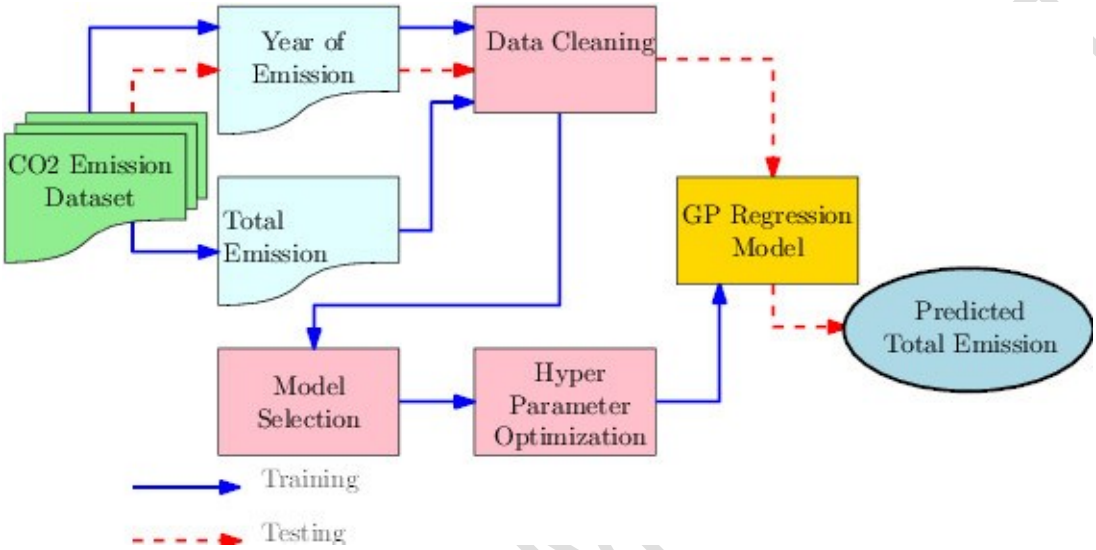
FOSSIL FUEL EMISSION PREDICTION SYSTEM USING MULTIPLE KERNEL GAUSSIAN PROCESS

Angel Deborah S[1], Rajalakshmi Sivanaiah[2*], S Milton Rajendram[3], Mirnalinee T T[4]
[1, 2] Assistant Professor, [3, 4] Professor
Department of Computer Science and Engineering,
Sri Sivasubramaniya Nadar College of Engineering, Chennai - 603110, Tamil Nadu, India
[*]Corresponding author: *rajalakshmis@ssn.edu.in*

GRAPHICAL ABSTRACT



ABSTRACT:

According to the United States Environmental Protection Agency, 81% of greenhouse gas emission is due to Carbon-dioxide. When fossil fuels, solid waste, trees, and wood products are burnt, Carbon-dioxide emission occurs. The concentration of Carbon-dioxide in the atmosphere is reduced when it is absorbed by plants as a part of the biological carbon cycle. The major sources of CO_2 emission are from fossil fuel such as coal, natural gas, oil, cement production, gas flares used in industrial plants and bunker fuels used in ships. Increase in CO_2 emission leads to the increase in global warming. Climate change, change in seasonal events and decrease in agricultural productivity are the major impacts of global warming. Hence it is important to study and analyse CO_2 emission based on the recorded data across the globe. Further the major source of emission must be detected, and alert messages must be sent to the pollution control boards in various countries to take necessary remedial actions. This work focuses on predicting the CO_2 emission in the near future in various countries based upon the "Fossil-Fuel CO_2 Emissions by Nation" dataset recorded by Carbon Dioxide Information Analysis Center (CDIAC), Oak Ridge National Laboratory. This system comprises of data cleaning, data normalization, optimization, and model building. The model built using Multiple Kernel Gaussian Process (MKGP) will predict the concentration of CO_2 that may be present in the atmosphere in upcoming years. Based on the prediction the major source of CO_2 emission is identified. We have studied the effects of Radial Basis Function kernel, Rational Quadratic, Periodic kernel and combinations of the kernels on India USA and China

dataset. We have proposed a Multi-Kernel Gaussian Process for predicting the fossil fuel emission (MKGP - FFE). It has been inferred that combination of kernels performed well when compared to individual kernels in most of the cases.

Keywords: Fossil fuel, CO_2 emission, global warming, CDIAC, data cleaning, data normalization, model building, machine learning techniques.

1. Introduction:

Global warming and the effects of fossil fuel emission are considered among the critical issues in the fields of science. This has raised the concern regarding the contributions of carbon dioxide (CO2) to global warming [1]. The United States Environmental Protection Agency pointed out that more than 81% of greenhouse gases are probably caused by the emission of Carbon-dioxide. The emission of Carbon-dioxide occurs when fossil fuels, solid waste, trees, and wood products are burnt. This emission has caused a negative impact across the globe. One of the major impacts is an increase in temperature, which has started affecting health in several societies across the world. Weather conditions are extremely increasing, in particular heat waves, floods, and storms, which results in an increasing loss of human life and injuries due to natural disasters caused by climate change. This increase in carbon dioxide affects determinants of health, such as quality and quantity of foods, water resources, and ecological disease control vector [2]. Energy consumption is considered as the main source of greenhouse emissions [3]. From the period 1970–2010 energy consumption for the Organization of the Petroleum Exporting Countries (OPEC) has risen by 685%, whereas the emissions of CO2 have risen by 440% as a result of burning fossil fuels within that period. Thus, there is a drastic increase in energy consumption and CO2 emissions of the OPEC countries.

The CO2 emission caused by burning of fossil fuels has raised the global temperature [4, 25]. The OPEC countries have contributed more to the world CO2 emissions in 2010. Such a contribution can significantly affect the use of energy in the future by OPEC countries. Global warming is one of the critical issues currently faced by the world. As global warming has become a serious threat, the policy makers and government officials throughout the globe are striving in creating a new template which takes into consideration energy conservation and reduction in emission of greenhouse gases [5]. In fact, by reducing greenhouse gas emission, global warming can be reduced. Reducing greenhouse gas emission, in particular CO2 emission, requires an accurate fossil fuel emission prediction system. In turn this can help the policy makers and officials to closely monitor and control fossil fuel emission [6]. The change of climate policies and providing a pointer for alternative energy sources requires a proper CO2 prediction system [7]. The formation of preventive measures for reducing CO2 emissions has encouraged past researchers to apply computational intelligent algorithms due to their supremacy over formal logic, mathematical programming [8], and statistical methods [9] for predicting the CO2 emission.

Our primary aim is to apply the Gaussian Process Multi-Kernel methods to predict the CO2 emission. As the dataset is small, most of the deep learning techniques will not perform well.

Whereas Multi-Kernel Gaussian Process performs better in a small dataset when compared to other machine learning and deep learning techniques. In this work we have analysed the performance of various kernels separately and combining 2 or more kernels together (multi-kernel) to understand it's working.

2. Related works:

Inspite of the drawbacks of the traditional methods, a non-homogeneous exponential equation and a linear equation was used by Meng et al. [10] to construct a model for the prediction of CO2 emissions related to energy related in China. Chen and Wang [11] used a hybrid of Fuzzy Regression and Back Propagation Neural Network (FRBPNN) to forecast the global concentration of CO2 thereby avoiding limitations of the traditional methods. This hybrid method was found to improve the accuracy of CO2 forecasting. Chen [6] applied a collaborative fuzzy neural network to increase the forecasting accuracy of the FRBPNN. Results obtained show that the collaborative fuzzy neural network performs better than the FRBPNN and statistical methods in the forecasting of global CO2. Bao and Hui [12] used the Grey model to construct a model for the forecasting of CO2 emissions in Shijiazhuang, China. The model constructed was used to project the CO2 emissions of Shijiazhuang from 2010 to 2020. In yet another study, the emission of CO2 related to energy in developing countries was predicted using the Grey model [13]. It was found that the Grey model was not effective with a large sample of data; it only requires small samples of observations to be robust [14]. Also, it was observed that the model lacks fitting ability and has a defect in nonlinear modelling [15]. This encouraged Tan and Zhang [15] to apply GA to enhance fitting ability of the Grey model and combined the genetic algorithm (GA) fitted Grey model into BPNN for enhancing its nonlinear approximation ability. The model was applied to predict energy load with enhanced performance. However, the BPNN is a gradient based algorithm that has the chance of being trapped in local minima; it has slow convergence rate, and the algorithm is highly dependent on parameter settings and it generates complex error surfaces with a many local minimum [16–17].

Fuzzy systems are short of the capability of learning input data; also, human language is used to represent the input and output of the systems. Hence, incomplete, or wrong rules cannot be handled finely by fuzzy systems. The GA eradicates previous knowledge of the problem if the population changes [18] and requires multiple parameter settings that weaken its robustness [19]. Studies on the forecasting of OPEC CO2 emission from petroleum consumption are insufficient in the literature, regardless of the increasing consumption of petroleum and emissions of CO2 by the OPEC countries. Cao uses component prediction method for flue gas of natural gas combustion based on Nonlinear Partial Least Squares Method. Nonlinear partial least squares (NPLS) with extended input could solve the nonlinear problems effectively [20]. The NPLS method with extended input based on Radial Basis Function Neural Network (RBFNN) is used for component prediction of flue gas. A near-infrared spectral dataset of flue gas of natural gas combustion is used for estimating the effectiveness of the NPLS method. The NPLS method with RBFNN is computationally inefficient and suffers from over-fitting.

Abbas applied the Adaptive Neuro-Fuzzy Inference System (ANFIS) model to predict CO2 emissions based on important input indicators like energy consumption and economic

growth [21]. The fuzzy rules through ANFIS were used to generalize the relationships of the input and output indicators in order to make a prediction of CO2 emissions. This was also found to be computationally inefficient and overfitting. Wilson and Adams introduced simple closed form kernels that can be used with Gaussian processes to discover patterns and enable extrapolation [22]. Importance of machine learning and deep learning algorithms in CO2 emissions are discussed in [23-24]. Miyazaki and Bowman used modified Environmental Kuznets Curve to postulate the coevolution of fossil fuel CO2 (FFCO2) and NOx emissions using Kalman filter and analysed the relationship between NOx and FFCO2 [26]. Ahmadi et.al. used artificial neural network approach Group Method of Data Handling (GMDH) to predict the CO2 emission based on the consumption of fossil fuels like coal, oil, natural gas and renewable energy sources [27]. Drawbacks of the previous studies and lack of work on the forecasting of OPEC CO2 emission from petroleum consumptions stirred the present research.

3. Gaussian process:

A Gaussian process (GP) is a collection of random variables, any infinite number of which have a joint Gaussian distribution. Using a Gaussian process, we can define a distribution over functions $f(x)$ *f(x)* in Equation 1

$f(x) \sim GP\big(m(x), k(x,x')\big)$(1)

where $x \varepsilon R^P$ is an arbitrary input variable, and the mean function $m(x)$ $k(x,x')$ as in Equation 3 are defined as $m(x) = E[f(x)]$

(2)

$$k(x,x') = cov(f(x), f(x')) \quad (3)$$

Any collection of function values has a joint Gaussian distribution denoted by Equation 4

$$[f(x_1), f(x_2), \ldots, f(x_N)]^T \sim N(\mu, K)) \quad (4)$$

where the *N* x *N* covariance matrix *K* has entries $K_{ij} = k(x_i, x_j)$, and the mean μ has entries $\mu_i = m(x_i)$. The properties of the functions- smoothness, periodicity, etc.. are determined by the kernel function. Mean function, μ(x), usually defined to be zero. Justified by manipulating the data. Covariance function(kernel), *k(x, x)* defines the prior properties of the functions considered for inference. The properties include stationarity and smoothness. Stationary process is a stochastic process whose joint probability distribution does not change when shifted in time. Consequently, parameters such as the mean and variance, if they are present, also do not change over time. In probability theory and statistics, smoothness of a density function is a measure which determines how many times the density function can be differentiated. Statistical parameters of the process do not change with time.

3.1. Covariance Kernel Functions

The heart of every Gaussian process model is a covariance kernel. The kernel k explicitly specifies the covariance between a pair of random function values at a pair of input points: *k(x,x') = cov(f(x),f(x'))*. The particular choice of covariance function determines the properties of sample functions drawn from the GP prior (e.g. smoothness, length scales,

amplitude etc). Therefore, it is an important part of GP modelling to select an appropriate covariance function for a particular problem. The are several types of covariance functions it includes the following:

3.1.1. Radial Basis Function Kernel (RBF)

The Radial Basis Function (RBF) kernel, also sometimes called the Gaussian, squared exponential function, or exponentiated quadratic kernel, is probably the most widely used kernel. The RBF kernel has become the default kernel for GPs. It has some nice properties like very smooth. It is universal and has a clear connection with Gaussian density. Gaussian process with an RBF kernel has the form as given in Equation 5,

$$k_{RBF}(x, x') = \sigma^2 exp(-(x - x')^2/2l^2) \quad (5)$$

To model the long term smooth rising trend, we use an RBF covariance term. It has two hyperparameters controlling the amplitude σ and characteristic length-scale l.

3.1.2. Rational Quadratic (RQ)

The squared exponential kernel assumes that the data are only varying at one particular length-scale. In different mechanisms underlying the data could be varying on different scales. It may often be unsure about the scales over which data are varying and indeed we may wish to account for infinitely many scales. The rational quadratic (RQ) kernel is a scale mixture (infinite sum) of squared exponential kernels with different length-scales. The RQ kernel is intended to model multi-scale data. This kernel is equivalent to adding together many SE kernels with different length scales. So, GP priors with this kernel expect to see functions which vary smoothly across many length scales. The rational quadratic (RQ) covariance function is mentioned in Equation 6

$$k_{RQ}(x, x') = \sigma^2(1 + (x - x')^2/2l^2)^{-\alpha} \quad (6)$$

It includes three hyperparameters values, magnitude σ, length scale l and shape parameter α. The RQ kernel is a scale mixture of squared exponential kernels with different length-scales. To model the medium-term irregularities a rational quadratic term is used.

3.1.3. Periodic (PE)

This kernel is mostly useful in combination with other covariance functions. For example, this kernel is used in combination with squared exponential (SE) and rational quadratic (RQ) kernels. This kernel gives rise to periodic functions. The Periodic Covariance function is given in Equation 7,

$$k(x, x') = \sigma^2 exp(-((x - x')^2/2b^2) - (-2sin^2((x - x'))/h^2)) \quad (7)$$

Periodic covariance function can be used with a period of one year to model the seasonal variation. It includes three hyperparameters values, magnitude *l,* decay-time for the periodic component *b* and smoothness of the periodic component *h*.

3.1.4. Matern (MA)

The Matern kernel is the second most popular kernel, after the squared exponential or RBF kernel. Smoothness properties of the squared exponential kernel is unrealistic and recommends the Matern kernel as an alternative. The kernel is given in Equation $8k(xi, xj) = 1/\Gamma(v)2v - 1\left(\left(\sqrt{2}v/l\right)d(xi, xj)\right)vKv\left(\left(\sqrt{2}v/l\right)d(xi, xj)\right)$ (8)

where $d(.,.)$ is the Euclidean distance, $K_v(.)$ is a modified Bessel function and $\Gamma(.)$ is the gamma function. It has an additional parameter which controls the smoothness of the resulting function. The smaller, the less smooth the approximated function is.

3.1.5. Spectral Mixture (SM)

It is closed form kernels for automatic pattern discovery and extrapolation. These spectral mixture (SM) kernels are derived by modelling the spectral density of the kernel (its Fourier transform) using a scale-location Gaussian mixture. SM kernels form a basis for all stationary covariances and can be used as a drop-in replacement for standard kernels, as they retain simple and exact learning and inference procedures.

4. Proposed System:

The architecture diagram of the proposed system Multi-Kernel Gaussian Process for fossil fuel emission (MKGP - FFE) to predict the CO2 emission in the near future in various countries based upon the Fossil-Fuel CO2 Emissions by Nation dataset[1] recorded by Carbon Dioxide Information Analysis Center (CDIAC), Oak Ridge National Laboratory is given in Figure 1. The system comprises data collection, data cleaning, model Selection, hyperparameter optimization, Cholesky decomposition and prediction stages. The data is collected from Carbon Dioxide Information Analysis Center (CDIAC), Oak Ridge National Laboratory. The features of the data are given in Table 1. The collected data is cleaned to eliminate all the missing values and erroneous values. The data is structured and given as input to train a machine learning model in the model selection stage. Model selection is the process of setting hyperparameters to the covariance functions of the Gaussian Process (GP). It is used to refine the predictions and give a better interpretation about the properties of the given data. The families of covariance kernel functions are, Radial Basis Function (RBF), Matern (MA), Rational Quadratic (RQ), Periodic (PE), etc. Each of these families have a number of free parameters, these values are determined using an optimization algorithm

[1] https://www.osti.gov/biblio/1394401

during the hyper parameter optimization stage. One such multivariate optimization algorithm is a Scaled Conjugate Gradient (SCG) algorithm. SCG is fully automated, it includes no

critical user-dependent parameters and thereby avoids a timeconsuming line search. GP makes useof inverse matrix computation which is computationally intense. One of the approaches to reduce the computation time to compute an inverse matrix is Cholesky's decomposition. The Cholesky decomposition of a symmetric, positive-definite matrix decomposes matrix **A** into the product of its lower triangular matrix **L** and its transpose. It is represented as shown in Equation 9,

$$A = LL^T \tag{9}$$

where L is called the Cholesky factor of A. Cholesky decomposition is done because it is faster and numerically stable when compared to directly inverting the matrix.

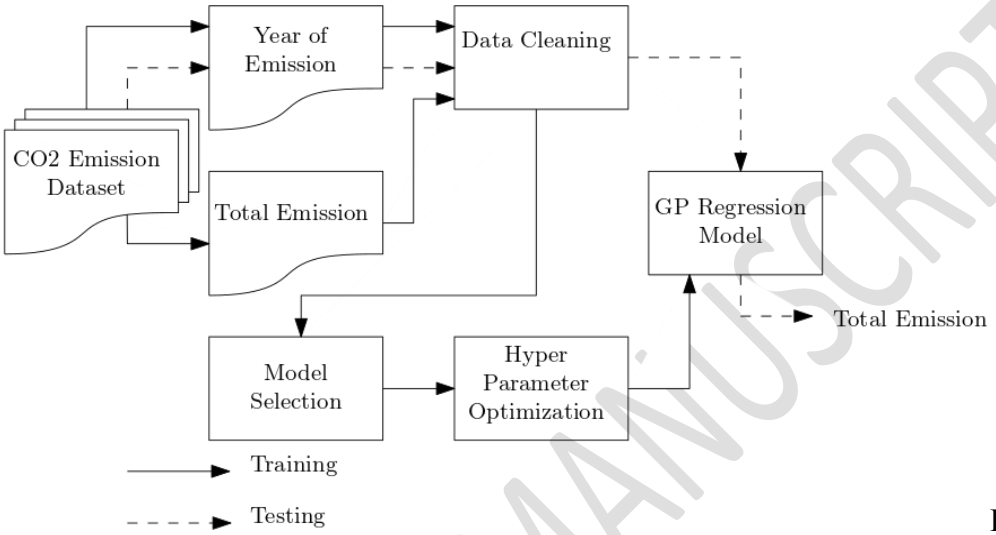


Figure 1.

Overall System Architecture Diagram

Once the hyperparameters of GP are optimized during hyper parameter optimization and Cholesky's decomposition, the learnt model is generated with multiple kernels selected during model selection stage. The learnt Multiple Kernel Gaussian Process (MKGP - FFE) model is used to predict the per capita CO2 emissions.

Table 1. Features of Fossil-Fuel CO2 Emissions Dataset

Nation	Emissions from liquid fuel consumption
Year	Emissions from gas fuel consumption
Total CO2 emissions from fossil-fuels and cement production	Emissions from cement production
Emissions from solid fuel consumption	Emissions from gas flaring
Emissions from bunker fuels	Per capita CO2 emissions (metric tons of carbon)

5. 5. Implementation:

The steps to implement MKGP - FFE regression are given in algorithm 1. Based on algorithm 1, algorithm 2 is derived to build the learnt MKGP model to predict the total emissions. Algorithm 1 takes the training inputs, training target, kernel functions, noise level and test data as input, based on the processing it produces predictive mean, variance, and log-marginal likelihood as output. The training inputs include features like Year, total CO2 emissions from fossil-fuels and cement production, emissions from solid fuel consumption, emissions from liquid fuel consumption, emissions from gas fuel consumption, emissions from cement production, emissions from gas flaring and emissions from bunker fuels. According to requirements kernel parameters are initialized. The algorithm 1 performs computation of kernel matrix, inverse matrix using Cholesky decomposition, predictive mean, predictive variance and log-marginal likelihood in a sequential way. Algorithm 2 is responsible for finding the optimal hyperparameters and thereby formulating the learnt MKGP model. It works based on the concept of maximizing the log marginal likelihood. The predictive mean of the MKGP model is the predicted per capita CO2 emission and the predictive variance of the MKGP is the uncertainty in the predictive per capita CO2 emission.**Algorithm 1. Multiple Kernel Gaussian Process (MKGP) Regression**

Input: X (training inputs) , $\mathbf{y}$ (training targets), k_{RBF} (RBF kernel function),k_{RQ} (RQ kernel function),k_{PE} (PE kernel function), σ_n^2 (noise level), $\mathbf{x}_*$ (test input)
Output: Predictive mean, variance and log-marginal likelihood
$\mathbf{k_{new}}(x, x') = \mathbf{k_{RBF}}(x, x') + \mathbf{k_{RQ}}(x, x') + \mathbf{k_{PE}}(x, x')$// Fixed Rule linear MKL
$K = matrix(\mathbf{k_{new}}(x, x'))$// Compute Kernel matrix
$L = \text{Cholesky}(K + \sigma_n^2 I)$ // Inverse of kernel matrix using Cholesky decomposition
$\alpha = L^T \backslash (L \backslash \mathbf{y})$ // Compute Predictive Mean
$\overline{f}_* = \mathbf{k}_*{}^T \alpha$
$\mathbf{v} = L \backslash \mathbf{k}_*$ // Compute Predictive Variance
$V[f_*] = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{v}^T \mathbf{v}$
$\log p(\mathbf{y}|X) = -\frac{1}{2}\mathbf{y}^T \alpha - \sum_i \log L_{ii} - \frac{n}{2} \log 2\pi$ // Compute log marginal likelihood
return f_* (mean), $V[f_*]$ (variance), $\log p(\mathbf{y}|X)$ (log-marginal likelihood)

Algorithm 2. MKGP for Emission Prediction

Input: Feature Vectors(x), Target Vectors
Output: Built n-kernel GP Regression model

$\mathbf{k_{new}}(x, x') \leftarrow \mathbf{k_{RBF}}(x, x') + \mathbf{k_{RQ}}(x, x') + \mathbf{k_{PE}}(x, x')$
Initialize GP regression model using $\mathbf{k_{new}}(x, x')$ (new-kernel) function, Feature Vectors and Target Vectors
/* Initialize hyperparameters: σ^2(Variance); l (Lengthscale) ; σ_w^2 (vector of the variances of the prior over input weights);σ_b^2 (variance of the prior over bias) */
$\sigma^2 \leftarrow$ Initialize variance value
$\sigma_w^2 \leftarrow$ Initialze vector of the variances of the prior over input weights
$\sigma_b^2 \leftarrow$ Initialize variance of the prior over bias
$l \leftarrow$ Initialize lengthscale value
while *log marginal likelihood* $\neq$ *maximum* **do**
  Update σ^2, l, σ_w^2, σ_b^2 using Optimizer
end
return σ^2,l,σ_w^2, σ_b^2, Built n-kernel GP regression model

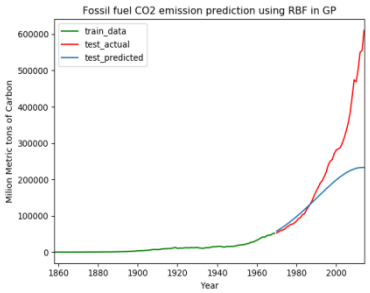
6. Results and Discussion:

The proposed system (MKGP-FFE) uses the data collected from Carbon Dioxide Information Analysis Center (CDIAC), Oak Ridge National Laboratory. There are 265 samples available in the training dataset and 120 samples available in the test dataset. The quality of the data is ensured by using appropriate data preparation techniques. Based on the Exploratory data analysis, particular kernels are chosen to build the MKGP model for three different countries

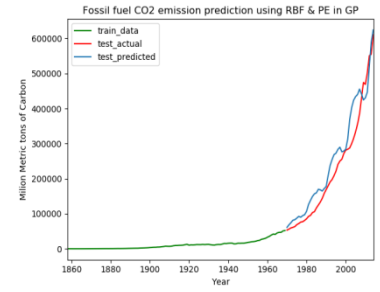
namely India, USA, and China. As data collected for different countries adheres to various patterns, different kernel functions are chosen to build an accurate GP model for the respective patterns. In some cases, the accuracy can be achieved using a single kernel, whereas in most cases the multiple kernel learning works better. The proposed work uses a fixed rule linear combination approach to perform multiple kernel learning. Based on the choice of the kernels, the parameter of the kernel varies. The inverse of the new kernel is found using Cholesky decomposition. The parameters are learnt using the Limited-Memory Broyden–Fletcher–Goldfarb–Shanno (LBFGS) algorithm.

From algorithm 2 it is clear that the optimal values of the parameter are found by maximizing the likelihood. Using the optimal values of parameter the computed mean and computed variance are found as in algorithm 1. The predictive mean of the MKGP model is the predicted per capita CO2 emission and the predictive variance of the MKGP is the uncertainty in the predictive per capita CO2 emission. The visual representation of the predicted and actual values is shown in Figure 2.
From the figure 2, it is observed that in the graphs based on India data, the test_predicted of RBF+PE+RQ is closer to test_actual whereas in the graphs based on USA data, the test_predicted of RBF+RQ is closer to test_actual. In case of China data, the test_predicted of a single kernel RBF is closer to test_actual.

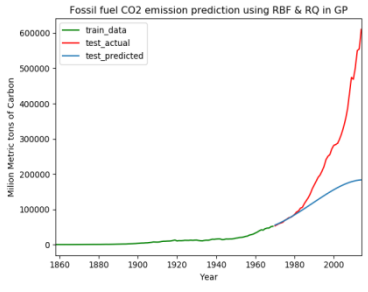
We have used the metrics like correlation coefficient *(r)* and log likelihood *(L)* to measure the performance of the Kernel methods in Gaussian Process. The formula to calculate r and L are given in Equations 10 and 11 respectively. Correction coefficient (r) is used to find the relationship between the parameters in the dataset. This is used to identify the level of correlation. If the r value is positive, then the parameters are positively correlated and if it is a negative value, then the parameters are negatively correlated. If the r value is 0, then there is no correlation between the parameters. It gives the dependency and contribution of the parameters or attributes in predicting the response or output variable. Log likelihood function (L) is an important measure used in point estimation of the response parameters in statistical models. Since we are predicting the CO2 emission value, we have used the log likelihood function as a metric to evaluate the model.



India



India



India

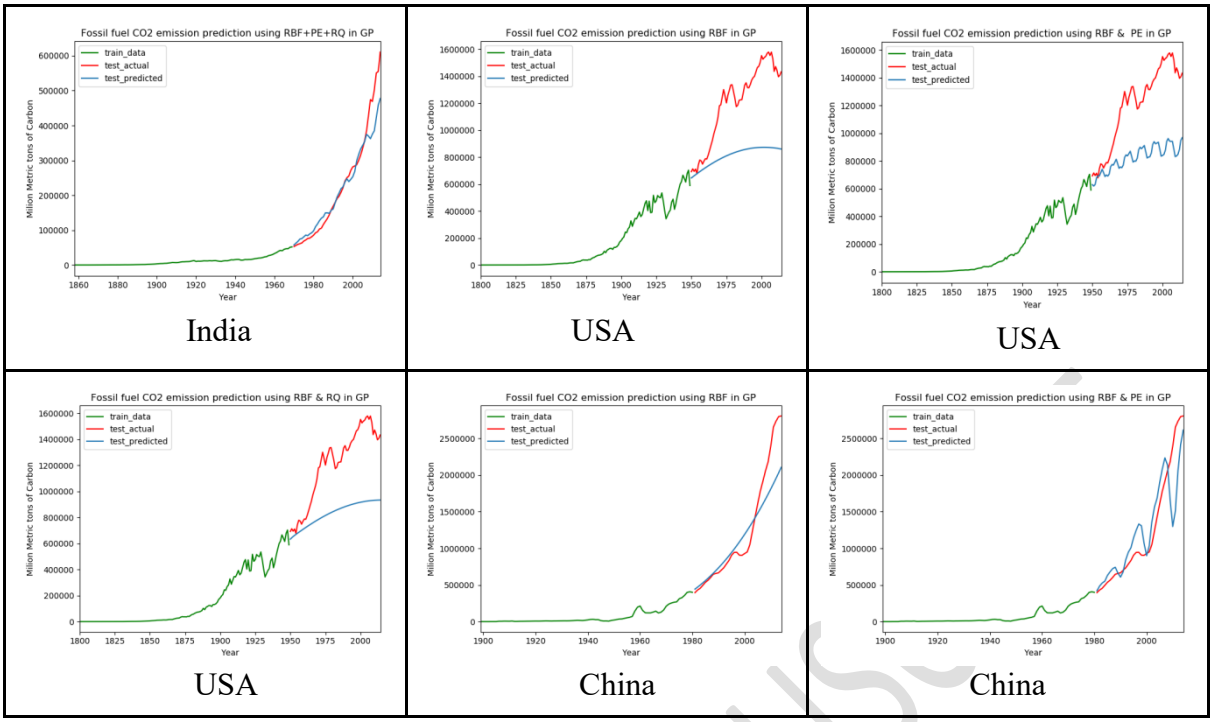


Figure2. Fossil Fuel CO2 emission prediction using different combinations of Single Kernel and Multiple Kernels in GP for India, USA, and China

The correlation coefficient and Log likelihood values obtained for various combinations of kernel functions using data of India, USA and China are given in the Tables 2 to 4.

$$r_{xy} = \frac{n\sum_{i=1}^{n}(xy) - \sum x \sum y}{\sqrt{(n\sum x^2 - (\sum x)^2)\,(n\sum y^2 - (\sum \quad y)^2)}} \quad (10)$$

$$L = \sum_{i=1}^{n} \ln f_i\,(y_i \mid \theta) \quad (11)$$

Table 2. Performance Evaluation for India

Metric	RBF	RBF+PE	RBF+RQ	RBF+PE+RQ
r	0.9216	0.9744	0.9263	0.9823
L	-102.73e6	-607.64e5	-504.23e5	-104.58e5

In Table 2 derived from India data, the results obtained show that the combination of all three kernels (RBF+RQ+PE) has better performance with highest correlation coefficient and highest log likelihood. Table3 derived based on the USA data, the results obtained show that the combination of RBF and RQ has better performance with the highest correlation coefficient.

Table 3. Performance Evaluation for USA

Metric	Two RBF	RBF+PE	RBF+RQ

r	0.9603	0.8749	0.9656
L	-138.92e5	-137.55e5	-45.02e5

Likewise table 4 for China, the results obtained shows that RBF has better performance with highest correlation coefficient and highest log likelihood.

Table 4. Performance Evaluation for China

Metric	**RBF**	**RBF+PE**
r	0.9634	0.8549
L	-453.47e7	-341.19e8

7. Conclusion and Future work:

In this work we have analyzed the importance of kernel methods in Gaussian process for small datasets. In real time scenario it is practically not possible to collect more critical data and so less amount of data is only available. Data augmentation techniques can be done to generate new data. Again, this may not provide good suggestions for critical data. Here Gaussian process was studied, and prediction was done on Fossil fuel emission data set using MKGP – FFE with the help of Radial Basis Function kernel, Rational Quadratic, Periodic kernel, and combination of kernels. The multiple kernel learning in GP has performed better for data from India and USA, whereas the single kernel itself has performed better for data from China. It has been found that RBF has been good enough in learning the long-term smoothness in data. The PE kernel has been proper in identifying the periodical variation exhibited by the data. Similar RQ kernel has contributed in identifying the haphazard behaviour in the data. When the data exhibited many properties like smoothness, periodicity and haphazardness, the linear combination of all the three kernels gave the best result. In future, problems can be further enhanced by using several other multiple kernel learning approaches to provide more generalized output.

8. References:

1. Wang, S. S., Zhou, D. Q., Zhou, P., & Wang, Q. W. (2011). CO2 emissions, energy consumption and economic growth in China: A panel data analysis. *Energy Policy*, *39*(9), 4870-4875.
2. Richardson, K., Steffen, W., Schellnhuber, H. J., Alcamo, J., Barker, T., Kammen, D. M., ... & Stern, N. (2009). *Synthesis Report. Climate change: global risks, challenges and decisions, Copenhagen, Denmark, 10-12 March, 2009*. University of Copenhagen.
3. Sari, R., & Soytas, U. (2009). Are global warming and economic growth compatible? Evidence from five OPEC countries?. *Applied Energy*, *86*(10), 1887-1893.
4. Mondor, E., & Tremblay, M. (2010). Global atmospheric change and animal populations. *Nature Education Knowledge*, *3*(2010), 23-36.
5. Adetutu, M. O. (2014). Energy efficiency and capital-energy substitutability: Evidence from four OPEC countries. *Applied Energy*, *119*, 363-370.

6. Chen, T. (2012). A collaborative fuzzy-neural system for global CO2 concentration forecasting. *International Journal of Innovative Computing, Information and Control*, *8*(11), 7679-7696.
7. Lin, C. S., Liou, F. M., & Huang, C. P. (2011). Grey forecasting model for CO2 emissions: A Taiwan study. *Applied Energy*, *88*(11), 3816-3820.
8. Sahab, M. G., Toropov, V. V., & Gandomi, A. H. (2013). A review on traditional and modern structural optimization: problems and techniques. *Metaheuristic applications in structures and infrastructures*, 25-47.
9. Yu, L., Wang, S., & Lai, K. K. (2008, October). A generalized Intelligent-agent-based fuzzy group forecasting model for oil price prediction. In *2008 IEEE International Conference on Systems, Man and Cybernetics* (pp. 489-493). IEEE.
10. Meng, M., Niu, D., & Shang, W. (2014). A small-sample hybrid model for forecasting energy-related CO2 emissions. *Energy*, *64*, 673-677.
11. Chen, T., & Wang, Y. C. (2011). A fuzzy-neural approach for global CO2 concentration forecasting. *Intelligent Data Analysis*, *15*(5), 763-777.
12. Bao, G. L., & Hui, H. Q. (2012). Grey forecasting model for CO2 emissions. In *Advanced Materials Research* (Vol. 518, pp. 1664-1668). Trans Tech Publications.
13. Meng, M., Niu, D., & Shang, W. (2014). A small-sample hybrid model for forecasting energy-related CO2 emissions. *Energy*, *64*, 673-677.
14. Zhang, Q. M., Chen, N., & Wang, F. (2009). Combined forecast model of gas load based on grey theory. In *Fuzzy Information and Engineering Volume 2* (pp. 1275-1281). Springer, Berlin, Heidelberg.
15. Tang, N., & Zhang, D. J. (2011). Application of a load forecasting model based on improved grey neural network in the smart grid. *Energy Procedia*, *12*, 180-184.
16. Zweiri, Y. H.; Whidborne, J. F.; Sceviratne, L. D., (2003). A three-term backpropagation algorithm. Neurocomputing, 50, 305–318.
17. Hagan, M. T., Demuth, H. B., & Beale, M. (1997). *Neural network design*. PWS Publishing Co..
18. Dehuri, S., & Cho, S. B. (2010). A comprehensive survey on functional link neural networks and an adaptive PSO–BP learning for CFLNN. *Neural Computing and Applications*, *19*(2), 187-205.
19. Yang, X. S., & Deb, S. (2014). Cuckoo search: recent advances and applications. *Neural Computing and Applications*, *24*(1), 169-174.
20. Cao, H., Yan, X., Li, Y., Wang, Y., Zhou, Y., & Yang, S. (2014). A component prediction method for flue gas of natural gas combustion based on nonlinear partial least squares method. *The Scientific World Journal*, *2014*.
21. Mardani, A., Streimikiene, D., Nilashi, M., Arias Aranda, D., Loganathan, N., & Jusoh, A. (2018). Energy consumption, economic growth, and CO2 emissions in G20 countries: application of adaptive neuro-fuzzy inference system. *Energies*, *11*(10), 2771.
22. Wilson, A., & Adams, R. (2013, February). Gaussian process kernels for pattern discovery and extrapolation. In International Conference on Machine Learning (pp. 1067-1075).

23. Deng, S., Deng, X., Chen, H., & Qin, Z. (2024). Estimating fossil CO2 emissions from COVID-19 post-pandemic recovery in G20: A machine learning approach. Journal of Cleaner Production, 442, 140875.
24. Al-Nefaie, A. H., & Aldhyani, T. H. (2023). Predicting CO2 Emissions from Traffic Vehicles for Sustainable and Smart Environment Using a Deep Learning Model. Sustainability, 15(9), 7615.
25. Wang, J., & Azam, W. (2024). Natural resource scarcity, fossil fuel energy consumption, and total greenhouse gas emissions in top emitting countries. Geoscience Frontiers, 15(2), 101757.
26. Miyazaki, K., & Bowman, K. (2023). Predictability of fossil fuel CO2 from air quality emissions. Nature communications, 14(1), 1604.
27. Ahmadi, M. H., Jashnani, H., Chau, K. W., Kumar, R., & Rosen, M. A. (2023). Carbon dioxide emissions prediction of five Middle Eastern countries using artificial neural networks. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 45(3), 9513-9525.