Validation of Stochastic Lagrangian Particle Dispersion Models Using Inter-Model Variability and Sensitivity Analysis: Case study on April 2025 sand-dust transport and hypothetical nuclear releases

Chow Yin Hei
R4 summer intern, The Hong Kong Observatory

Supervisors: Kwan Kam Lun SEO(R)41, Kong Yu Chau AgSSO(R)4

Agenda

- Background & motivation
- Methodology
- Case study 1: April 11 sand-dust storm over China
    - Similarity among model predictions
    - Validation against real measurements
- Case study 2: Hypothetical nuclear leak
- Conclusion

Background & motivation

- Flexpart is a Lagrangian particle dispersion model for simulating the turbulent transport of substance in the atmosphere over long distances. It performs importance sampling of the underlying distribution of dispersed species by introducing particles as mesoscale entities that follow the turbulent flow.
- Given initial condition of the release, Flexpart calculate particle trajectories according **input meteorological fields.** The air concentration of the released species is then calculated by interpolating the particles. Loss processes like dry deposition, cloud scavenging, and radioactive decay are modelled by adjusting the weight of each particle.
- Flexpart version 11 came out in October 2024. Substantial changes are made since version 9, in terms of the coordinate system used, below cloud scavenging schemes, interpolation scheme, and parallelization. Version 9 is currently used by HKO in preparing for nuclear emergencies, thus the **validation of newer versions is of interest for potential adoption**.

Methodology

Two cases are studied for version 9, 10 and 11

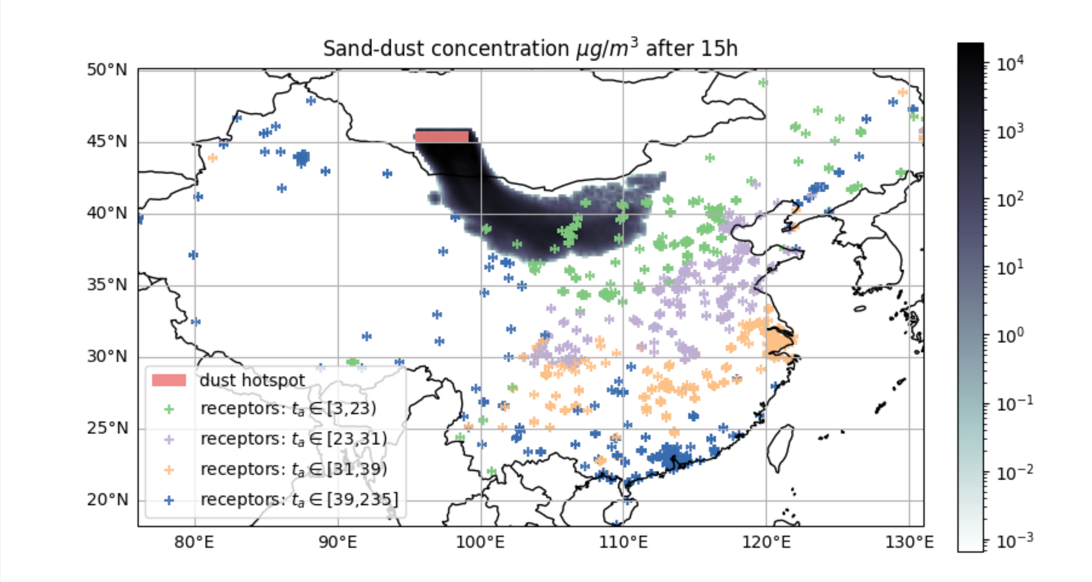
1. **(Real-world.)** April 11 2025 sand-dust storm event initializing at Gobi desert and affecting Northern and Southern China.

2. **(Hypothetical.)** A nuclear release scenario at Hanbit nuclear power plant May 11, 2025. The hypothetical dispersion of Cs-137 radioactive species is modelled to test consistency among the versions.

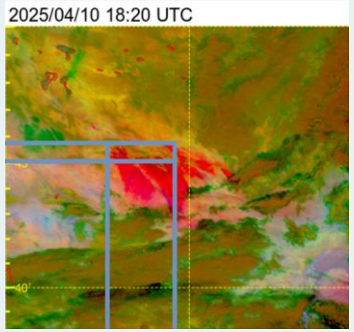
Case	Real-world	Hypothetical
Release time (UTC)	20250410 15:00	20250511 00:00
Release duration	14 h	3 h
Release location	Dust hotspot identified in Gobi desert	Hanbit nuclear power plant (point source)
Release height	0 to 600 m	0 to 50 m
Species	Sand-dust	Cs-137
Mass emitted	1.1×10^{11} kg	5 kg
Particle number	1M	0.1M
Meteorological forcing fields	GFS forecast data every 12 h	GFS forecast data at release time
Simulation time	10 days	10 days

Case 1. April 11 2025 sand-dust storm originating from Gobi desert

- **(Initial condition.)** The range of latitude, longitude and time of sand-dust production is determined by satellite images. A heuristic release profile consisting of a uniform initial distribution is used.
- **(Ground-truth.)** Model predictions are benchmarked against of openly available PM10 time-series data provided by China National Environmental Monitoring Centre (CNEMC).
- After data cleaning, hourly PM10 time-series at 1046 locations remains. Sampling locations cover the affected area.



(Animated gif in .pptx version – time-evolution of sand-dust transport)

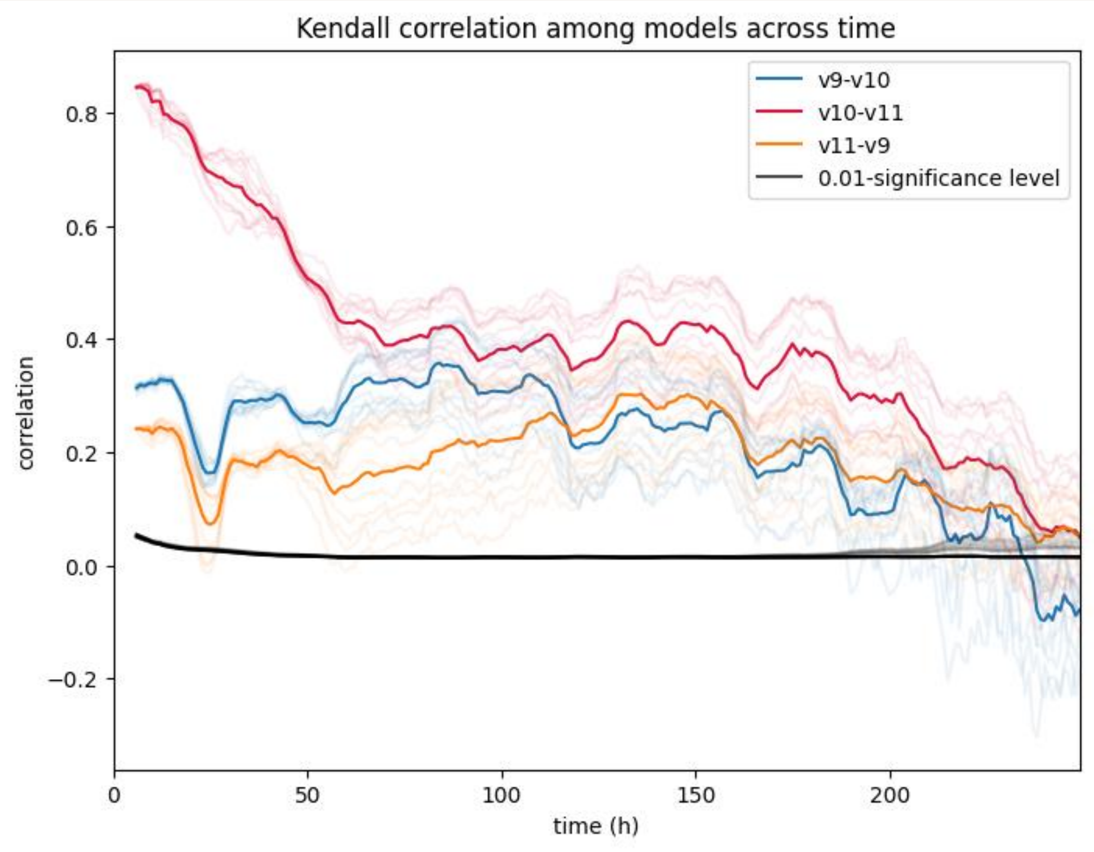


Comparison in absence of ground-truth

Correlation among model predictions

- Kendall correlation is measured pairwise for the three models. To avoid overestimation, the support of each measurement is the set of grid points on which at least one of the pair gives non-zero concentration

- Version 10 and 11 maintains the highest correlation across the 10-day simulation period. This aligns with the more significant changes in version 10 update.



Distance in fluctuation

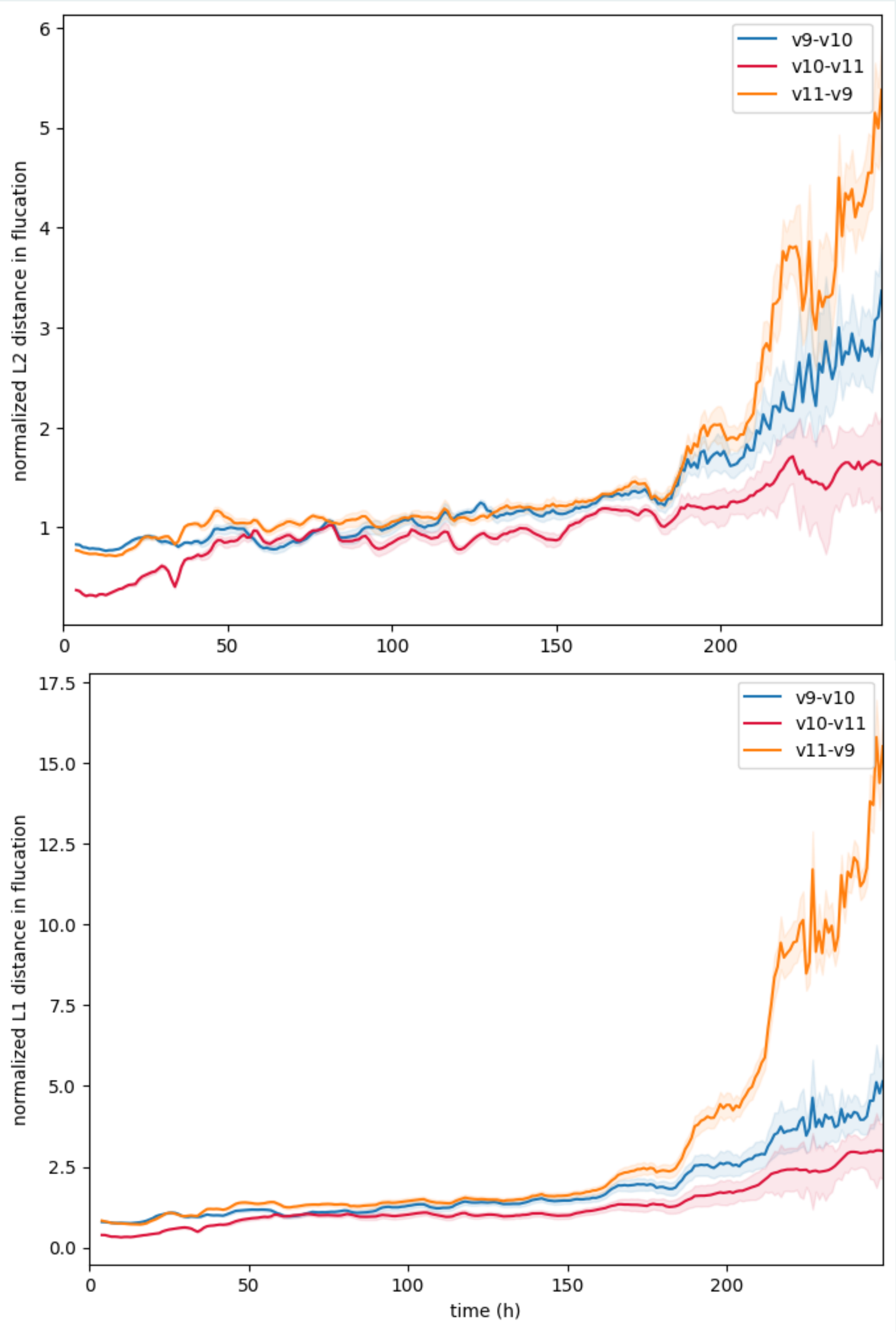
Noting that concentration is non-negative, we define for all processes x, y

$$\delta x := \frac{x}{\langle x \rangle_t}$$
$$d^p(x, y) := \frac{||x - y||_p}{\sqrt{||x||_p ||y||_p}}$$

as oppose to conventional fluctuations and normalized L_p distances. The quantity of interest is

$$d(\delta x, \delta y) \in [0, +\infty) \ ,$$

which can be interpreted as a geometrically normalized distance between concentration processes x, y output by Flexpart.



Validation of turbulent mixing over the model domain – well-mixed criterion

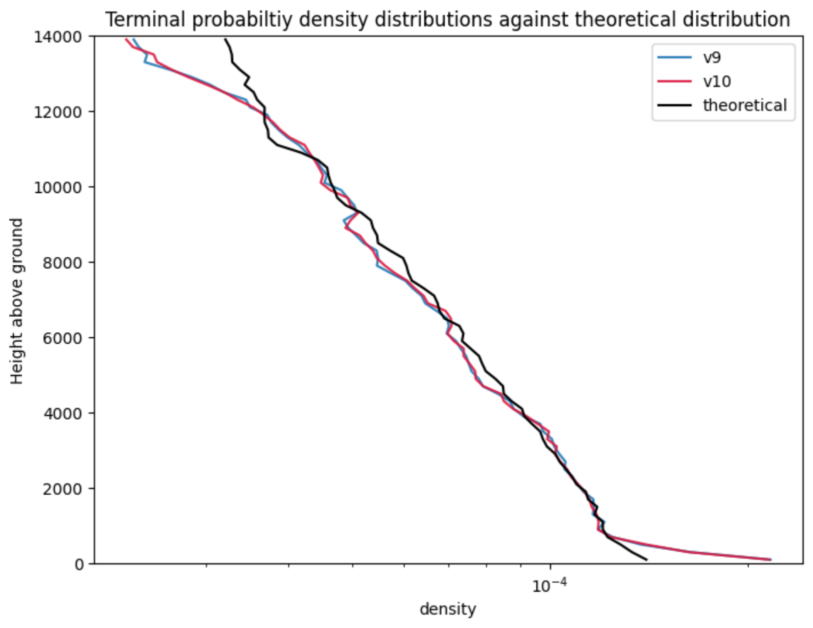
- In current operations at HKO, Flexpart uses a computationally cheaper Gaussian turbulence scheme, which doesn’t enforce well-mixed criterion a priori.

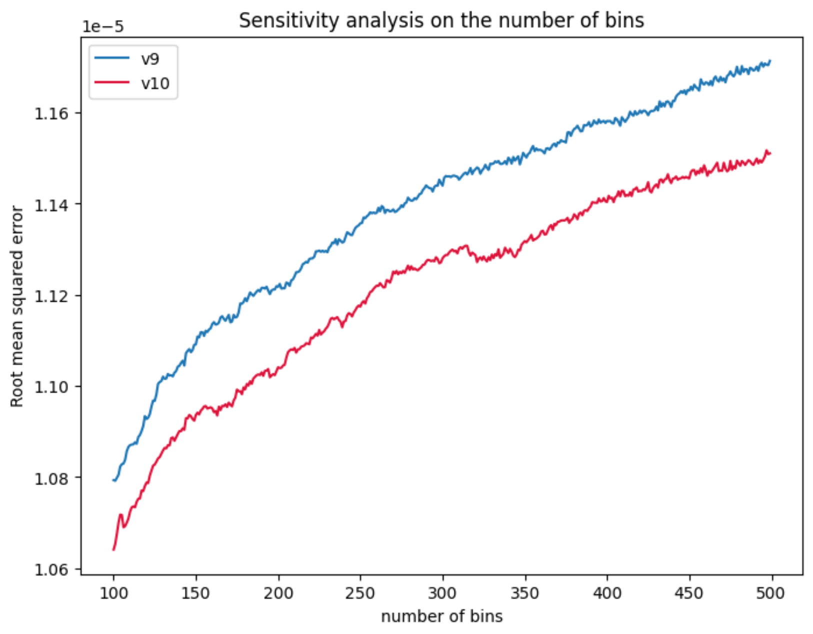
- This test is to validate whether the vertical turbulent transport generated is still physically realistic over the same model domain used for modelling the sand-dust storm event.

- Well-mixed criterion states that a species of **passive** particle uniformly distributed within a turbulent flow in both position and velocity will remain that way in the absence of **sources** and **sinks.** It presents a challenge to particle dispersion models in general.

- Air tracers are distributed according to the air density over the entire model domain (The GFS data used spans from ground to at least 15 km). The terminal distribution is then compared to the air density at that time.

- Future direction: Reanalysis data should be used

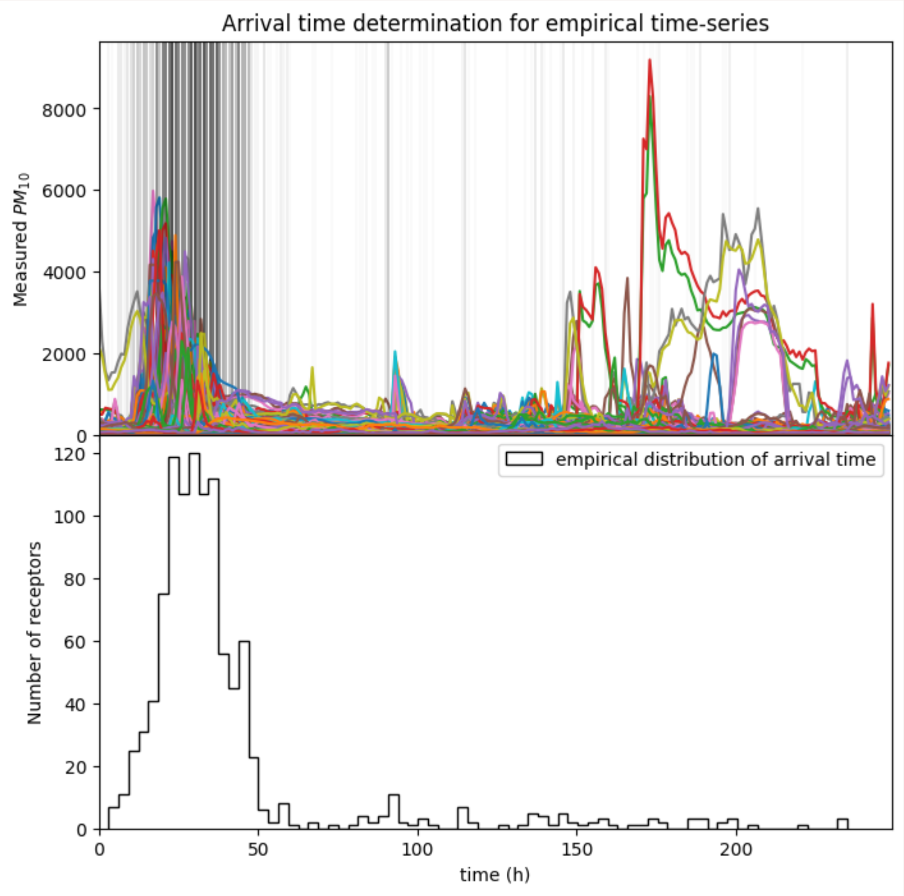


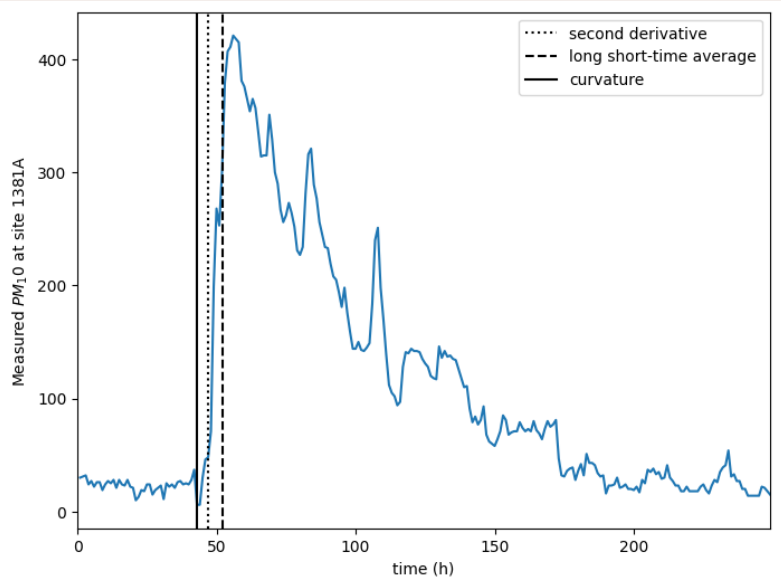


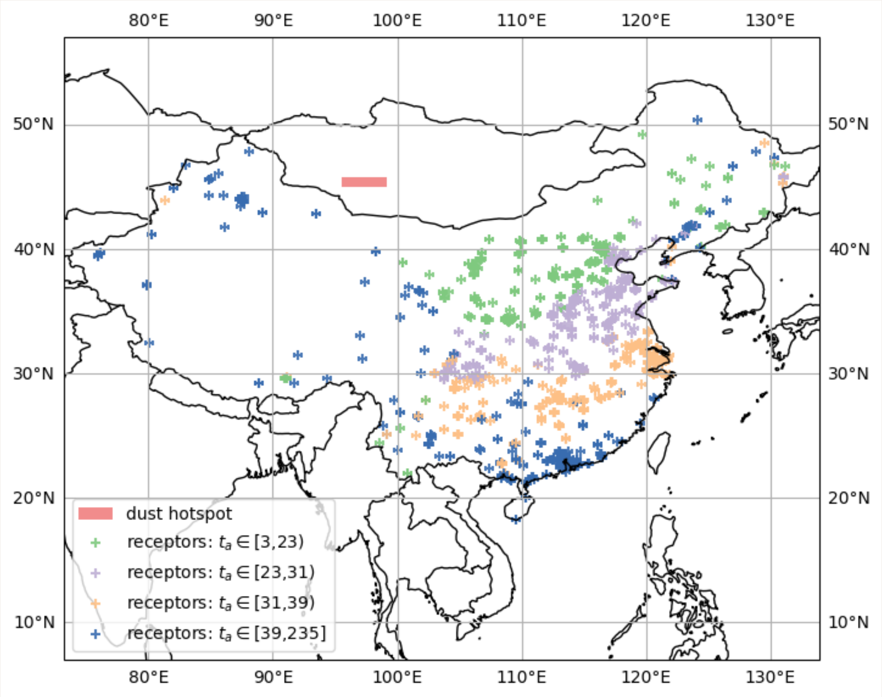
Performance metric: Arrival time

Distribution and visualization of arrival time testing set

- Three schemes for determining arrival time is considered, all based on a quantile q of the time-series': curvature, long short-time average, or second derivative

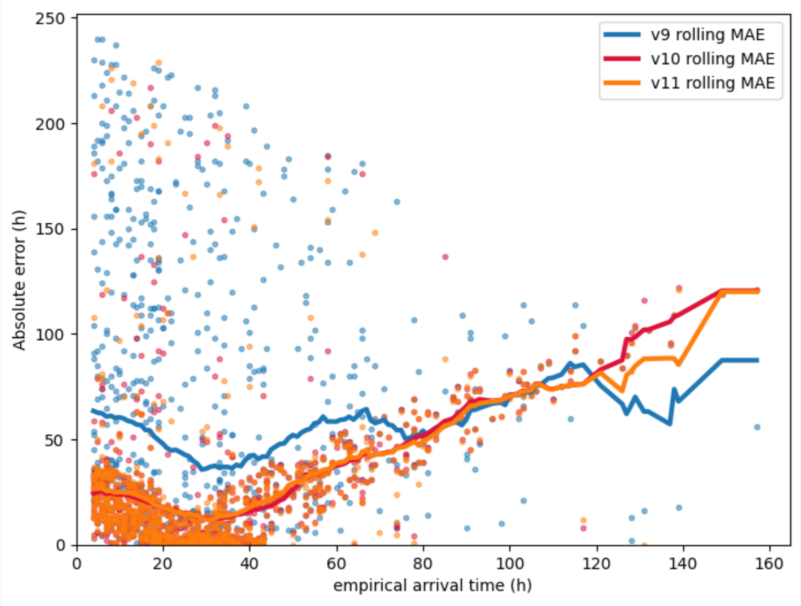


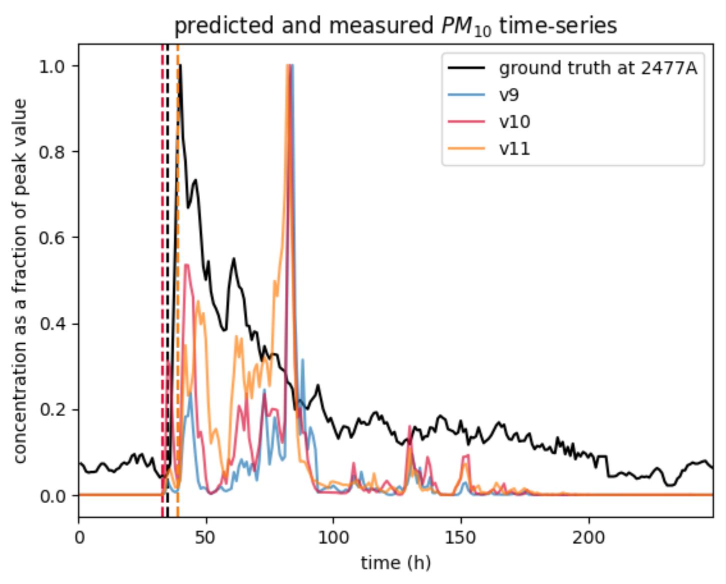




Determination of cutoff and the performance (MAE) sensitivity on it

- The performance metric (arrival time) is considered ill-defined if there're no obvious peak observed in the measured PM10 time-series
- Receptors are therefore ranked by the variance on the arrival time determined over different choices of model parameter q.
- Top 20 receptors are considered to have ill-defined arrival time and is not used in evaluating the model performance



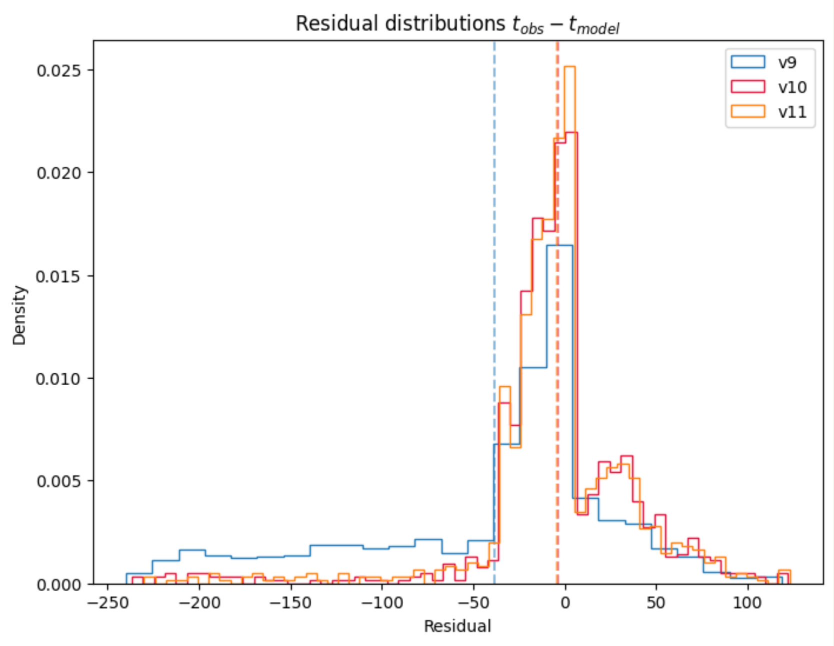


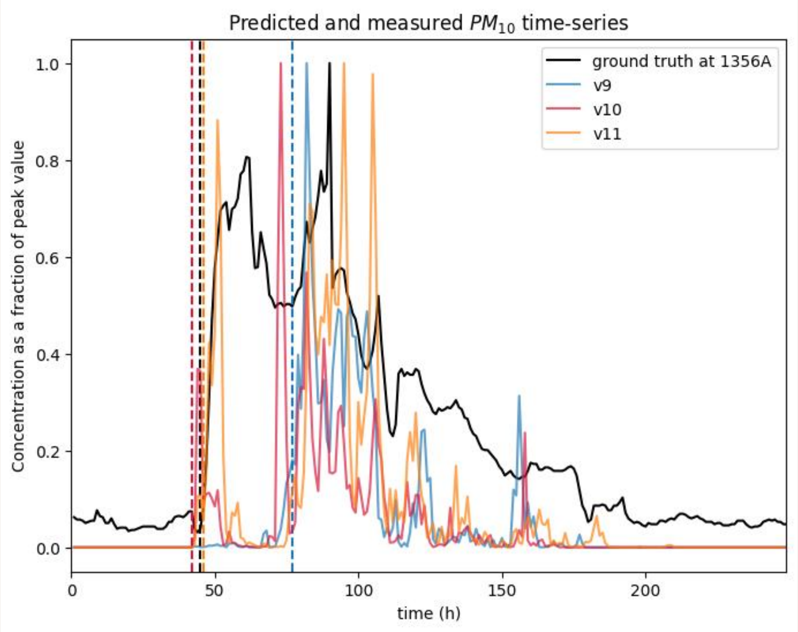
Marginal case at cutoff. Location: 111.6217, 26.2081

Performance among models

Model	n	MAE	RMSE	bias
Fp9	1026	51.40	77.75	35.44
Fp10	1026	26.11	42.20	4.12
Fp11	1026	26.06	42.62	4.35

Table 1: Performance in arrival time prediction. n is the support, or the total number of model-observation pairs used, obtained after removing 20 receptors. Units of MAE, RMSE and bias is in hour.

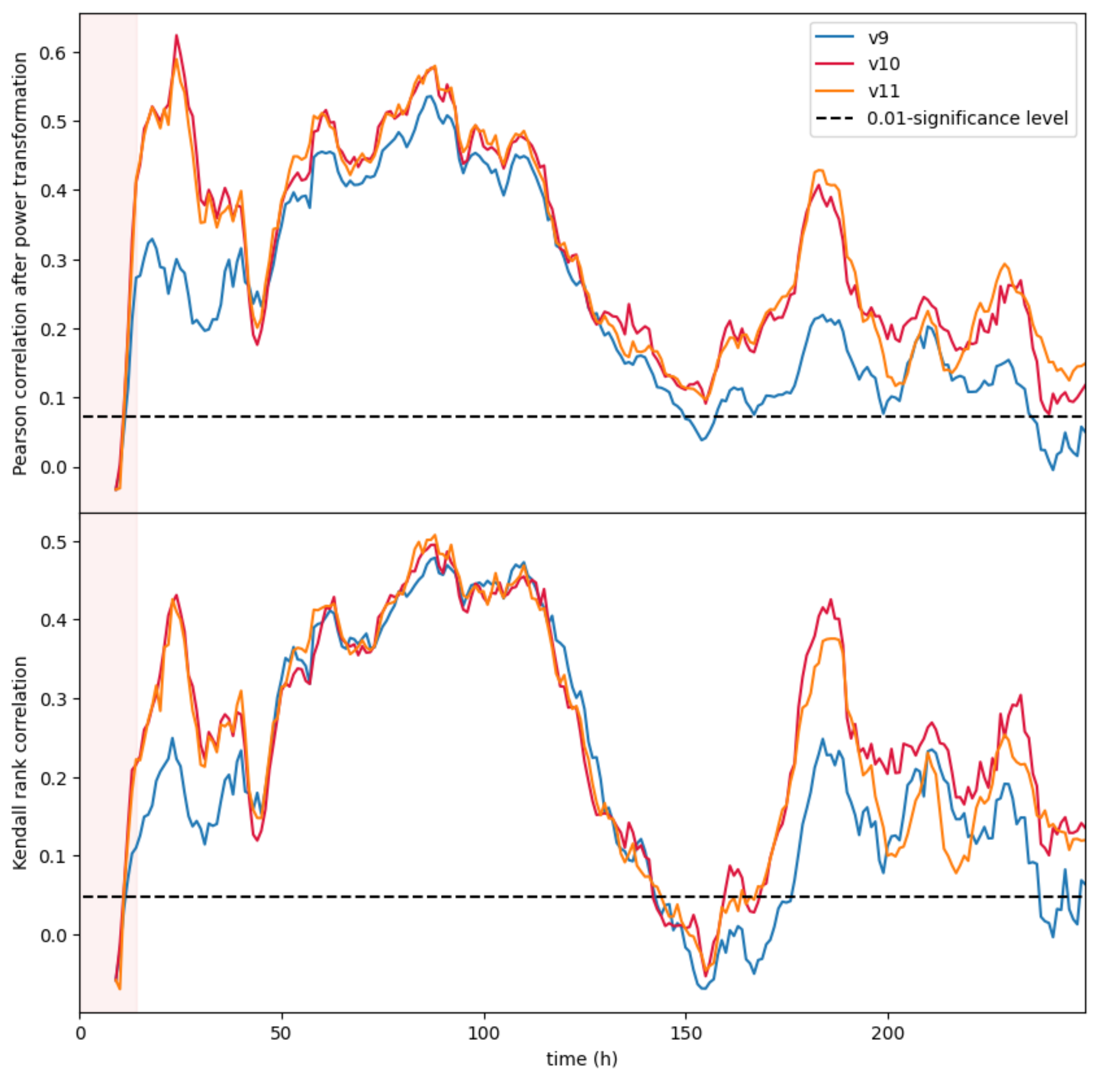




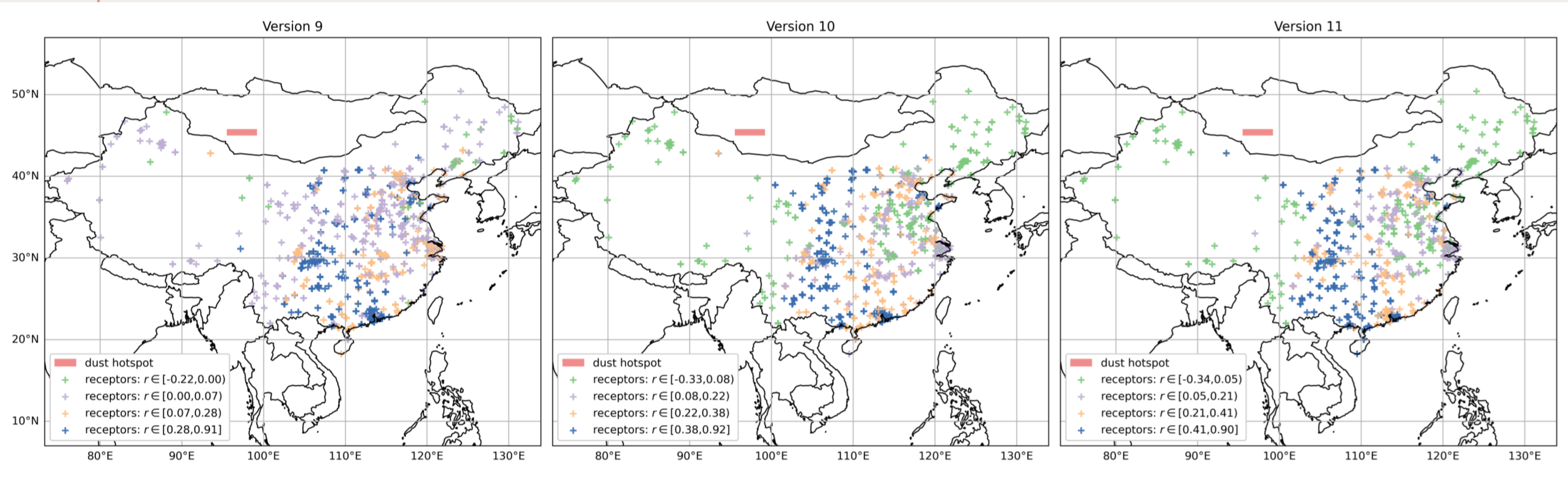
Hong Kong:
Arrival time for Hong Kong is 43-45 h after start of release

Performance metric: Correlations with ground-truth

Correlations with ground-truth across simulation time

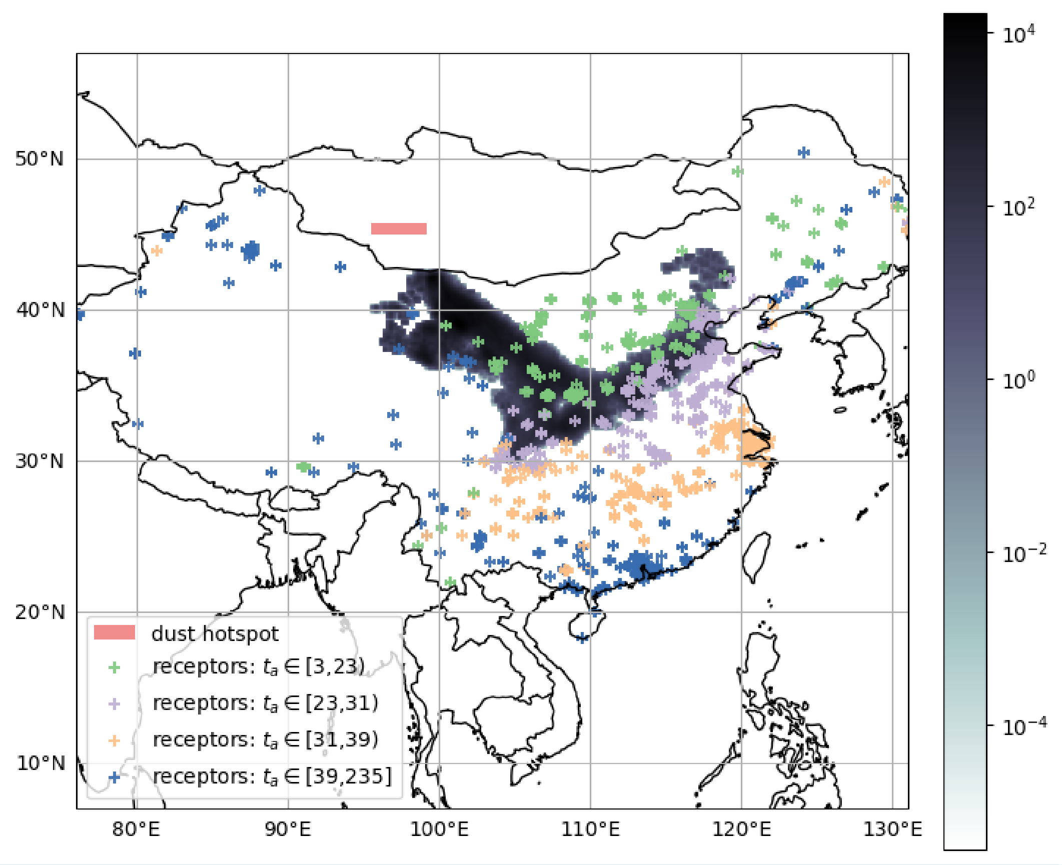


Location-dependence of Pearson correlations

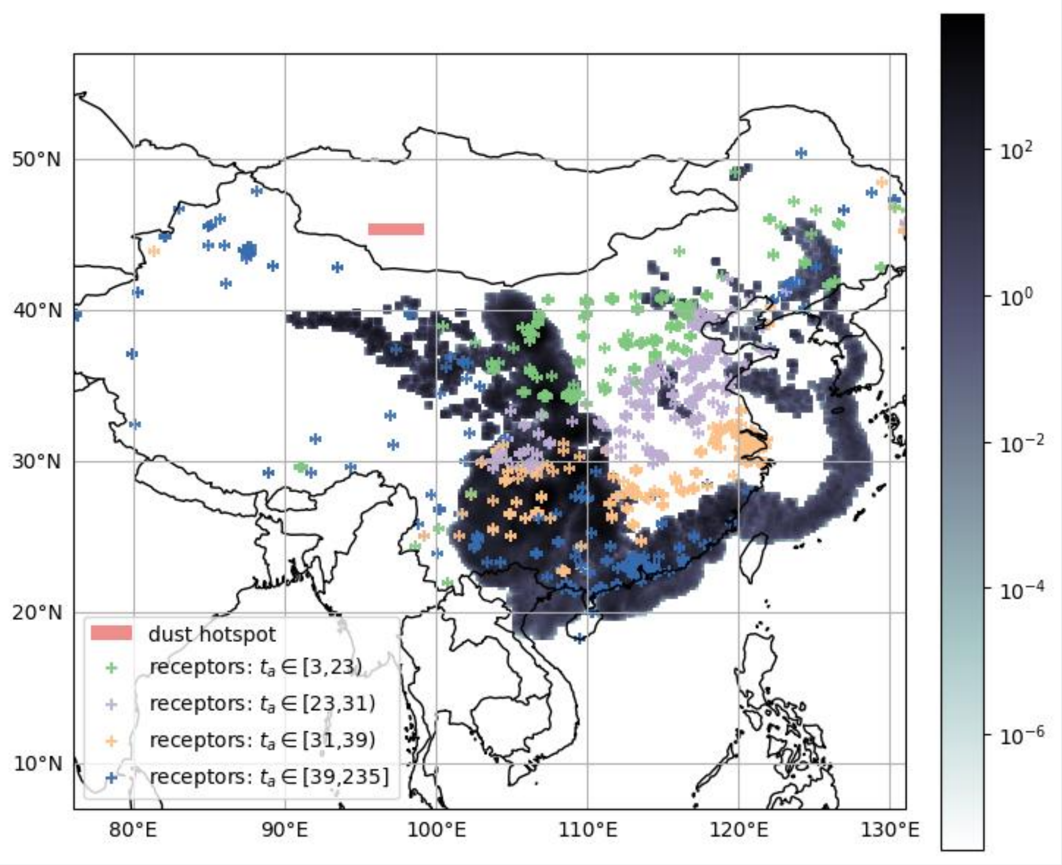


Mean: 0.16 0.24 0.24
Median: 0.07 0.22 0.21

(For reference)



24 h



48 h

Unit is in μg/m^3

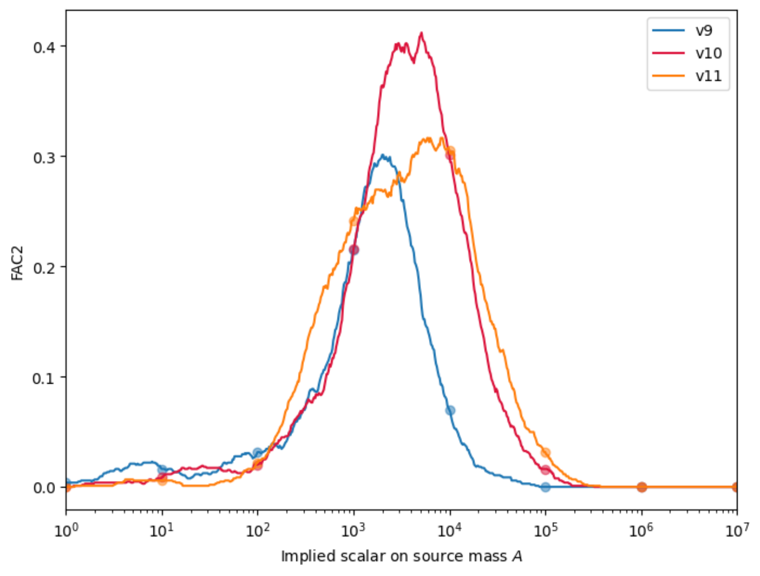
Performance metric: Estimation of peak PM10 value

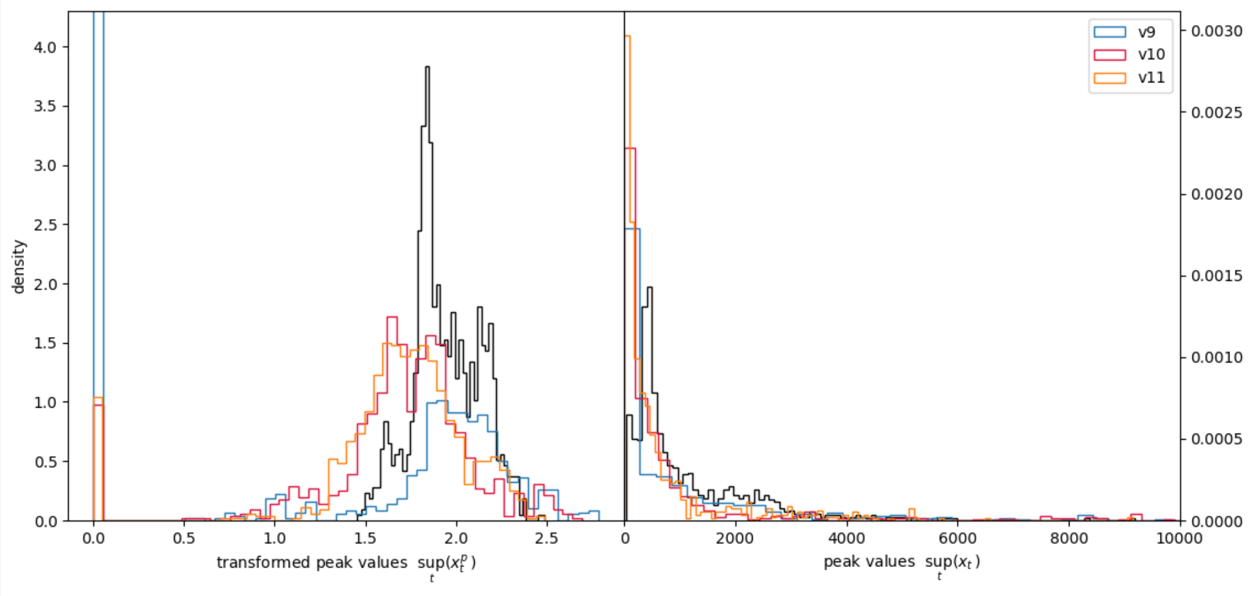
Adjustments on source mass

- Assume that peak values prediction, or concentration output of Flexpart in general, is linear in initial released mass.
- Tune the orders of magnitude against FAC2 over all 1046 receptors
- Compare modelled and observed peak PM10

Model	Fp9	Fp10	Fp11
FAC2	0.22	0.30	0.30
implied ratio	1000	10000	10000
MAE	1589.91	930.59	936.81
RMSE	3028.33	1811.48	1425.35
median FB	-0.93	-0.99	-1.02

Table 2: Model performance in peak PM_{10} estimation. Implied ratio refers to the factor of overestimation on the initial release mass inferred from the modelled peak values, and units of MAE and RMSE are in hour. All models show severe underestimation, and thus no conclusion is made other than that one of the assumptions on linearity or the initial condition might be unrealistic.

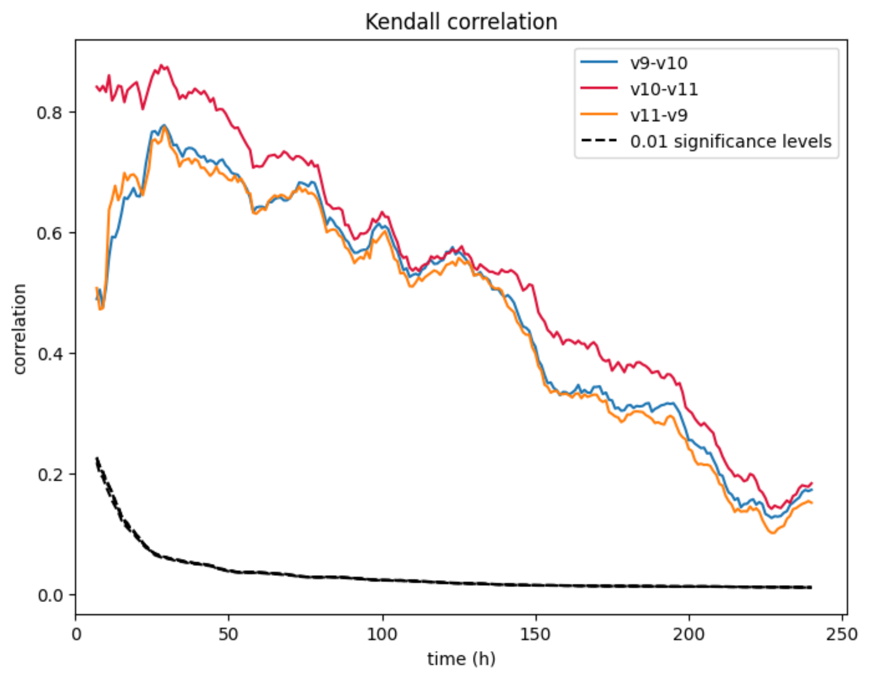


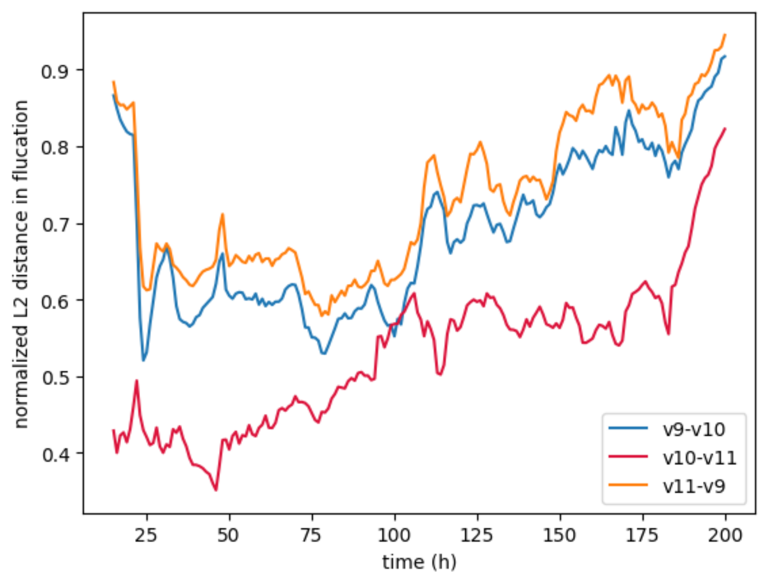


Case 2. Hypothetical nuclear accident

Consistent result in inter-model similarity is obtained

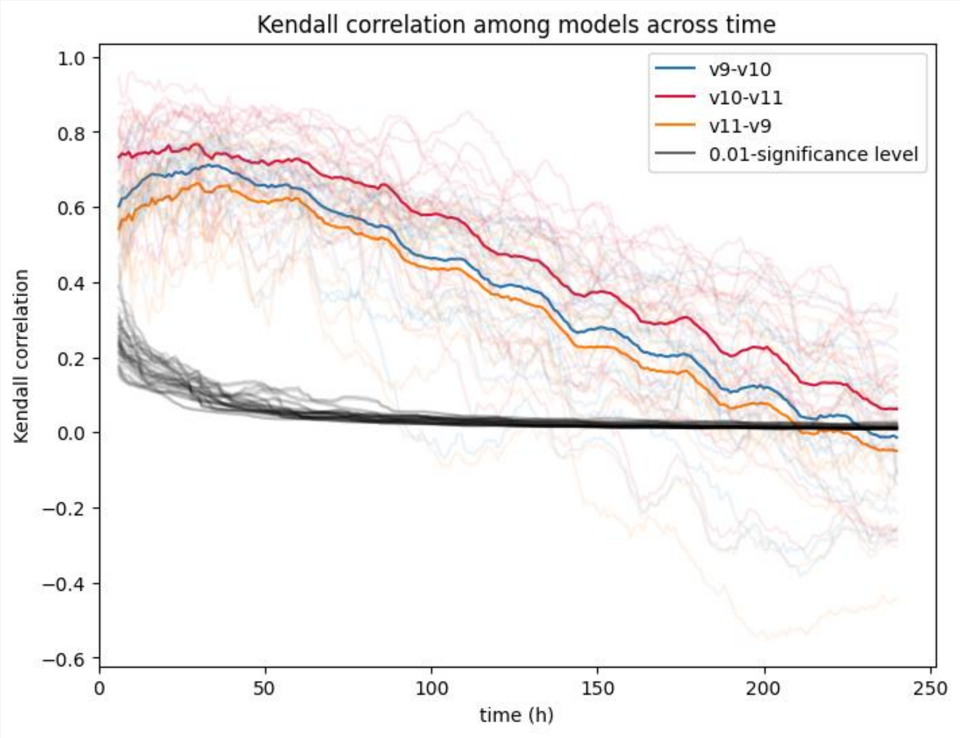
- The normalized Lp distance is the largest between version 9 and 11, and smallest between version 10 and 11. Version 10 and 11 is the most correlated pair.
- These align with what is observed in modelling sand-dust storm, and therefore, suggest that the phenomenon is not an artifact of release locations, release duration, released species, and weather conditions.

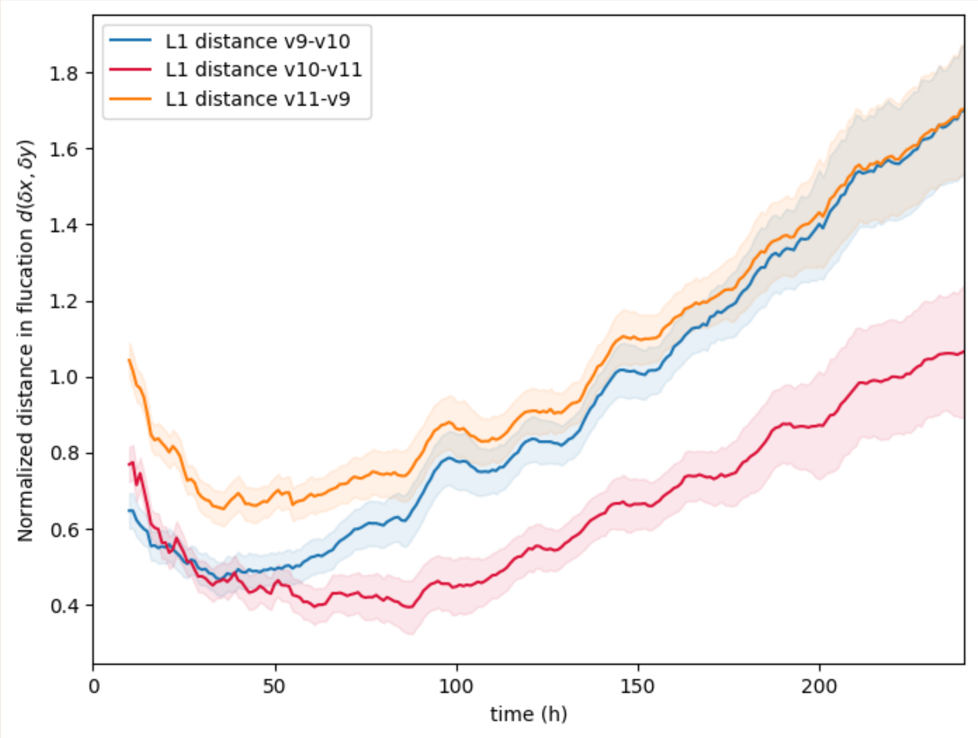




(Extra slide.) Measurement repeated for 18 nuclear power plants around the globe

- Models maintain a significant correlation for around 8-9 days on average





Conclusion

- **(Inter-model similarity.)** Flexpart 9 is the odd-one-out. Difference between version 10 and 11 is the least while that between version 9 and 11 is the greatest.
- **(Predictive power.)** Both Flexpart 10 and 11 better predict arrival times than Flexpart 9, but between themselves, they share comparable performances.
- All three models show positive spatial-correlation with observation, but the level of correlation is similar. Flexpart 9 has near zero median temporal-correlation with observation.
- All three versions are unable to predict peak PM10 values under heuristic tuning of the release mass. A more robust test is needed removing the assumption of linearity, or by fine-tuning the initial condition.
- **(Realistic diffusion.)** Well-mixed criterion is better reproduced by Flexpart version 10 than by 9, suggesting more realistic convection and diffusion.

Future directions

- **(Method assumption.)** An alternative choice of proxy for sand-dust air concentration is the PM10/PM2.5 ratio, since sand-dust is estimated to have a mean diameter of 6μm. A bad proxy may be the reason for the bad alignment in absolute values, and removing the PM2.5 component could potentially help reduce the noise in the current benchmark.
- More accurate initial condition may improve model performance. Varies the distribution within the current release box and see the variability in 1. RMSE of arrival times; 2. FAC2 of peak values; 3. time or space-averaged correlations with real data. High variability suggest a high sensitivity of the performance metric on the initial condition, and that a further fine-tuning of it is worth the computation time.
- Reanalysis data should be used to test well-mixed criterion
- **(Model assumption.)** Test if the use of skewed turbulence schemes and adaptive time-stepping improve or worsen model performance in terms of the abovementioned metrics.

Reference

1. Filonchyk, M. (2022). Characteristics of the severe March 2021 Gobi Desert dust storm and its impact on air pollution in China. *Chemosphere, 287*, 132219. https://doi.org/10.1016/j.chemosphere.2021.132219
2. Kai, K., Kawai, K., Ito, A., Aizawa, Y., Minamoto, Y., Munkhjargal, E., & Davaanyam, E. (2021). Dust hotspot in the Gobi Desert: A field survey in April 2019. *SOLA, 17*, 130–133. https://doi.org/10.2151/sola.2021-023
3. Li, H., Yang, L., Cao, Y., Li, X., & Fang, S. (2025). Validation and sensitivity analysis of the FLEXPART model using the Kincaid experiment data. *International Journal of Advanced Nuclear Reactor Design and Technology, 7*(2), 110–119. https://doi.org/10.1016/j.jandt.2025.05.002
4. Stohl, A., Sodemann, H., Eckhardt, S., Frank, A., Seibert, P., Wotawa, G., Morton, D., Arnold, D., & Harustak, M. (n.d.). The Lagrangian particle dispersion model FLEXPART version 9.3 [Manuscript submitted for publication]. Norwegian Institute for Air Research. Retrieved from https://www.flexpart.eu/fp_downloads/68_flexpart9.3.pdf
5. Pisso, I., Sollum, E., Grythe, H., Kristiansen, N. I., Cassiani, M., Eckhardt, S., Arnold, D., Morton, D., Thompson, R. L., Groot Zwaaftink, C. D., Evangeliou, N., Sodemann, H., Haimberger, L., Henne, S., Brunner, D., Burkhart, J. F., Fouilloux, A., Brioude, J., Philipp, A., Seibert, P., & Stohl, A. (2019). The Lagrangian particle dispersion model FLEXPART version 10.4. *Geoscientific Model Development, 12*(12), 4955–4997. https://doi.org/10.5194/gmd-12-4955-2019
6. Bakels, L., Tatsii, D., Tipka, A., Thompson, R., Dütsch, M., Blaschek, M., Seibert, P., Baier, K., Bucci, S., Cassiani, M., Eckhardt, S., Groot Zwaaftink, C., Henne, S., Kaufmann, P., Lechner, V., Maurer, C., Mulder, M. D., Pisso, I., Plach, A., Subramanian, R., Vojta, M., & Stohl, A. (2024). FLEXPART version 11: Improved accuracy, efficiency, and flexibility. *Geoscientific Model Development, 17*(21), 7595–7627. https://doi.org/10.5194/gmd-17-7595-2024

Reference (cont)

7. NERIS Platform. (2018). *Proceedings of the Fourth NERIS Workshop: Adapting nuclear and radiological emergency preparedness, response and recovery to a changing world (25–27 April 2018, Dublin, Ireland)*. NERIS. ISBN 978-2-9552982-2-0

8. Saunier, O., Korsakissok, I., Didier, D., Doursout, T., & Mathieu, A. (2020). Real-time use of inverse modeling techniques to assess the atmospheric accidental release from a nuclear power plant. *Radioprotection*, 55(2), 107–115. https://doi.org/10.1051/radiopro/2020044

Appendix

Vertical profile across time