Limitations of Parametric Group Method of Data Handling and Empirical Improvements for the Application of Rainfall Modelling

Ronald William Lake  
Charles Sturt University

Saeed Shaeri  
Charles Sturt University

STMLD Senevirathna (✉ lsenevirathna@csu.edu.au)  
Charles Sturt University  https://orcid.org/0000-0003-0880-3601

Research Article

Keywords: Ensemble Empirical Mode Decomposition, Group Method of Data Handling (GMDH), Ivakhnenko, Machine Learning, Rainfall, Unscented Kalman Filter, Wavelet Transform, Wavelet Packet Transform

Posted Date: April 22nd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1495426/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Limitations of Parametric Group Method of Data Handling and Empirical Improvements for the Application of Rainfall Modelling

Ronald William Lake*, Saeed Shaeri*, STMLD Senevirathna*

*Faculty of Business, School of Computing, Mathematics and Engineering, Charles Sturt University,
Bathurst, New South Wales, Australia

*Corresponding author (lsenevirathna@csu.edu.au)

Abstract

This study introduces parametric Group Method of Data Handling (GMDH) into regional NSW, Australia, for the first time with the intent to assess the impact of climate change through rainfall modelling of publicly available data. The modelling presented the opportunity to investigate improvements to GMDH for rainfall modelling, with an empirical assessment undertaken. State variable distribution, their classification within the context of fuzzy, and the need to integrate Zadeh’s principle of incompatibility into the GMDH modelling format are all assessed. The mathematical foundations of GMDH are discussed within the heuristic framework of data partitioning, partial description synthesis and the limitations of least squares coefficient determination, Gödel’s incompleteness theorem and the necessity for an external criterion in the selection procedure for the Ivakhnenko polynomial. Methods for modelling improvement include the potential for hybridisation with least square support vector machines (LSSVM), the application of Kalman filters for parameter estimation, and the combination with signal processing techniques; ensemble empirical mode decomposition (EEMD), wavelet transformation (WT), and wavelet packet transformation (WPT) being investigated as is the implementation of enhanced GMDH (eGMDH) and fuzzy GMDH (FGMDH). The inclusion of exogeneous data is also discussed and whether application presents within the GMDH modelling paradigm. The study concludes with recommendations made to enhance the potential for future rainfall modelling study success.

Keywords

Ensemble Empirical Mode Decomposition, Group Method of Data Handling (GMDH), Ivakhnenko, Machine Learning, Rainfall, Unscented Kalman Filter, Wavelet Transform, Wavelet Packet Transform
Introduction

Rainfall trend analysis is an active research area that includes environmental, agricultural, and engineering studies. By reviewing historical rainfall data that extends back for over 100 years without discontinuities, it provides an opportunity for trend identification that may point to a changing climate. The Intergovernmental Panel on Climate Change (IPCC, 2014) advise that statistically significant increases in the number of heavy rainfall events have occurred since 1951 across more regions than the converse. The (Australian) Commonwealth Scientific and Industrial Research Organisation (CSIRO) and the Bureau of Meteorology (BoM) released the State of Climate Report (CSIRO & BoM, 2020); In Australia, there has been an increase in the intensity of heavy rainfall events. Extreme rainfall events up to and including 60-minute duration have increased in frequency by at least 10% across different regions of the country, with northern Australia experiencing them with greater prevalence (p. 8). The CSIRO and BoM (2020); extreme rainfall events of short duration impose the added potential of flash flooding placing communities at risk.

As a means of attempting to quantify the existence of trends in rainfall data for the purpose of making a prediction on future rainfall events, a selection of 6 local government areas (LGAs) within the Central West of New South Wales (NSW) Australia were chosen. The historical rainfall datasets recorded by the Bureau of Meteorology (BoM, 2020) are publicly available and for this subset, the data extends back 135 years in at least one instant. Covering the LGAs of Forbes, Lachlan, Bland, Parkes, Cowra, and Weddin, the adjoining LGA of Cabonne being purposely excluded as the geomorphology is quite disparate, in turn influencing noticeable meteorological differences, an example being rainfall within the 800-1200 mm range per year. Contrast the far western plains which receive only around 400 mm per year (Office of Environment & Heritage, 2014). This choice for exclusion is supported by the Central West and Orana – climate change snapshot (2014), there is considerable rainfall variability across the region. The complex interactions between weather patterns within the region, large-scale climate influences such as El Niño Southern Oscillation (ENSO), the topography of the Blue Mountains and the Great Dividing Range coupled with the circa 80km straight line distance from the East coast (Office of Environment & Heritage, 2014). Much of the central west experience annual mean precipitation within the 400–800 mm range. Cashen (2011) quantifying the effects of climate change being superimposed over the Australian climate influences of ENSO and the Indian Ocean Dipole (IOD) – like ENSO another atmospheric phenomenon, is the primary driver for concentrating on the chosen LGAs. ENSO in its two forms – La Niña delivering an
increase in rainfall, and El Niño – a reduction in rainfall, combined with positive IOD – rainfall decrease, negative IOD – rainfall increase, oscillate about the typical rainfall trend (Cashen, 2011).

The analysis of the monthly rainfall data was performed under the umbrella of machine learning, specifically the Group Method of Data Handling (GMDH), a polynomial neural network (PNN). To the best of the author’s knowledge, no published studies identified have used GMDH within a regional Australian context for time series rainfall and temperature trend analysis. The study was undertaken across 2020 and 2021 encompassing both temperature and rainfall data. GMDH Shell version 3 (GMDH, 2022) was used for the modelling with user access to influence regressor distribution or external criterion not permitted. The results illustrated a noticeable disparity between the accuracy of the temperature modelling when comparing the rainfall modelling which Cox et al. (2019) details is intrinsically noisy featuring large scatter with no obvious trend. The consistency of the results and the coefficient of determination being used as a reference for acceptance. The objective of this paper is to study an established collection of methods that, when combined with GMDH deliver modelling improvement. Many of these improvements to GMDH were not utilised for rainfall modelling, but the results illustrate the potential if applied within the rainfall domain. This paper brings together these methods of improvement that to the author’s knowledge have not been presented collectively in a single paper. Each section details a mathematical background and guidance for its implementation. The aim being to present a paper that is both rigorous and informative. The composition of parametric GMDH is introduced, the formation of combinatorial algorithms (COMBI) and multilayered iterative algorithms (MIA) (Madala & Ivakhnenko, 1994). Data modelling specifically within the context of rainfall provide guidance on the inconsistency of the standard GMDH results. Hybridising with Least Square Support Vector Machines (LSSVM), the application of Unscented Kalman filters for polynomial parameter determination and three signal processing techniques for the hybridisation with GMDH. Variations of GMDH are also discussed within the context of enhanced and fuzzy prior to suggesting additional methods that may prove beneficial for modelling accuracy. It is to be noted that within this paper the term standard GMDH applies to the COMBI and MIA algorithms.

Analysis of Parametric GMDH

GMDH is machine learning, a branch of artificial intelligence (Wakefield, 2021), introduced in 1966 by Dr A. G. Ivakhnenko, an algorithm that would allow the development of high order regression-type polynomials (Farlow, 1981). Ivakhnenko in designing GMDH employed a heuristic and perceptron approach, the latter being a feature of artificial neural networks (ANN) (Anastasakis & Mort, 2001). GMDH algorithms are grouped broadly into two
categories: Parametric and non-parametric. COMBI and MIA fall within the former category, where the input data is either exact or possessing noise of low variance (Anastasakis & Mort, 2001). These algorithms describe Anastasakis and Mort (2001) models the full range of “possible input variable combinations”, selecting the best model that has been generated from the complete set of models according to the external criterion (p. 4).

Ivakhnenko et al. (1983) define COMBI algorithms as a complete mathematical inductive method as no potential models will be passed prior to consideration. COMBI algorithms organise the models through gradual term increments from 1 to \( m \), where \( m \) is the number of arguments whilst the external criterion will specify the optimum solution that exists between models that exhibit the same degree of complexity. It will produce a minimum value within the “plane of complexity vs selection criteria”, thus corresponding to the non-physical optimum model. The primary distinction between the COMBI algorithm and the MIA is in the number of layers. The structure of the multilayer algorithms is according to Anastaskais and Mort (2001) very comparable to multilayer feedforward neural networks, however the distinction lies in the number of layers and neurons which are objectively allocated by the external criterion in compliance with the incompleteness theorem. The principal theory behind GMDH and the GMDH algorithms is framed by four heuristics (Ivakhnenko, 1970).

1) Collect a dataset ideally representative of the object sought, example for the research undertaken, monthly rainfall data covering the six LGAs.

2) Partition the dataset into two subsets, the first is deemed the training set by which the polynomial coefficients are determined. The second dataset forms the testing set for use with the external criterion allowing separation of the embedded metadata into two divergent categories: - helpful or unhelpful. The partitioning is undertaken automatically by the GMDH algorithmic process, requiring no user input. Test data will usually compose 33% of the total dataset, with the algorithm selecting where to draw the data from (Dorn et al., 2019).

3) Produce a set of elementary functions (often quadratic polynomials) delivering increasing complexity through an iterative process where a range of different models are produced.

4) Apply the external criterion for selection of the optimised model, this procedure being based upon Gödel’s incompleteness theorem, which states “under certain conditions in any language there exist true but unproveable statements” (Uspensky, 1994, p. 241). The implication here being that for the model that is most representative of the system to be found, a comparative analysis is required with the external criterion. He et al. (2008) state that the data used within the training set and the external criterion being
mutually exclusive. The data not used within the training set for estimating the parameters and creating
the model forms the testing set data, which is used by the external criterion for evaluating and selecting
the model of best quality. He et al. (2008) emphasises the significance of optimised cooperation between
the external criterion and the division of the dataset, the latter though is not an option with GMDH Shell
3. The reader is encouraged to refer to their work for a comprehensive analysis.

The foundational mathematical process that underpins GMDH theory is the Volterra functional series which is
represented in discrete analogue form as the Kolmogorov – Gabor polynomial (Anastasakis & Mort, 2001). A
nonlinear multivariate high-order polynomial possessing the ability to describe time series, both present and in
the past (Gilbar, 2002). This capability is distinct from a Taylor series that only possesses the ability to describe
a specific moment in time. This means the former captures dynamic time representation; the latter is specifically
static.

\[ Y = (x, t) = a_0 + \sum_{i=1}^{M} a_i x_i + \sum_{i=1}^{M} \sum_{j=1}^{M} a_{ij} x_i x_j + \sum_{i=1}^{M} \sum_{j=1}^{M} \sum_{k=1}^{M} a_{ijk} x_i x_j x_k + \ldots \]  

(1)

The model output response is designated \( Y, x = (x_1, x_2, x_3, \ldots, x_m) \) the vector of input variables also referred to as
regressors where \( x_m \in \mathbb{R}^m \), and \( a = (a_0, a_1, a_2, \ldots, a_m) \) the vector of coefficients or weights, and \( M \) is the
number of regressors. Müller et al. (2007) details that the GMDH algorithm utilises an inductive approach framed
by the self-organisation principle. The inductive approach is unbounded with the regressors randomly shifted and
activated allowing for the closest match to the dependent variables to be selected Madala (1991). Self-organising
is according to Green et al. (1988) a non-parametric process in terms of there being an \textit{a priori}. The idea of the
existence of a unique model with optimum complexity which can be determined through self-organisation forms
the foundation of the inductive approach. GMDH delivers the output through the construction of analytic functions
formed by quadratic polynomials in a feed-forward network structure. The polynomial coefficients are derived
from pairs of regressors through a regression technique based upon the ordinary least squares method (OLS). A
cascade of quadratic-polynomials commonly referred to as partial descriptions (PD) reside in each of the neurons
Madala and Ivakhnenko (1994). The quadratic polynomials being of the form:

\[ y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_1^2 + a_4 x_2^2 + a_5 x_1 x_2 \]  

(2)

In defining a non-linear system encompassing rainfall data with the multilayer algorithm considered first:
\[ y = f(x_1, x_2, ..., x_n) \] (3)

where \( x_1, x_2, ..., x_n \) and \( y \) form the state system variables. The input variables are termed the learning set describes Jiřina and Jiřina jr (2013), being normally divided into “two disjoint subsets”, the first being the training set, the second being the validation set (p. 455). During the learning process, the PD coefficients are formed from elements within the training set, the elements from the validation set expose the magnitude of the error allowing for PD retention or disposal. The function \( f \) links these variables through the relationship defined by the Kolmogorov – Gabor polynomial. Nikolay and Iba (2003) highlight the fact that the Weierstrass theorem illustrates that these polynomials are a format that is universal for “non-linear function modelling” as they can represent approximately any continuous function defined on a compact set to a degree of precision that is arbitrary, within the context of an average squared residual (ASR), provided the number of terms is sufficient. The series is not infinite in practice, instead truncated due to design decisions (p. 1528). Each GMDH active neuron houses a PD, accepting two inputs then producing a single output. Within the first network layer, Ivakhnenko (1971) and Dorn et al. (2012) define the neuron number as:

\[ m = C^n_2 = \frac{n^2 - n}{2} \] (4)

where \( m \) is the neuron (PD quantity) within the first layer, and \( n \) is the number of inputs selected for that layer.

From each neuron (PD), the output takes the form:

\[ h_{11} = f_{11}(x_1, x_2) \]
\[ h_{12} = f_{12}(x_1, x_3) \]
\[ ...... \]
\[ h_{1m} = f_{1m}(x_{n-1}, x_n) \] (5)

Training of the first layer is undertaken with a comparative analysis of neuron outputs against the external criterion. The external criterion is formed from a vector of regressors not used within the training data set, i.e., a selection criterion. Removal of neurons delivering the least favourable results takes place thereby only retaining those neurons (PD) that best fit the criterion. The selected neurons form a subset of the original set of neurons within the first layer.

\[ [\hat{h}_{11}, \hat{h}_{12}, ..., \hat{h}_{1m}] \subset [h_{11}, h_{12}, ..., h_{1m}]; \]
\[ \hat{f}_{11}, \hat{f}_{12}, \ldots, \hat{f}_{1m} \subset [f_{11}, f_{12}, \ldots, f_{1m}] \quad (6) \]

noting that \( \hat{m} < m \) where \( \hat{m} \) are the selected neurons from the first layer.

A second layer of neurons is now formed designated as \( p \) with the \( \hat{m} \) selected neurons from the first layer sampled in pairs forming the new layer \( p = C^\hat{m}_2 \). In essentially the same format as eq. (5), the outputs from \( p \) can be defined as:

\[
\begin{align*}
    h_{21} &= f_{21}(\hat{h}_{11}, \hat{h}_{12}) \\
    h_{22} &= f_{22}(\hat{h}_{11}, \hat{h}_{13}) \\
    & \ldots \ldots \\
    h_{2p} &= f_{2p}(\hat{h}_{1(\hat{m}-1)}, \hat{h}_{1\hat{m}}) \\
\end{align*}
\quad (7)
\]

Neuron selection and the addition of new layers continues until the optimum model is reached, with the criterion of regularity being based upon the mean square error (MSE) with Figure 1 illustrating. The minimal value of the MSE corresponding with optimal model complexity.

![Fig 1. Selection of the optimised model](image)

The poorest performing neurons are removed so only those that fit the criterion are retained, with the best performing often referred to as the Ivakhnenko polynomial. The selection process tests the training data set with the test data against the external criterion. The output from the MSE for each data selection as defined by Ivakhnenko (1971) and Dorn et al. (2012) is: -
MSE\(_{(hln)}\) = \(\frac{1}{s} \sum_{t=1}^{s} (h_{ln}(t) - y(t))^2\) \hspace{1cm} (8)

where \(h_{ln}(t)\) is the output from the partial description within neuron \(n\) in the layer \(l\) at time \(t\), \(s\) refers to the number of elements within the sample space of input variables and \(y(t)\) is the network model output that is desired. The GMDH training procedure incorporates the addition of layers, the determination of coefficients associated with each partial description, and the elimination of neurons that produce the poorest results. Outputs that are retained from one layer form the basis of inputs for next layer. A halt to the training process occurs when the number of neurons within the current layer falls to one following application of the external complement, or when a subsequent layer does not improve the overall network performance. Should that occur, that last layer is then removed. Preservation of the neuron from the previous layer that delivered the best performance is retained whilst all other neurons within that layer are discarded. Pham and Liu (1994) describe the trimming process, that of removing all neurons from all layers that do not contribute to the final output. Figure 2 illustrates MIA algorithm architecture. The bold lines illustrate the active connections between the inputs and neurons, and between neurons and neurons. All selected neurons also appear in bold, having met the requirements of the selection criterion.

Fig 2. GMDH Multilayer Algorithm

Each neuron within each layer accepts two inputs, producing a single output as described by Ivakhnenko (1971) and Dorn et al. (2012). The inputs are of the form \(x_i\) and \(x_j\) which are supplied to a PD with output \(y\). The linear coefficients are \((a_0, ..., a_5)\). The PD where \((x_i, x_j) \subset x ; i \neq j\) (Madala & Ivakhnenko, 1994) is of the form of eq. (2)
or represented in matrix format as

\[ h = \mathbf{x} a \]  \hspace{1cm} (9)

where \( \mathbf{x} \) is the row vector of pre-processed inputs \( x_i \) and \( x_j \) for one PD, conveyed by

\[ \mathbf{x} = (1, x_i, x_j, x_i^2, x_j^2, x_i x_j) \]  Pham and Liu (1994), and \( \mathbf{a} \) is a column vector containing elements of coefficients that have not yet been determined \( \mathbf{a} = (a_0, \ldots, a_5)^T \). For time series rainfall forecasting, the assumption is made that there are \( s \) elements within the sample space selected by the algorithm:

\[
\begin{align*}
    h(1) &= f(x_1(1), x_2(1), \ldots, x_n(1)) \\
    h(2) &= f(x_1(2), x_2(2), \ldots, x_n(2)) \\
    \vdots \\
    h(s) &= f(x_1(s), x_2(s), \ldots, x_n(s))
\end{align*}
\]  \hspace{1cm} (10)

A time series PD will deliver an output at time \( t \) supplied by two inputs \( x_i(t) \) and \( x_j(t) \)

\[
    h(t) = a_0 + a_1 x_i(t) + a_2 x_j(t) + a_3 x_i^2(t) + a_4 x_j^2(t) + a_5 x_i(t) x_j(t)
\]  \hspace{1cm} (11)

where \( t = 1, 2, \ldots, s \).

If we now let \( h(t) = y(t) \), (12) can be portrayed as

\[ y = \mathbf{X} \mathbf{a} \]  \hspace{1cm} (12)

where

\[
\mathbf{x} = \begin{pmatrix}
    1 & x_i(1) & x_j(1) & x_i^2(1) & x_j^2(1) & x_i(1)x_j(1) \\
    1 & x_i(2) & x_j(2) & x_i^2(2) & x_j^2(2) & x_i(2)x_j(2) \\
    \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
    1 & x_i(s) & x_j(s) & x_i^2(s) & x_j^2(s) & x_i(s)x_j(s)
\end{pmatrix}
\]  \hspace{1cm} (13)

and

\[ \mathbf{y} = (y(1) \ y(2) \ \ldots \ \ y(s))^T \]  \hspace{1cm} (14)

From (13), the matrix of coefficients for the partial descriptions are calculated using the LSM: -
$a = (X^TX)^{-1}X^Ty$  \hspace{1cm} (15)

Green et al. (1988) GMDH can produce exceedingly complex models. As an example, if the model has $r$ layers based upon $m$ regressors, the final polynomial representing the model would have degree $2^r$ with the number of terms being

$$\left(\frac{2^r + m}{m}\right)$$  \hspace{1cm} (16)

So, with four layers and six regressors within the input vector, the GMDH polynomial presents a degree of 16 with 74,613 terms. The partial description coefficients are determined by the regression method of least squares (Anastasakis & Mort, 2002). The data selected that forms the training set is used for the determination of the coefficients (Duffy et al., 1975). When the regressors within the input vector are well defined, the coefficient estimates will be accurate, but if the regressors are ill-defined, then the coefficients will be subject to bias (Anastasakis & Mort, 2001). Real world data though is normally ill-defined and thus the coefficient estimates for the partial descriptions will be subject to bias. Ivakhnenko and Zholnarskij (1992) identifies the method of instrumental variables as a potential replacement for least squares delivering estimates with less bias.

The average error in estimating the PD coefficients is illustrated mathematically (Madala & Ivakhnenko, 1994).

Let $b$ denote the desired value and $y$ the estimated output value for the PD under consideration. $N_A$ is the set of training data. The output errors are given by

$$e_q = y_q - b_q; q \in N_A$$  \hspace{1cm} (17)

For the input vector, the total squared error is

$$E = \sum_{q \in N_A} e_q^2$$  \hspace{1cm} (18)

This is the average error minimised in coefficient estimation.

Each layer that forms the neural network details Madala and Ivakhnenko (1994) contains groups of neurons that link to neurons within the following layer. The PD coefficients at each neuron being estimated through minimising error $E$. The selection criterion is used to assess whether each neuron in each layer will be accepted or rejected. If accepted, the neuron is switched on allowing its output to form an input in the following layer. The process continues throughout the network until the most suitable output representative of the model is found by the
selection criterion. The PD most frequently used is the second order polynomial, but there are alternatives (Anastasakis & Mort, 2001). Additional mathematical expressions that can be used within Parametric GMDH include linear and cubic equations (Farzi, 2008). MIA can also utilise transfer functions details Kondo and Ueno (2012) within each neuron simultaneously with the function that delivers the smallest mean square error being selected for that given neuron. Transfer functions include –

\[ z = \frac{1}{1 + e^{-y}} \]  
\[ z = e^{-y^2} \]  
\[ z = \tan(y) \]

GMDH Limitations

The exclusion of essential regressors initiates noise that impairs model performance (Anastasakis & Mort, 2001). From studies undertaken by Green et al. (1988), the problem appears to be caused by collinearity. Even with independent regressors within the input vector, the PDs formed within the first and subsequent iterations were not mutually independent. Green et al. (1988) further emphasised that this results in a selection of regressors being excluded. Overfitting can also be a problem, and when combined with multiple neuron layers instability delivering poor prediction quality results (Green et al., 1988). Biased estimates of PD coefficients being a result of the application of OLS is an additional shortcoming (Anastasakis & Mort, 2001). There is an assumption that the observed output values and the estimated output values should be reflected in a Gaussian distribution meaning that the use of linear regression is justified for PD parameter determination. The reality though is this assumption is frequently violated meaning that OLS is not a suitable method (Anastasakis & Mort, 2001). Standard GMDH will also fail in the presence of fuzzy input data, meaning a suitably modified GMDH could be appropriate (Anastasakis & Mort, 2001). The limitations listed may not seem considerable, but their implication can be significant to the functionality of standard GMDH. For this reason, modified versions of GMDH are now explored. These formats will be shown to deliver improvements over standard GMDH that are worth investigating further, particularly to test their suitability for rainfall modelling.
Hybridising GMDH

Accurate rainfall prediction is a challenge when using a singular model, so the introduction of a hybrid that combines two models has the potential to deliver performance that exceeds the capabilities of each of the composing models (Parviz et al., 2021). One example used for time series forecasting is hybridising least square support vector machines (LSSVM) with GMDH. In research undertaken by Samsudin et al. (2011) for time series forecasting, the hybrid LSSVM GMDH model delivered more accurate results due to its robust nature and ability to model non-linear data. Support vector machines (SVM) like GMDH fall under the umbrella of ML. They map the vector of regressors into a designated high dimensional feature space $Z$ via a non-linear mapping procedure selected a priori. A hyperplane is constructed within this space that ensures high generalisation ability of the network (Cortes & Vapnik 1995). The technique of support vector networks detail being originally developed for limited application in separating training data without errors. The situation often arose where this ideal was not possible, so by extension SVM were deemed a new class of ML with comparable power and universality to neural networks. Further development saw the introduction of the least squares’ variant of SVM designated as LSSVM (Cortes & Vapnik 1995). The differences being the SVM employs equality constraints whereas the LSSVM uses inequality constraints and implementing the system of linear least squares as its loss function (Samsudin et al., 2010a). The LSSVM also offers good convergence combined with high precision whereas the SVM uses quadratic program solvers that are more challenging to use (Samsudin et al., 2010a).

Analogous to GMDH, the LSSVM predictor detailed in Samsudin et al. (2010b) is trained by employing a set of historic time series regressors which deliver a single output as the goal. Detailing the mathematical approach presented by Samsudin et al. (2011); take a given training set composed of $n$ data points $\{x_i, y_i\}_{i=1}^n$, input data $x_i \in \mathbb{R}^n$, $p$ being the total number of data patterns. Output $y_i \in \mathbb{R}$.

The SVM approximation of the function is presented in the form

$$ y(x) = \mathbf{w}^T \Phi(x) + b $$

with $\Phi(x)$ the high dimensional feature spaces being non-linearly mapped from the input space $x$. In LSSVM for estimating the function, the formulation of an optimisation problem is presented (Suykens et al., 2002)

$$ \min f(w, e) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{\gamma}{2} \sum_{i=1}^n e_i^2 $$

The following constraints apply
To obtain the solution, the Lagrange is constructed

\[ L(w, b, e, \alpha) = J(w, e) - \sum_{i=1}^{N} \alpha_i (w^T \phi(x_i) + b + e_i - y_i) \]  

(24)

With Lagrange multipliers \( \alpha_i \). The conditions that apply for optimality are equations 38 - 41

\[ \frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^{N} \alpha_i \phi(x_i) \]  

(25)

\[ \frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{N} \alpha_i = 0 \]  

(26)

\[ \frac{\partial L}{\partial e_i} = 0 \rightarrow \alpha_i = \gamma e_i \]  

(27)

\[ \frac{\partial L}{\partial \alpha_i} = 0 \rightarrow w^T \phi(x_i) + b + e_i - y_i = 0 \]  

(28)

for \( i = 1, 2, ..., n \).

Following the elimination of \( e_i \) and \( w \) the solution is presented by the linear equations

\[ \begin{pmatrix} 0 \\ 1 \end{pmatrix} \phi(x_i)^T \phi(x_j) + \gamma^{-1} I \begin{pmatrix} b \\ a \end{pmatrix} = \begin{pmatrix} 0 \\ y \end{pmatrix} \]  

(29)

with \( y = [y_1; ...; y_n] \), \( 1 = [1; ...; 1] \), \( \alpha = [\alpha_1; ...; \alpha_n] \)

Utilising Mercer’s theorem which sets the conditions for which a function can be considered a Kernel (Bhattacharyya, 2018), the Kernel function is defined as being

\[ K(x_i, x_j) = \phi(x_i)^T \phi(x_j), \ i, j = 1, 2, ..., n \]  

(30)

The LSSVM model for estimating functions is now illustrated as

\[ y(x) = \sum_{i=1}^{n} \alpha_i K(x_i, x_j) + b \]  

(31)

noting that \( \alpha_i, b \) are the solutions of the linear system

Samsudin et al. (2011) notes that selecting the Kernel function \( K(., .) \) has several options. \( K(x_i, x_j) \) is the Kernel function, with its value equal to the inner product of two vectors \( x_i \) and \( x_j \) within the feature space \( \phi(x_i) \) and \( \phi(x_j) \), such that \( K(x_i, x_j) = \phi(x_i) \ast \phi(x_j) \).

Examples of Kernel functions include
Linear \[ K(x_i, x_j) = x_i^T x_j \]

Polynomial \[ K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \quad \gamma > 0 \]

Radial Basis Function \[ K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \quad \gamma > 0 \]

Sigmoid \[ K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r) \] (32)

Kernel parameters $\gamma$, $r$, and $d$ require careful selection given they implicitly define the spatial framework of the high dimensional feature space $\emptyset(x)$ thereby controlling final solution complexity. Construction of the hybrid LSSVM with GMDH follows the procedure of Samsudin et al. (2011). The complete dataset is to be normalised prior to being separated into two disjoint sets covering training and testing. Sotelo (2017) optimisation occurs through minimisation of the decision vector $w$ with the scale of the input data influencing the optimal hyperplane. The recommendation is made that the data be standardised with a mean of zero and a variance of one. Figure 3 illustrates the structure of the SVM. Most real-world data is nonlinearly separable meaning the hyperplane is not represented by a straight line. To overcome this problem, Kernel functions are used to allow transformation of the nonlinear data for linear presentation at higher dimensions (Ampadu, 2021).

Fig 3. Structure of the Support Vector Machine

The steps taken for hybridisation are (Samsudin et al., 2011)

1) Separate the normalised data into training and testing sets
2) Using the GMDH MIA, combinations of all input state variables \((x_i, x_j)\) are generated, with the totality of independent variables being \(C_M^2\). The regression polynomial is constructed with an approximation of the output given by equation (9). The linear vector of coefficients \(A\) for each PD determined by OLS.

3) The output of each neuron \(x'\) is assessed against the external criterion with the smallest MSE selected for the formation of a binary input \(\{x_1, x_2, \ldots, x_M, x'\}\) with \(M = M + 1\), for a neuron within the next hidden layer.

4) In the GMDH output layer from neurons within the hidden layers, these outputs form inputs \(\{x_1, x_2, \ldots, x_M, x'\}\) for the LSSVM. The minimal MSE from the LSSVM will receive selection as the output model.

5) The minimal MSE obtained from the LSSVM for the test data set extracted at each layer during the current iteration is compared against the minimal value from the previous iteration. In the case where there is an improvement, steps 2 to 4 are repeated else the iterations cease with the knowledge that the network is now complete. Determination of the final layer signifies that only one node with the best performance will realise selection. When this occurs the remaining nodes with the output layer are discarded, thus delivering the hybrid Group Least Squares Support Vector Machine (GLSSVM) model.

The Kalman Filter and GMDH

In its original form, the Kalman Filter (KF) algorithm employs a dynamics model that describes its status and its expected status at the following step provided the system is linear, with any process disturbance combined with error of measurement being additive and Gaussian (Pasek & Kaniewski, 2021). Most real-World systems which include rainfall data are within the non-linear category, and in the case of monthly rainfall data, non-Gaussian. To overcome this incompatibility, the Extended Kalman Filter (EKF) was developed for handling non-linear functions which are locally approximated with linear equations through Taylor expansion (Pasek & Kaniewski, 2021). Monthly rainfall data is highly non-linear potentially introducing significant errors in linearisation as the first term only of the Taylor series is utilised (Pasek & Kaniewski, 2021). The EKF also requires computation of the Jacobian Matrix at each time step. If the initial conditions are poorly known or there are significant measurement errors, major errors in state vector estimation could result leading potentially to divergence of the filter (Kraszewski & Czopik, 2017). As a means of overcoming these problems, the Unscented Kalman Filter (UKF) was developed by Julier and Uhlmann (Julier & Uhlmann, 1997). To implement the UKF, a set of sigma points are chosen with known mean and covariance. At each point a nonlinear function is applied yielding a set
of transformed points. The transformed point statistics can be calculated to provide an estimate of the nonlinearly
transformed mean and covariance (Julier & Uhlmann, 2004). In combining the UKF with GMDH, the UKF is
utilised for parameter estimation for each GMDH neuron PD (Luzar et al., 2011). GMDH determines the
parameters for each PD separately, and it is this procedure that allows the UKF to be utilised (Mrugalski, 2013).
The main advantage of applying UKF for estimation of PD parameters is in the generation of an asymptotically
stable GMDH model (Mrugalski, 2013).

The design of a Kalman Filter is normally based upon an analytical model of a dynamic process or system
\[ \mathbf{X}_{k+1} = F(\mathbf{x}_k, \mathbf{p}) \]
with \( \mathbf{x}_k \) the unknown state vector for the dynamic process taken at time step \( k \) (Pan et al., 2020). \( F(\cdot) \)
is the dynamic model of the system parameterised by \( \mathbf{p} \) a covariance vector. \( \mathbf{y}_k = H(\mathbf{X}_k, \mathbf{p}) \)
is the observable variables model belonging to the dynamic system, with \( \mathbf{y}_k \) the observation vector composed of variables that can
be measured (Pan et al., 2020). \( H(\cdot) \) is the state to measurement matrix which maps between \( \mathbf{x}_k \) and \( \mathbf{y}_k \). Additive
process noise affects the degree of accuracy of \( F \) which is modelled by \( \mathbf{Q} \), the covariance matrix. \( \mathbf{R} \) the observation
noise covariance matrix models the uncertainty properties of system observations (Pan et al., 2020). The UKF is
initialised by making an estimate of the unknown state variables \( (\hat{\mathbf{x}}_0) \), supplied with various measurements \( (\mathbf{y}_k) \).
The UKF state variables \( (\hat{\mathbf{x}}_0) \) are recursively estimated through covariance minimisation
\[ \hat{\mathbf{P}}_k = \text{cov}(\mathbf{x}_k - \hat{\mathbf{x}}_k) \]
\( \hat{\mathbf{P}}_k \)
is a diagonal matrix with elements representing the uncertainty estimate of \( \hat{\mathbf{x}}_k \) (Pan et al., 2020). Through
implementation of the UKF algorithm, Masoumnezhad et al. (2016) minimisation of the mean squared estimate
errors with an increase in the convergence rate in addition to a stability improvement against corruption by noise
of the experimental data. For a detailed mathematical explanation of the parameter estimation of GMDH dynamic
neurons (e.g., Mrugalski, 2013). Figure 4 illustrates the integration of the UKF into GMDH for PD parameter
estimation. No literature to date has been found where this approach has seen application for rainfall modelling.
It has however been used with success in the identification of various dynamic systems (Luzar et al., 2011), and
for the successful design of a robust fault detection system (Mrugalski, 2013).
Signal Processing Approaches

In research undertaken by Moosavi et al. (2017) hybridising wavelet transform (WT) and wavelet packet transform (WPT) with GMDH (WGMDH and WPGMDH respectively) delivered improved results over GMDH alone for the purpose of runoff forecasting. Of these two the WPT hybrid delivered the best results. Both the WT and WPT improve the performance of the GMDH modelling by decomposing the original dataset into components (Moosavi et al., 2017). A study undertaken by Moosavi (2019) predicting rainfall within the context of natural disasters, GMDH was hybridised with Ensemble Empirical Mode Decomposition (EEMD), and again with WT, and WPT forming (WTGMDH) and (WPTGMDH) respectively. In all three cases, the rainfall modelling with GMDH delivered an improvement once hybridised with these signal processing approaches. The WPTGMDH preformed best, followed by EEMD-GMDH, then WTGMDH. All these hybrids outperformed the standard GMDH without signal processing assistance. In further research undertaken by Moosavi et al. (2021), the combination of GMDH with the same three separate signal processing approaches; EEMD, WT, and WPT delivered an improvement in the performance of GMDH for groundwater level forecasting in all three cases. Given these finding it seems prudent to investigate further the potential for improving the modelling of monthly rainfall data with GMDH by pre-processing the data with these signal processing approaches. An example of the monthly rainfall modelling undertaken during 2020 – 2021 using GMDH is illustrated in Figure 5 covering the town of Forbes from 1995 till 2021.
The BoM (2020) data profile is formed by the grey graph whilst the blue graph illustrates the GMDH model with the red being the prediction across 12 months into the future. The coefficient of determination returned a value of 0.246. All rainfall modelling undertaken returned similar unsatisfactory results. There is no clear trend depicted within the actual data nor can past events be a reliable means of predicting future events. The focus of this empirical assessment is to consider whether pre-processing the rainfall data might potentially improve the modelling outcome, thereby providing a greater degree of confidence in the rainfall prediction.

Ensemble Empirical Mode Decomposition (EEMD)

Empirical mode decomposition (EMD) was introduced by Huang et al. (1996). It is a method that is applicable for both nonlinear and non-stationary time series with Srikanthan et al. (2011) noting that the EMD method makes no prior assumption as to the form of the time series prior to analysis. The EMD can adaptively decompose any complex dataset into a series of intrinsic mode functions (IMF) (Huang et al., 1996). The IMFs are independent and in combination with a residual can effectively be summed together to reform the original series (Srikanthan et al., 2011). EMD does however experience mode mixing which occurs when at a given frequency a fluctuation may separate across two IMFs. Specifically, mode mixing often occurs from intermittent signals (Wu & Huang, 2009). The EMD algorithm employs a data driven adaptive iterative approach meaning that there is difficulty in avoiding mode mixing without subjectively contemplating the likely form of any signal for extraction before undertaking analysis (Srikanthan et al., 2011). To overcome the mode mixing problems, a noise - assisted data analysis (NADA) is proposed, the Ensemble EMD (EEMD) providing definition of the true IMF components “as the mean of an ensemble of trials, each consisting of the signal plus white noise of finite amplitude” (Wu & Huang,
Looking more closely initially at the EMD method then introducing white noise for the EEMD approach. Using rainfall data $X(t)$ as an example and the approach of Wu and Huang (2009), decomposition of the data in terms of IMFs $g_j$

$$x(t) = \sum_{j=1}^{m} g_j + r_m$$  \hspace{1cm} (33)

with $r_m$ the residue of data $x(t)$, following extraction of $m$ IMFs. The IMFs are oscillatory functions, simple in structure with both amplitude and frequency that vary. IMFs have the following properties

1. Across the full length of each IMF when comparing the number of extrema and the number of zero-crossings, their difference must be either zero, or at most one. This is not to be confused with the numbers from $X(t)$ which could be significantly different.

2. At any random location selected within the data, the envelope defining the local maxima and the envelope defining the local minima, their mean is zero.

Local extrema are only used by the EMD through a sifting process, so for any data,

a. All local extrema, both maxima and minima are to be identified, connecting all extrema with a cubic spline thus forming two envelopes.

b. The first component $A$ is obtained as the difference between the data and local mean of the envelope dyad.

c. Let $A$ now be treated as the data repeating steps a and b as often as necessary to deliver symmetric envelopes with respect to zero mean, with the final $A$ redefined as $g_j$.

The sifting is complete when the residue $r_m$ depicts a monotonic function preventing extraction of any further IMFs. (Wu & Huang, 2009). The gradient of the residue illustrates the trend within the data which is very useful when viewing historic rainfall data. Before incorporating the additional details of the EEMD, it is useful to review a few important EMD properties. Wu and Huang (2009):

1. Being based upon local data characteristics, the adaptive data analysis method effectively captures nonlinear, nonstationary oscillations with greater effect.

2. Any white noise only series or fractional Gaussian, EMD is equivalent to a dyadic filter bank.

3. Intermittent data can compromise the dyadic property

4. A reference scale that is uniformly distributed could be obtained through the addition of noise enabling the compromised dyadic property to be repaired by EMD.
5. There is no correlation between corresponding IMFs depicting differing noise series. As such, cancellation of the means of corresponding IMFs presenting different white noise series is likely.

With this information clear and present, the EEMD features:

1. The targeted data is augmented with a white noise series.
2. The data featuring the additional white noise is decomposed into IMFs.
3. Steps 1 and 2 are continually repeated with the addition of a different series of white noise on each occasion.
4. The ensemble means pertaining to the corresponding IMFs based upon the decompositions are obtained as the final result.

In utilising the EEMD, the decomposition effects being that the augmented white noise series cancel out in the final mean of the corresponding IMFs (Wu & Huang, 2009). The mean IMFs are positioned within the natural dyadic filter windows thereby significantly reducing the potential of mode mixing whilst preserving the dyadic property (Wu & Huang, 2009). The IMFs are more effective in the isolation of physical processes across a range of time scales owing to mutual orthogonality (Molla et al., 2005). The implication of implementing EEMD into the GMDH rainfall modelling is based upon the historic rainfall data signal being decomposed into a set of IMFs with a residual prior to processing by GMDH. The process that GMDH takes is through the implementation of supervised learning, and with IMFs being supplied as input data, that process and the formation of a function to represent the data enabling the capacity to make a prediction of future rainfall events, should in theory be improved. Figure 6 illustrates a flowchart of the EEMD-GMDH hybrid.
Wavelet Transform (WT)

Like EEMD, the WT provides a mechanism for signal decomposition. Wavelets are functions that meet a set of mathematical conditions with the ability to represent data and or other functions (Graps, 1995). They have the capacity to analyse non-stationary time series across a range of frequencies Daubechies (1990). Through decomposition of a time series into time-frequency space, determination of the dominant modes of variability can be achieved, and their time invariance. Wavelet transforms are available in both continuous and discrete forms and both types have been used in modelling of hydrometeorological studies (Torrence & Compo 1998). The idea that is fundamental behind WT is in their ability to perform signal analysis at different resolutions which precipitates their most important feature, that of multiresolution decomposition (Lubis et al., 2017). Özger et al. (2012) employed the continuous WT (CWT) in their study of drought forecasting whilst Alfa et al. (2019) utilised the discrete WT (DWT) in a hybrid with GMDH for their drought forecasting study. Moosavi et al. (2021) employed the DWT as a hybrid with GMDH in their research on groundwater level modelling. Providing a brief mathematical overview with reference to Addison (2018) and Lambers (2006), (Refer as an example A wavelet tour of signal processing by Stéphane Mallet, 1999 for an in-depth mathematical explanation).

A function $\psi(x)$ in satisfying these conditions
1. \[ E = \int_{-\infty}^{\infty} |\psi(t)|^2 dt < \infty \]

2. \[ C_\psi = \int_{-\infty}^{\infty} \frac{|\hat{\psi}(f)|^2}{f} df \equiv C_\psi < \infty \]

is a wavelet with condition 1 possessing finite energy. If \( \hat{\psi}(f) \) is the Fourier transform of \( \psi(t) \), then condition 2 is a necessary requirement. The implication being that \( \hat{\psi} = 0 \). This is the admissibility condition, and \( C_\psi \) the admissibility constant.

A wavelet family is a group of functions acquired by translating and dilating a wavelet graph. The mother wavelet \( \psi(x) \) is composed of functions \( \psi_{a,\tau}(x) \) such that

\[ \psi_{a,\tau}(x) = \frac{1}{\sqrt{a}} \psi \left( \frac{x-\tau}{a} \right) \]  

(34)

with \( \tau \) is the translation or centre of \( \psi_{a,\tau} \) and \( a \) is the dilation parameter.

A function \( f(x) \) with CWT, as introduced by Morlet in 1984 being defined as

\[ Wf(a,\tau) = \int_{-\infty}^{\infty} f(x) \psi_{a,\tau}(x) \, dx \]  

(35)

with inverse transform

\[ f(x) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{[a]} \frac{1}{|a|^2} Wf(a,\tau) \psi_{a,\tau}(x) \, da \, d\tau \]  

(36)

The wavelet transform of the function \( f \) is the convolution with the conjugate wavelet (Mallat, 1999):

\[ Wf(\tau,a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \psi^{*} \left( \frac{t-\tau}{a} \right) dt \]  

(37)

where the mother wavelet \( \psi(t) = \pi^{-\frac{1}{4}} e^{-\frac{t^2}{2}} (e^{-i\omega_0 t} - e^{-\omega_0^2 t}) \)  

(38)

with \( \omega_0 = \sqrt{\frac{2}{ln2}} \) and \( i = \sqrt{-1} \) (Russell and Han, 2016).

It is important to note that if using the CWT for pre-processing the graph of the monthly rainfall data before supplying this information to GMDH, the CWT stores changes in \( f(x) \) allowing both compression and removal of noise.

In the case of the DWT the values of \( a \) and \( \tau \) are limited. As the signal is represented discretely, numerical computations are required for the integrals that define the coefficients. In taking a mother wavelet, to obtain an
orthogonal family of wavelets select $a = a_0^m$ and $\tau = n\tau_0$, with $m$ and $n$ integers, $a_0$ is a dilation parameter $> 1$, and $\tau_0$ a translation parameter $> 0$. The DWT is now presented as

$$Wf(m, n) = \langle \psi_{m,n}, f \rangle = \int_{-\infty}^{\infty} \psi_{m,n}(x)f(x)dx$$  \hspace{1cm} (39)

with

$$\psi_{m,n}(x) = 2^{-m/2}\psi\left(\frac{x-n2^m}{2^m}\right)$$  \hspace{1cm} (40)

The inverse transform being given by

$$f(x) = \sum_{m,n} \psi_{m,n}(x)Wf(m, n)$$  \hspace{1cm} (41)

It is important to recognise that the integral defining $Wf(m, n)$ exists across an unbounded interval, but effectively it exists across a finite interval provided the mother wavelet has compact support, hence a numerical approximation is easy. Moosavi et al. (2021) detailed the importance of selecting the best WT structure, choosing the mother wavelet and the most effective level of decomposition. The decomposed data is then supplied to GMDH for processing with the optimum model delivered. Figure 7 illustrates a flowchart featuring the integration of the WT with GMDH.

**Fig 7.** Flowchart of Hybridised WT-GMDH Algorithm
Wavelet Packet Transform (WPT)

The WPT provides an improvement on the WT by raising the degree of resolution of high frequency signals precipitating an enhanced high frequency time-frequency localisation effect (Yan et al., 2021). The WPT prevents loss of time-frequency information, the signal placed within a domain where simultaneous analysis in both time and frequency can occur (Wickerhauser, 1991; Gokhale and Khanduja, 2010). Wavelet packet decomposition (WPD) involves passing the signal through a greater number of filters when compared to the WT (Gokhale & Khanduja, 2010). In general terms, a WPT is a square integrable function with a mean of zero and compact support, across both time and frequency (Zha et al., 2008). A brief mathematical outline by Zha et al. (2008) is provided. In describing wavelet packets by a collection of functions \( \{\varphi_k\}_{k=0}^{\infty} \) obtained from:

\[
\varphi_{2k}(x) = \sqrt{2} \sum_{n} h_k(n) \varphi_k(2x - n) \tag{42}
\]

\[
\varphi_{2k+1}(x) = \sqrt{2} \sum_{n} g_k(n) \varphi_k(2x - n) \tag{43}
\]

noting that the discrete filter \( h_k(n) \) and \( g_k(n) \) are quadrature mirror filters (p. 405). The function \( \varphi_0(x) \) is identifiable with the scaling function, and \( \varphi_1(x) \) with the mother wavelet. The wavelet packet basis due to its orthonormality and completeness guarantee retention of the original signal information. The inverse transform of the wavelet packets can be expressed as:

\[
\varphi_k(x) = \sum_n [h_k(n) \varphi_{2k}(x - 2n) + g_k(n) \varphi_{2k+1}(x - 2n)] \tag{44}
\]

The potentially significant component of the WPT specifically within this paper and the context of rainfall modelling, is its ability to recursively decompose high frequency signal components (Zha et al., 2008). The WPT constructs a tree formation multiband extension of the WT. This facilitates the ability of the WPT to recursively divide the entire frequency band for noise detection and removal. By sifting the signal of components that can be classed as noise within the context of isolating the intense rainfall events, it is hypothesised that the capacity of the hybridised WPTGMDH to produce models with improved coefficient of determination is possible. Such could potentially highlight trends within the rainfall data with greater clarity and prediction significance. Figure 8 illustrates the hybrid WPTGMDH modelling process.
Improved GMDH Algorithms

MIA and COMBI algorithms share the same approach of determining the PD coefficients, that of OLS. This regression approach can be subject to the production of coefficients with biased estimates which is detrimental to model accuracy. Structurally, the MIA is superior to the COMBI algorithm by virtue of possessing multiple layers of neurons which make processing time more expedient. One method for improving the performance of the MIA algorithm can be found within the neuron pruning process (Buryan, 2013). In the MIA, pruning of neurons based upon the external criterion can result in the loss of neurons that may ultimately have been useful. Synthesising the neuron outputs within a given layer promotes ill performing regression within the next layer (Buryan, 2013). The complexity of the network can exceed levels desirable owing to the quadratic polynomials that form the PDs when the vector of regressors is highly nonlinear (Buryan & Onwubolu, 2011). To overcome these shortcomings, the enhanced MIA GMDH (eGMDH) or alternatively (eMIA-GMDH) algorithm is proposed, detailed by Buryan and Onwubolu (2011) from work undertaken by Buryan (2006).

Enhanced MIA-GMDH

The improvements that feature encompass the following (Buryan & Onwubolu, 2011)

1. Within the original GMDH, eq. (4) \( r = 2 \) for all layers, with eGMDH application applies only within layer one. All subsequent layers can utilise different values of \( r \).

2. Utilising OLS for PD coefficient determination is not limited to quadratic polynomials, instead including a further seven types which modify eq. (2)
   a. Harmonic based on the cosine function
b. Radical

c. Inverse polynomial

d. Natural Logarithm

e. Exponential

f. Arc Tangent

g. Rounded Polynomial presenting integer coefficients

3. The pruning of layers is semi-randomised. Upon selecting a subset of the best performing neurons in each layer, the balance being selected at random.

4. Implementation of coefficient rounding and thresholding as a means of stabilising the regression process by rejection of coefficients that lay either above or below pre-set thresholds.

5. Neuron inputs are not binary limited

From experiments undertaken by Buryan (2013), the enhanced MIA-GMDH delivered improved results when compared to the standard MIA-GMDH algorithm for both Mackay-Glass time series predictions, and for JPY/USD exchange rate predictions for both daily and monthly time frames. In a study undertaken by Unwubolu et al. (2007) covering weather forecasting in Fiji, the eGMDH delivered improved results when compared to a PNN and an enhanced PNN (ePNN) for temperature modelling, but it was less successful for rainfall modelling against the alternatives. The study concluded with an unclear finding as to why this was the case.

**Enhanced GMDH using Modified Levenberg-Marquardt Method**

Pournasir et al. (2013) proposed an enhanced GMDH algorithm utilising a modified Levenberg-Marquardt (LM) method which incorporated the use of singular value decomposition (SVD) for the first time as an improved means for the first guess. It is noted that the process of combining the SVD output into LM for the initial guess was not detailed within their paper. The LM algorithm facilitate swift convergence for solving nonlinear systems and singularity problems (Pournasir et al., 2013). The experiment results showed that the enhanced GMDH using the LM algorithm outperformed standard GMDH in delivering results with high inventory control accuracy (Pournasir et al., 2013). In fitting a parameterised mathematical model to a set of data points (within the focal context of the initiator of this report – rainfall modelling) through minimisation of an objective function, which within the context of GMDH is the selection criterion – the MSE, the problem of least squares arises. Minimising the objective with respect to the parameters may be possible with speed provided the solution fits a linear matrix.
equation (Gavin, 2020). If, however this is not the case and the fit function is not linear within its parameters, then
the least squares problem is in need of an iterative algorithmic solution (Gavin, 2020). The algorithmic process
involves reducing the sum of the squares of the errors that exist between the function representative of the model,
and the data points through a sequence of carefully selected model parameter updates (Gavin, 2020). The LM
algorithm merges two numerical minimisation procedures: the gradient descent (GD) method and the Gauss-
Newton (GN) method. In the case of the former, by updating the parameters in the direction of steepest descent,
the sum of the squared errors is reduced. For the latter, the assumption is made that the least squares function is
locally quadratic in the parameters, and through this assumption the sum of the squared errors is reduced thereby
allowing the minimum of the quadratic to be found (Gavin, 2020). When the parameters are far removed from
their optimal value, the LM behaves akin to the GD method, and when the parameters are close to their optimal
target, the LM acts like the GN method (Gavin, 2020). In providing a mathematical explanation of the integration
of the LM algorithm with GMDH (and SVD for the initial guess), some background terminology is first
introduced. With a given data set \(d(t_i, y_i)\) and a model function \(\Phi(x; t_i)\), the difference of the functions is obtained
with \(r_i(x) = \Phi(x; t_i) - y_i\) noting that \(y_i\) is the y component of data point \(t_i\) (Croeze et al., 2012). Thus, defining
the objective function (cost function) as it applies in problems of least squares:

\[
f(x) = \frac{1}{2} \sum_{i=1}^{m} r_i^2(x) \tag{45}
\]

Minimising \(f(x)\) allows the parameters to be found that match most accurately the model to the observed data.
Each residual \(r_i\) is a smooth function from \(\mathbb{R}^n\) to \(\mathbb{R}\). The residual vector of \(m\) components is given by:

\[
r(x) = (r_1(x), r_2(x), \ldots, r_m(x))^T \tag{46}
\]

This allows eq. (45) to be rewritten using the residual vector:

\[
f(x) = \frac{1}{2} \|r(x)\|^2 \tag{47}
\]

The gradient of the residual vector \(r(x)\) is required when calculating the gradient of the objective function \(f(x)\).
The Jacobian \(J(x)\) is a matrix with elements all \(\nabla r_i(x)\). Both the gradient and the Hessian of the objective can be
expressed explicitly in terms of the Jacobian:

\[
\nabla f(x) = \sum_{i=1}^{m} r_i(x) \nabla r_i(x) = J(x)^T r(x) \tag{48}
\]

\[
\nabla^2 f(x) = \sum_{i=1}^{m} \nabla r_i(x) \nabla r_i(x)^T + \sum_{i=1}^{m} r_i(x) \nabla^2 r_i(x) = J(x)^T J(x) + \sum_{i=1}^{m} r_i(x) \nabla^2 r_i(x) \tag{49}
\]
It is noted that it is a requirement of the Hessian Matrix to be positive definite for least-squares problems. In the instance when both the residual and the solution are extremely close, $\nabla^2 f(x)$ approximation can be made using only the first term. This approximation is utilised in both the GN and LM methods (Croeze et al., 2012; Lourakis, 2005). Developed by Levenberg (1944) and Marquardt (1963) as a method for solving nonlinear least squares problems that brings together the methods of GD and GN (Lourakis, 2005). Given that both methods complement each other in their advantages, Levenberg proposed the algorithm (Ranganathan, 2004)

$$x_{i+1} = x_i - (H + \lambda I)^{-1} \nabla f(x_i) \quad (50)$$

noting that $H$ is the Hessian matrix being evaluated at $x_i$ and $\lambda$ is the damping parameter.

The disadvantage of this algorithm being, should $\lambda$ be large, then $H$ is not used. This problem was solved by Marquardt when the identity matrix in eq. (50) was replaced by the diagonal of $H$ which delivered the LM algorithm (Ranganathan, 2004).

$$x_{i+1} = x_i - (H + \lambda \text{diag}[H])^{-1} \nabla f(x_i) \quad (51)$$

It is noted that the inverse is often achieved by implementing SVD (Ranganathan, 2004).

An equal but alternative representation is presented utilising the Jacobian, as detailed by Manfre (2021):

$$(f^T J + \lambda I) \alpha_i = f^T [\hat{y}_i - f(a_n)] \quad (52)$$

with $\hat{y}_i$ the known output, and the Jacobian Matrix (Transtrum et al., 2011) is the matrix of the residuals with respect to the parameters (p. 1).

Step size can be calculated by rearranging eq. (67) to deliver

$$\alpha_i = (f^T J + \lambda I)^{-1} f^T [f(x, a) - \hat{y}_i] \quad (53)$$

The integration of LM with GMDH is presented in pseudocode (Transtrum et al., 2011; Mulashani et al., 2021).

1. Input rainfall data into GMDH
2. Determine the initial information for the LM-GMDH structure – initial point $x_0$, damping parameter $\lambda$,
   $\lambda_1, \lambda_1$ for damping term adjustment
3. Estimate the parameter weights
4. Evaluate the Jacobian at the initial parameter value and the residuals
5. Calculate the metric $\Gamma = J^T J + \lambda I$, the objective function $f(x) = \frac{1}{2} |r(x)|^2$, and $\nabla f(x) = J^T r$
6. Calculate the new residuals \( r_{new} \) at the location provided by \( x_{new} = x - \Gamma^{-1}\nabla f(x) \), then calculate the objective function at the new location \( f(x)_{new} = \frac{1}{2} |r(x)_{new}|^2 \).

7. If \( f(x)_{new} < f(x) \), accept the step, assign \( x_{new} \) to \( x \) and set \( r = r_{new} \) then assign \( \lambda = \frac{1}{\lambda_{down}} \) else reject the step, retain the old parameter guess \( x \) and residuals, setting \( \lambda = \lambda_{up} \).

8. Convergence assessment. If there is convergence, return \( x \) as the best fitting parameters. If there is no convergence but the step was accepted, calculate the Jacobian at the new parameter values and return to step 5.

9. Assess the PDs within the current layer against the external criterion.

10. Retain the best performing neurons.

11. If the current layer is the final layer, terminate, else return to step 2.

---

**Fuzzy GMDH**

In non-fuzzy or standard GMDH, the residuals between the observed output values and estimated output values are assumed to be Gaussian thereby allowing parameter estimation through linear regression (Anastasakis & Mort, 2001). This assumption though does not always hold meaning OLS is not appropriate (Anastasakis & Mort, 2001). It follows that real world systems majorly adhere to Zadeh’s principle of incompatibility thus the modelling procedure and theory that is more appropriate is fuzzy (Anastasakis & Mort, 2001). In dealing with the fuzzy phenomenon and fuzzy data Tanaka et al. (1989) suggest utilising possibility measures to describe the fuzzy system equation. Variations between observed and model values are normally categorised as measurement errors within a standard regression model (Tanaka et al., 1989). Within the context of fuzzy, the system parameters are considered responsible, correspondingly the deviations are reflected in possibilistic linear systems (Tanaka et al., 1989). In fuzzy GMDH (FGMDH), the architecture remains unchanged, but the PDs contain fuzzy parameters which to be found require possibilistic linear regression (Anastasakis & Mort, 2001). A mathematical description as detailed by Hayashi and Tanaka (1990); Fuzzy number \( M, \mu_M: \mathbb{R} \to [0,1] \), satisfies:

\[ M_\lambda = \{ x : \mu_M(x) \geq \lambda \} \rightarrow \text{closed interval} \]  

\[ \exists x \text{ such that } \mu_M(x) = 1 \]  

\[ \mu_M(\lambda x_1 + (1 - \lambda)x_2) \geq \mu_M(x_1) \wedge \mu_M(x_2) \text{ with } \lambda \in [0,1] \]  

An alternative to linear or quadratic polynomials that form the PDs are orthogonal polynomials (Zaychenko & Zaychenko, 2019). Their orthogonality precipitates faster coefficient determination when compared with non-
orthogonal polynomials, and the coefficients are not dependent upon the initial polynomial degree (Zaychenko &
Zaychenko, 2019). Chebyshev’s polynomial approximation is a form of regression that minimise autocorrelation
between the model response and the sampling locations (Nakajima, 2006). Chebyshev orthogonal polynomials
are especially well suited to equally spaced sample points (Nakajima 2006), which is the case with the BoM
rainfall data. A mathematical representation of the general case of Chebyshev’s orthogonal polynomials can be
found within Zaychenko and Zayets (2001). PDs featuring trigonometric polynomials are also an option with
FGMDH with a mathematical outline provided within Zaychenko and Zaychenko’s (2019) paper. The advantage
that FGMDH offers over standard MIA GMDH is the zero requirement to use OLS for determining the PD
parameters. Fuzzy GMDH can be used with both crisp and fuzzy regressors which extends its application for
potential rainfall modelling. Zaychenko and Zaychenko (2019) advise high accuracy of results when modelling
with FGMDH for forecasting financial processes. Shi et al. (2020) found from their research that FGMDH
identified regional economic bottlenecks within China with high recognition accuracy when compared to standard
GMDH. Panchal et al. (2014) in their study of rainfall – runoff modelling using fuzzy logic returned a coefficient
of determination of 0.988. They however did not combine it with GMDH, but the illustration shows the gains that
can be achieved with the fuzzy approach. No papers have been located to date that uses FGMDH for rainfall
modelling.

Neuro-Fuzzy GMDH

Neuro-fuzzy is derived from the application of Gaussian radial basis functions (GRBF) as PDs (Anastasakis &
Mort, 2001). Radial basis functions (RBF) are univariate functions that provide a mechanism to approximate
multivariate functions through linear combinations of terms (Buhmann, 2010). In considering the GRBF as fuzzy
production rules, Mamdani (1976) explored a fuzzy reasoning rule: Let $x_1 = A_{k1}$ and $x_2 = A_{k2}$ then output $y =
\omega_k$. Letting the Gaussian membership function $A_{kj}$ of the $k$th fuzzy rule ($k = 1, \ldots, A$) in the domain of the $j$th
input variables $x_j (j = 1, 2)$ be defined as:

$$A_{kj}(x_j) = \exp\left(-\frac{(x_j - a_{kj})^2}{b_{kj}}\right)$$  (57)

with parameters $a_{kj}$ and $b_{kj}$ being given for each rule. Let $\omega_k$ be a real number from the conclusion of the $k$th
rule suggesting model output $y$. The degree of compatibility from the proposition is:

$$\mu_k = \prod_{j=1}^{2} A_{kj}(x_j)$$  (58)
with the model output defined

\[ y = \sum_{k=1}^{K} \mu_k w_k \]  
(59)

The GRBF \( \leftrightarrow y \rightarrow \) a neural network composed of three layers (Poggio & Girosi, 1990). The neuro-fuzzy GMDH (NFGMDH) is formed through implementation of this fuzzy reasoning model as the PDs Mucciardi (1972, as cited in Nagasaka et al., 1995). NFGMDH follows the same paradigm as standard GMDH. Within the context of the GRBF as PDs, the inputs from the \( \beta \)th model and \( \rho \)th layer form the output variables of the \((\beta - 1)\)th and \( \beta \)th model within the \((\rho - 1)\)th layer (Nagasaka et al., 1995; Najafzadeh, 2015). The mathematical expression for ascertaining \( y^{\rho\beta} \) is illustrated as:

\[ y^{\rho\beta} = f\left(y^{\rho-1,\beta-1}, y^{\rho-1,\beta}\right) = \sum_{k=1}^{K} \mu^\rho_w^\beta w^\rho_w^\beta \]  
(60)

\[ \mu_k^{\rho\beta} = \exp\left\{-\frac{(y^{\rho-1,\beta-1} - a_k^{\rho\beta})^2}{b_k^{\rho\beta} - b_k^{\rho\beta}}\right\} \]  
(61)

with \( \mu_k^{\rho\beta} \) and \( w_k^{\rho\beta} \) as the \( k \)th Gaussian function with its associated weight parameter respectively. Additionally, \( a_k^{\rho\beta} \) and \( b_k^{\rho\beta} \) are the Gaussian parameters that feature for the \( i \)th input variable supplied by the \( \beta \)th model and \( \rho \)th layer (Nagasaka et al., 1995; Najafzadeh, 2015). The complete model output is illustrated with:

\[ y = \frac{1}{M} \sum_{m=1}^{M} y^{\rho\beta} \]  
(62)

Through each iteration of the network model construction, the error parameter is:

\[ E = \frac{(y^* - y)^2}{2} \]  
(63)

with \( y^* \) the predicted output value (Nagasaka et al., 1995; Najafzadeh, 2015).

In terms of successful outcomes, Nagasaka et al. (1995) reported from their research in utilising NFGMDH for modelling grinding characteristics with GRBF PDs an improved result when compared with standard GMDH. Yousefpour and Ahmadpour (2011) from their research into air pollution predictions with NFGMDH found improved results when compared to those obtained from multilayer perceptron (MLP) neural network. Miyagishi et al. (2010) found an improvement when compared to the Kalman filter for temperature prediction when the seasonal change was moderate. No papers were located where NFGMDH has been used to model rainfall data.
Exogenous Data

Exogenous data was used within the research study undertaken by Moosavi (2019) investigating the impact of rainfall on natural disasters. GMDH was utilised with three signal processing approaches: EEMD, WT, and WPT, in addition to exogenous data. Two distinct datasets were used for the prediction of rainfall, one without exogenous data, and one with. The exogenous data were evaporation, minimum and maximum temperature, and humidity. All data was on a monthly timescale. Modelling rainfall with one-to-four-month forecasts using just GMDH delivered poor performance. With inclusion of the exogenous data, the GMDH modelling was an improvement on the modelling that excluded the exogenous data, but overall was still poor. Hybridising GMDH with each of the signal processing techniques in turn improved the rainfall modelling, which further improved when exogenous data was included. Exogenous data was used by Mendoza et al. (2020) for their research study into rainfall in Ecuador. The modelling was undertaken using a dynamic harmonic regression framework where global climate signals formed the exogeneous data. Their results were favourable illustrating greater reliability and robustness against limitations within the data. In a study undertaken by Sarveswararao and Ravi (2020) covering ATM cash demand forecasting, GMDH was supplied with a dummy exogenous variable in the form of day of the week. Supplied with this information, the GMDH modelling delivered an improvement in the symmetric mean absolute percentage error (SMAPE) when compared to the data modelling without inclusion of the dummy exogenous variable. From the evidence presented, the GMDH paradigm is accepting of the inclusion of exogenous data which evidently improves the success of the model.

Discussion

As a mechanism for enhancing the useability and possibly the capacity of standard GMDH to model (within the context of the initiating study) rainfall data, this empirical study considers alternative platforms for further investigation. Gaining an appreciation and deeper understanding of the GMDH algorithm could potentially allow for greater modelling success. Luzar and Witczak (2014) utilised MATLAB for implementing the algorithmic idea of GMDH. Advantages of this approach include but are not limited to state vector dimensions, choice of multiple selection criterions, and stopping criteria. Onwubolu (2014) details the implementation of GMDH in C, and Onwubolu (2016) the implementation of the GMDH paradigm into MATLAB is detailed. Dag and Yozgatligil (2016) investigated short term forecasting with a GMDH R-package running within the R-workspace. This option allows the implementation of different transfer functions precipitating the selection that delivers the smallest prediction MSE. The Hilbert-Huang Transform (HHT), although not specifically within the scope of this paper is
worth mentioning as the HHT is a combination of EMD and Hilbert spectral analysis (HSA) (Huang & Shen, 2014). The HHT is potentially suitable for the analysis of nonlinear and nonstationary data (Huang & Shen, 2014). Bowman et al. (2013) discuss the HHT package that is available for the R programming language. As GMDH can also be implemented with the R platform, the potential for pre-processing the rainfall data with HHT cannot be denied. This paper suggests that this combination would be a viable option for further investigation.

Conclusion

The GMDH hierarchical structure of the COMBI and MIA algorithms has been discussed with an investigation of state variable distribution, their classification, and the synthesis of PDs. The limitations of OLS in determining PD coefficients, the inherent potential for biased estimates, the significance of fuzzy input data, and the integration of Gödel’s incompleteness theorem initiating the requirement of an external criterion. The methods for modelling improvement covered hybridising with LSSVM, which was shown to be successful for time series forecasting, although modelling of rainfall data appears not to have yet taken place. In implementing LSSVM, normalised data is a requirement, which is distinct from how data is provided to standard GMDH. This provides an opportunity to test the application of normalised data for both standard GMDH and when hybridised and by allowing for further research in utilising LSSVM with GMDH for rainfall modelling and forecasting. The integration of Kalman filters into the GMDH paradigm for the formation of PD parameters was detailed with success in this application in the modelling of dynamic systems and fault detection. This pairing has not yet been applied to rainfall modelling and forecasting which leaves open the opportunity for further research in this area. By hybridising GMDH with three signal processing techniques; EEMD, WT, and WPT, delivered an improvement for rainfall prediction within the context of natural disasters with WPTGMDH delivering the largest improvement. It would be interesting to see the outcome of further rainfall modelling within the Australian context using these signal processing techniques, particularly if it covered the same six LGAs in the first author’s original study.

Enhanced MIA-GMDH received exposure to rainfall modelling but delivered results that were deemed unsatisfactory, quite distinct from the success in the temperature modelling. This paper recommends that this would be an ideal area for further research encompassing rainfall modelling through hybridisation with one or more of the applications already discussed. The hybridising of the LM algorithm with GMDH for determining the PD parameters was found to be very successful for the high degree of accuracy associated with inventory control. This pairing has not yet been found to have received application of rainfall modelling, so that provides an opportunity for further research in this area. It would be an advantage to see to what benefit the modelling has for
rainfall forecasting when the PD parameters are not within the confines of OLS. Fuzzy GMDH finds favour when modelling real world systems given their inherently fuzzy nature resulting in adherence to Zadeh’s principle of incompatibility. The PD parameters within FGMDH require determination by possibilistic linear regression or they take an alternative form of orthogonal polynomials. FGMDH can see application of modelling with both crisp and fuzzy input vector regressors thereby widening the base of application. Fuzzy logic as distinct from FGMDH was used with great success in a study of rainfall-runoff modelling returning a coefficient of determination of 0.988. This paper highlights the potential for further research into rainfall modelling and forecasting with the application of FGMDH. NFGMDH which utilises GRBFs as PDs found success in terms of an improvement in previous studies that modelled griding characteristics and for air pollution predictions, when compared to standard GMDH and an MLP. No studies have been found where NFGMDH has been applied to rainfall modelling, but the opportunity exists for further research in this area. The inclusion of exogenous data delivered an improvement in the modelling for each of the cited studies. This paper considers it a worthwhile proposition to include exogenous data in all the modelling avenues discussed, as the evidence suggests an improvement in the model would be achieved.

Statements & Declarations

Ethical Approval
Not applicable

Consent to Participate
Not applicable

Consent to Publish
all authors agreed with the content and that all gave explicit consent to submit and that they obtained consent from the university.

Authors Contributions
All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Ron Lake with the supervision of STMLD Senevirathna and Saeed Shaeri. The first draft of the manuscript was written by Ron Lake and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.
Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Availability of data and materials

Not applicable

References


http://dx.doi.org/10.1098/rsta.2017.0258


https://doi.org/10.35940/ijrte.d7402.118419


Understanding of Support Vector Machine (SVM) (ai-pool.com)


Microsoft Word - Research Report.doc (gmdhsoftware.com)


The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis on JSTOR


[https://doi.org/10.1142/8804](https://doi.org/10.1142/8804)


[AR5 Synthesis Report: Climate Change 2014 — IPCC](https://doi.org/10.1142/8804)


[heuristic.pdf (gmdh.net)](https://heuristic.pdf)


[Estimating the coefficients of polynomials in parametric GMDH algorithms by the improved instrumental variables method (researchgate.net)](https://researchgate.net)


View of Vol 1 No 01 (2017): Journal of Applied Geospatial Information (JAGI). (polibatam.ac.id)

(PDF) An unscented Kalman filter approach to designing GMDH neural networks: Application to the tunnel furnace (researchgate.net)


(PDF) A GMDH TOOLBOX FOR NEURAL NETWORK-BASED MODELLING (researchgate.net)


Comparison of Inductive Versus Deductive Learning Networks (wolfram.com)


Advances in the linguistic synthesis of fuzzy controllers - ScienceDirect


The Interesting World of Non-Linear Regressions | by Diego Manfre | Towards Data Science


An Algorithm for Least-Squares Estimation of Nonlinear Parameters on JSTOR

Robust GMDH-type neural network with unscented Kalman filter for non-linear systems - M.
Masoumnezhad, A. Jamali, N. Nariman-zadeh, 2016 (sagepub.com)

Local rainfall modelling based on global climate information: A data-based approach.

neurofuzzy GMDH. *Researchgate.*

TEMPERATURE PREDICTION FROM REGIONAL SPECTRAL MODEL BY NEUROFUZZY GMDH
(researchgate.net)

https://doi.org/10.1155/DDNS/2006/45348

https://doi.org/10.1007/s11269-016-1507-3

Moosavi, V. (2019). Prediction of rainfall as one of the main variables in several natural disasters. In
Pourghasemi H., Rossi M. (eds) Natural Hazards GIS-Based Spatial Modeling Using Data
Mining Techniques. Advances in Natural and Technological Hazards Research, vol 48. Springer,
Cham.
https://doi.org/10.1007/978-3-319-73383-8_8

https://doi.org/10.1007/s11053-020-09799-w


(PDF) An unscented Kalman filter in designing dynamic GMDH neural networks for robust fault detection (researchgate.net)


https://doi.org/10.1016/j.energy.2021.121915


Retrieved from the University of Sydney


http://dx.doi.org/10.1016/j.oceaneng.2015.01.014


note_num_cheby.pdf (paulklein.ca)


Poggio, T., & Girosi, F. (1990). Regularization algorithms for learning that are equivalent to multilayer networks. *American Association for the Advancement of Science, 247*(4945), 978 – 982. [Regularization Algorithms for Learning that are Equivalent to Multilayer Networks on JSTOR](https://www.jstor.org/article/27721885)


Page | 45

(PDF) ATM Cash demand forecasting in an Indian Bank with chaos and deep learning
(researchgate.net)


D:\VIJAY JI\ARF_2020 Journals\2 (arfjournals.com)


AM221_lecture9.pdf (harvard.edu)


Effect of Feature Standardization on Linear Support Vector Machines | by Dave Sotelo | Towards Data Science


Retrieved from the University of Sydney

Page | 46


Phys. Rev. E 83, 036701 (2011) - Geometry of nonlinear least squares with applications to sloppy models and optimization (aps.org)


CiteSeerX — SELF-ORGANIZING DATA MINING FOR WEATHER FORECASTING (psu.edu)


https://doi.org/10.1016/0304-3975(94)90222-4


A guide to machine learning algorithms and their applications | SAS


