}^{3} a_j^{(m_k)}, \qquad s_j \in \{0, 1, 2, 3\}.$$

The activation pattern is also encoded by the integer bitmask

$$f_j = a_j^{(m_1)} + 2a_j^{(m_2)} + 4a_j^{(m_3)}, \qquad f_j \in \{0, \dots, 7\}.$$

Thus, f_j records the exact combination of ADMs that exceed the probability trigger level in cell j, whereas s_j records only how many ADMs are active.

3.3.1. TMR decision logic

The decision logic applied to multi-model action information is inspired by the triple modular redundancy (TMR) concept used in safety-critical technical systems. In TMR architectures, three independent subsystems perform the same task, and a voter implements a majority rule to identify faulty behaviour and select the effective output (Lyons and Vanderkulk, 1962). By analogy, the three ADMs can be viewed as redundant simulators. This motivates three planning modes: a permissive OR rule, a majority rule, and a conservative AND rule. In the implementation these correspond to the union, consensus, and intersection overlay modes, respectively (see Table 2).

The OR/union mode is the most inclusive and identifies all cells supported by at least one ADM. The majority/consensus mode suppresses single-model activations and retains cells supported by at least two

Algorithm 2 Computation of ADM-wise relative-frequency differences and scalar inter-model metrics

Require: Grid cells $G = \{1, \ldots, N_G\}$;
set of ADMs $\mathcal{M} = \{\text{RIMPUFF}, \text{DIPCOT}, \text{LASAT}\}$;
ordered model pairs $\mathcal{P} \subseteq \mathcal{M} \times \mathcal{M}$;
exceedance-frequency fields $p_j^{(m)}$ for each $j \in G$ and $m \in \mathcal{M}$.

Ensure: For each pair $(m_a, m_b) \in \mathcal{P}$: signed difference field $d_j^{(m_a,m_b)}$, absolute difference field $|d_j^{(m_a,m_b)}|$, and scalar metrics MB, RMSE, FB, and r.

for all $(m_a, m_b) \in \mathcal{P}$ **do**
  Define comparison domain

$$S_{ab} \leftarrow \left\{ j \in G : p_j^{(m_a)} > 0 \vee p_j^{(m_b)} > 0 \right\}, \qquad n_{ab} \leftarrow |S_{ab}|.$$

▷ exclude zero–zero cells

  for all $j \in S_{ab}$ **do**
    Compute signed relative-frequency difference

$$d_j^{(m_a,m_b)} \leftarrow p_j^{(m_a)} - p_j^{(m_b)}.$$

    Compute absolute difference

$$|d_j^{(m_a,m_b)}| \leftarrow \left| p_j^{(m_a)} - p_j^{(m_b)} \right|.$$

  end for
  Compute $\bar{p}_{ab}^{(m_a)}$ and $\bar{p}_{ab}^{(m_b)}$ over S_{ab}. ▷ pair-specific model means
  Compute $\text{MB}^{(m_a,m_b)}$, $\text{RMSE}^{(m_a,m_b)}$, $\text{FB}^{(m_a,m_b)}$, and $r^{(m_a,m_b)}$ using the definitions above.
end for
return Difference fields and scalar inter-model metrics for all model pairs.

Table 2
Logical rules for combining ADM activation signals in the TMR-inspired decision logic.

Decision logic	Overlay mode	Condition	Interpretation
OR rule	Union	$s_j \geq 1$	At least one ADM activates the action
Majority rule	Consensus	$s_j \geq 2$	At least two ADMs activate the action
AND rule	Intersection	$s_j = 3$	All three ADMs activate the action

Table 3
Colour scheme for model combinations in the overlay plots, expressed using the activation vector $(a_j^{(\text{RIMPUFF})}, a_j^{(\text{DIPCOT})}, a_j^{(\text{LASAT})})$ and the corresponding bitmask f_j.

Colour	Activation vector	f_j	Interpretation
White	(0, 0, 0)	0	None; no ADM active
Red	(1, 0, 0)	1	RIMPUFF only
Blue	(0, 1, 0)	2	DIPCOT only
Green	(0, 0, 1)	4	LASAT only
Magenta	(1, 1, 0)	3	RIMPUFF + DIPCOT
Cyan	(1, 0, 1)	5	RIMPUFF + LASAT
Yellow	(0, 1, 1)	6	DIPCOT + LASAT
Dark grey	(1, 1, 1)	7	All three ADMs

ADMs. The AND/intersection mode is the most restrictive and retains only cells where all three ADMs activate the action.

3.3.2. RGB representation

With three ADMs, a bitmask-based colour scheme provides a compact visual encoding of which models trigger an action in each grid cell. The colour assigned to a cell is determined by the activation mask f_j. Table 3 gives the colour mapping used in the overlay plots.

In contrast to standard additive colour mixing, the combination of all three active ADMs is represented by dark grey rather than white, while white is reserved for cells where no model triggers the action. This choice improves the legibility of map labels and avoids confusion between full three-model agreement and the background.

4. Results

The plotting conventions in all maps use Cartesian axes labelled in kilometres with the release point at (0, 0) (marked by a St. Andrew's cross). Frequency maps show normalized frequencies in 10% intervals, while overlay maps use the RGB scheme defined in Table 3. Concentric circles and radial lines with angle labels are added in light grey to provide approximate spatial references; together, these rays and rings partition the area around the site into sectors and distance bands that facilitate interpretation of the action areas. All ADM panels are plotted with identical x/y limits so that spatial extents are directly comparable across models. Sea cells are retained in the maps for geographic context but are excluded from any planning interpretation when they fall outside inhabited areas. Results for evacuation are shown here.

4.1. Thresholded frequency analysis

Frequency maps quantify, per cell, the fraction of statistical members in which the intervention level is exceeded (Section 3.1). The three ADMs produce distinct spatial patterns of proposed evacuation areas, even when all runs use identical source term and meteorological inputs.

4.1.1. Borssele (flat coastal terrain)

At Borssele the three ADMs agree visually only in the immediate near field; at larger distances the location and extent of proposed evacuation areas differ markedly. For the 72 h prognosis the maximum normalized evacuation frequencies observed are approximately: LASAT 100.0%, DIPCOT 71.9%, RIMPUFF 70.2%. LASAT produces relatively compact high-probability regions near the source, whereas RIMPUFF and DIPCOT exhibit broader, more scattered low-probability activations (< 10%), Fig. 2. The annual aggregation tends to obscure prevailing wind directions in the low-frequency contours; nevertheless the dominant southwesterly–northeasterly axis at Borssele remains visible in the > 40% contours of all three ADMs and aligns with the local climatology. Sea cells are retained in the maps because offshore receptors (shipping lanes, offshore wind installations) are part

Algorithm 3 Construction of thresholded multi-model overlay from exceedance frequencies

Require: Grid cells $G = \{1, \ldots, N_G\}$;
ordered models $(m_1, m_2, m_3) = (\text{RIMPUFF}, \text{DIPCOT}, \text{LASAT})$;
exceedance frequencies $p_j^{(m_k)}$ for each model m_k and cell j;
global probability threshold $T \in [0, 1]$;
overlay mode $\text{mode} \in \{\text{union}, \text{consensus}, \text{intersection}\}$.
Ensure: For each selected cell j, a bitmask $f_j \in \{1, \ldots, 7\}$ encoding which models exceed the probability threshold.

for all $j \in G$ **do**
  $f_j \leftarrow 0$
  $s_j \leftarrow 0$ ▷ initialize bitmask and model-support count
  if $p_j^{(m_1)} > T$ **then**
    $f_j \leftarrow f_j$ OR 1
    $s_j \leftarrow s_j + 1$ ▷ set bit for RIMPUFF
  end if
  if $p_j^{(m_2)} > T$ **then**
    $f_j \leftarrow f_j$ OR 2
    $s_j \leftarrow s_j + 1$ ▷ set bit for DIPCOT
  end if
  if $p_j^{(m_3)} > T$ **then**
    $f_j \leftarrow f_j$ OR 4
    $s_j \leftarrow s_j + 1$ ▷ set bit for LASAT
  end if
  if $s_j = 0$ **then**
    continue ▷ no model exceeds the probability threshold
  end if

  $\text{keep}(j) \leftarrow \text{False}$
  if mode = union **then**
    $\text{keep}(j) \leftarrow (s_j \geq 1)$ ▷ retain all non-zero model combinations
  else if mode = consensus **then**
    $\text{keep}(j) \leftarrow (s_j \geq 2)$ ▷ retain cells supported by at least two ADMs
  else if mode = intersection **then**
    $\text{keep}(j) \leftarrow (s_j = 3)$ ▷ require all three ADMs active
  end if
  if $\text{keep}(j)$ **then**
    Store f_j as the class label for cell j
  end if
end for
return $\{f_j : j \in G, \text{keep}(j) = \text{True}\}$ ▷ bitmasks encoding selected model combinations

of the planning domain; however, the new post-processing module also supports an optional land-mask for applications focused solely on terrestrial countermeasures.

To emphasize the more likely action areas, frequency maps can be displayed only for cells above a user-defined threshold, Fig. 3. The choice of threshold is driven by planning considerations rather than by a fundamental statistical criterion. For single-ADM maps we use a 10% probability cutoff, which at Borssele yields a radial extent of the highlighted area (approximately 30 km) comparable to historical planning zones for the FKA scenario class and consistent with conservative emergency-planning practice. The apparent left–right symmetry of some low-frequency bands is a plotting artefact of the annual aggregation over many wind directions; it does not imply an absence of prevailing-wind structure. In the coastal case, the most frequent sectors are still aligned with the dominant transport directions, and the sea-facing sectors remain mostly inactive once the 10% planning filter is applied.

4.1.2. Cattenom (low-mountain terrain)

Frequency maps are more asymmetric than at Borssele; LASAT again tends to produce compact, high-frequency features, while RIMPUFF and DIPCOT give smoother, more extended patterns. Absolute frequencies are generally similar or somewhat lower than at Borssele for some ADMs, but low-frequency activations can extend further into particular sectors Fig. 4.

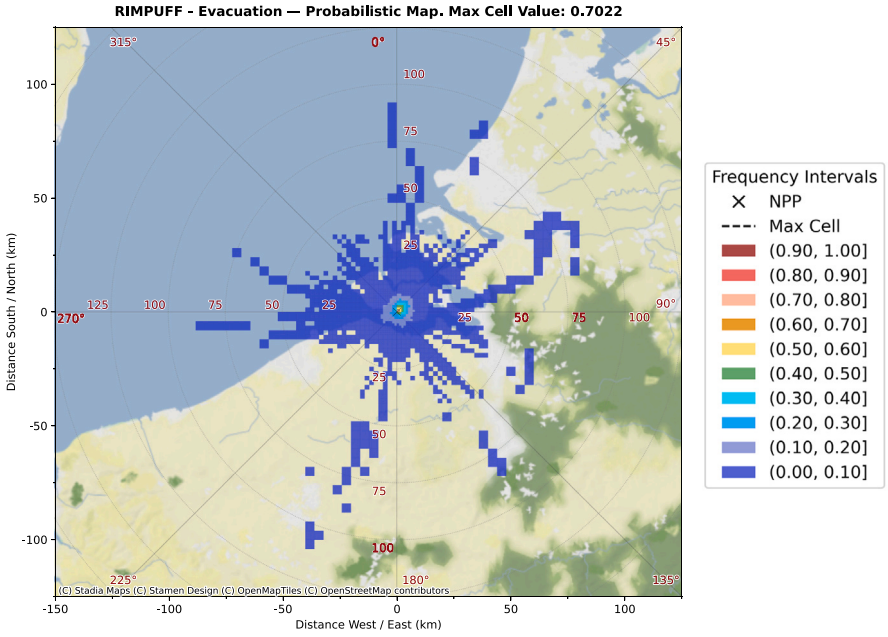
4.1.3. Leibstadt (mountainous terrain)

The strongest ADM dependence appears at Leibstadt. LASAT shows detailed structures conforming to valleys and flow channels; RIMPUFF and DIPCOT give smoother, more diffuse fields. Anisotropy is pronounced and the spatial structure of action areas is markedly model-dependent Fig. 5.

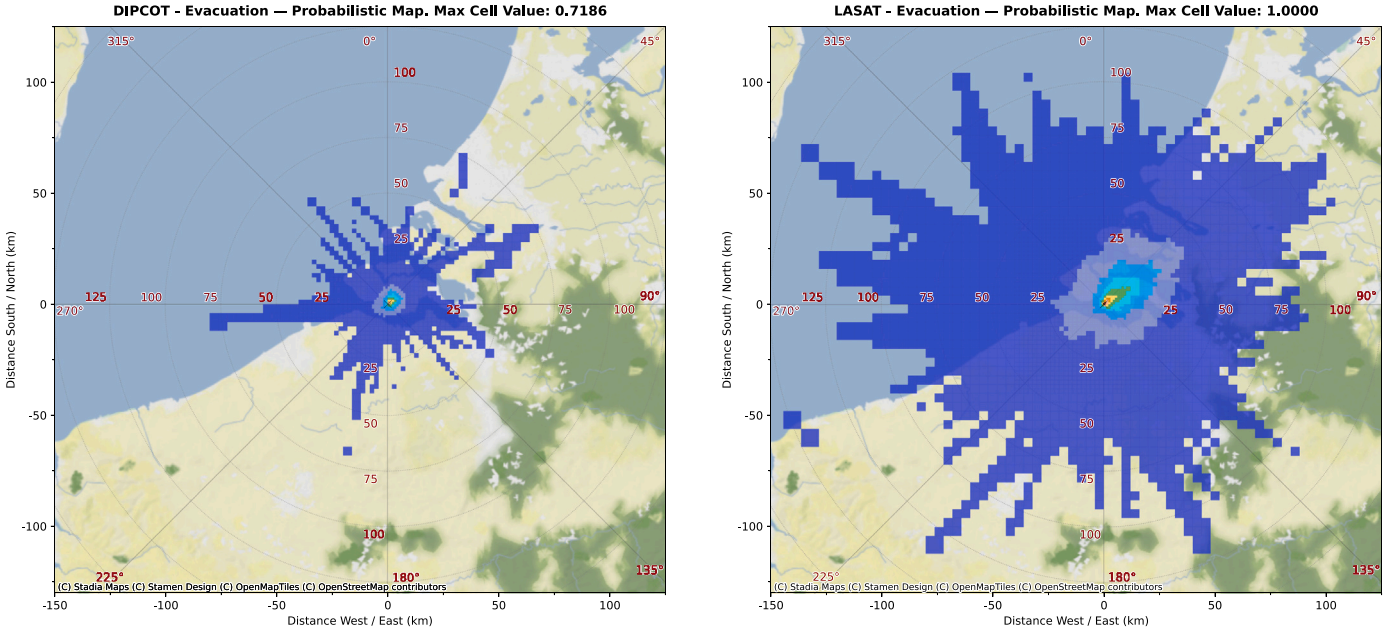
4.2. Model difference results

Pairwise difference fields confirm that discrepancies involving LASAT are largest. For Borssele, the maximum negative differences, corresponding to higher LASAT exceedance frequency, reach approximately −29.2% and −24.3% in the comparisons with RIMPUFF and DIPCOT, respectively. In contrast, the maximum positive differences, corresponding to higher RIMPUFF or DIPCOT exceedance frequency, are smaller, reaching approximately 4.1% and 3.3%. Differences between RIMPUFF and DIPCOT are substantially smaller, with extrema of approximately −4.9% and 8.7%. Regions where the difference is close to zero identify cells that are relatively robust to ADM choice, whereas large-magnitude differences highlight locations where the selected ADM would strongly affect emergency-planning outputs (Fig. 6).

The scalar metrics support the visual interpretation of the difference fields (Table 4). Across all three sites, the RIMPUFF–DIPCOT pair



(a)



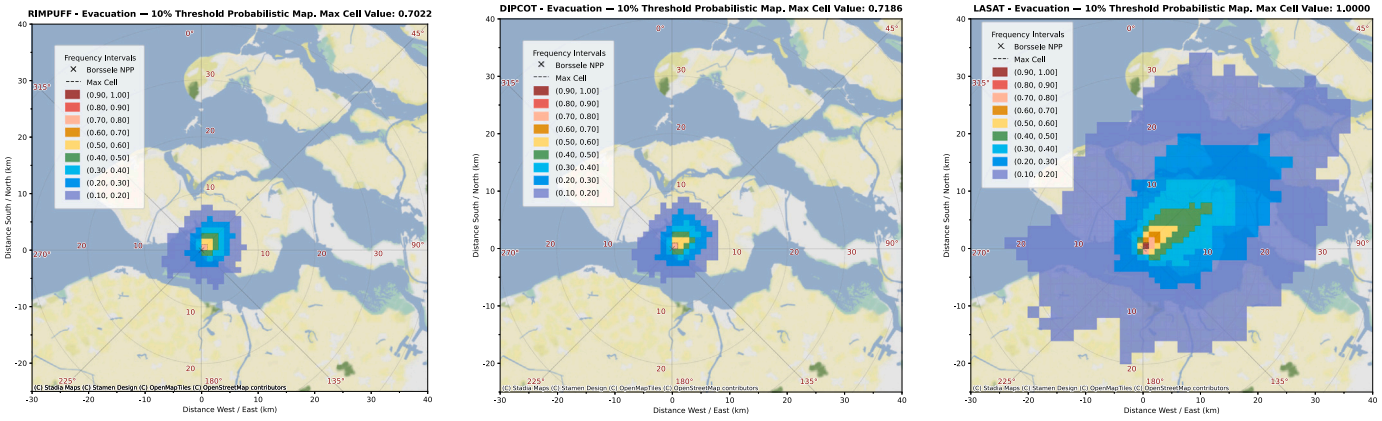
(b) (c)

Fig. 2. Normalized evacuation frequency at Borssele (72 h prognosis). Panels: (a) RIMPUFF, (b) DIPCOT, (c) LASAT. All panels use identical axis limits and the same colour scale so that the spatial extent can be compared directly. Panel titles report the maximum normalized frequency for each ADM (approx. RIMPUFF 70.2%, DIPCOT 71.9%, LASAT 100.0%). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

shows the strongest agreement: MB and FB are close to zero, RMSE remains low, and the spatial pattern correlation is consistently high ($r \geq 0.988$). This indicates that RIMPUFF and DIPCOT produce very similar evacuation exceedance-frequency fields, with only small differences in both mean frequency and spatial structure.

Pairs involving LASAT show systematically larger discrepancies. For all sites, MB and FB are negative for the ordered pairs RIMPUFF–LASAT and DIPCOT–LASAT, indicating that LASAT produces higher exceedance frequencies on average over the pair-specific comparison domain. This LASAT-related bias is strongest at Borssele, where FB is approximately −1.31 and RMSE is about 0.09 for both LASAT pairs. Cattenom shows a similar but slightly weaker LASAT-related bias, whereas Leibstadt exhibits the smallest LASAT-related normalized bias (FB ≈ −0.44) and the lowest RMSE among the LASAT comparisons.

The pattern correlations further distinguish magnitude differences from spatial structure differences. Although LASAT pairs have larger MB, FB, and RMSE, their correlations remain positive, indicating that the broad spatial patterns are not completely unrelated. However, the correlations are clearly lower for LASAT pairs than for RIMPUFF–DIPCOT, especially at Borssele ($r = 0.773$–0.790), showing that LASAT



(a) (b) (c)

Fig. 3. Evacuation frequency at Borssele (72 h prognosis) with a 10% threshold applied. Panels: (a) RIMPUFF, (b) DIPCOT, (c) LASAT. Only cells where the normalized evacuation frequency exceeds 10% are shown. The panels use identical axis limits and the same colour scale as Fig. 2. The typical radial extent is about 30 km; elliptical patterns reflect local wind directions. Background graphics (axes, rings) as in Fig. 2.

differs not only in mean exceedance frequency but also in spatial distribution. The larger comparison domains n_{ab} for LASAT pairs, particularly at Borssele and Cattenom, further indicate that LASAT contributes additional non-zero exceedance cells. Overall, the metrics confirm that RIMPUFF and DIPCOT are mutually consistent, whereas LASAT is the main contributor to inter-model variability.

Across sites, the difference plots and scalar metrics confirm that model choice can substantially affect both the frequency and spatial extent of proposed actions. Pairwise comparisons are therefore useful for identifying where ADM choice matters most. However, they do not provide a systematic way to combine information from all three ADMs in operational planning. This motivates the multi-model overlay analysis in the next subsection, where agreement and divergence among all ADMs are assessed within a single spatial framework.

4.3. Multi-model results

Overlay maps condense per-run model concurrence information using the RGB scheme described in Section 3.3.2. We emphasize that the per-model RGB overlays are intended as planning-stage diagnostics rather than as operational products for first responders. In an actual emergency, responder-facing maps would normally be based on an aggregated TMR-style decision (OR, majority or AND rule), potentially combined with additional decision-support criteria, while the RGB overlays remain available to planners and analysts to document model agreement, identify single-model outliers and prioritize locations for further deterministic investigation.

Applying the three logical rules yields the expected ordering of planned footprints: OR (≥ 1 ADM) is most extensive, majority (≥ 2 ADMs) reduces the footprint substantially, and AND (= 3 ADMs) produces the most compact areas.

The OR rule yields the largest planned action areas and is conservative with respect to potential under-protection but demanding in terms of resources, and it is most prone to including potential false-positive activations in the sense of isolated, single-ADM signals. The majority and AND rules result in smaller planned areas and may be more consistent with limited resources but risk missing rare scenarios where only one ADM signals an action. These logical combinations are applied both to the unthresholded overlays and, in conjunction with thresholds on normalized frequencies, to derive alternative sets of multi-model-based action areas.

Within the majority-rule maps, the three possible two-model combinations — RIMPUFF + LASAT, RIMPUFF + DIPCOT and DIPCOT+ LASAT — carry different implications. RIMPUFF + LASAT activations typically reflect agreement between a puff model and a terrain-resolving Lagrangian model, whereas DIPCOT + LASAT activations represent internal Lagrangian consensus. RIMPUFF + DIPCOT activations, by contrast, signal agreement between models with more similar turbulence representations. Comparing the spatial distribution of these three combinations therefore helps to isolate which physical and numerical choices in the ADMs drive robust features and where LASAT's more complex treatment of turbulence and terrain is decisive.

At all sites the OR footprint is broad (Fig. 7), whereas majority and AND criteria reduce the spatial extent markedly. Due to limited space in each cell, only a single RunID is displayed within a fully bordered cell when there is exactly one activation. When more than one ADM triggers the action, a RunID is written inside the cell only if all participating ADMs activate the action in the same run. Single-ADM isolated activations, especially at large distances, are identifiable in overlays and are natural candidates for deterministic follow-up as potential false-positive cases (Section 5).

The three two-model combinations should also be interpreted separately in the planning context: R+L highlights the strongest contrast between a puff-style and a fully Lagrangian treatment, R+D isolates the effect of replacing the Gaussian puff core by a hybrid particle reconstruction, and D+L is the most useful pair for identifying differences between two Lagrangian formulations with different terrain and turbulence handling. This is the right granularity for planners because it shows not only whether the models disagree, but also which physical representation is driving the disagreement.

From a decision-maker perspective, the RGB overlay is therefore best read as a diagnostic rather than as a direct instruction to evacuate. It shows where the footprint is robust across models, where a majority is sufficient, and where a single-model signal should be inspected more carefully. In that sense, the overlay is a planning tool for ranking confidence, not a crisis-response automation.

For Cattenom and Leibstadt most majority-rule evacuation cells correspond to three-model agreement (dark grey), with two-model concentrated in peripheral sectors (Fig. 8).

The AND footprint at Leibstadt is notably anisotropic with a pronounced east–west orientation, contrasting with the more circular patterns observed at Borssele and Cattenom — an effect consistent with the local topography and valley-controlled wind regimes (Calpini et al., 2011; Meteoblue, 2024; Topographic-map, 2024) (Fig. 9).

From a planning perspective, focusing on AND-based footprints and using majority logic selectively in peripheral sectors may offer a practical compromise between robustness and operational feasibility. However, this strategy risks omitting rare, single-model signals;

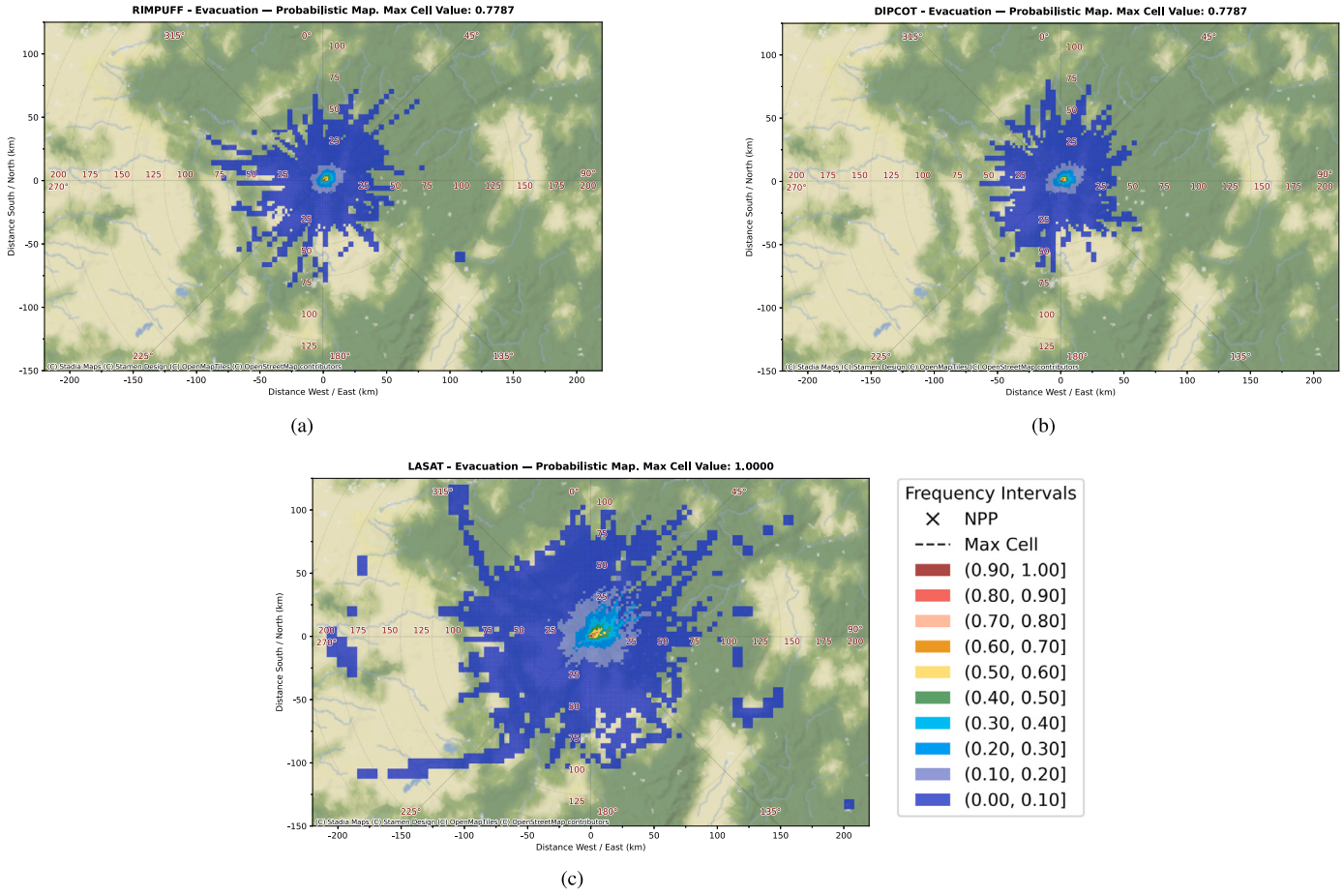


Fig. 4. Normalized evacuation frequency at Cattenom (72 h prognosis). Panels: (a) RIMPUFF, (b) DIPCOT, (c) LASAT. Conventions as in Fig. 2. Note increased asymmetry relative to Borssele, reflecting terrain and local wind climatology.

Table 4
Scalar inter-model comparison metrics for evacuation exceedance-frequency fields. Metrics are computed over the pair-specific union domain S_{ab}, where at least one of the two ADMs has a non-zero exceedance frequency. MB and FB are signed with respect to the ordered pair (m_a, m_b); positive values indicate higher exceedance frequencies for m_a.

Site	Pair (m_a, m_b)	n_{ab}	MB	RMSE	FB	r
Borssele	RIMPUFF–DIPCOT	1426	−0.0004	0.0112	−0.0104	0.9884
Borssele	RIMPUFF–LASAT	3079	−0.0612	0.0924	−1.3155	0.7730
Borssele	DIPCOT–LASAT	3066	−0.0613	0.0912	−1.3096	0.7896
Cattenom	RIMPUFF–DIPCOT	2127	−0.0042	0.0133	−0.0785	0.9907
Cattenom	RIMPUFF–LASAT	3114	−0.0574	0.0843	−0.8986	0.8934
Cattenom	DIPCOT–LASAT	3106	−0.0547	0.0798	−0.8349	0.9137
Leibstadt	RIMPUFF–DIPCOT	2324	−0.0010	0.0202	−0.0131	0.9897
Leibstadt	RIMPUFF–LASAT	2788	−0.0366	0.0698	−0.4392	0.9464
Leibstadt	DIPCOT–LASAT	2634	−0.0379	0.0678	−0.4267	0.9480

combining logical rules with frequency thresholds (Section 4.1.1) provides a straightforward mechanism to balance coverage and resource constraints.

To focus on higher-confidence regions, the overlay logic can be combined with frequency thresholds. For multi-model overlays we adopt a slightly stricter 15% per-ADM cutoff, because in the overlay each grid cell can be supported by up to three independent ADMs and a tighter per-model condition keeps the overall false-positive rate comparable to the 10% single-model case. Both thresholds are user-configurable in the post-processing module. For example, multi-model evacuation overlays with a 15% threshold retain only cells with normalized evacuation frequency ≥ 15% in at least one ADM. This removes most low-frequency single-ADM cells and yields more compact areas dominated by LASAT and multi-model agreement (Fig. 10), particularly around the core sectors at each site. In combination, the AND rule highlights high-confidence core areas, majority logic captures broader consensus regions, and thresholding ensures that rare, low-probability activations including potential false positives, do not dominate the footprint.

5. Deterministic case study

The overlay data provide an explicit link from statistical diagnostics to individual deterministic simulations: for any selected cell the overlay file lists the RunIDs of statistical members that produced the activation pattern of interest. These RunIDs can be used to re-run JRODOS deterministically with the full emergency chain for the corresponding date/time and thus produce time-resolved dose and meteorological

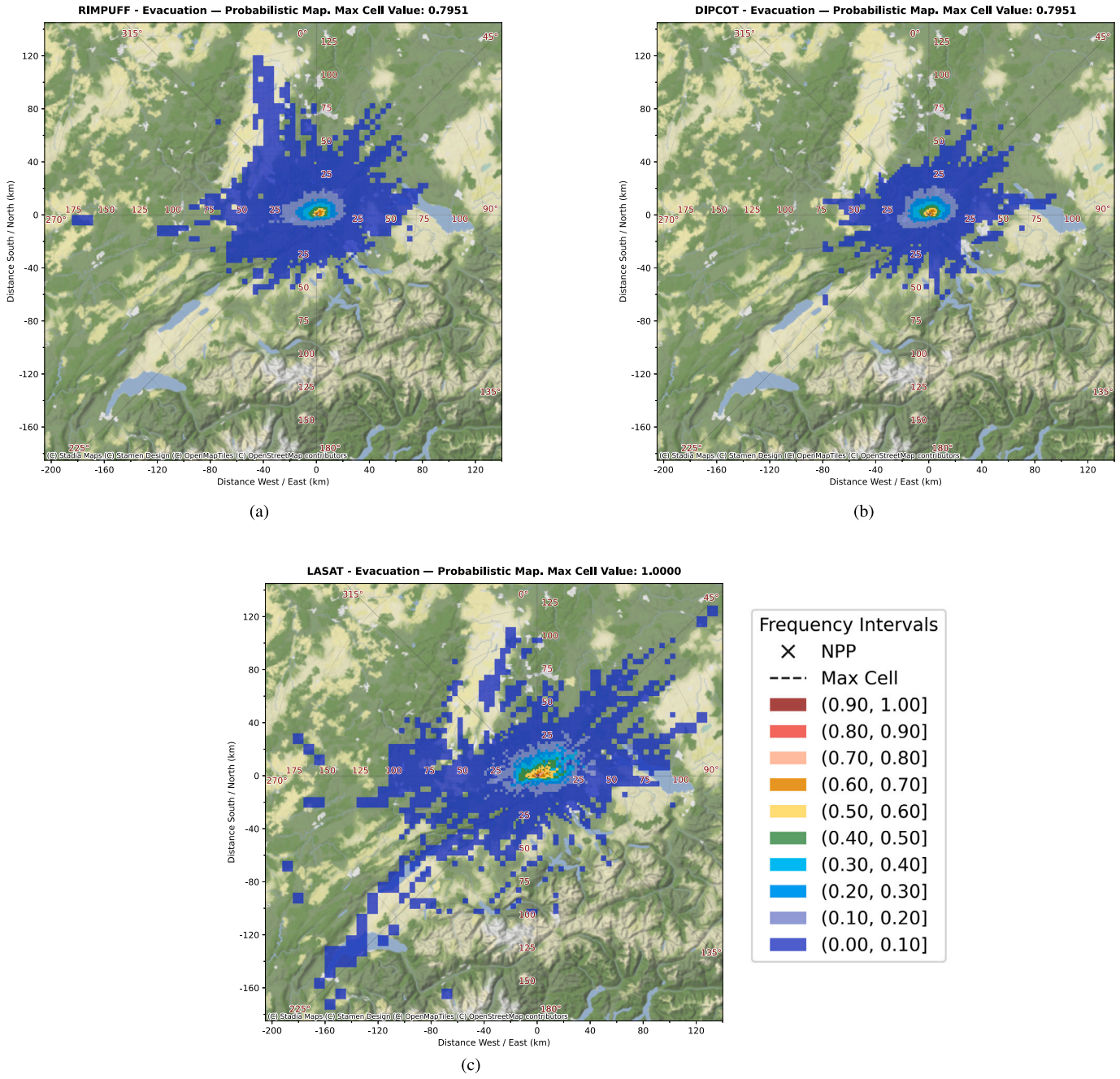


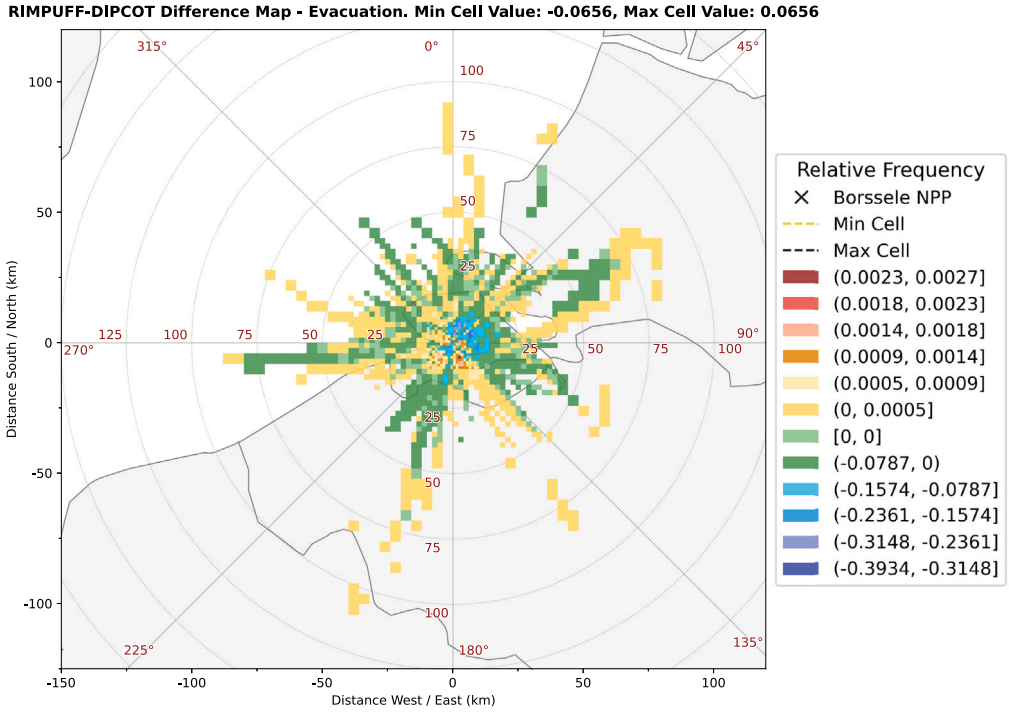
Fig. 5. Normalized evacuation frequency at Leibstadt (72 h prognosis). Panels: (a) RIMPUFF, (b) DIPCOT, (c) LASAT. Conventions as in Fig. 2. LASAT displays valley-aligned, high-detail structures; Gaussian/hybrid models yield smoother, more diffuse fields. The strongly anisotropic patterns reflect the influence of complex terrain and prevailing wind directions.

fields. Such targeted deterministic runs allow process-level interpretation (wind turns, precipitation scavenging, plume residence time) and help to assess whether isolated or low-confidence statistical signals are physically plausible or are model-specific artefacts.

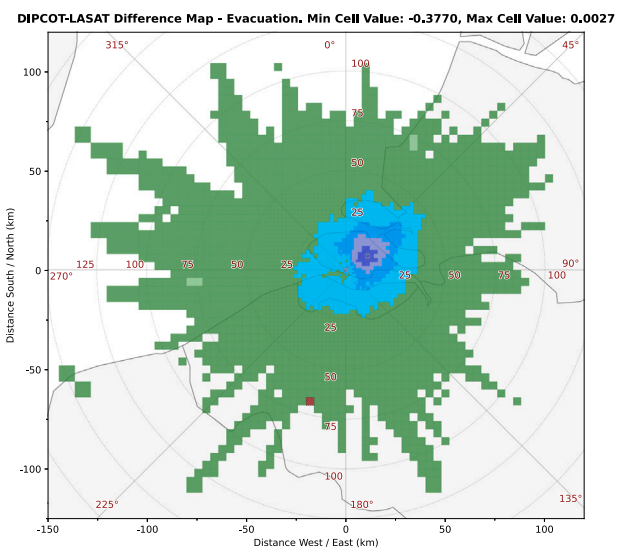
We illustrate the workflow with the Borssele example Fig. 11. The RGB overlay flagged an isolated evacuation activation at approximately 200 km from the release point occurring only for RIMPUFF in one statistical member (RunID 310), while LASAT and DIPCOT did not activate evacuation at that cell. In Fig. 11 (a) the target cell (JRODOS cell 3522) is the upper of the two highlighted brown squares; the lower square marks its immediate southern neighbour and is shown for context only. Fig. 12 (a) displays the potential effective dose at the end of the 24 h prognosis with the JRODOS default eight-bin logarithmic colour bar spanning 1–1000 mSv (class boundaries at 1, 3, 10, 30, 100, 300 and 1000 mSv). For cell 3522 the potential effective dose is 96.1 mSv, which falls into the 30–100 mSv class, whereas all eight neighbouring cells remain below 1 mSv and are plotted in the lowest class. The evacuation recommendation is therefore a sharply isolated single-cell feature rather than part of a coherent high-dose plume. In the corresponding early-intervention dose field (Fig. 12 (b)) the same cell records 103 mSv, just above the 100 mSv evacuation threshold, while neighbouring cells remain far below threshold.

Key deterministic results for cell 3522 (RunID 310, RIMPUFF) Fig. 12:

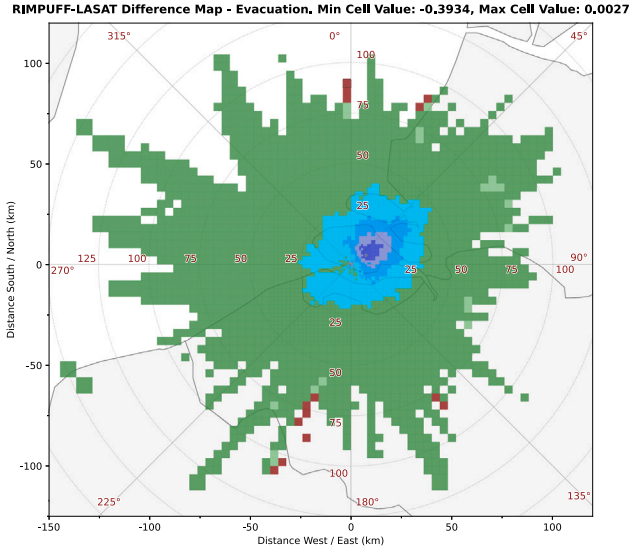
- Potential effective dose at 24 h: 96.1 mSv (just below the 100 mSv evacuation threshold). Fig. 12(a)



(a)



(b)



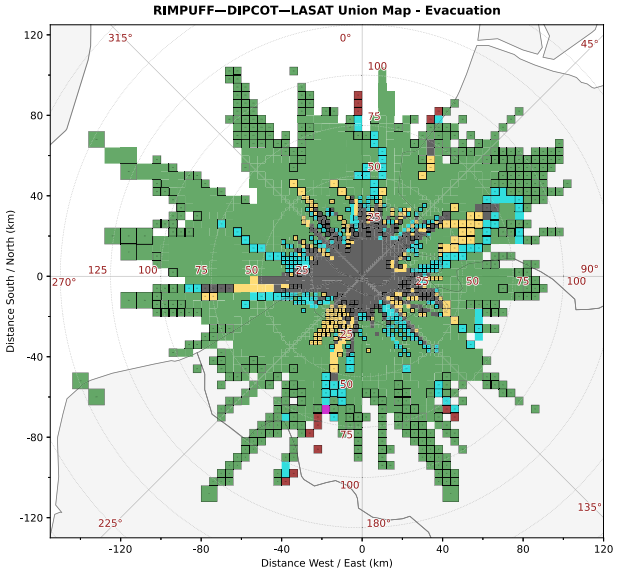
(c)

Fig. 6. Pairwise difference maps (Borssele, 72 h prognosis). Panels: (a) RIMPUFF − DIPCOT, (b) DIPCOT − LASAT, (c) RIMPUFF − LASAT. Diverging colour scale: red = positive (first model higher), blue = negative (second model higher), light green ≈ 0 (identical frequencies). Colour scale and panel titles indicate global extrema (min/max differences). Large negative values in (b) and (c) indicate regions where LASAT predicts the action substantially more often than the other ADMs. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

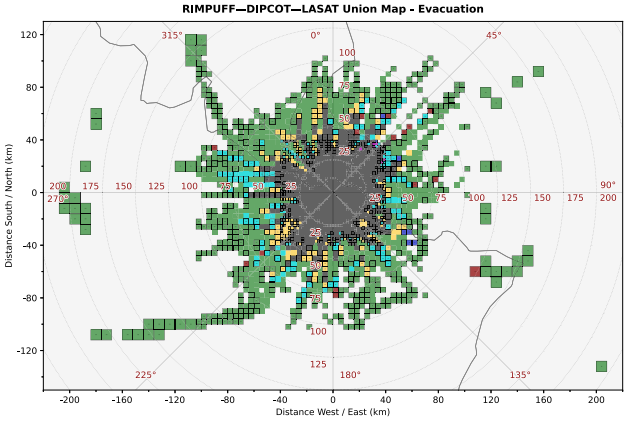
- Early (intervention) dose: 103 mSv (slightly above the 100 mSv evacuation threshold) Fig. 12(b)
- Meteorological diagnostics show a nocturnal wind turn that shifts the plume back northwards and increases accumulation over the cell Fig. 12(c).

For the same date and cell, deterministic runs with LASAT and DIPCOT produced substantially lower doses (end-of-period potential effective doses approximately 1.23 mSv and 0.211 mSv, respectively; intervention doses approximately 2.56 mSv and 0.225 mSv, respectively), and neither model recommended evacuation. Hence the evacuation signal at 200 km is specific to RIMPUFF for this run and represents a marginal exceedance in that model only. From an emergency-planning perspective, such an isolated, single-model activation can be regarded as a potential false positive in the operational sense that an evacuation would be planned based on one ADM, while the other ADMs and the broader ensemble do not support it.

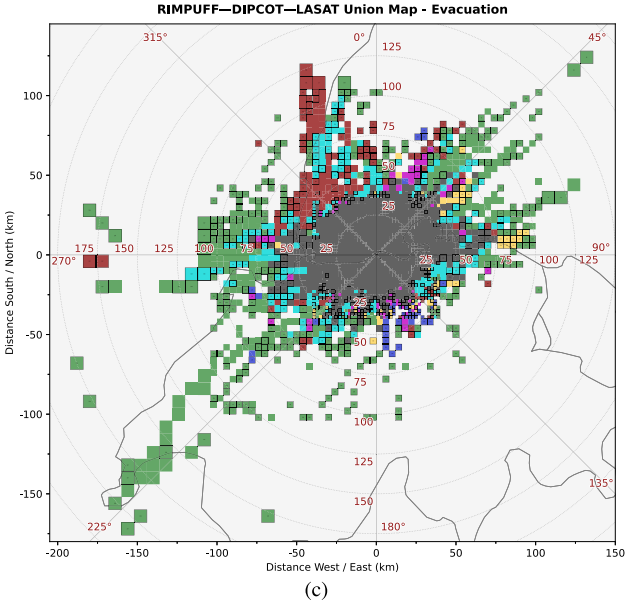
This case demonstrates the value of the deterministic follow-up workflow: overlay diagnostics identify rare, single-model activations (candidates for deeper inspection), and targeted deterministic runs reveal whether those activations arise from plausible meteorological



(a)



(b)



(c)

Fig. 7. OR-rule multi-model evacuation overlay (72 h prognosis). Panels: (a) Borssele, (b) Cattenom, (c) Leibstadt. Colours encode ADM combinations (red = RIMPUFF only; green = LASAT only; blue = DIPCOT only; yellow/magenta/cyan = two-model combinations; dark grey = all three ADMs; white = none). The OR rule includes any cell activated by at least one ADM, producing broad footprints. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

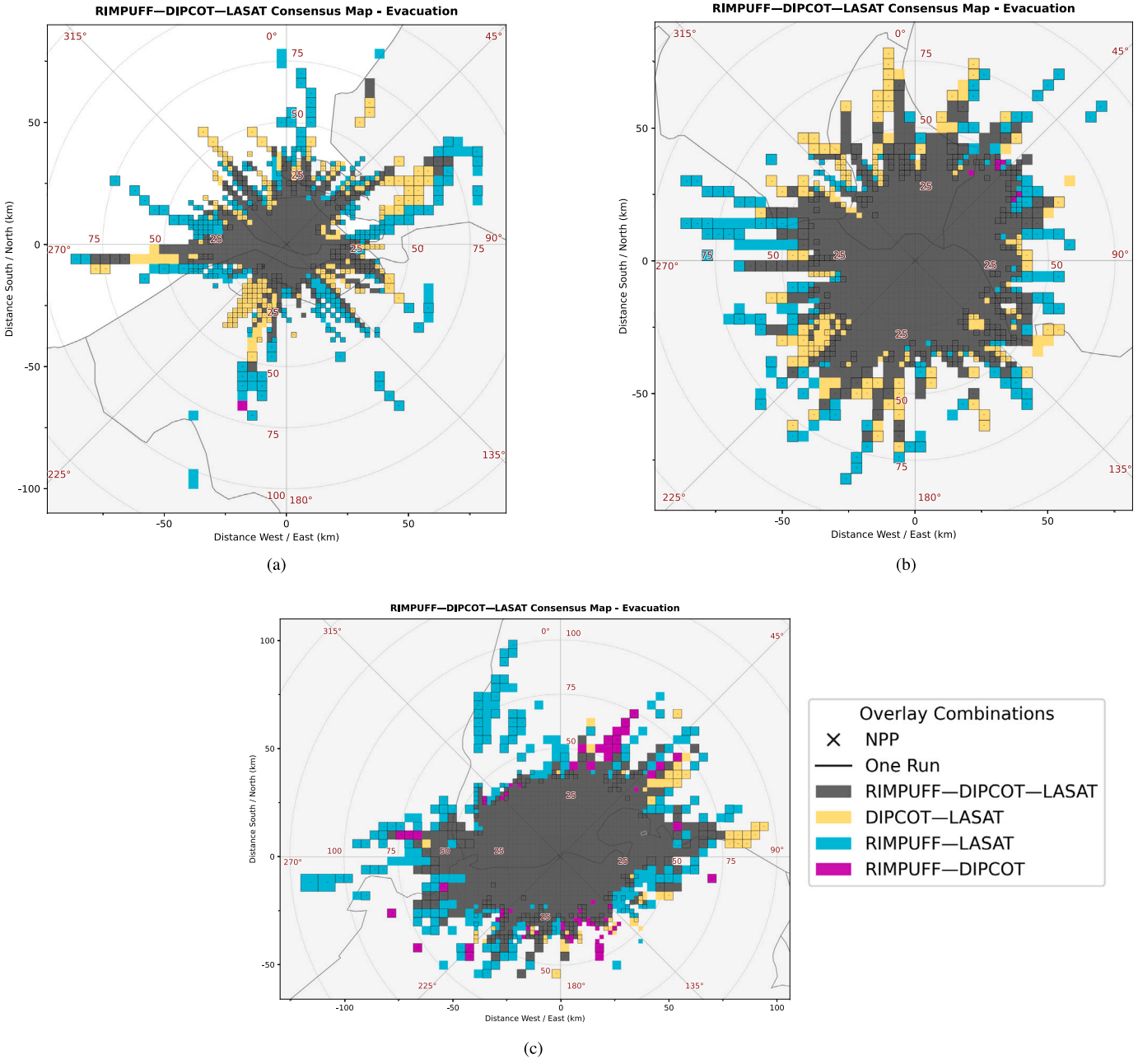
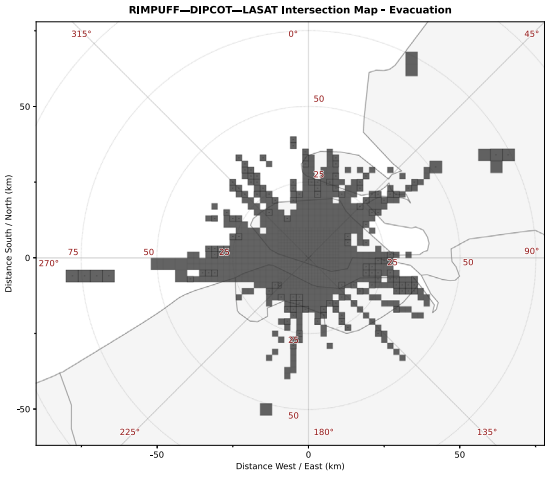


Fig. 8. Majority-rule (at least 2 ADMs) evacuation overlays (72 h prognosis). Panels: (a) Borssele, (b) Cattenom, (c) Leibstadt. Colours as in Table 3. Majority voting substantially reduces the footprint relative to OR; dark grey indicates robust three-model agreement and dominates central regions, while two-model and single-model activations are typical in peripheral sectors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

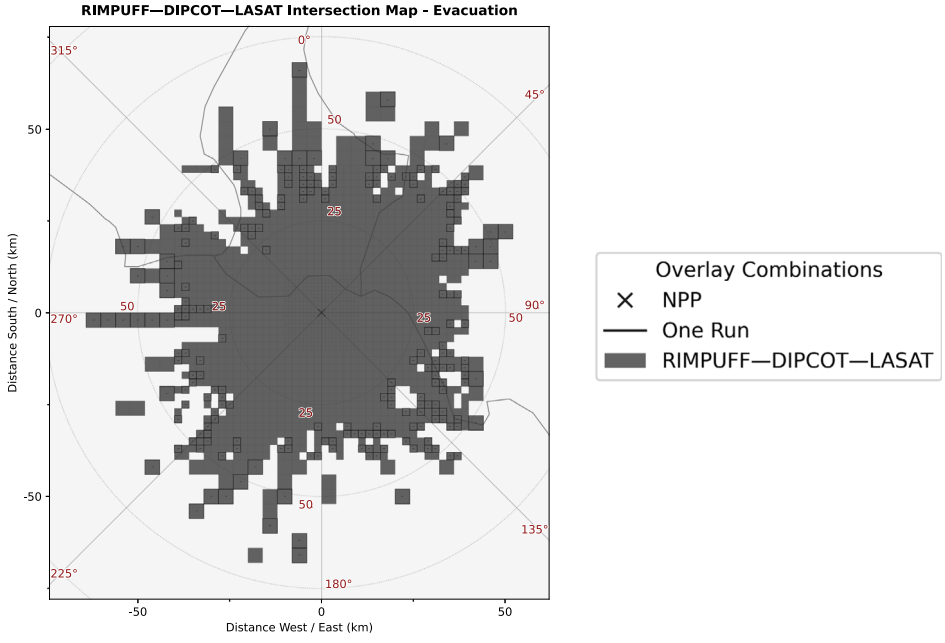
processes (here, a wind turn) or are more likely artefacts. In practice, such deterministic triage can help planners prioritize further analysis, identify potential false-positive evacuation signals, and avoid unnecessary measures triggered by marginal, model-specific exceedances. As a result, flagging cells by overlay pattern and then performing a small set of deterministic follow-ups is an efficient and interpretable procedure to evaluate isolated activations in large collections of statistical runs.

6. Discussion

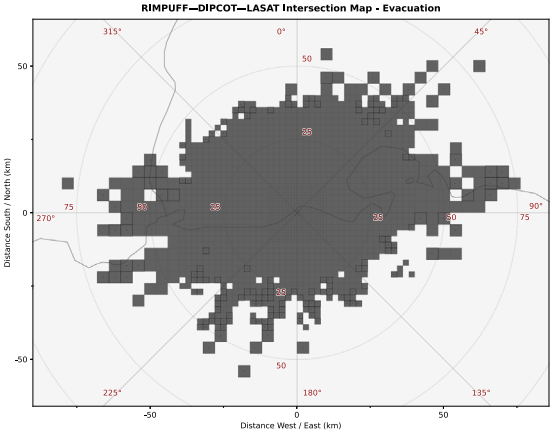
The results show that the choice of ADM within the JRODOS LSMC has a substantial influence on JRODOS-proposed action areas, even for flat terrain and more so for complex terrain. LASAT generally produces higher frequencies and more compact high-probability areas near the source, whereas RIMPUFF and DIPCOT yield broader patterns with lower maxima. Although DIPCOT and LASAT are both Lagrangian-particle implementations, three methodological choices explain most of the divergence between them. First, LASAT advects particles with an explicit stochastic turbulence scheme that resolves the velocity covariance matrix near the source, while DIPCOT relies on similarity-theory-derived turbulence statistics, leading to comparatively broader and lower near-source concentrations in DIPCOT. Second, the kernel used to reconstruct concentrations from particles differs (a Gaussian kernel in DIPCOT versus a finer-grain occupancy estimator in LASAT), which affects the spatial sharpness of high-probability areas. Third, LASAT incorporates terrain-following advection and obstacle handling



(a)



(b)



(c)

Fig. 9. AND-rule (=3 ADMs) evacuation overlays (72 h prognosis). Panels: (a) Borssele, (b) Cattenom, (c) Leibstadt. AND areas are compact and indicate cells where all three ADMs concur. The Leibstadt AND footprint shows a pronounced east–west orientation, reflecting topographically controlled flow.

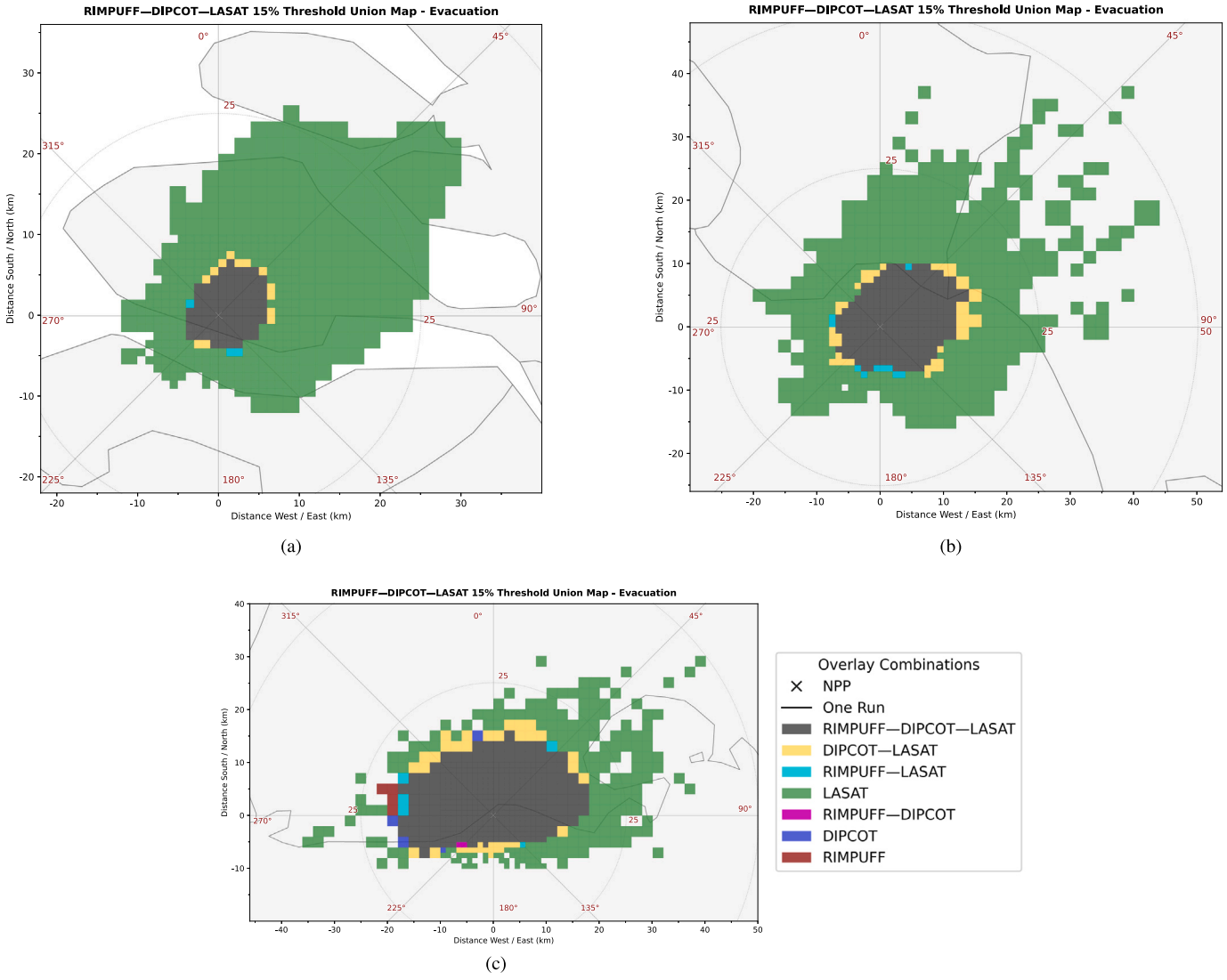


Fig. 10. Multi-model evacuation overlays with a 15% frequency threshold (72 h prognosis). Panels: (a) Borssele, (b) Cattenom, (c) Leibstadt. Only cells with normalized evacuation frequency ≥ 15% in at least one ADM are shown. The threshold removes low-frequency single-ADM cells and yields more compact, higher-confidence action areas, particularly where LASAT and multiple ADMs contribute.

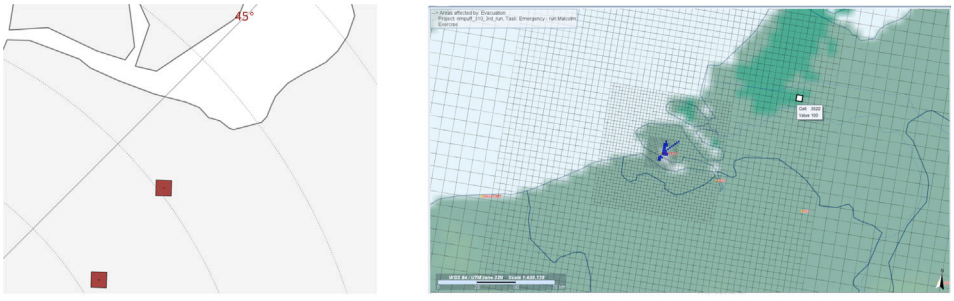


Fig. 11. Borssele — deterministic JRODOS evacuation map for RIMPUFF, RunID 310 (24 h prognosis, 08:34 5 Nov to 08:34 6 Nov 2024). Panel (a) shows the location of cell 3522, approximately 200 km from the release point; panel (b) shows the corresponding evacuated area. Cell 3522 is the upper highlighted cell marked as evacuation by RIMPUFF.

more directly than DIPCOT, which becomes the dominant effect at Cattenom and Leibstadt. RIMPUFF differs from both because puff dispersion is governed by stability-class-based spread and by a coarser representation of terrain influence, producing smoother but more extended footprints. These differences are consistent with previous work emphasizing model and input uncertainties in atmospheric dispersion modelling (Tomasi et al., 2019; Ulimoen et al., 2022; Korsakissok et al., 2020; Sørensen et al., 2020; Giovannini et al., 2020) and underline the need to account for model diversity when using JRODOS for emergency planning.

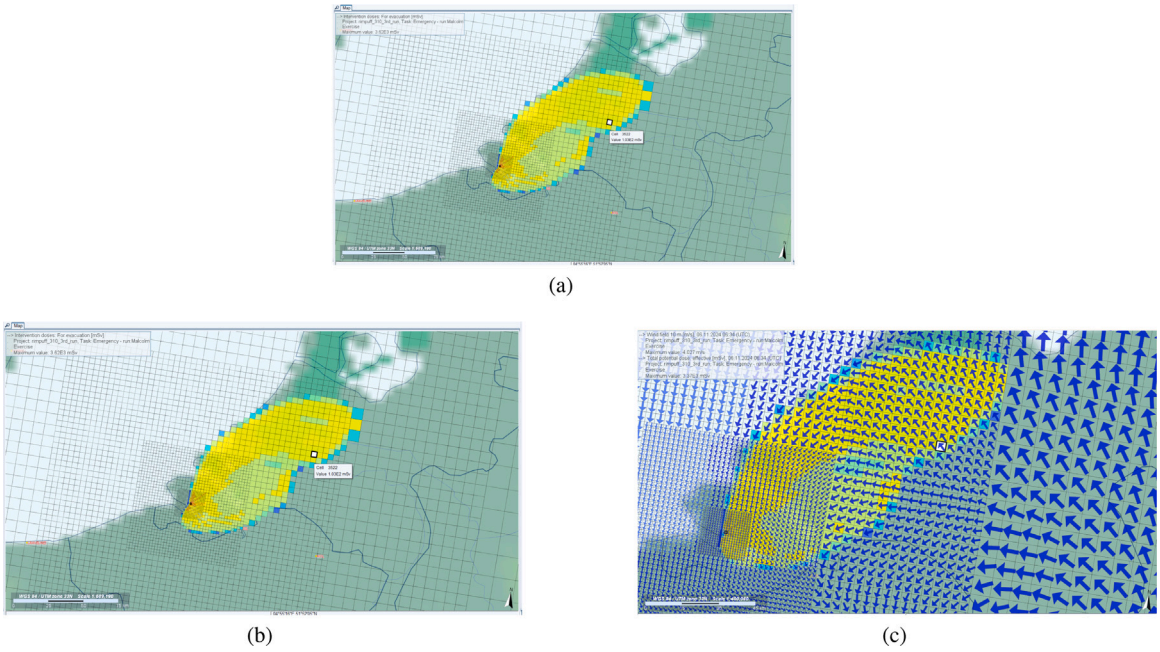


Fig. 12. Borssele RunID 310 showing cell 3522 for RIMPUFF — (a) potential effective dose (end of 24 h prognosis), cell records 96.1 mSv, slightly below the 100 mSv evacuation threshold, (b) intervention early-dose field, cell records 103 mSv, slightly above threshold, and (c) wind vectors at approximately 05:04 indicating the wind turn responsible for returning plume mass over the cell.

The RGB overlay approach and associated data provide a compact view of model agreement and disagreement, distinguishing robust (three-model) from partially robust (two-model) and single-model action regions. In combination with a simplified TMR-style logic, this enables transparent comparison of OR-, majority- and AND-based planning strategies. OR-based areas are conservative but extensive; AND-based areas are compact but may miss rare cases; majority-based areas offer an intermediate option that may be attractive where resources are limited. This is precisely why the overlays are useful for planners: they expose confidence structure, not just binary alert/no-alert output.

Frequency and overlay plots with thresholds link model predictions to operational constraints by excluding very low-frequency activations and highlighting site-specific sector patterns instead of circular zones. Fixed thresholds and simple logical rules cannot capture all real-world considerations (such as population, infrastructure and societal acceptability), but they provide a structured starting point that can be embedded in broader decision-support processes.

The deterministic case study illustrates how isolated, single-ADM activations flagged by the overlays can be examined through targeted JRODOS runs. In the example considered, a marginal exceedance of the evacuation threshold in one model at large distance is not reproduced by the other ADMs. Such analyses can help assess the plausibility of isolated recommendations, identify potential false positives in the ensemble, and, where appropriate, avoid unnecessary actions.

Opportunities for methodological extension include broader source term definitions, multi-year statistics to assess seasonal robustness, explicit seasonal/diurnal/precipitation stratification of the annual ensemble, and integration of additional dispersion models (e.g. FLEXPART (Stohl et al., 2005)) and meteorological data providers to expand the model space. A current limitation of the present study is the absence of a direct comparison with measured deposition or air-concentration data: the F6.DWR-FKA scenario is a generic worst-case source term for which no observational counterpart exists. A natural extension is therefore to repeat the workflow with a well-documented accident, such as the Fukushima Daiichi release, and to compare the resulting probabilistic maps with publicly available airborne and deposition monitoring datasets. This would allow a quantitative skill assessment of each ADM in a realistic setting and would provide an external benchmark for the multi-model planning approach proposed here. Future work could also investigate the integration of the multi-model indicators developed here into probabilistic and multi-criteria decision-analytic frameworks for radiological protection, explicitly incorporating factors such as exposure, population distribution and logistical constraints. In addition, the per-RunID structure of the overlay files allows the same statistical members to be re-aggregated by season, by time of day and by precipitation regime without re-running the dispersion simulations; producing seasonal- and diurnal-resolved frequency maps and quantifying the role of wet deposition in the action-area extent are immediate next steps that we plan to address with the existing 2024 ensemble.

7. Conclusion

This study investigated how three atmospheric dispersion models integrated in the JRODOS LSMC — RIMPUFF, DIPCOT and LASAT — influence JRODOS-proposed evacuation areas for a severe accident scenario at three NPP sites with contrasting terrain. Annual ensembles of simulations were generated with the JRODOS statistical module and post-processed using frequency analysis, spatial differences and multi-model overlays.

The results confirm that ADM choice can substantially alter both the extent and orientation of proposed evacuation areas, particularly in complex terrain. A multi-model overlay framework, combined with TMR-inspired OR, majority and AND rules and user-defined frequency thresholds, was shown to provide site-specific action areas that reflect local meteorology and model diversity, and to make explicit the trade offs between conservative and resource-efficient planning strategies. A deterministic case study further demonstrated how rare, single-model signals identified in the overlays can be examined using targeted JRODOS runs to assess their plausibility.

Taken together, the statistical and deterministic tools presented here constitute a planning-oriented extension of JRODOS: they combine the existing statistical module with a dedicated post-processing workflow to support the design and review of feasible, locally adapted emergency plans that explicitly account for meteorological variability and model diversity.

CRediT authorship contribution statement

Ramy-Badr Ahmed: Formal analysis, Writing – original draft, Conceptualization, Methodology, Software, Visualization, Data curation, Writing – review & editing . **Thomas Schichtel:** Supervision, Conceptualization, Writing – review & editing. **Dmytro Trybushnyi:** Supervision, Conceptualization, Software, Visualization, Writing – review & editing. **Wolfgang Raskob:** Supervision, Validation, Investigation. **Sadeeb S. Ottenburger:** Supervision, Investigation, Resources, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Research Link Provided

JRODOS Multi-Model Atmospheric Dispersion Pipeline for Nuclear Emergency Planning (Original data) (Mendeley Data)

References

Ahmed, R.-B., 2026. JRODOS Multi-Model Atmospheric Dispersion Pipeline for Nuclear Emergency Planning. http://dx.doi.org/10.17632/kw76hfx5rd.3, Mendeley Data, V3, https://data.mendeley.com/datasets/kw76hfx5rd/3.

Andronopoulos, S., Davakis, E., Bartzis, J., Kovalets, I., 2010. RODOS meteorological pre-processor and atmospheric dispersion model DIPCOT: a model suite for radionuclides dispersion in complex terrain. Radioprotection 45 (5), S77–S84.

Asselt, E.D., Twenhöfel, C.J., Duranova, T., Smetsers, R.C., Bohunova, J., Müller, T., 2020. Facilitating the decision-making process after a nuclear accident: Case studies in the Netherlands and Slovakia. Integr. Environ. Assess. Manag. 17 (2), 376–387.

Calpini, B., Ruffieux, D., Bettems, J.-M., Hug, C., Huguenin, P., Isaak, H.-P., Kaufmann, P., Maier, O., Steiner, P., 2011. Ground-based remote sensing profiling and numerical weather prediction model to manage nuclear power plants meteorological surveillance in Switzerland. Atmos. Meas. Tech. 4 (8), 1617–1625.

Davakis, E., Andronopoulos, S., Vlachogiannis, D., Venetsanos, A., Bartzis, J., Nychas, S., 2001. Validation of the demokritos dispersion modelling system based on the Indianapolis experiment. Int. J. Environ. Pollut. 16 (1–6), 88–100.

Davakis, E., Nychas, S., Andronopoulos, S., Bartzis, J., 2003. Validation study of the dispersion Lagrangian particle model DIPCOT over complex topographies using different concentration calculation methods. Int. J. Environ. Pollut. 20 (1–6), 33–46.

Dyer, L.L., Astrup, P., 2012. Model evaluation of RIMPUFF within complex terrain using an ^{41}Ar radiological dataset. Int. J. Environ. Pollut. 48 (1–4), 145–155.

Ehrhardt, J., Shershakov, V., Zheleznyak, M., Mikhalevich, A., 1996. RODOS: decision support system for off-site emergency management in Europe. Nucl. Eng. Des. 145, 1087–1096.

EU-Council, 2014. Laying down basic safety standards for protection against the dangers arising from exposure to ionising radiation. Off. J. Eur. Union 2013/59/EURATOM (COUNCIL DIRECTIVE 2013/59/EURATOM).

Giovannini, L., Ferrero, E., Karl, T., Rotach, M.W., Staquet, C., Trini Castelli, S., Zardi, D., 2020. Atmospheric pollutant dispersion over complex terrain: Challenges and needs for improving air quality measurements and modeling. Atmosphere 11 (6), 646.

Hernández-Ceballos, M.A., Sangiorgi, M., García-Puerta, B., Montero, M., Trueba, C., 2020. Dispersion and ground deposition of radioactive material according to airflow patterns for enhancing the preparedness to N/R emergencies. J. Environ. Radioact. 216, 106178.

IAEA, 2016. Preparedness and response for a nuclear or radiological emergency. http://dx.doi.org/10.61092/IAEA.3DBE-055P.

IAEA, 2017. Fundamental safety principles. http://dx.doi.org/10.61092/IAEA.HMXN-VW0A.

IAEA, 2026. International nuclear and radiological event scale (INES). https://www.iaea.org/resources/databases/international-nuclear-and-radiological-event-scale.

Ievdin, I., Trybushny, D., Zheleznyak, M., Raskob, W., 2010. RODOS re-engineering: aims and implementation details. Radioprotection 45 (5), S181–S189.

Institute for Thermal Energy Technology and Safety (ITES/RESIS), 2019. RODOS User Guide, Version 4.0. Technical Report, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany.

Janicke, L., 1994. The embedding of the Lagrangian dispersion model LASAT into a monitoring system for nuclear power plants. In: Gryning, S.-E., Millán, M.M. (Eds.), Air Pollution Modeling and Its Application X. Springer US, Boston, MA, pp. 405–411.

Korsakissok, I., Bedwell, P., Berge, E., Hamburger, T., 2019. Guidelines for the Use of Ensemble Calculations in an Operational Context, Indicators to Assess the Quality of Uncertainty Modelling and Ensemble Calculations, and Tools for Ensemble Calculation in Emergency Response. Technical Report, CONFIDENCE Project, https://www.meenas.eu/wp-content/uploads/2023/05/D9.5_Guidelines-for-the-use-of-ensemble-calculations_28112019.pdf.

Korsakissok, I., Périllat, R., Andronopoulos, S., Bedwell, P., Berge, E., Charnock, T., Geertsema, G., Gering, F., Hamburger, T., Klein, H., Leadbetter, S., Lind, O., Pázmándi, T., Rudas, C., Salbu, B., Sogachev, A., Syed, N., Tomas, J., Ulimoen, M., de Vries, H., Wellings, J., 2020. Uncertainty propagation in atmospheric dispersion models for radiological emergencies in the pre- and early release phase: summary of case studies. Radioprotection 55, S57–S68.

Landman, C., Päsler-Sauer, J., Raskob, W., 2014. Introduction. In: The Risks of Nuclear Energy Technology: Safety Concepts of Light Water Reactors. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 311–313.

Leung, W.H., Ma, W.M., Chan, P.K., 2018. Nuclear accident consequence assessment in Hong Kong using JRODOS. J. Environ. Radioact. 183, 27–36.

Leutbecher, M., Palmer, T.N., 2008. Ensemble forecasting. J. Comput. Phys. 227 (7), 3515–3539.

Lyons, R.E., Vanderkulk, W., 1962. The use of triple-modular redundancy to improve computer reliability. IBM J. Res. Dev. 6 (2), 200–209.

Meteoblue, 2024. Meteoblue wetterdaten basel. https://www.meteoblue.com/de/wetter/archive/export.

Meutter, P.D., Delcloo, A.W., 2022. Uncertainty quantification of atmospheric transport and dispersion modelling using ensembles for CTBT verification applications. J. Environ. Radioact. 250, 106918.

Mikkelsen, T., Larsen, S.E., Thykier-Nielsen, S., 1984. Description of the risø puff diffusion model. Nucl. Technol. 67 (1), 56–65.

NOMAD, 2024. NOMADS at ncep.noaa.gov. https://nomads.ncep.noaa.gov/.

Päsler-Sauer, J., 2010. Comparison and validation exercises of the three atmospheric dispersion models in RODOS. Radioprotection 45 (5), S89–S96.

Raskob, W., Dechy, N., Donovan, A., Gallego, E., Nanba, K., Romão, X., Tanzi, C.P., Wiens, M., Ceballos, M.A.H., 2020. Fukushima Daiichi Accident in 2011. Technical Report, Publications Office of the European Union, http://dx.doi.org/10.2760/571085.

Raskob, W., Landman, C., Trybushnyi, D., 2017. JRODOS: Real-time online decision support system for nuclear emergency management. https://www.eu-alara.net/images/stories/pdf/program17/Session3/14%20jrodos-2017-alara.pdf.

Salpadimos, N., Karfopoulos, K., Potiriadis, C., Andronopoulos, S., Seimenis, I., 2024. Atmospheric dispersion software intercomparison exercise and sensitivity of results. HNPS Adv. Nucl. Phys. 30, 124–130.

Sørensen, J.H., Bartnicki, J., Buhr, A.M.B., Feddersen, H., Hoe, S.C., Israelson, C., Klein, H., Lauritzen, B., Lindgren, J., Schönfeldt, F., Sigg, R., 2020. Uncertainties in atmospheric dispersion modelling during nuclear accidents. J. Environ. Radioact. 222, 106356.

Steinhauser, G., Brandl, A., Johnson, T.E., 2014. Comparison of the chernobyl and fukushima nuclear accidents: A review of the environmental impacts. Sci. Total Environ. 470–471, 800–817.

Stohl, A., Forster, C., Frank, A., Seibert, P., Wotawa, G., 2005. Technical note: The Lagrangian particle dispersion model flexpart version 6.2. Atmos. Chem. Phys. 5 (9), 2461–2474.

Tomasi, E., Giovannini, L., Falocchi, M., Antonacci, G., Jiménez, P.A., Kosovic, B., Alessandrini, S., Zardi, D., Monache, L.D., Ferrero, E., 2019. Turbulence parameterizations for dispersion in sub-kilometer horizontally non-homogeneous flows. Atmos. Res. 228, 122–136.

Topographic-map, 2024. Leibstadt topographic map, elevation, terrain. https://en-ca.topographic-map.com/map-lfx2f3/Leibstadt/?center=47.73994%2C8.30045&zoom=9.

Ulimoen, M., Berge, E., Klein, H., Salbu, B., Lind, O.C., 2022. Comparing model skills for deterministic versus ensemble dispersion modelling: The fukushima daiichi NPP accident as a case study. Sci. Total Environ. 806, 150128.

UNSCEAR, 2000. Sources and Effects of Ionizing Radiation. United Nations Scientific Committee on the Effects of Atomic Radiation. Technical Report, United Nations, https://www.unscear.org/unscear/uploads/documents/unscear-reports/UNSCEAR_2000_Report_Vol.I.pdf.

Vries, H.D., Geertsema, G., Korsakissok, I., Périllat, R., Scheele, R., Tomas, J., 2019. Published Sets of Probability Maps of Threshold Exceedance for Scenarios Provided to WP4, WP5 & WP6. Technical Report, CONFIDENCE Project.

Walter, H., Gering, F., Arnold, K., Gerich, B., Heinrich, G., Welte, U., 2016. RODOS-Based Simulation of Potential Accident Scenarios for Emergency Response Management in the Vicinity of Nuclear Power Plants. Technical Report, Bundesamt für Strahlenschutz, https://inis.iaea.org/records/s8dmm-vw463.