



Title	Study on Hybrid Approach Incorporating Machine Learning Algorithms and Chemical Transport Model for Estimating Air Quality in Japan and Thailand
Author(s)	Thongthammachart, Tin
Citation	大阪大学, 2022, 博士論文
Version Type	
URL	<a href="https://hdl.handle.net/11094/89623">https://hdl.handle.net/11094/89623</a>
rights	
Note	やむを得ない事由があると学位審査研究科が承認したため、全文に代えてその内容の要約を公開しています。全文のご利用をご希望の場合は、<a href="https://www.library.osaka-u.ac.jp/thesis/#closed">大阪大学の博士論文について</a>をご参照ください。

*The University of Osaka Institutional Knowledge Archive : OUKA*

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

## Abstract of Thesis

Name (THONGTHAMMACHART Tin)	
Title	Study on Hybrid Approach Incorporating Machine Learning Algorithms and Chemical Transport Model for Estimating Air Quality in Japan and Thailand (日本とタイにおける大気質推計のための機械学習と化学輸送モデルを組み込んだハイブリッド型手法に関する研究)
<p><b>Abstract of Thesis</b></p> <p>Chapter I is an introduction to the dissertation which describes the rationale, novelty, and objectives of the study. A statistical technique, referred to land use regression (LUR), has typically been applied for an estimation of air pollutant concentrations in a particular area. However, the conventional approach with linear regression bears the disadvantages of limited temporal resolution and difficulty in capturing non-linearity between predicted and predictor variables. Chemical transport model (CTM) and machine learning (ML) algorithms can be applied to overcome the limitations of the conventional approach. Although the hybrid LUR model incorporating ML algorithm and CTM were used for air quality assessment, it has not been applied for assessing air quality that is affected by long-range transport and local pollution, e.g., Japan. In addition, it has not been studied in developing countries due to the limited number of air quality monitoring stations. Therefore, this hybrid approach requires validation to ensure that it is effectively applicable for air pollutant predictions. The main objective of this dissertation is to develop and evaluate the hybrid LUR approach which incorporates ML algorithm and CTM for air pollutant predictions in Japan and Thailand.</p> <p>Chapter II presents the air pollutant information, the description of the modelling system, the LUR technique, the description of machine learning, and statistical evaluation methods that were applied to this study.</p> <p>Chapter III is to develop and evaluate the hybrid LUR model incorporating ML algorithm and CTM for NO<sub>2</sub> and PM<sub>2.5</sub> assessment in Japan. Random forests (RF), as an ML algorithm, was used to build the LUR model. This model was termed a land use random forests (LURF) model. The LURF model was conducted in two conditions by inclusion and exclusion of CTM for investigating NO<sub>2</sub> and PM<sub>2.5</sub> predictabilities between with-CTM and without-CTM models. Moreover, the air pollutant predictabilities from the with-CTM LURF model were compared with those from the LUR model built with linear regression and CTM to investigate the performance between non-linear and linear models. As the results, the with-CTM LURF model performs higher NO<sub>2</sub> and PM<sub>2.5</sub> predictive accuracy than the without-CTM LURF model and the LUR model built with linear regression. Accordingly, the hybrid approach is advantageous in accurately estimating daily NO<sub>2</sub> and PM<sub>2.5</sub> levels.</p> <p>Chapter IV is to enhance the NO<sub>2</sub> and PM<sub>2.5</sub> predicting performances of the hybrid LUR models by using Light Gradient Boosting Machine (LightGBM). The LightGBM is a newly developed ML that is highly efficient and capable of handling large-scale data. NO<sub>2</sub> and PM<sub>2.5</sub> predictabilities from the LightGBM model was compared with those from the hybrid LUR models built with RF and Extreme Gradient Boosting Machine (XGBoost) algorithms. The advantages of the three ML algorithms were discussed and provided in this chapter. The findings demonstrate that the LightGBM model processed faster and achieved more accurate NO<sub>2</sub> and PM<sub>2.5</sub> predictions than the RF and XGBoost models.</p> <p>Chapter V is to evaluate and implement the hybrid LUR model for air quality assessment in developing country. The hybrid LUR model developed from LightGBM was used to predict the PM<sub>2.5</sub> levels and illustrate its spatiotemporal distributions in central region of Thailand. The outcome reveals that the hybrid approach accurately estimates the daily ambient PM<sub>2.5</sub> level, which is influenced by transboundary and local pollutions.</p> <p>Chapter VI summarizes the important conclusions that the hybrid modelling approach can be effectively applicable for air pollutant predictions in both developed and developing countries. Moreover, this chapter describes the limitations of the study together with the recommendations for future studies. For instance, the modelling approach can be utilized for risk assessment studies on the short-term (acute) effects of large cohorts of residents, including those living away from air quality monitoring stations.</p>	

## 論文審査の結果の要旨及び担当者

氏名 ( THONGTHAMMACHART Tin )		
	(職)	氏名
論文審査担当者	主査 教授	近藤 明
	副査 教授	東海 明宏
	副査 准教授	嶋寺 光

## 論文審査の結果の要旨

大気汚染暴露による疾病を評価するために、地域の大気汚染物質濃度の時空間分布の推定が必要です。この推定には、土地利用回帰 (LUR) と呼ばれる統計的手法が一般的に良く用いられているが、推定精度が不十分であることが指摘されている。それは、LUR モデルが、予測変数と予測変数の間の非線形性を表現できないという欠点を有しているからである。本研究は、化学輸送モデル (CTM) および機械学習 (ML) アルゴリズムを LUR モデルに組み込んだハイブリッド LUR モデルを開発し、この欠点を克服し、大気汚染物質濃度の時空間分布の推定精度の向上を目指したものである。

はじめに、ML アルゴリズムと CTM を組み込んだハイブリッド LUR モデルに関して、ML アルゴリズムの組み込みの有無、CTM の組み込みの有無の異なる条件下で、日本の  $NO_2$  および  $PM_{2.5}$  の時空間濃度分布推定を実施している。ここでの推定では、ML アルゴリズムとしてランダムフォレスト (RF) アルゴリズムを利用している。計算結果は、RF アルゴリズムと CTM を組み込んだハイブリッド LUR モデルが、日本の  $NO_2$  および  $PM_{2.5}$  の日変動濃度を正確に推定することが出来ることを明らかにしている。

次に、最新の ML アルゴリズムである Light Gradient Boosting Machine (LightGBM)、従来の ML アルゴリズムである RF および Extreme Gradient Boosting Machine (XGBoost) アルゴリズムを用いて、ハイブリッド LUR モデルを構築し、日本の  $NO_2$  および  $PM_{2.5}$  の時空間濃度分布推定を実施している。計算結果は、LightGBM モデルが RF および XGBoost モデルよりも計算速度が速く、より精度よく  $NO_2$  および  $PM_{2.5}$  の日変動濃度を正確に推定することが出来ることを明らかにしている。

さらに、日本とは気象条件および大気汚染条件が異なり、かつモニタリングデータが少ないタイ中部地域に対して、LightGBM アルゴリズムと CTM を組み込んだハイブリッド LUR モデルを適応し、 $PM_{2.5}$  の時空間濃度分布推定を実施している。計算結果は、近隣諸国からの越境汚染と地域汚染の影響を受けるタイ中部地域の  $PM_{2.5}$  の日変動濃度を正確に推定することが出来ることを明らかにしている。

以上のように、本論文は環境エネルギー工学、特に環境科学に寄与することが大きい。

よって本論文は博士論文として価値あるものと認める。