Estimation Method of Covariance Matrix in Atmospheric Inversion of CO_2 Emissions

Yubin Han[1,2,a], Ge Ren[2,3,b], Shiqi Zhao[2,c], Kailun Du[2,d], Hong Lin[1,2,3,e*]

[1]School of Computer and Artificial Intelligence, Zhengzhou University, 450001 Zhengzhou, China
[2]Zhengzhou Institute of Metrology, 450001 Zhengzhou, China
[3]National Institute of Metrology, 100029 Beijing, China

Abstract: Atmospheric inversion of CO2 Emissions is based on the correction of prior carbon dioxide flux inventories using concentration monitoring data and atmospheric transport models to obtain posterior carbon dioxide flux. In atmospheric inversion studies, fixed covariance functions are commonly used to generate covariance matrices, and the hyperparameters in the covariance functions are empirically estimated. In this study, we design and implement an ideal experiment based on meteorological data from the central urban area of Zhengzhou, using WRF-STILT to generate sensitivity matrices and construct real carbon emission inventories and prior inventories. Based on the real carbon emission inventories and sensitivity matrices of monitoring stations, simulated observation concentration values are generated. Firstly, based on the observed concentration values, sensitivity matrices of monitoring stations, prior inventories, and constructed covariance matrices, the values of hyperparameters are determined based on maximum marginal likelihood estimation. Then, the influence of different prior covariance functions on the inversion results is tested, and it is found that the prior covariance matrix generated by the balgovind covariance function is most suitable for the experimental data.

1 Introduction

In recent years, global climate change has become increasingly evident, with the greenhouse effect emerging as a major contributing factor, drawing growing attention. Carbon dioxide, a primary greenhouse gas, stands as one of the key factors driving climate change[1]. Despite some progress globally in mitigating climate change, current emission levels fall far short of the substantial reductions required to meet the ambitious global climate goals, posing ongoing challenges in addressing climate change. Human activities are the primary cause of the rapid increase in atmospheric carbon dioxide concentration[2]. Human activities are predominantly concentrated in urban areas, which serve as major emission sources of carbon dioxide[3]. With the continuous growth of the global population, this situation is expected to intensify in the future.

Estimating carbon dioxide emissions in urban areas is crucial for formulating and implementing relevant regulations and scientifically addressing climate issues. Traditionally, estimation methods based on emission inventories have been employed. These methods involve monitoring and recording the types and quantities of carbon emission sources in the study area, multiplied by the corresponding emission factors to generate emission data inventories. However, the inventory-based approach has limitations in terms of low spatiotemporal resolution and accuracy. In recent years, methods based on atmospheric inversion for estimating carbon flux have been developed[4]. These inversion methods, optimized based on monitored atmospheric carbon dioxide concentrations and atmospheric transport models, offer higher spatiotemporal resolution and lower uncertainty. The inversion, conducted within a Bayesian framework, requires building covariance matrices to represent the uncertainty and correlation between the concentration observation vector and the prior inventory vector[5]. In traditional inversion methods, a fixed covariance function is typically chosen to generate the covariance matrix, and certain parameters, such as variance and correlation distance, are set based on prior inventories and empirical knowledge. This approach leads to significant uncertainty in the generated covariance matrix.

Similar issues exist in some Bayesian framework-based machine learning methods, such as Gaussian process regression. However, in the field of machine learning, clear solutions have been developed for such problems, namely determining the values of covariance function parameters through maximum marginal likelihood estimation[6]. Additionally, modern GPU parallel computing capabilities accelerate relevant matrix operations, making the algorithm feasible in terms of time costs. Hence, this paper employs maximum marginal likelihood estimation to determine the values of

[a]aakagi@qq.com, [b]reng@nim.ac.cn, [c]469600057@qq.com,
[d]dkl1439@163.com; [e]*Corresponding author's email: elinhong@nim.ac.cn;

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

parameters for the prior covariance function in the inversion algorithm. The study also tests the impact of using different covariance functions on inversion results, aiming to identify the most suitable covariance function and parameter settings for minimizing posterior flux uncertainty.

2 Inversion Methodology [7]

The relationship between the carbon dioxide flux random vector $\boldsymbol{f}$ and the observation random vector $\boldsymbol{\mu}$ can be expressed as:

$$\boldsymbol{\mu} = \boldsymbol{Hf} + \epsilon \tag{1}$$

where $\boldsymbol{H}$ is the sensitivity matrix representing the sensitivity of monitoring sites to carbon dioxide flux in the inversion area, obtained through WRF-STILT combined with atmospheric transport data reverse simulation, and ϵ is the error term representing the sum of various errors, including model transport errors and concentration observation errors.

According to the Bayes' theorem, the maximum a posteriori (MAP) estimate:

$$\boldsymbol{f} = \boldsymbol{f^b} + \boldsymbol{BH^T}(\boldsymbol{HBH^T} + \boldsymbol{R})^{(-1)}(\boldsymbol{\mu} - \boldsymbol{Hf^b}) \tag{2}$$

From the expression of the maximum a posteriori, it can be observed that the inversion algorithm can be considered as a correction to the prior flux.

3 Methods of covariance matrix estim -ation

3.1 Maximum marginal likelihood estimation

Typically, a hierarchical approach can be employed to consider Bayesian models[6]. If uncertainties exist in setting certain parameters, they can be treated as hyperparameters, which can then be determined through model selection. In the context of the inversion problem, the parameters in the likelihood function covariance matrix $\boldsymbol{R_\theta}$ and the parameters in the prior covariance matrix $\boldsymbol{B_\theta}$ are considered as hyperparameters denoted by $\boldsymbol{\theta}$.

In a hierarchical consideration of the Bayesian model, the first layer involves obtaining the first type of maximum likelihood for the flux $\boldsymbol{f}$ to be inverted. The second layer involves obtaining the second type of maximum likelihood for the hyperparameters $\boldsymbol{\theta}$. The choice of hyperparameters in the second layer can impact the distribution of the inverted flux $\boldsymbol{f}$ in the lower layer.

In the first type of maximum likelihood, the posterior expression for the parameterized flux $\boldsymbol{f}$ is given by:

$$p(\boldsymbol{f}|\boldsymbol{\mu},\boldsymbol{\theta}) = \frac{p(\boldsymbol{\mu}|\boldsymbol{f})p(\boldsymbol{f}|\boldsymbol{\theta})}{p(\boldsymbol{\mu}|\boldsymbol{\theta})} \tag{3}$$

In the second type of maximum likelihood, the p osterior expression for the hyperparameters $\boldsymbol{\theta}$ is give n by:

$$p(\boldsymbol{\theta}|\boldsymbol{\mu}) = \frac{p(\boldsymbol{\mu}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{\mu})} \tag{4}$$

We determine the values of hyperparameters $\boldsymbol{\theta}$ through maximum marginal likelihood estimation in the first type of maximum likelihood, which then acts as the likelihood function in the second type of maximum likelihood estimation.

An illustrative explanation of why maximum marginal likelihood estimation automatically balances model fitting and model selection is presented graphically. The graph depicts the marginal likelihood functions for three models of varying complexity, showcasing their distributions for all possible data under ideal conditions. A simpler model may fit only a small amount of data but, due to the integral of the marginal probability density function equating to 1, it can have a higher marginal likelihood value for the data it fits. Conversely, a more complex model, while potentially fitting more data, does not necessarily have a higher marginal likelihood value, making it less likely to be selected. This balance between data fitting and model complexity forms an equilibrium in the selection of models.

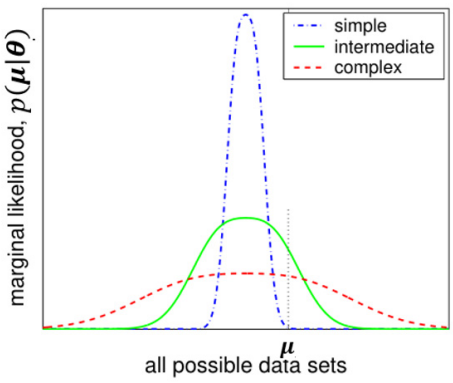


Fig. 1. model selection

The expression for the maximum marginal likelihood function can be derived, and its negative logarithm is calculated:

$$\mathcal{L}(\boldsymbol{\theta}) = \frac{1}{2} ln|\boldsymbol{D_\theta}| + \frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{Hf^b})^{\mathrm{T}}\boldsymbol{D_\theta^{-1}}(\boldsymbol{\mu} - \boldsymbol{Hf^b}) + C \tag{5}$$

Therefore, the maximum marginal likelihood estimation problem is transformed into the minimum value, that is, when the minimum value $\boldsymbol{L}(\boldsymbol{\theta})$ is taken, it corresponds to $\boldsymbol{L}(\boldsymbol{\theta})$ the optimal value of the parameter in the covariance matrix, and the concentration observation data can be considered as much as possible by inversion according to the hyperparameters obtained by the maximum marginal likelihood estimation $\boldsymbol{\theta}$, and at the same time, the inversion model is too complex and overfitting can be avoided.

Therefore, the problem of maximum marginal likelihood estimation is transformed into finding the minimum value of $\boldsymbol{L}(\boldsymbol{\theta})$. When $\boldsymbol{L}(\boldsymbol{\theta})$ is minimized, the corresponding $\boldsymbol{\theta}$ represents the optimal values for the parameters in the covariance matrix. Inversion can then be carried out based on the hyperparameters obtained from the maximum marginal likelihood estimation, considering concentration observation data as much as possible while avoiding the occurrence of overfitting due to overly complex inversion models.

3.2 Optimization problem and time cost

According to Section 3.1, the problem of hyperparameter estimation is transformed into an optimization problem, i.e., finding the parameter values when a given function takes extreme values. To address this optimization problem, the NLopt library is chosen for solving[8]. NLopt provides various optimization algorithms, and in this paper, the ISRES algorithm is selected. This algorithm is an improved random search evolutionary strategy that uses a heuristic approach to avoid local optima.

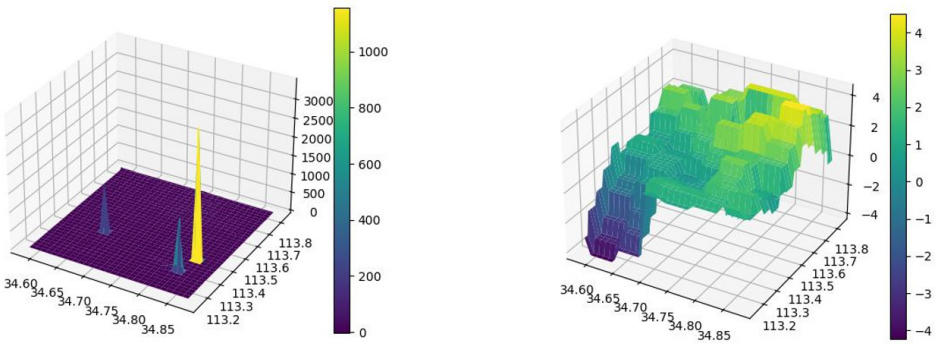
Optimization algorithms typically require multiple evaluations of the objective function when solving optimization problems. When the scale of the inversion region is large, evaluating the objective function involves large matrix operations. The computational cost on the CPU becomes prohibitively high, so leveraging the parallel computing capabilities of modern GPUs is essential to make the algorithm feasible in terms of time costs. Accelerating matrix operations using NVIDIA GPUs can be achieved through the CUDA interface provided by NVIDIA or cuBLAS (CUDA Basic Linear Algebra Subprograms library).However, these methods can be complex to program. Since machine learning also involves large matrix operations, machine learning libraries often provide pre-packaged CUDA interfaces for matrix operations. In this paper, the PyTorch machine learning library is used[9], utilizing its CUDA interface to accelerate relevant matrix operations.

4 Experiment setup

4.1 Experimental data

Sensitivity matrices $\boldsymbol{H}$ for each monitoring site are generated using WRF-STILT combined with atmospheric transport data[10]. Based on the true emission inventory $\boldsymbol{f}$ and the sensitivity matrices $\boldsymbol{H}$ for each monitoring site, along with a randomly generated error term, concentration observation data for each monitoring site are generated.

Since the purpose of this study is to find and verify the estimation method of the prior covariance matrix in atmospheric inversion, the real emission inventory and the prior inventory with large differences in values are constructed by using an ideal experimental method, real emissions and prior emissions as shown in Figure 2, the unit is μmol·m^{-2}·s^{-1}.



(a)Real emissions (b)Prior emissions

Fig. 2. true emissions and prior emissions

4.2 Prior covariance function

To test the impact of covariance matrices generated using different covariance functions on inversion results, the covariance function used for constructing the observation vector covariance matrix $\boldsymbol{R}$ is fixed to the gamma-exponential covariance function. The covariance functions used for constructing the prior flux covariance matrix $\boldsymbol{B}$ are listed in Table 1:

Table 1. Covariance Functions

Covariance function	expression
Exponential covariance function	$Cov = \sigma^2 e^{-\frac{r}{l}}$
Gamma exponential covariance function	$Cov = \sigma^2 e^{-\left(\frac{r}{l}\right)^{\gamma}}$
Balgovind covariance function	$Cov = \sigma^2\left(1+\frac{r}{l}\right)$
Square exponential covariance function	$Cov = \sigma^2 e^{-\frac{r^2}{2l^2}}$

5 Results

5.1 Covariance function selection results

Different covariance functions are tested, and inversion is performed based on the hyperparameter values obtained from maximum marginal likelihood estimation. Since this is an ideal experiment, the root mean square error (RMSE) between the posterior flux and the true flux inventory is calculated, as shown in Table 2:

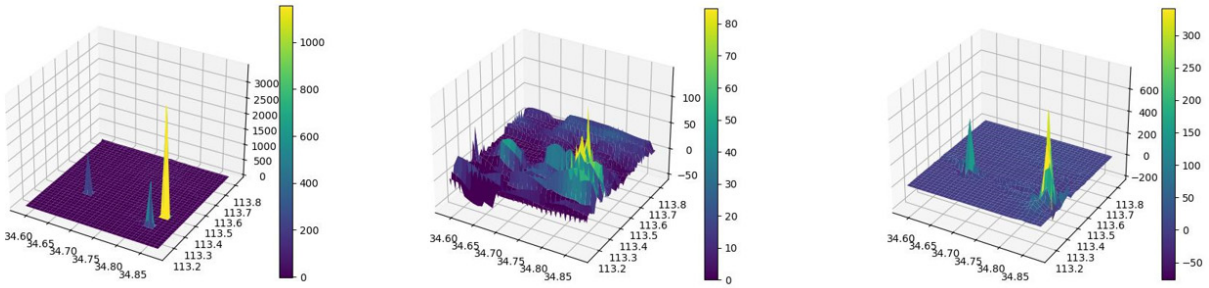
Table 2. Experimental results

Prior flux covariance matrix $\boldsymbol{B}$	RMSE/μmol·m^{-2}·s^{-1}
Balgovind covariance function	97.6086
Exponential covariance function	103.618
Gamma exponential covariance function	110.271
Square exponential covariance function	98.2326

The experimental results indicate that when the Balgovind covariance function is used to generate the prior flux covariance matrix ***B***, the RMSE of the inverted posterior flux is minimized, with the minimum RMSE value being 97.6086.

5.2 Optimization effect of inversion results

Two sets of inversion experiments are conducted using the same observation data, sensitivity matrix, and prior inventory. One set uses the Balgovind covariance function to generate the prior flux covariance matrix ***B*** for inversion, while the other set uses an empirically estimated prior covariance matrix. The corresponding posterior flux inventories are obtained, and the results are shown in the figures. Figure 3(a) represents the true flux inventory, Figure 3(b) shows the posterior inventory corresponding to inversion using an empirically estimated prior covariance matrix, and Figure 3(c) shows the posterior inventory corresponding to inversion using the prior covariance matrix obtained through maximum marginal likelihood estimation. It can be observed that the inversion based on the covariance function obtained through maximum marginal likelihood estimation results in a smaller RMSE and more accurate estimates for high-value points of emission sources.



(a) Real emissions (b) RMSE= 118.701 (c) RMSE= 97.6086

Fig. 3. comparison of experimental results

6 Conclusion

This study utilized maximum marginal likelihood estimation to determine the parameter values in the covariance function of the atmospheric inversion algorithm. Additionally, the use of GPU acceleration for matrix operations made the algorithm feasible in terms of time costs. The inversion effects were tested by employing different covariance functions to generate the prior covariance matrix. The final results indicate that when the Balgovind covariance function is used for the prior flux covariance matrix ***B***, the inversion results are optimal, exhibiting the smallest RMSE values and more accurate estimates for high-value points.

The research presented a universal approach for determining parameter values in the covariance function when using atmospheric carbon dioxide concentration observations and atmospheric transport models to invert carbon dioxide flux. The feasibility of this method was verified through an ideal experiment. Furthermore, the study identified the most suitable covariance function for generating the prior covariance matrix in this experiment. The research area was limited to a specific urban region, and future experiments could consider using larger urban areas for validation.

Overall, the study contributes to the understanding of parameter estimation in covariance functions for atmospheric inversion algorithms and provides a practical approach for determining these parameters in the context of carbon dioxide flux inversion.

Acknowledgments

This study was supported by the Fundamental Research Funds for Central Public Welfare Scientific Research Institutes sponsored by the National Institute of Metrology, P.R. China [Grant No. AKYZD2207-2] and the Fundamental Research Funds for Central Public Welfare Scientific Research Institutes sponsored by National Institute of Metrology, P.R. China [Grant No. AKYZZ2307]

References

1. Solomon, S., Plattner, G.K., Knutti, R., Friedlingstein, P.(2009) Irreversible climate change due to carbon dioxide emissions. Proceedings of the National Academy of Sciences.,106:1704-1709. https://doi.org/10.1073/pnas.0812721106
2. Etheridge, D. M., Steele, L. P., Langenfelds, R. L., Francey, R. J., Barnola, J.M., Morgan, V. I.(1996) Natural and anthropogenic changes in atmospheric CO2 over the last 1000 years from air in Antarctic ice and firn. Journal of Geophysical Research.,101 (D2): 4115–4128. https://doi.org/10.1029/95JD03410
3. Satterthwaite, D. (2008). Cities' contribution to global warming: notes on the allocation of greenhouse gas emissions. Environment and Urbanization., 20(2): 539-549.https://doi.org/10.1177/0956247808096127

4. Lauvaux T., et al.(2016)High-resolution atmospheric inversion of urban CO2 emissions during the dormant season of the Indianapolis Flux Experiment (INFLUX), J. Geophys.Res. Atmos., 121:5213–5236.doi:10.1002/2015JD024473.
5. Wu, L., Bocquet, M., Chevallier, F., Lauvaux, T., Davis, K. (2013)Hyperparameter estimation for uncertainty quantification in mesoscale carbon dioxide inversions. Tellus B.,65(1):p.20894. https://doi.org/10.3402/tellusb.v65i0.20894
6. Rasmussen, C.E., Williams, C.K.I.(2006) Model Selection and Adaptation of Hyperparameters. In: Rasmussen, C.E.(Eds.), Gaussian Processes for Machine Learning.the MIT Press, Cambridge.pp.105-128.
7. Tarantola, A.(2004) The Least-Squares Criterion. In: Mosegaard ,K.(Eds.), Inverse problem theory and methods for model parameter estimation. SIAM, Philadelphia.pp.57-75.
8. Johnson,S.G.The NLopt nonlinear-optimization package.http://github.com/stevengj/nlopt.
9. Paszke, A. et al., (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. In Advances in Neural Information Processing Systems 32.http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.
10. Lin, J.C., Gerbig, C., Wofsy, S. C., Andrews, A. E., Daube, B. C., Davis, K. J., and Grainger, C. A.(2003) A near-field tool for simulating the upstream influence of atmospheric observations: The Stochastic Time-Inverted Lagrangian Transport (STILT) model. Journal of Geophysical Research-Atmospheres., 108: 4493.doi:10.1029/2002JD003161.