

Research Article

Forecasting Time Series Movement Direction with Hybrid Methodology

Salwa Waeto,^{1,2} Khanchit Chuarkham,³ and Arthit Intarasit^{1,2}

¹Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Pattani Campus, Pattani 94000, Thailand

²Centre of Excellence in Mathematics, Commission on Higher Education, Ratchathewi, Bangkok 10400, Thailand

³Faculty of Commerce and Management, Prince of Songkla University, Trang Campus, Trang 92000, Thailand

Correspondence should be addressed to Arthit Intarasit; a.intarasit@gmail.com

Received 30 November 2016; Revised 26 March 2017; Accepted 4 April 2017; Published 31 July 2017

Academic Editor: Dejian Lai

Copyright © 2017 Salwa Waeto et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Forecasting the tendencies of time series is a challenging task which gives better understanding. The purpose of this paper is to present the hybrid model of support vector regression associated with Autoregressive Integrated Moving Average which is formulated by hybrid methodology. The proposed model is more convenient for practical usage. The tendencies modeling of time series for Thailand's south insurgency is of interest in this research article. The empirical results using the time series of monthly number of deaths, injuries, and incidents for Thailand's south insurgency indicate that the proposed hybrid model is an effective way to construct an estimated hybrid model which is better than the classical time series model or support vector regression. The best forecast accuracy is performed by using mean square error.

1. Introduction

Time series modeling and forecasting are a challenge for describing dynamic phenomena and pattern behavior of the time series. In recent years, the issue of accurate Thailand's south insurgency trends has been receiving more attention. There are many research papers that studied the unrest in southern Thailand. According to the database of Deep South Watch [1], Jitpiromsri and Mccargo [2] and Jitpiromsri [3] reported the trends of Thailand's south insurgency using diagram for comparing the monthly number of the unrest incidents. By applying a polynomial least-square regression, they provided the forecasting model for describing the unrest incidents in the south of Thailand. This polynomial is not indeed fitting the monthly number of the unrest incidents as well.

In this study, we would like to identify patterns and trends of Thailand's south insurgency and to evaluate the accuracy of model for modeling and forecasting. By doing this, we use the traditional regression models such as Autoregressive

(AR), Moving Average (MA), Autoregressive Moving Average (ARMA), and Autoregressive Integrated Moving Average (ARIMA). These models are also called the Box-Jenkins models.

In general, time series data of Thailand's south insurgency can be categorized as nonstationary time by using Box-Jenkins methodology. Then an estimated model of time series data of Thailand's south insurgency can be obtained by support vector regression (SVR). We aim to combine ARIMA and SVR for making an adequately estimated model in order to forecast time series of Thailand's south insurgency.

This paper is organized as follows. Section 2 provides some backgrounds of mathematical theories related to time series modeling and forecasting and SVR. The detail of proposed hybrid model is explained in Section 3. Section 4 gives experimental results obtaining the proposed hybrid model with the first difference in time series of Thailand's south insurgency. Finally, the main conclusions are summarized in Section 5.

2. Background and Mathematical Theory

2.1. Autoregressive Integrated Moving Average Modeling. Three basic methods for forecasting time series are naïve model, exponential smoothing model, and ARIMA model. The first two models relate to a random walk as the formulation of the model. In this section, ARIMA model will be reviewed.

An autoregressive model of order p abbreviated as AR(p) model is

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + w_t = \sum_{i=1}^p \phi_i y_{t-i}, \quad (1)$$

where y_t is stationary, ϕ_1, \dots, ϕ_p are constants ($p \neq 0$), and w_t is a white noise series with zero mean and variance σ_w^2 . AR(p) model of (1) predicts the current value y_t by the p past function $y_{t-1}, y_{t-2}, \dots, y_{t-p}$ which explains y_t as a linear combination of $y_{t-1}, y_{t-2}, \dots, y_{t-p}$.

The Moving Average model of order q abbreviated as MA(q) model is

$$y_t = \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q} + w_t = \sum_{j=1}^q \theta_j w_{t-j}, \quad (2)$$

where y_t is stationary, ϕ_1, \dots, ϕ_q are constants ($q \neq 0$), and w_t is a Gaussian white noise series with mean zero. MA(q) model of (2) explains the current value y_t by a linear combination of the q white noise $w_{t-1}, w_{t-2}, \dots, w_{t-q}$.

Autoregressive Moving Average model abbreviated as ARMA(p, q) model developed by Box and Jenkins [4] is defined by the combined autoregressive and the Moving Average model. It has the form

$$y_t = \delta + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \theta_j w_{t-j} + w_t. \quad (3)$$

According to the original Box-Jenkins methodology, an integrated process is the stationary process obtained by differenced a nonstationary process. The stationary ARMA(p, q) process after being differenced d times is denoted by ARIMA(p, d, q):

$$\Delta^d y_t = \delta + \phi_1 \Delta^d y_{t-1} + \cdots + \phi_p \Delta^d y_{t-p} + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q} + w_t, \quad (4)$$

where Δ^d denoted d th difference time series [5]. These models are as foundation model for time series forecasting.

2.2. Box-Jenkins Methodology. Plots of autocorrelation function (acf) and partial autocorrelation function (pacf) are the main tools in order to identify parameters for AR, MA, ARMA, and ARIMA models. AR(p) is used to obtain an estimated model for time series when the acf exhibits tendency to die down quickly, either by an exponential decay or by a damped sine wave whereas the pacf exhibits tendency to show spike (significant autocorrelation) for lags up to p and then will die down immediately.

Opposite to AR(p), MA(q) is used to obtain an estimated model of time series when the acf exhibits tendency to die down quickly, either by an exponential decay or by a damped sine wave whereas the pacf exhibits tendency to show spike (significant autocorrelation) for lags up to p and then will die down immediately.

A mixed process ARMA(p, q) is suggested when either the acf or the pacf exits tend to show spike for lags up, respectively, to p and q and then die down quickly, either by an exponential decay or by a damped sine wave. Proceeding diagnostic checking to identify p and q for the mixed process ARMA(p, q) which is able to fit to times series is the best performance [6].

This identification as described in this section will be important to diagnose a model of our study.

2.3. Hybrid Models. In recent years, the forecasting model used in the literature can be classified into three categories: statistical models, artificial intelligence model (AI), and hybrid model.

Statistical models are known as time series models including naïve model, AR model, MA model, ARIMA model, exponential smoothing, and generalized autoregressive conditional heteroskedasticity (GARCH) volatility which aim to utilize time series analysis to identify the pattern of time series and provide the future value based on the obtained pattern.

ARIMA model is known as Box-Jenkins model [4] which includes AR and MA models identified by Box-Jenkins methodology. These models are based on the assumption that the time series under study are stationary and linear which means that the relationship between the input and output series is linear.

AI models are the second kinds of forecast time series, practically artificial neural networks (ANNs), genetic algorithm (GA), and supported vector machine (SVM). AI models can capture nonlinear pattern and improved forecast performance.

Many of the literatures introduce a hybrid model in order to capture the linear and nonlinear characteristics in time series. Wang et al. [7] reported that using a statistical model alone or using an AI model alone are not adequate in making forecasts for stock price time series.

2.4. Hybrid Methodology. A hybrid model is described by a combination of models with mixed methodology for formulation. Many literatures suggested that time series consists of linear L_t and nonlinear N_t as in the form

$$y_t = L_t + N_t. \quad (5)$$

An estimated model of (5) is formulated as follows: using linear statistic model to obtain an estimated model of linear component L_t denoted by \hat{L}_t and after that modeling the residual $y_t - \hat{L}_t$ which contains only the nonlinear relationship to obtain an estimated model of nonlinear component N_t denoted by \hat{N}_t .

Zhang [8] utilized the hybrid model by introducing the estimated model of (5) in the form $\hat{y}_t = \hat{L}_t + \hat{N}_t$,

where \widehat{L}_t is prescribed by ARIMA model and \widehat{N}_t is prescribed by feedforward neural networks model. Modified Zhang's hybrid approach with estimated \widehat{N}_t by support vector machine (SVM) model can be found in many literatures, for example, De Oliveira and Ludermir [9], while Aladag et al. [10] estimated \widehat{N}_t by Elman's recurrent neural networks (ERNN) model and applied to Canadian Lynx data.

2.5. Supported Vector Regression. Let the dot product space \mathfrak{R}^d be our data universe with vectors $\mathbf{x} \in \mathfrak{R}^d$ as objects. Let S be a sample set such that $S \subset \mathfrak{R}^d$. Let $f : \mathfrak{R}^d \rightarrow \mathfrak{R}$ be the target function. Let $D = \{(\mathbf{x}, y) \mid \mathbf{x} \in S \text{ and } y = f(\mathbf{x})\}$ be the training set.

The regression problem is to find the best approximate model $\tilde{f} : \mathfrak{R}^d \rightarrow \mathfrak{R}$ for the true underlying function f mapping input \mathbf{x} to output y by using D such that $\tilde{f}(\mathbf{x}) \equiv f(\mathbf{x})$.

The regression problem is classified as linear or nonlinear type. For the linear regression model, the best approximate model \tilde{f} can be obtained from the set of possible functions with the following set of specifications:

$$\{\tilde{f} \mid \tilde{f}(\mathbf{x}) = \boldsymbol{\omega}^T \mathbf{x} + b, \boldsymbol{\omega} \in \mathfrak{R}^d, b \in \mathfrak{R}\}, \quad (6)$$

where $\boldsymbol{\omega}$ is a weight vector and b is a constant.

Generally, in order to describe nonlinear relationship between input and output, the SVR allied $\Phi : \mathfrak{R}^d \rightarrow \mathfrak{F}$ transform the nonlinear regression problem in the lower dimension input space \mathfrak{R}^d into a linear regression problem in a high dimension feature space \mathfrak{F} . In the new space \mathfrak{F} , a linear model \tilde{f} is formulated, which represents a nonlinear model in the original space:

$$\tilde{y} = \tilde{f}(\mathbf{x}, \boldsymbol{\omega}) = \langle \boldsymbol{\omega}, \Phi(\mathbf{x}) \rangle + b, \quad (7)$$

where $\langle \cdot, \cdot \rangle$ denotes the dot product in \mathfrak{F} . Linear SVR model \tilde{f} in (6) is obtained from (7) by using the identity function $\Phi(\mathbf{x}) \rightarrow \mathbf{x}$.

Performing SVR to fit linear regression \tilde{f} to the training data by estimate $\boldsymbol{\omega}$ and b in (7) as minimization of the following regularized function:

$$\underset{\boldsymbol{\omega}, b}{\text{minimize}} \quad R(C) = \frac{1}{2} \|\boldsymbol{\omega}\|^2 + C \sum_{i=1}^{\ell} L_2^{\varepsilon}(\tilde{f}(\mathbf{x}_i), y_i), \quad (8)$$

where both C and ε are user-given parameters and $L_2^{\varepsilon}(\tilde{f}(\mathbf{x}), y)$ is quadratic ε -insensitive loss function defined by $L_2^{\varepsilon}(y, f(\mathbf{x})) = |y - f(\mathbf{x})|_{\varepsilon}^2$.

The following two propositions related to the formulation of an estimated model. These propositions are modified from [11, 12] for our study.

Proposition 1. *Given a regression training set $D = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_{\ell}, y_{\ell})\} \subseteq \mathbb{R}^n \times \mathbb{R}$ the optimal support vector regression model is computed by $\tilde{f}^*(\mathbf{x}) = \langle \boldsymbol{\omega}^*, \mathbf{x} \rangle - b^*$, where*

the parameters $\boldsymbol{\omega}^$ and b^* solved the following optimization problem:*

$$\begin{aligned} \underset{\boldsymbol{\omega}, b}{\text{minimize}} \quad & R(C) = \frac{1}{2} \|\boldsymbol{\omega}\|^2 + \frac{C}{2} \sum_{i=1}^{\ell} (\xi_i^2 + \widehat{\xi}_i^2), \\ \text{subject to} \quad & y_i - \tilde{f}(\mathbf{x}_i) \leq \varepsilon + \xi_i, \\ & \tilde{f}(\mathbf{x}_i) - y_i \leq \varepsilon + \widehat{\xi}_i, \\ & \xi_i, \widehat{\xi}_i \geq 0, \\ & \forall i = 1, \dots, \ell \end{aligned} \quad (9)$$

hold with $\tilde{y} = \tilde{f}(\mathbf{x}) = \langle \boldsymbol{\omega}, \mathbf{x} \rangle + b$.

The constant C is called the penalty constant which is trade-off between margin maximization and the minimization of the slack variables.

Proposition 2. *Given a regression training set $D = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_{\ell}, y_{\ell})\} \subseteq \mathbb{R}^n \times \mathbb{R}$ the optimal support vector regression model is computed by $\tilde{f}^*(\mathbf{x}) = \boldsymbol{\omega}^* \cdot \mathbf{x} - b^*$, where*

$$\begin{aligned} \boldsymbol{\omega}^* &= \sum_{i=1}^{\ell} (\alpha_i^* - \tilde{\alpha}_i^*) \mathbf{x}_i, \\ b^* &= \frac{1}{\ell} \sum_{i=1}^{\ell} \boldsymbol{\omega}^* \cdot \mathbf{x}_i - \left(y_i + \varepsilon + \frac{\alpha_i^*}{C} \right) \end{aligned} \quad (10)$$

and α^* are the parameters solved by the following dual quadratic optimization problem:

$$\begin{aligned} \underset{\boldsymbol{\alpha}, \tilde{\boldsymbol{\alpha}}}{\text{minimize}} \quad & \sum_{i=1}^{\ell} y_i (\alpha_i - \tilde{\alpha}_i) - \varepsilon \sum_{i=1}^{\ell} (\alpha_i + \tilde{\alpha}_i) \\ & - \frac{1}{2} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \left((\alpha_i - \tilde{\alpha}_i) (\alpha_j - \tilde{\alpha}_j) \mathbf{x}_i \cdot \mathbf{x}_j + \frac{1}{C} \delta_{ij} \right), \end{aligned} \quad (11)$$

$$\text{subject to} \quad \sum_{i=1}^{\ell} (\alpha_i - \tilde{\alpha}_i) = 0, \quad \alpha_i \geq 0, \tilde{\alpha}_i \geq 0, \quad \forall i = 1, \dots, \ell.$$

The parameter $\boldsymbol{\omega}^*$ is obtained by α_i^* and $\tilde{\alpha}_i^*$ which satisfied optimization (11). The parameter b^* is solved as follows: obtain $\xi_i = \alpha_i^*/C$ from (11) and substitute $\tilde{f}^*(\mathbf{x}) = \boldsymbol{\omega}^* \cdot \mathbf{x} - b^*$ and $\xi_i = \alpha_i^*/C$ in the constraint $\tilde{f}(\mathbf{x}_i) = y_i + \varepsilon + \widehat{\xi}_i$; then solve for b_i^* . Define b^* as the average of b_i^* .

The optimal regression model is obtained by substituting $\boldsymbol{\omega}^*$ into b^* and into $\tilde{f}^*(\mathbf{x}) = \boldsymbol{\omega}^* \cdot \mathbf{x} - b^*$ where we have the following lemma.

Lemma 3. *The optimal regression model is*

$$\begin{aligned} \tilde{f}^*(\mathbf{x}) &= \sum_{i=1}^{\ell} (\alpha_i^* - \tilde{\alpha}_i^*) \mathbf{x}_i \cdot \mathbf{x} - \frac{1}{\ell} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} (\alpha_i^* - \tilde{\alpha}_i^*) \mathbf{x}_i \\ &\quad \cdot \mathbf{x}_j - \left(y_j + \varepsilon + \frac{\alpha_j^*}{C} \right), \end{aligned} \quad (12)$$

where the coefficient $(\alpha_i^* - \tilde{\alpha}_i^*)$ is nonzero as support vector. The optimal regression model $\tilde{f}^*(\mathbf{x})$ depends only on the support vectors.

3. Formulation of the Proposed Model

In this section, we want to formulate the proposed model. We begin by using the hybrid models that combine several models in order to reduce the risk of using an inappropriate model, obtain the results that are more accurate than the previous one, and improve overall forecasting performance.

Assume that (y_t) is the under-study time series based on the assumption of linear and stationary time series. Then, we use the Box-Jenkins methodology to check behavior of (y_t) . After this step, we can get a suitable model of AR(p) or MA(q) or ARIMA(p, d, q) in order to estimate \tilde{y}_t^A . By fitting under-study time series with (2) or (3) or (4), we can get \tilde{y}_t^A in the form $g^\Delta(a_{t-1}, a_{t-2}, \dots, a_{t-p}, w_{t-1}, w_{t-2}, \dots, w_{t-q})$. According to Lemma 3, perform SVR for under-study time series in order to evaluate \tilde{y}_t^S from (12). This model is a function of its past N values in the form $g^*(\beta_{t_1}, \beta_{t_2}, \dots, \beta_{t_N}, b^*)$ with $t_1 \leq \dots \leq t_{N-1} \leq t_N$.

Consider a time set $\{t_1, \dots, t_k, t_{k+1}, \dots, t_N\}$ with $y_{t_1} = a$ and $y_{t_N} = b \neq a$. There is only one single time point (necessarily from time t_1 to time t_N) precisely on time $\tilde{y} \approx \tilde{y}_{t^*}^S$ satisfying $e_{t^*}^S < e_{t^*}^A$ for all $t^* \in \{t_1, \dots, t_k\}$ and $\tilde{y} \approx \tilde{y}_{t^\Delta}^A$ satisfying $e_{t^\Delta}^A < e_{t^\Delta}^S$ for all $t^\Delta \in \{t_{k+1}, \dots, t_N\}$. The proposed hybrid model \tilde{y} is the estimated model defined by setting $\tilde{y} \approx \tilde{y}_{t^*}^S$ for all $t^* \in \{t_1, \dots, t_k\}$ and vanishing otherwise and setting $\tilde{y} \approx \tilde{y}_{t^\Delta}^A$ for all $t^\Delta \in \{t_{k+1}, \dots, t_N\}$ and vanishing otherwise. The proposed hybrid model can be extended to include two or more time intervals.

The under-study time series (y_t) is initially modeled by the proposed hybrid model as follows:

$$y_t = \tilde{y}_t + e_t = (\tilde{y}_{t^*}^S + \tilde{y}_{t^\Delta}^A) + e_t, \quad (13)$$

where e_t is residuals of the time series model in the time t that is as obtained from (13),

$$e_t = y_t - \tilde{y}_t = (y_{t^*} - \tilde{y}_{t^*}^S) + (y_{t^\Delta} - \tilde{y}_{t^\Delta}^A) = e_{t^*}^S + e_{t^\Delta}^A, \quad (14)$$

where $e_{t^*}^S$ and $e_{t^\Delta}^A$ are residuals of the under-study time series model in the time t of the estimated model of $\tilde{y}_{t^*}^S$, respectively, to $e_{t^\Delta}^A$.

4. Application of the Proposed Hybrid Model to Thailand's South Insurgency Movement Direction Forecasting

4.1. Data Set. In this research, we are interested in studying the unrest in the four southern provinces of Thailand, particularly in Pattani, Yala, Narathiwat, and parts of Songkla. We consider the monthly number of deaths, injuries, and incidents in these provinces. At the time of working research, we can get the latest data from Deep South Watch (DSW) [1] and Deep South Coordination Center (DSCC) [13]. By

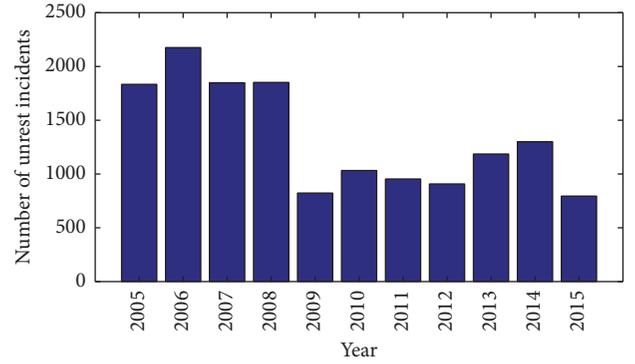


FIGURE 1: Number of unrest incidents in the four southern provinces of Thailand (Pattani, Yala, Narathiwat, and Songkla) from 2005 to 2015.

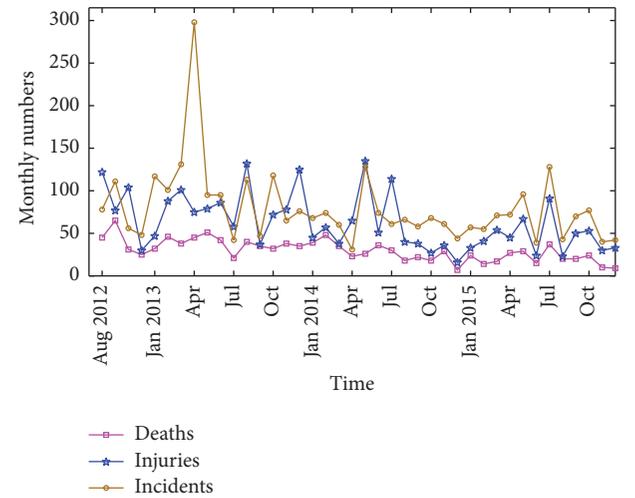


FIGURE 2: Monthly number of deaths, injuries, and incidents for unrest in the four southern provinces of Thailand.

using the proposed hybrid model, our aim is to formulate an estimate model for the trend of the number of deaths, injuries, and incidents in these regions.

Figure 1 illustrates a diagram of the number of unrest incidents in the four southern provinces from 2005 to 2015. This diagram presents a high frequency of the number of unrest incidents with a small fluctuation in the first period (2005 to 2008), a decreasing frequency of the number of unrest incidents in the middle period (2009 to 2012), and an increasing frequency of the number of unrest incidents from 2013 to 2014, the lowest frequency in 2015.

The data series of our study consists of 40 months of deaths, injuries, and incidents in the four southern provinces of Thailand from September 2012 to December 2015.

Figure 2 presents three graphs describing three data series of monthly number for deaths, injuries, and incidents. It shows that the graph of deaths is in the bottom for all periods of time, while the graph of injuries is in the middle between the graphs of deaths and incidents in almost all periods of time. Moreover, the graph of incidents is in the top in almost all periods of time.

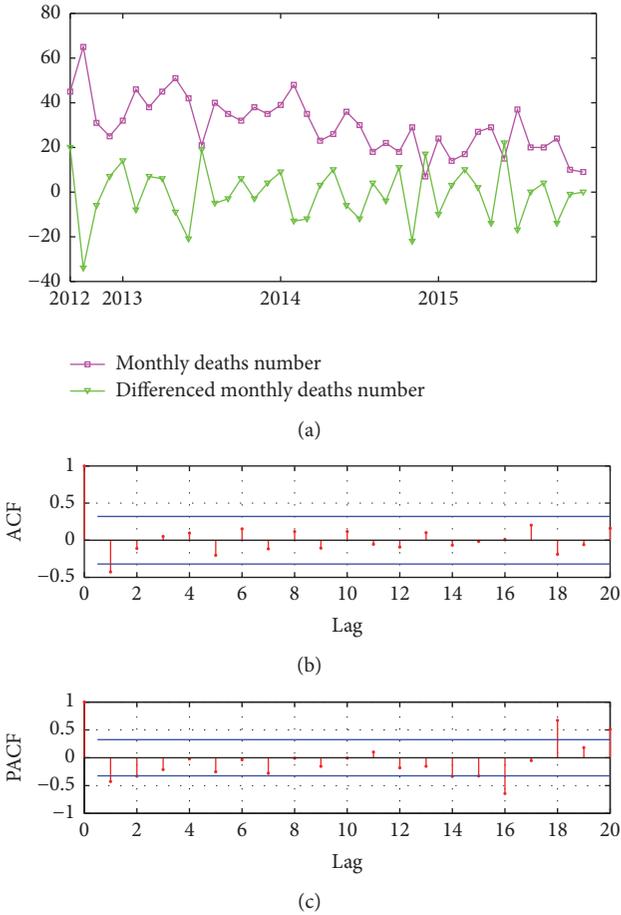


FIGURE 3: (a) Monthly number of deaths is plotted against its first differenced series, acf (b) and pacf (c) plots for the first difference in monthly number of deaths.

From Figure 2, we can see that the number of incidents is not necessary to be equal to the sum of numbers of deaths and injuries. Sometimes, there is an unrest incident; no deaths or no injures occurs. Or there are high numbers of deaths and injuries in some incidents.

Monthly numbers of injuries and incidents are apparently stationary. A candidate model for monthly number of two data series can be determined by plotting of acf and pacf. However, the monthly number of deaths exhibits a linear trend in the mean since it has a clear downward slope.

Figure 3 shows comparing of monthly number of deaths plotted against its first differenced series for monthly number of deaths (a) and plotting of acf (b) and pacf (c). The data series of injures plotted against its first differenced series is shown in Figure 4 and the data series of incidents plotted against its first differenced series is shown in Figure 5.

Plotting of the first differenced series (Figures 3, 4, and 5) shows that it looks like a stationary process, although plotting acf and pacf of series of deaths, injuries, and incidents cannot clearly identify parameter for constructing an estimated model formulated by the ARIMA model.

The acf for the first difference in monthly number of deaths tends to die down quickly whereas the pacf tends to

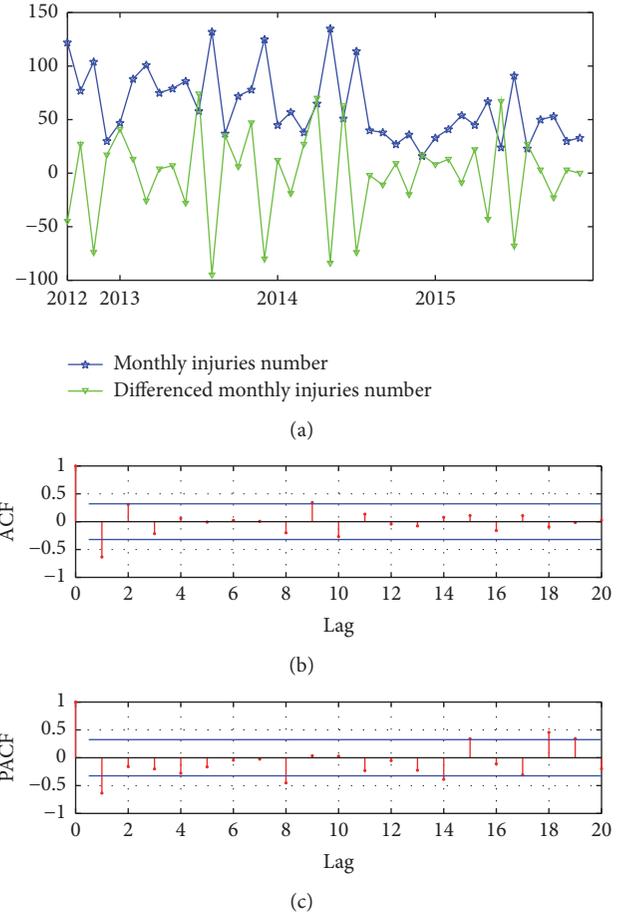


FIGURE 4: (a) Monthly number of injuries is plotted against its first differenced series, acf (b) and pacf (c) plots for the first difference in monthly number of injuries.

show spike for lags up to 1 which ignores significant spikes in each plot when it is outside the limits. This suggests that the first difference in monthly number of deaths can be a model as an AR(1).

Similarly, the first differenced series of injures and incidents can be a model as an AR(1). After checking of residual in diagnosis stage, this indicates that ARMA(2, 3) is a candidate model for formulating an estimate model for the first difference in monthly number of deaths and injuries. MA(1) is also a candidate model for the first difference in monthly number of incidents. With notation of ARIMA(p, d, q), ARIMA(2, 1, 3) is an estimated model for monthly number of deaths and injuries and ARIMA(0, 1, 1) for monthly number of incidents.

Table 1 reports mean square error (mse) of three estimated models for monthly number of deaths, injuries, and incidents formulated by ARIMA, SVR, and hybrid. The mean square error of the formal model is calculated by choosing the best trajectory: 1×10^6 trajectories simulated by ARIMA for each series.

Plotting a convergent of mean square error is calculated from monthly number and an estimated model with 2,500,

TABLE I: Some reports of mean square error for fitting and forecasting the series.

Time series	Model		
	ARIMA*	SVR	Hybrid
Deaths series	9.4383	7.0882	0.7922
Injuries series	28.0352	20.4161	0.9921
Incident series	41.8077	31.7669	1.469

Note. ARIMA(2,1,3) for monthly number of deaths and injuries and ARIMA(0,1,1) for monthly number of incidents.

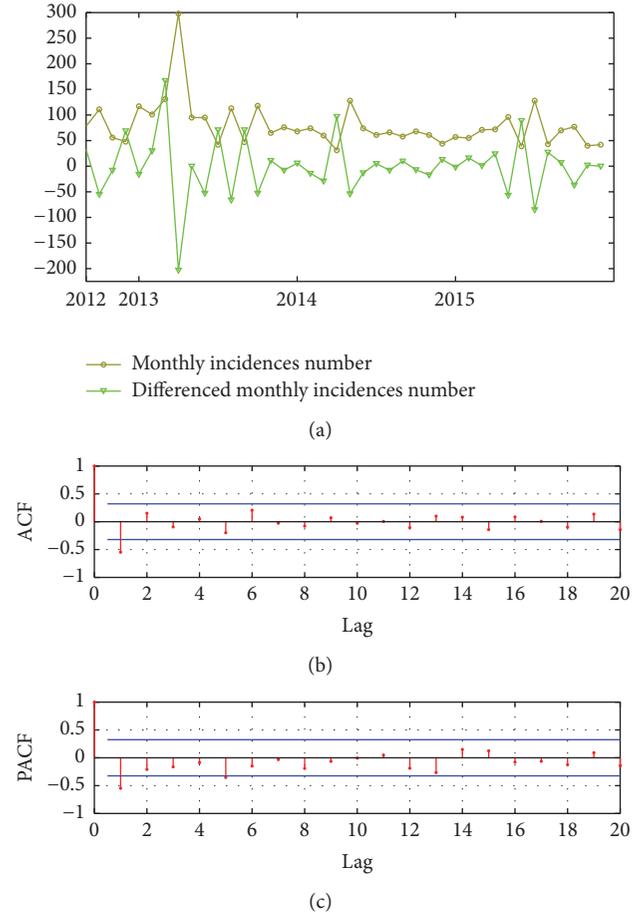


FIGURE 5: (a) Monthly number of incidents is plotted against its first differenced series, acf (b) and pacf (c) plots for its first difference in monthly number of incidents.

5,000, ..., 1×10^6 trajectories for monthly number of deaths, injuries, and incidents illustrated in Figure 6.

Setting $\varepsilon = 0.0025$, $c = 150000$, $\gamma = 3.25$, and $b = 2.75$ for SVR model and using ARIMA(2, 1, 3) model in order to select from the best trajectory from 1×10^6 trial trajectories, then both models are combined in order to formulate an estimated model for monthly number of deaths: $\tilde{y}_t = \tilde{y}_t^S + \tilde{y}_t^A$, where

$$\begin{aligned} \tilde{y}_t^A = & -1.079 - 0.046\Delta y_{t-1} - 0.751\Delta y_{t-2} - 0.915w_{t-1} \\ & - 0.915w_{t-2} - w_{t-3} + w_t, \end{aligned} \quad (15)$$

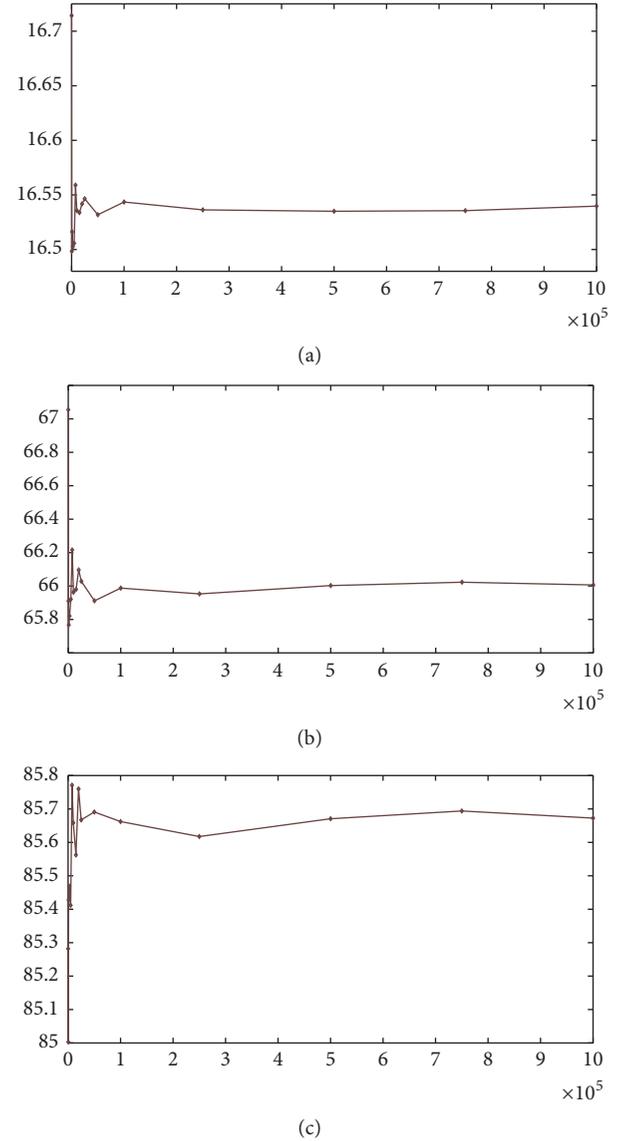


FIGURE 6: Fitting performance for monthly number of deaths (a) and injuries (b) with ARIMA(2, 1, 3) and for monthly number of incidents (c) with ARIMA(0, 1, 1).

$t \in \{35, \dots, 40\}$, and $\tilde{y}_t^S = \sum_{i=1}^{35} \beta_i^* K(\mathbf{x}_i, \mathbf{x}) + b^*$ where $t \in \{1, \dots, 36\}$, $b^* = -0.7115$, and β_i^* , $i = 1, \dots, 35$, where it has specified a value.

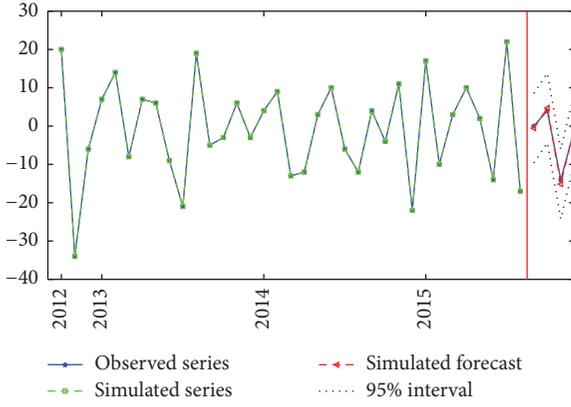


FIGURE 7: The actual, fitted, and forecasted series by hybrid model for series of deaths.

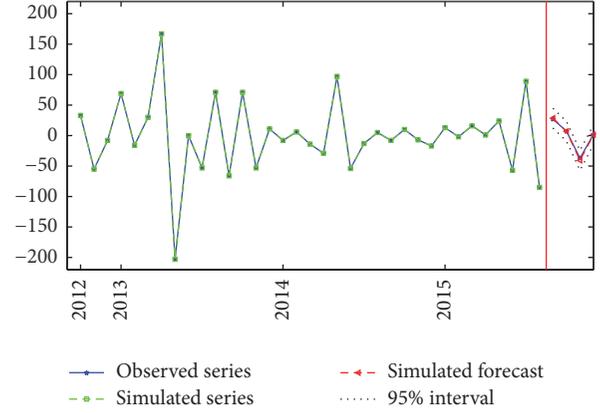


FIGURE 9: The actual, fitted, and forecasted series by hybrid model for number of incidents.

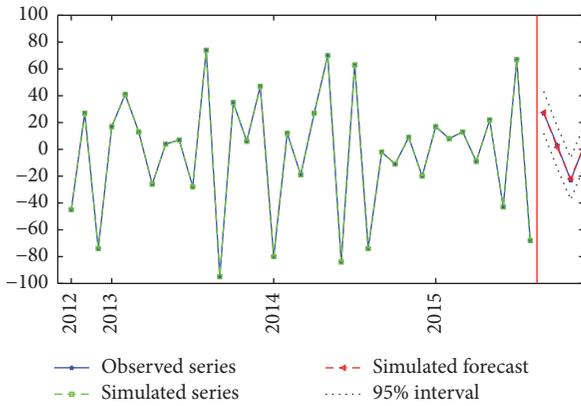


FIGURE 8: The actual, fitted, and forecasted series by hybrid model for series of injuries.

Predictive performance of SVR-ARIMA(2, 1, 3) hybrid model for monthly number of deaths and injuries, respectively, is shown in Figure 7.

In the same way, for monthly number of injuries, setting $\epsilon = 0.025$, $c = 350000$, $\gamma = 2.755$, and $b = 0.00125$ for SVR model and using ARIMA(2, 1, 3) model in order to select from the best trajectory from 1×10^6 trial trajectories, both models are combined in order to formulate an estimated model for monthly number of injuries: $\tilde{y}_t = \tilde{y}_t^S + \tilde{y}_t^A$, where

$$\begin{aligned} \tilde{y}_t^A = & -5.851 - 1.560\Delta y_{t-1} - 0.773\Delta y_{t-2} - 0.618w_{t-1} \\ & - 0.618w_{t-2} - w_{t-3} + w_t, \end{aligned} \quad (16)$$

$t \in \{35, \dots, 40\}$, and $\tilde{y}_t^S = \sum_{i=1}^{35} \beta_i^* K(\mathbf{x}_i, x) + b^*$, where $t \in \{1, \dots, 36\}$, $b^* = -3.0611$, and β_i^* , $i = 1, \dots, 35$, where it has specified a value.

Predictive performance of SVR-ARIMA(2, 3) hybrid model for difference monthly number of injuries is shown in Figure 8.

For monthly number of incidents, set $\epsilon = 0.025$, $c = 200000$, $\gamma = 1.555$, and $b = 0.725$ for SVR model and use ARIMA(0, 1, 1) model to select from the best trajectory

from 1×10^6 trial trajectories. Then these two models are combined in order to formulate an estimated model for monthly number of incidents: $\tilde{y}_t = \tilde{y}_t^S + \tilde{y}_t^A$, where

$$\tilde{y}_t^A = -0.193 - w_{t-1} + w_t, \quad (17)$$

$t \in \{36, \dots, 40\}$, and $\tilde{y}_t^S = \sum_{i=1}^{35} \beta_i^* K(\mathbf{x}_i, x) + b^*$, where $t \in \{1, \dots, 35\}$, $b^* = -1.4121$, and β_i^* , $i = 1, \dots, 35$, where it has specified a value.

Predictive performance of SVR-ARIMA(0, 1, 1) hybrid model for monthly number of incidents is shown in Figure 9.

5. Conclusions

The hybrid SVR-ARIMA model has been investigated to formulate time series model of monthly number of Thailand's south insurgency in this study. In particular, we consider the first difference in monthly number of deaths, injuries, and incidents in Pattani, Yala, Narathiwat, and Songkla provinces in 40 months from September 2012 to December 2015. According to the hybrid methodology, the SVR-ARIMA(p, d, q) model is obtained by combining ARIMA(p, d, q) and SVR model. Plotting of autocorrelation and partial autocorrelation indicates that the first difference in monthly number of deaths, injuries, and incidents is linear and stationary.

The test results of the estimated model are obtained from the proposed hybrid model and compared with the estimated model of the AR(p), MA(q), ARIMA(p, d, q), and SVM models. This presents the fact that the proposed hybrid model performs better than the remaining models. For time series of Thailand's south insurgency, SVR-ARIMA(2, 1, 3) is the estimated model for monthly number of deaths and injuries and SVR-ARIMA(0, 1, 1) is the estimated model for monthly number of incidents. In particular, SVR-ARIMA(2, 1, 3) consists of two components: the first component uses the SVR model in order to formulate the estimated model for historical data and the second component uses the ARIMA model in order to formulate the estimated model for the unseen value in the short future.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors gratefully acknowledge the Deep South Coordination Center (DSCC) and Deep South Watch (DSW) for providing the data. This research was supported by grant funds from the Centre of Excellence in Mathematics, the Commission on Higher Education, Thailand.

References

- [1] Deep South Watch (DSW). Deep South Incident Database for the Thailand. <http://www.deepsouthwatch.org/dsid>.
- [2] S. Jitpiromsri and D. Mccargo, "The Southern Thai Conflict Six Years On: Insurgency, Not Just Crime," *Contemporary Southeast Asia*, vol. 32, no. 2, pp. 156–183, 2010.
- [3] S. Jitpiromsri, *An Inconvenient Truth about the Deep South Violent Conflict: A Decade of Chaotic, Constrained Realities and Uncertain Resolution*, Center for Conflict Studies and Cultural Diversity (CSCD), Deep South Watch (DSW), Prince of Songkla University, 2010.
- [4] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, Calif, USA, 1976.
- [5] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, Prentice Hall, Englewood Cliffs, NJ, USA, 3rd edition, 1994.
- [6] D. Asteriou and S. G. Hall, *Applied Econometrics*, Palgrave Macmillan, China, 2nd edition.
- [7] J.-J. Wang, J.-Z. Wang, Z.-G. Zhang, and S.-P. Guo, "Stock index forecasting based on a hybrid model," *Omega*, vol. 40, no. 6, pp. 758–766, 2012.
- [8] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.
- [9] J. F. L. De Oliveira and T. B. Ludermir, "A hybrid evolutionary system for parameter optimization and lag selection in time series forecasting," in *Proceedings of the 3rd Brazilian Conference on Intelligent Systems, BRACIS 2014*, pp. 73–78, Sao Paulo, Brazil, October 2014.
- [10] C. H. Aladag, E. Egrioglu, and C. Kadilar, "Forecasting nonlinear time series with a hybrid methodology," *Applied Mathematics Letters*, vol. 22, no. 9, pp. 1467–1470, 2009.
- [11] L. Hamal, *KnowledgeDiscovery with Support Vector Machines*, John Wiley & Sons, Hoboken, NJ, USA, 2009.
- [12] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines*, Cambridge University Press, Cambridge, UK, 2000.
- [13] Deep South Coordination Center (DSCC). Database of the violent event and victims in southern Thailand. <http://dsrd.pn.psu.ac.th/webnew/index.php/database.html>.




Hindawi

Submit your manuscripts at
<https://www.hindawi.com>

