DE LA RECHERCHE À L'INDUSTRIE

Synthetic Data And Deep Neural Networks For Atmospheric Dispersion Modelling In Urban Areas

HARMO 20

20th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes

M. Mendil, S. Leirens, P. Armand, C. Duchenne

June, 15, 2021

www.cea.fr

Overview

1.1 Scope of the Study

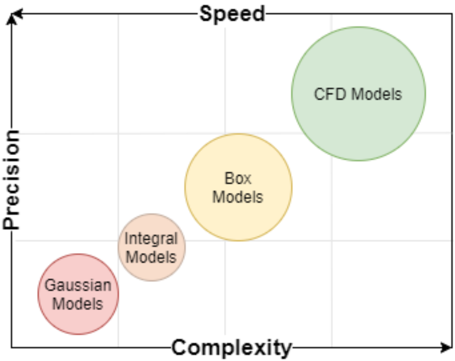
- Unexpected pollution emissions in urban areas : accidental (*e.g.* chemical accidents) or malicious (*e.g.* hostile fire)

FIGURE – Philadelphia Refinery Explosion (Forbes, 2019)

- Emergency crisis intervention from the authorities to protect the population and the environment

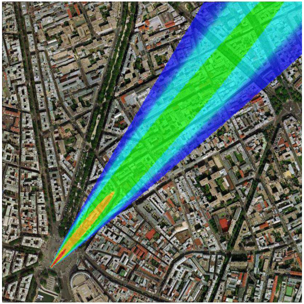
- Need for **fast** and **accurate** pollution models to estimate exposure risks and provide recommendations to decision makers

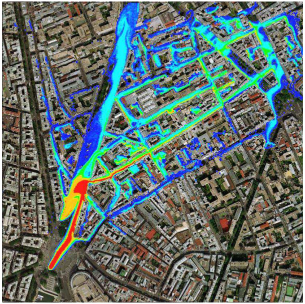
1.2 Air Pollution Models



- Several families of models
- Trade-off between precision and complexity

1.3 Pollution in urban areas



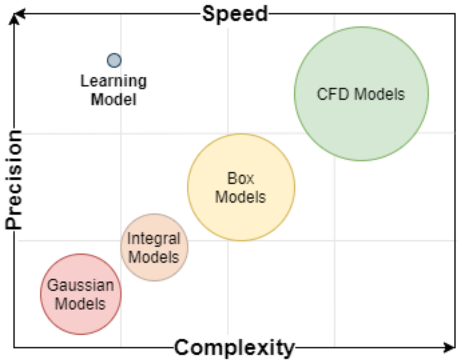


Dispersion predicted with Gaussian (left) and CFD (right) models

(P. Armand, C. Duchenne, and L. Patryl, ITM 2015, France)

- Large model sensitivity in urban areas
- CFD-level accuracy is required for risk assessment

1.3 Contribution



- Feasibility of a transport and dispersion learning model that is :
  - Fast
  - Precise
  - Usable for any urban area

2.1 Machine Learning

"A computer program is said to learn from experience *E* with respect to some class of tasks *T* and performance measure *P*, if its performance at tasks in *T*, as measured by *P*, improves with experience *E*" (Tom Mitchell, 1997)

2.1 Machine Learning

"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E" (Tom Mitchell, 1997)

- The task T : how the machine should process the inputs
  - predict the concentration field subsequent to an accidental release

2.1 Machine Learning

"A computer program is said to learn from experience *E* with respect to some class of tasks *T* and performance measure *P*, if its performance at tasks in *T*, as measured by *P*, improves with experience *E*" (Tom Mitchell, 1997)

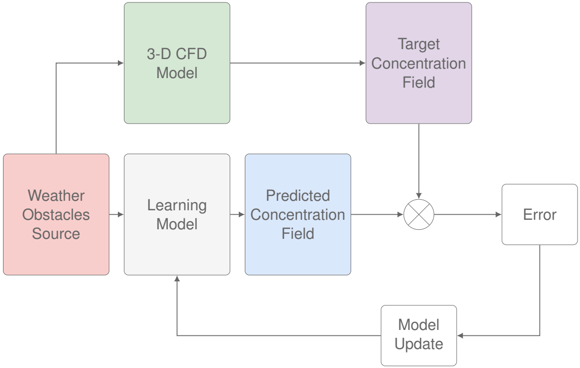
- The task *T* : how the machine should process the inputs
  - predict the concentration field subsequent to an accidental release
- The performance measure *P* : quatitative performance at accomplishing the task *T*
  - mean squarred error between predicted and target concentration values

2.1 Machine Learning

"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E" (Tom Mitchell, 1997)

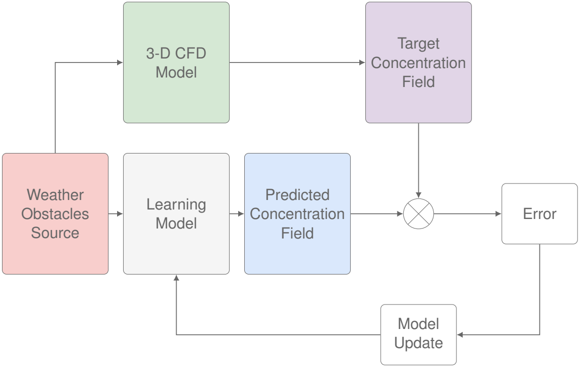
- The task T : how the machine should process the inputs
  - predict the concentration field subsequent to an accidental release
- The performance measure P : quatitative performance at accomplishing the task T
  - mean squarred error between predicted and target concentration values
- The experience E : how the algorithm experiences the data it learns from
  - learning useful transport and dispersion properties from a database of various concentration fields

2.1 The Learning Process (1/2)



- Model inputs : weather, obstacle map, emission source
- Learning model : parametric function of the inputs
- Predicted model outputs : time-integrated concentration field
- Target outputs : ground truth time-integrated concentration field

2.1 The Learning Process (2/2)



- Learning : iterative model update to minimize the prediction error
- Generalization : learned model must perform well on new, previously unseen inputs

2.2 Synthetic Data (1/2)

- Deep learning algorithms requires a large quantity of data
  - Real data from real size or small-scall experiments
    - ☺ High accuracy
    - ☹ Slow and expensive
  - Synthetic data from computer simulations
    - ☻ Model-dependent accuracy
    - ☺ Cheap, relatively fast, flexible

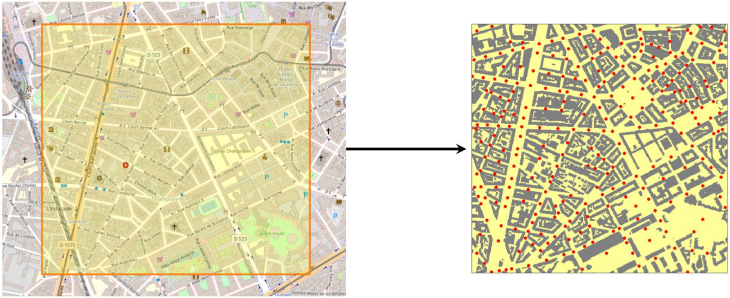
2.2 Synthetic Data (1/2)

- Deep learning algorithms requires a large quantity of data
  - Real data from real size or small-scall experiments
    - ☺ High accuracy
    - ☹ Slow and expensive
  - Synthetic data from computer simulations
    - ☻ Model-dependent accuracy
    - ☺ Cheap, relatively fast, flexible

- Synthetic data generated by Parallel Micro-SWIFT-SPRAY (PMSS)
  - 3-D atmospheric transport and dispersion simulator
  - Lagrangian particle dispersion model
  - High time and space resolution
  - Accounts for the presence of obstacles

2.2 Synthetic Data (2/2)

- Training Dataset : city of Grenoble (France)
  - Approx. 15000 PMSS simulations
  - 500 × 500 grid of 2 m space resolution
  - 274 different hypothetical emissions sources
  - 54 different stationary weather conditions



2.3 Learning Model Architecture

- No Free Lunch Theorem : there is no single best learning model suited for all problems
- Problem-related criteria
  - Pre/post-processing : scaling, centering, vectorization
  - Multilayer perceptron : integrated concentration regression
  - Encoder/decoder blocks : (space) dimension reduction

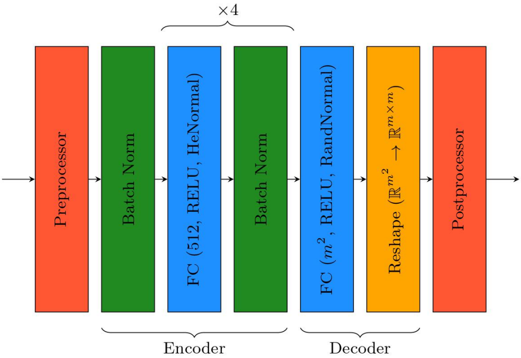
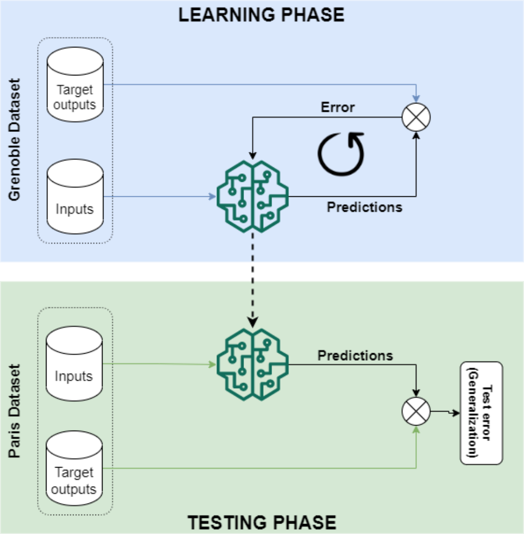


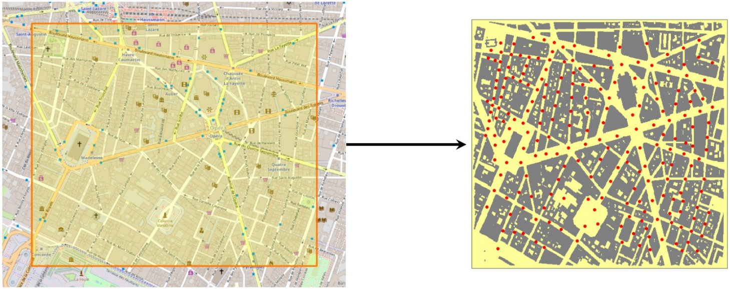
FIGURE – Learning Model Architecture

3.1 Generalization Test



3.1 Test Dataset

- Test Dataset : city of Paris (France)
  - Approx. 12000 PMSS simulations
  - 600 $\times$ 500 grid of 2 m space resolution
  - 222 different hypothetical emissions sources
  - 54 different stationary weather conditions



3.2 Learning Performance

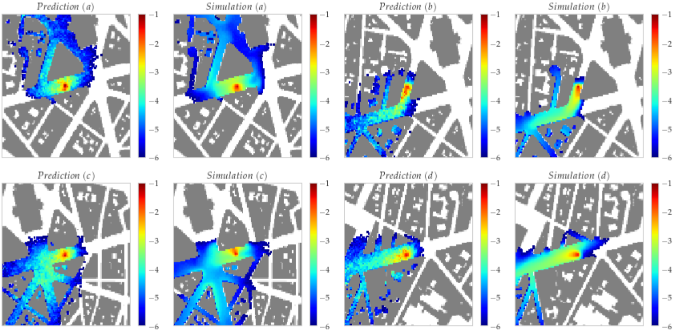


FIGURE – Predictions *vs.* Ground Truth (PMSS simulations) : integrated concentration fields (over 2 hours) in Paris (log scale)

- Average mean squared error : 0.96
  - Accurate dispersion modelling in street intersections
- Fast execution time $\approx$ 0.75 ms per prediction

Conclusions

- Need for fast and accurate models of accidental/malicious air pollution in urban areas
- First learning model of air transport and dispersion usable in any urban area
- The trained model is precise and enables fast predictions
- Next step : joint prediction of horizontal and vertical pollution distributions

Thanks for your attention

Commissariat à l'énergie atomique et aux énergies alternatives
Centre de Grenoble | 17 Avenue des Martyrs, 38054 Grenoble
T. +33 (0)4 38 78 44 00 | F.

DRT
LETI