This manuscript has been published in final form at https://doi.org/10.1016/j.envres.2019.05.044

# Extending the spatial scale of land use regression models for ambient ultrafine particles using satellite images and deep convolutional neural networks

Kris Y. Hong[a], Pedro O. Pinheiro[b], Laura Minet[c], Marianne Hatzopoulou[c], Scott Weichenthal[a,*]

[a] McGill University, Department of Epidemiology, Biostatistics and Occupational Health, Montreal, QC, Canada
[b] Element AI, Montreal, Canada
[c] University of Toronto, Toronto, Canada

## Abstract

We paired existing land use regression (LUR) models for ambient ultrafine particles in Montreal and Toronto, Canada with satellite images and deep convolutional neural networks as a means of extending the spatial coverage of these models. Our findings demonstrate that this method can be used to expand the spatial scale of LUR models, thus providing exposure estimates for larger populations. The cost of this approach is a small loss in precision as the training data are themselves modelled values.

## 1. Introduction

Land use regression (LUR) models are commonly used to estimate spatial variations in outdoor air pollution concentrations. However, the spatial coverage of these models is often limited by a lack of detailed traffic/geographic information system (GIS) data outside city boundaries. As a result, populations living outside city limits are often excluded from population-based studies of environmental pollutants even though spatial patterns in pollutant concentrations [1] and land use/built environment/ traffic characteristics can be similar to those within city boundaries.

We recently developed LUR models for ultrafine particle (< 0.1 µm, UFP) concentrations in Montreal (2011–2012) [2] and Toronto (2010–2011) [3], Canada and these models have subsequently been applied in several epidemiological investigations [4,5,6]. Given the availability of large Canadian Cohorts, many more subjects could have been included in these studies had exposure estimates been available on a broader spatial scale. This is particularly important for studies of rare outcomes (e.g. childhood cancers) where the number of cases can be small even if large numbers of people are followed over time. In this study, we examine a novel method of expanding the spatial scale of land use regression models by pairing LUR model estimates with satellite images with subsequent analyses using deep convolutional neural networks (CNN). The end result is a deep learning model that takes satellite images as input and returns continuous estimates of outdoor UFP concentrations. As information on many built environment characteristics can be captured in digital images (e.g. land use, road length, building density, etc.), this approach may offer a cost-effective method of extending the spatial scale of LUR models in the absence of GIS predictor data.

## 2. Methods

A database of approximately 112,000 random latitude-longitude points were generated across Montreal (n = 48,823) and Toronto (n = 63,565) and existing land use regression models [2,3] were used to estimate UFP concentrations at each point. Briefly, these land use regression models were developed using mobile monitoring data collected during the summer and winter months and included parameters for land use characteristics, proximity to sources such as airports and rail, and temperature and wind speed parameters to adjust for temporal variations in UFP concentrations between monitoring routes [2,3]. Next, satellite images were downloaded from Google Maps for each latitude-longitude point using the ggmap [7] package in the R statistical computing environment [8] at zoom levels 18 (covering approximately 250 x 250-m) and 19 (approximately 125 x 125-m). All images were saved at a resolution

of 256 x 256 x 3 to maintain a reasonable training time.

Since the original latitude-longitude coordinates were randomly generated, some coordinates could be close together and corresponding satellite images could overlap. This could lead to an inflation of model performance if two images very close together were included in both the model development and test sets (i.e. if the model were tested on images very similar to those on which it was trained). To overcome this problem, all latitude-longitude points were assigned a six-digit geohash code corresponding to their location on a rectangular grid of geohash cells covering approximately 600 x 600-m each (i.e. images very close to each other would almost always have the same geohash code). The database was then split into training (80%), validation (10%), and test sets (10%) such that the three sets were disjoint by geohash cells (i.e. satellite images overlapped within sets but rarely between sets). This ensured that information learned by the model from locations in the training set could also be generalized to new locations such as those in the validation and test sets. These data (i.e. UFP estimates and satellite images) were then used to train new deep convolutional neural networks to extend the spatial scales of the original LUR models using only satellite images. It is standard practice to use training, validation, and test sets when developing new deep learning models. Model training takes place on the "training set" and the "validation set" is used during model training to evaluate different hyperparameters used in training the deep learning model (i.e. learning rate, batch size, optimizer). The "test set" acts as a completely external dataset that plays no role in training model parameters or selecting hyperparameters. Having both validation and test sets acts as an extra guard against over-fitting.

LUR model estimates were extended to regions with land use/traffic characteristics similar to those in the original model area; similarity was based on a visual inspection of satellite images as GIS data were not available to conduct a quantitative comparison of land use characteristics. For example, if a given portion of the extended model area was similar in appearance to any region in the original model area it was included; this primarily included sub-urban commercial and residential areas with roadways and neighbourhoods comparable to those found in the original model area. As land use and traffic characteristics in the extended model areas were similar to those in the original model area, extending LUR model estimates to these regions is no different that predicting exposures for unmonitored regions within city boundaries. To further check this assumption, final model estimates were visually compared to land use features in satellite images in the extended model area to identify potential problem areas.

CNN-based models with two input images (i.e. two zoom levels) were trained to predict spatial variations in outdoor UFP concentrations on a continuous scale. The zoom level 18 and 19 images provided the model with context about the surrounding area and high-resolution information about the immediate area, respectively. Each input was fed into a convolutional base for feature extraction with an input size of 256 x 256 x 3 (i.e. width x height x colour channels), followed by a 2D global average pooling operation. The outputs from the two CNNs were then concatenated and mapped to a linear activation layer for prediction (i.e. the linear activation function provided continuous predictions of outdoor UFP concentrations). Since land features captured in satellite imagery are rotation-invariant (i.e. they retain the same landscape features regardless of their orientation: roads looks like roads, parks look like parks, etc.), we used data augmentation to increase the generalizability of our model by randomly flipping the two paired satellite image inputs in the training set in the vertical or horizontal direction (or neither or both).

Models were initialized with ImageNet weights, and were trained with a batch size of 64 images for up to 100 epochs (i.e. model training is an iterative process and 1 epoch is 1 iteration). Callback functions were used during model training to decrease the learning rate by a factor of 0.1 if the

validation accuracy did not improve for 10 epochs, and also to prematurely stop model training if the validation accuracy did not improve for 20 epochs. The model with the lowest validation root mean square error (RMSE) was retained. Separate models were trained for Montreal and Toronto using data from each city as this approach was superior to models trained using data from both cities combined (data not shown). Optimal model configuration was based on a systematic evaluation of two convolutional base architectures: Xception [9] and ResNet50 [10] in combination with either the RMSProp [11] or Nadam [12] optimizers using a learning rate of 0.002. A detailed leaderboard was maintained, tracking the performance of different combinations of model architectures and hyperparameters and the model that achieved the lowest RMSE on the validation dataset was selected as the final model. All analyses were conducted using the Keras package [13] in R [8] and Python with a Lambda Quad Workstation (Lambda Labs, San Fransisco, CA) containing 4 Nvidia Titan Xp GPUs. On average, model training took approximately 4-min/epoch.

Population centre shapefiles for the Greater Montreal and the Greater Toronto/Hamilton Area were downloaded from Statistics Canada, and a grid of latitude-longitude coordinates was generated inside the population centres at a 0.002 decimal degree resolution (approximately 150-m apart). Next, satellite images were downloaded for each latitude-longitude point and the CNN-based models were applied to the satellite images to predict UFP concentrations at each latitude-longitude location.

## 3. Results

Final CNN-based models used the Xception base and the Nadam optimizer with a learning rate of 0.002. CNN-based models provided reliable estimates of LUR model predictions in both cities (Fig. 1). Specifically, the slopes of linear regression models between LUR model predictions and CNN-based models in the test set were close to one in both Montreal (slope = 1.00, 95% CI: 0.99, 1.01) and Toronto (slope = 1.02, 95% CI: 1.01, 1.03) but the Toronto model was more precise (Toronto: RMSE = 1366 particles/cm$^3$, $R^2$ = 0.90; Montreal: RMSE = 3755 particles/cm$^3$, $R^2$ = 0.86). Intercept values were positive for Montreal (intercept = 242 particles/cm$^3$, 95% CI: -79, 563) and negative for Toronto (intercept = -461 particles/cm$^3$, 95% CI: -645, -277) suggesting that CNN-based models slightly underestimated LUR-based model predictions for Montreal and overestimated LUR-based model predictions for Toronto.

Fig. 2 illustrates the extended spatial scale of CNN-based model predictions compared to the original LUR models. In Montreal, the CNN-based model likely overestimated UFP concentrations in a small agricultural area north of Montreal island. In this region, the fields were plowed in regular rectangular shapes resembling roadways. This error would not impact population-based exposures (as there are no residences in these areas) but highlights the need to visually inspect model predictions to identify possible errors. In general, CNN-based models clearly identified high-traffic roadways and built-up areas as important predictors of UFP concentrations with lower values in residential areas with more green space. Moreover, when we mapped the spatial distribution of model residuals (i.e. the difference between LUR and CNN model predictions) there was no clear spatial trend in model errors across either city (Fig. 3).

## 4. Discussion and conclusions

Our findings suggest that satellite images and deep convolutional neural networks can be used to extend the spatial scale of LUR models to regions with missing GIS data. This approach requires visual inspection of built environment/land use characteristics in the extended area to avoid extrapolating beyond the original model parameters, and may be particularly useful for studies of rare diseases when it is important to include as many people as possible in epidemiological analyses. The cost of this approach is a decrease in the precision of exposure estimates as the training data are themselves modelled values; therefore, the total error is a combination of the original LUR model

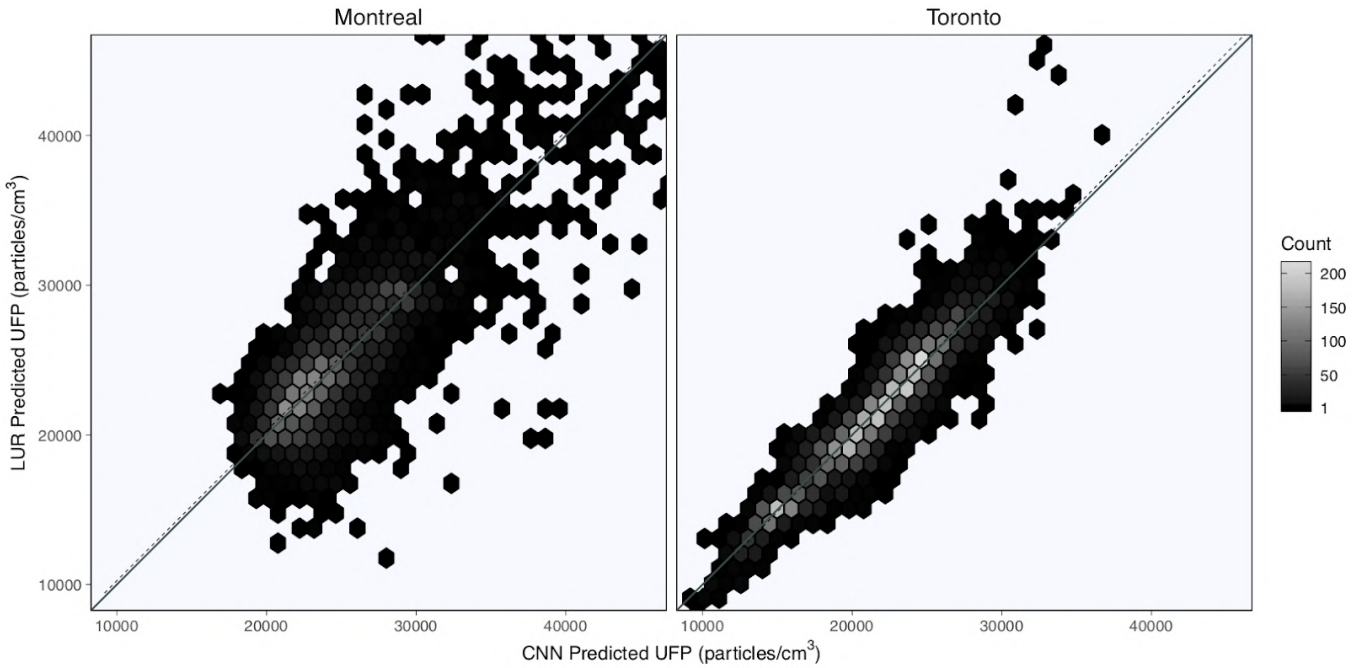


**Fig. 1.** Comparison of LUR-Predicted and CNN-Predicted UFP concentrations (particles/cm$^3$) in Montreal and Toronto, Canada.

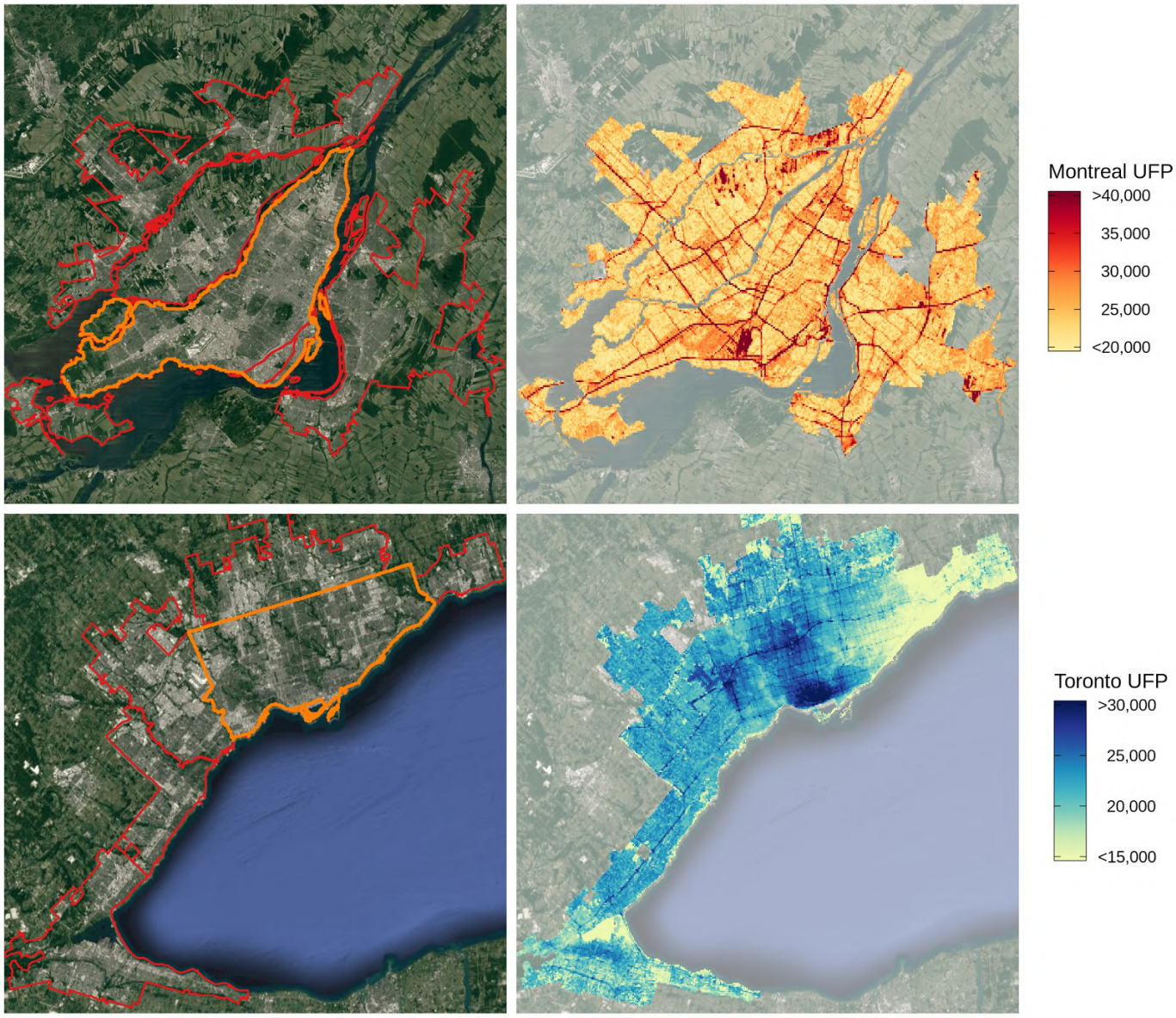


**Fig. 2.** CNN-based model predictions for UFP concentrations (particles/cm$^3$) using satellite images. The orange outlines in the left column indicate the spatial range of the original LUR models, and the red outlines indicate the extended spatial range of the CNN models.

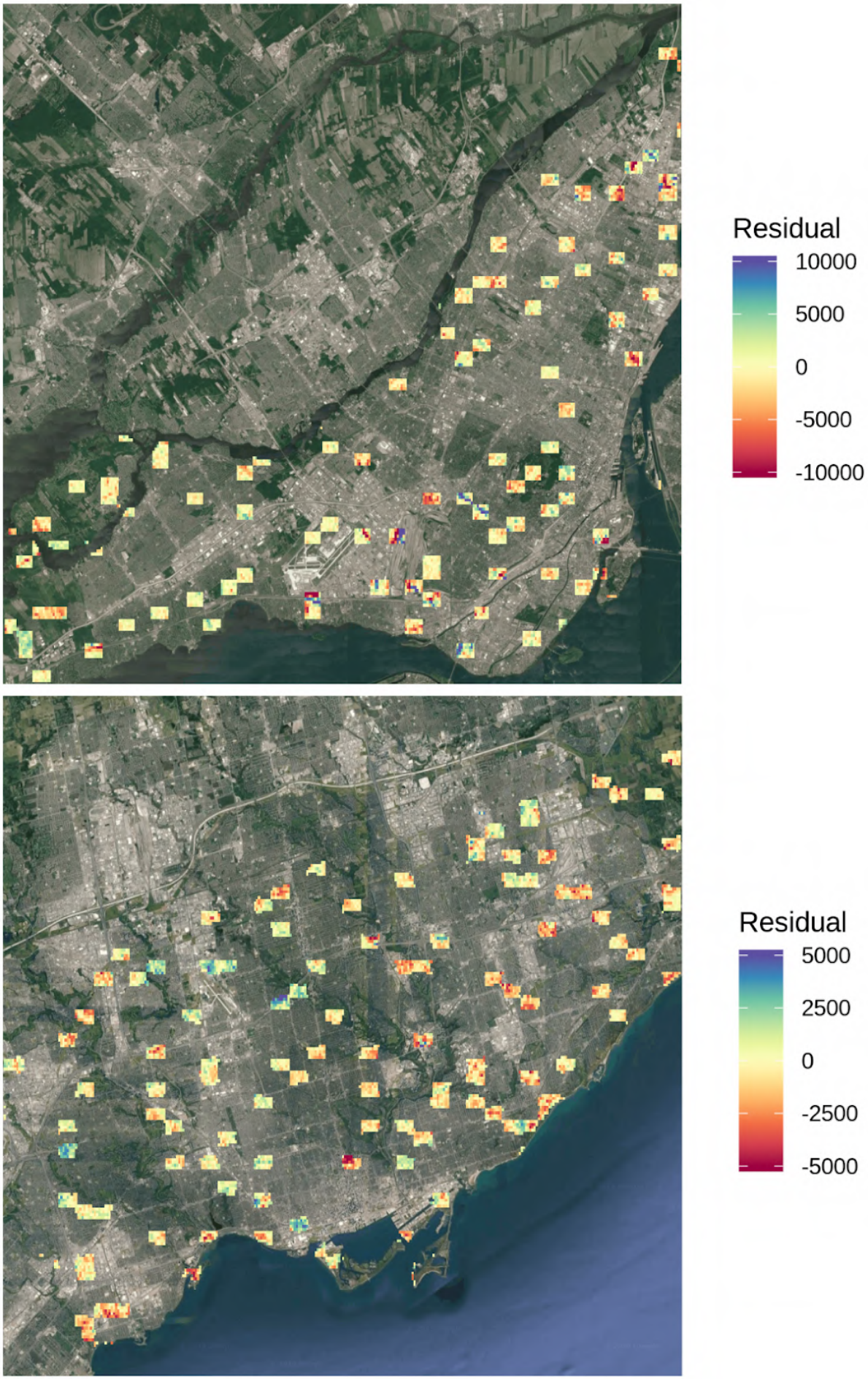


**Fig. 3.** Maps of residuals (LUR-CNN predictions) for spatial hold-out samples in the Montreal (top) and Toronto (bottom) test sets. Each rectangular cluster of tiles represents one geohash cell. Residuals are in units of particles/cm$^3$.

error plus the additional uncertainty contributed by the deep learning model. In some cases, however, the additional uncertainty contributed by the deep learning model may be rather low as was the case for UFPs in Toronto where the RMSE values was less than 1500 particles/$cm^3$ compared to LUR model[3] uncertainty of approximately 14,000 particles/$cm^3$.

One limitation of this study is that we did not have monitoring data outside the original study areas to validate CNN-based model predictions. However, the "boundaries" of the original LUR model areas have no real meaning and areas inside and outside formal city limits overlap extensively in terms of land use/traffic characteristics. The more important concern is extrapolating to land use/traffic characteristics not monitored in the development of the original LUR models: if the variance of land use/traffic characteristics is similar between regions, predicting UFP concentrations in areas outside city limits is no different than predicting values for unmonitored locations within city limits. As an additional check on this assumption, we compared satellite images and modelled estimates in extended areas and identified a region north of Montreal where the CNN-based model likely overestimated UFP exposures (possibly because plowed fields resembled roadways). As noted above, this error would not impact population-based exposures (as this was not a residential area) but does highlight the need to verify model predictions based on satellite images when similar data are not available in the training set. In general, visual inspection of the expanded study area is straightforward for city-scale models but may become tedious with large multi-city models. In these cases, it may be useful to identify metrics that can be used to quantitatively identify overlapping land use characteristics between study areas if GIS data are available.

A second limitation of CNN-based models is that their application will result in non-differential exposure measurement error that will bias effect estimates toward the null. This bias will be slightly greater than for original LUR models given the additional error contributed by the deep learning models as noted above. Therefore, the utility of this approach may be context specific and the benefits of increased sample size need to be weighed against this loss in precision. Finally, as satellite images were downloaded after the development of the original LUR models, it is possible that some features may have changed over time and this would contribute to differences between CNN and LUR-based estimates. However, major features (e.g. highways, rail yards) relevant to UFPs did not change location over this time.

In conclusion, our findings suggest that satellite images and deep convolutional neural networks can be used to extend the spatial scale of LUR models. This approach offers a cost-effective means of estimating exposures for larger numbers of people in studies of within-city spatial variations in environmental exposures as long as built environment and land use characteristics overlap between the original LUR study area and the extended region of interest. This approach comes at the cost of decreased precision in exposure estimates, and this limitation should be weighed against the benefits of increased spatial scale.

## References


1. Hakey, S., Sforza, P., Pierson, M., 2019. Using mobile monitoring to develop hourly empirical models of particulate air pollution in a rural Appalachian community. Environ. Sci. Technol. 53, 4305–4315.
2. Weichenthal, S., Van Ryswyk, K., Goldstein, A., Bagg, S., Shekkarizfard, M., Hatzopoulou, M., 2016. A land use regression model for ambient ultrafine particles in Montreal, Canada: a comparison of linear regression and a machine learning approach. Environ. Res. 146, 65–72.
3. Weichenthal, S., Van Ryswyk, K., Goldstein, A., Shekarrizfard, M., Hatzopoulou, M., 2016. Characterizing the spatial distribution of ambient ultrafine particles in Toronto, Canada: a land use regression model. Environ. Pollut. 208, 241–248.
4. Lavigne, E., Donelle, J., Hatzopoulou, M., Van Ryswyk, K., van Donkelaar, A., Martin, R.V., Chen, H., Stieb, D.M., Gasparrini, A., Crighton, E., Yasseen Iii, A.S., Burnett, R.T., Walker, M., Weichental, S., 2019 Feb 20. Spatiotemporal variations in ambient ultrafine particles and the incidence of childhood asthma. Am. J. Respir. Crit. Care Med.
5. Weichenthal, S., Lavigne, E., Valois, M.F., Hatzopoulou, M., Van Ryswyk, K., Shekarrizfard, M., Villeneuve, P.J., Goldberg, M.S., Parent, M.E., 2017. Spatial variations in ambient ultrafine particle concentrations and the risk of incident prostate cancer: a case-control study. Environ. Res. 156, 374–380.
6. Bai, L., Weichenthal, S., Kwong, J.C., Burnett, R.T., Hatzopoulou, M., Jerrett, M., van Donkelaar, A., Martin, R.V., Van Ryswyk, K., Lu, H., Kopp, A., Chen, H., 2019. Associations of long-term exposure to ultrafine particles and nitrogen dioxide with increased incidence of congestive heart failure and acute myocardial infarction. Am. J. Epidemiol. 188, 151–159.
7. Kahle D, Wickham H. ggmap: Spatial Visualization with ggplot2. The R Journal, 5(1), 144-161.
8. R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria 3-900051-07-0.
9. Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions. ArXiv, 1610. 02357vol. 3, 2017. Arxiv.org/abs/1610.02357v3.
10. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. ArXiv, 1512. 03385vol. 1, 2015. Arxiv.org/abs/1512.03385.
11. Hinton, G., Srivastava, N., Swersky, K., 2016. Lecture 6a Overview of Mini-Batch Gradient Descent (Coursera Lecture slides).
12. Dozat, T., 2016. Incorporating nesterov momentum into adam. In: ICLR Workshop, 2013-2016.
13. Chollet, F., 2015. Keras: deep learning library for theano and tensorflow. https://keras.io.