Short-Term Forecasting of Ocean Current Patterns Based on HF Radar Observation Data Using Grid-Based Spatial ARIMA Model

Ratchanont Pongto[1], Nopparat Wiwattanaphon[1], Peerapon Lekpong[1], Siam Lawawirojwong[2]
Siwapon Srisonphan[3], Kerk F. Kee[4] & Kulsawasd Jitkajornwanich[1,*]

[1]Dept. of Computer Science, Faculty of Science, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand
[2]Geo-Informatics & Space Technology Development Agency (PO), Bangkok 10210, Thailand
[3]Dept. of Electrical Engineering, Faculty of Engineering, Kasetsart University, Bangkok 10900, Thailand
[4]College of Media & Communication, Texas Tech University, Lubbock, TX 79409, USA
57050313@kmitl.ac.th, 57050254@kmitl.ac.th, 57050297@kmitl.ac.th, siam@gistda.or.th,
fengspsr@ku.ac.th, kerk.kee@gmail.com, kulsawasd.ji@kmitl.ac.th*

***Abstract*—Marine natural disasters have direct impacts on countries as well as their residents living on and near the coast. Warning and monitoring system can aid in reducing the loss of lives in the event of a disaster. HF (high frequency) radar, an IoT-enabled ocean surface current monitoring system, implementation is one of the first attempts towards achieving this goal. Although HF systems can monitor sea current patterns in terms of speed and direction for each of the pixels of the coverage area, it fails to predict future values, which are essential to many applications such as oil-spill trajectory prediction, water quality control and management, and optimized sea navigation. In this paper, we propose a model, called the grid-based spatial ARIMA (auto-regressive integrated moving average), to estimate the forecast values. As a result, the full potential of the HF systems can be utilized. The method considers not only observations of POI (point of interest), but also its neighboring pixels when predicting future values. The proposed method is implemented and compared with other existing approaches, including baseline, kNN, traditional ARIMA model, and LSTM (long short-term memory) techniques. The experimental results showed that our approach outperformed other methods in V comp prediction (with RMSEs of 6.23265) where a configuration of (p,d,q) as (2,0,1) and a historical dataset of 1 day and 7 hours prior was found to be the best combination.**

Keywords-ocean surface current; ARIMA; HF radar; spatio-temporal data mining.

I. INTRODUCTION

Marine natural disasters have direct impacts on countries as well as their residents living on and near the coast. Warning and monitoring system can aid in reducing the loss of lives in the event of a disaster. HF (high frequency) radar station implementation is one of the first attempts towards achieving this goal. For example, 18 HF coastal radar stations were installed along the Gulf of Thailand in order to monitor marine natural disasters as well as relevant risks by retrieving various parameters of sea surface, such as wave data and height, surface current data and direction, and current acceleration, covering the sea body 10-60 km away from the coast [1]. The HF systems record ocean surface current information on an hourly basis, including the volume of current body mass and the direction of the current.

Although HF systems can monitor sea surface patterns in terms of speed and direction for each pixel of the coverage area, it fails to predict future values, which are essential to many applications such as oil-spill trajectory prediction [2], water quality control and management, and sea navigation optimization. In this paper, we propose a model, called grid-based spatial ARIMA (auto-regressive integrated moving average), to estimate the forecast values so that the full potential of HF systems can be utilized. The traditional ARIMA model is adopted and modified to better suit our data set configuration with which spatial and temporal dimensions of data are incorporated. The data sets are stored and processed by using a cluster of servers, Windows Server 2012, MS SQLServer, Apache2 and python scripts, and the results are visualized by using JSON, Node.js, Google Maps APIs as well as HTML, CSS, Javascript and PHP.

The organization of this paper is as follows. Section 2 discusses related work in spatio-temporal mining and prediction. Section 3 describes HF radar data properties and characteristics, and it demonstrates the proposed methodology followed by Section 4, which presents the experimental results. Finally, we conclude our work and discuss some future work in Section 5.

II. RELATED WORK

Cheng et al. [3] surveyed several applications of spatio-temporal data mining and knowledge discovering (STDMKD) with an emphasis on wildfire forecasting. In their research, a wildfire forecasting system is implemented based on wildfire

Longitude	Latitude	Ucomp	Vcomp	VectorFlag	UStdDev	VStdDev	Covariance
100.1022264	12.2539686	10.533	-6.536	0	4.95	3.86	-17.18
100.1206079	12.253999	-1.258	-14.391	0	3.47	2.65	-7.76
100.1389894	12.2540281	-5.002	-15.802	0	3.32	2.92	-8.32
100.1573709	12.254056	-7.461	-15.855	0	4.96	4.9	-23.1
100.1757525	12.2540825	-8.445	-15.344	0	6.06	6.56	-38.59
100.194134	12.2541078	-6.314	-12.525	0	5.35	6.51	-33.31
100.230897	12.2541544	0.148	-4.985	0	5	6.24	-28.66
100.2492786	12.2541758	-1.812	-9.258	0	5.26	7.99	-39.39
100.2676601	12.2541959	-2.031	-7.061	0	4.14	6.13	-23.25
100.2860416	12.2542147	-1.534	-3.541	0	2.55	2.62	-5.4
100.3044232	12.2542322	-4.322	-4.143	0	999	999	999

Figure 1. An example of TUV file and corresponding reported grids from HF radar

events occurred in a given area in Canada using three different methods, namely, ISTIFF, ARIMA, and STIFF, which are then evaluated and compared against each other. The result shows that ISTIFF has the highest accuracy.

According to [4], to accommodate large-scale spatio-temporal data analysis, the concepts of spatio-temporal data mining and related tools need to be fully understood. A long list of applications can be benefited from spatio-temporal data mining technique such as ecology system and environmental management security, transportation, geography, spread of decease, and climate and weather. The main takeaway from this work is that domain knowledge plays a very important role in tuning and perfecting algorithms for maximum efficiency.

In [5], not only were spatio-temporal data models proposed, indexing and querying methods were also developed. Indexing and querying help improve data retrieval performance as well as mining process efficiency as a whole. Spatio-temporal concepts are quite different from concepts that are only spatial or temporal in nature. For example, the spatial indexing focuses on static object while temporal indexing focuses on time series without the object shape bounding. Thus, the indexing for spatio-temporal data (e.g., data of moving objects) need to incorporate both the dimensions of time and space. There are two main approaches presented in this research; both are the variation of R-tree and grid-based indexing.

Spatio-temporal data involve describing the position as well as the status of objects at different time intervals. Spatio-temporal data can be found in many applications, such as traffic monitoring, environmental management, and weather forecasting. These applications collect data at different places and times in various formats, which in turns lead to challenges in processing and analyzing “big spatio-temporal data.” Some examples of large-scale spatio-temporal data include GPS information of a moving car, movement of a hurricane, and historical patterns of coastal changes. The spatio-temporal data can capture changes of spatial features (e.g., shape) at location *x* over time interval *t*. More examples of spatio-temporal applications in other domains are listed as follows:

- *Meteorology:* weather data, tornado and rainstorm movement, temperature, pressure, and drought at a given area.
- *Biology:* animal movement, mating behavior, migration and extinction.
- *Plant Science:* change in quality of harvest soil, land use management.
- *Forestry:* growth of wild forest fire, tree cutting and planting planning.
- *Medical:* development of cancer based on treatment history of patients, embryo monitoring.
- *Geophysics:* volcanic earthquake development.
- *Ecology:* causal relationships in environmental change, pollution monitoring.
- *Transportation:* traffic monitoring, vehicle control and tracking.

III. METHODOLOGY

In this section, we first describe our coastal radar data properties, followed by our grid-based spatial ARIMA model based on traditional ARIMA time series analysis.

A. *Coastal Radar Data Description*

The two main characteristics of ocean surface current are direction and velocity, which were used in the analysis in this paper. The data set was provided by Coastal Radar Data Center at Geo-Informatics & Space Technology Development Agency (Public Organization), Ministry of Science & Technology of Thailand (http://coastalradar.gistda.or.th). The data set was collected from January 2014 to October 2015. It is in TUV format, where the spatial information is embedded in the text file. The data were reported hourly and stored in a fixed grid format. There is a total of 15,799 files (hours) with 3,081 grid locations (*lat/lon*) observed in our data set. Each measurement (row) represents an average speed and a direction of the sea surface current over a 2x2 km boundary centered at the reported grid (pixel). A TUV file contains 21 columns based on WGS84 reference system and UTC + 0 (Atlantic/Reykjavik) time zone. The example of TUV file and corresponding study area are shown in Figure 1. As shown in the TUV file, many variables are captured by the HF radar system. Out of 21 attributes, only six of them are used in this research: *Longitude* in degrees (LOND), *Latitude* in degrees (LATD), *Ucomp* (VELU: current vector velocity eastern (X-axis) component in cm/s), *Vcomp* (VELV: current vector velocity northern (Y-axis) component in cm/s), overall *Velocity* (VELO) and *Direction* (HEAD) [9]. The *time* attribute is also used, which was indicated in the filename. These seven measures are selected because they represent direction and velocity, the two main ocean current observation characteristics that are needed for other significant GIS applications, including oil-spill trajectory tracking, search and rescue, water quality control and management, and optimized sea navigation or even hydrological disaster prediction.

B. *ARIMA Time Series Analysis*

ARIMA (auto-regressive integrated moving average) method, developed by George E.P. Box and Gwilym M. Jenkins in 1970, was used to analyze time series data for this paper. It is considered one of the most popular tools for short-term prediction; its RMSE rate is typically lower than other methods (such as trend analysis, exponential smoothing, and multiple regressions). Also, it is less complicated in modeling forecast system as multi-tier equations. Because ARIMA estimate weights for predicting future values based on previous time-point data, it is important to eliminate trends and seasonality existed in the data, which can be done in various ways, such as taking log derivatives and calculating means. The data are required to be "stationary," a non-biased form that is ready for analyzing the movement of the data at a desired point in space and surrounding areas. The ARIMA model consists of three main parts: 1) auto regressive model (AR), 2) moving average model (MA), and 3) integrated process (I).

1) Auto regressive model (AR)

Auto-regressive model describes how y_t observation value can be determined by $y_{t-1},\ldots,y_{t-p}$ (or previous p observation values). The *AR(p)* process can be written in an equation with rank p as follows [6]:

$$AP(p): x_t = \mu + \emptyset_1 x_{t-1} + \emptyset_2 x_{t-2} + \cdots + \emptyset_p x_{t-p} + \varepsilon_t \quad (1)$$

2) Moving average model (MA)

Moving average (MA) model estimates observed y_t value by focusing on the observation different or error values from $\varepsilon_{t-1},\ldots,\varepsilon_{t-q}$ (or preceding q tolerance). The *MA(q)* or moving system can be written as en equation in term of q as follows [6]:

$$MA(q): x_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \cdots - \theta_q\varepsilon_{t-q} \quad (2)$$

3) Integrated Process (I)

The integrated process depicts a time difference between current data and historical data in the last d periods. This process is necessary when the given

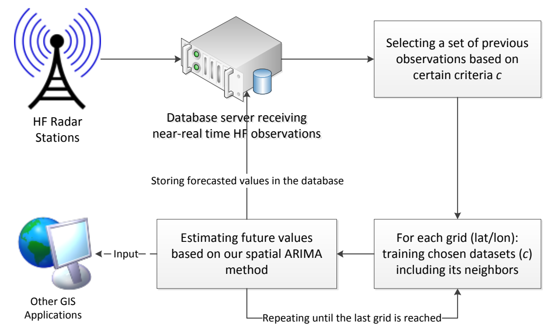

Figure 2. An overview of our methodology

time series are not stationary as ARIMA model only supports stationary dataset. In case of non-stationary data, it must be converted to stationary before ARIMA model can be applied. The process *I(d)* can be described as follows [6]:

$$\Delta_d x_t = \Delta_{d-1}(x_t - x_{t-1}) \quad (3)$$

Auto-regressive moving average model (ARMA) combines auto regressive (eq. 1) and moving average processes (eq. 2) together. It considers auto regressive system with p observations with moving average of q, with an assumption of data being stationary. The ARMA model can be expressed as an equation as follows:

$$y_t = \delta + \emptyset y_{t-2} + \cdots + \emptyset y_{t-p} + \varepsilon_t - \theta_1\varepsilon_{t-1} \quad (4)$$

ARIMA (auto-regressive integrated moving average model) model, on the other hand, consists of three processes mentioned above (eqs. 1-3), which can be defined as an equation in terms of *(p, d, q)* as follows:

$$\Delta_d y_t = \delta + \emptyset\Delta_d y_{t-1} + \emptyset\Delta_d y_{t-2} + \cdots + \emptyset\Delta_d y_{t-p} + \varepsilon_t - \ldots - \theta_q\varepsilon_{t-q} \quad (5)$$

Our spatial ARIMA model is based on the ARIMA model as opposed to the ARMA model. While the traditional ARIMA uses a single POI (lat/lon) when calculating future values, the proposed grid-based spatial ARIMA model utilizes the neighboring pixels as well as the coordinate point of interest, so the reliability of results can be enhanced. This is based on the assumption that a pixel and its surrounding ones should have similar movement patterns. We believe that this approach is a methodological innovation. An overview of our method is shown in Figure 2.

Additionally, according to [7], their ARIMA analysis is solely determined by the previous t time points when predicting $t+1$ value. However, considering only the previous values may not be sufficient for capturing the real and actual patterns of moving currents in the ocean. Thus, in our method,

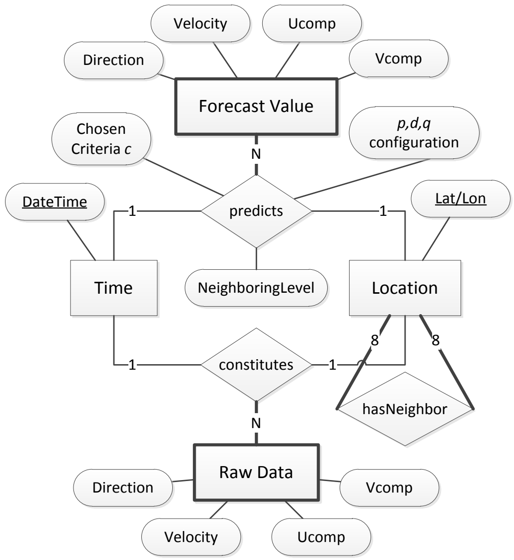

Figure 3. Conceptual storage structure of our raw and predicted values

we chose various ranges of data covering before and after the time point of interest for the experiments. The dataset is stored in a Microsoft SQL Server database with spatial features for easy neighboring pixels retrieval. The conceptual design of our database is shown in Figure 3.

IV. EXPERIMENTAL DESIGN & RESULTS

In our experiments, we estimated and compared the results with other existing methods (baseline, kNN, traditional ARIMA, and LSTM—with ~1,021 sites) using the same selected datasets. Each of the predicted values (i.e., U comp, V comp, velocity, direction) of these four data points is evaluated using RMSE and accuracy. Five main factors are considered: *p*, *d*, *q*, neighboring pixel level, as well as a chosen range of historical data (i.e., the last 7, 15, 24 hours within the last 1, 7, 10, and 15 days). In order to fully represent the complex characteristics of the complete data set. A small subset of data of four sites may not be realistic enough for modeling complex spatio-temporal phenomena. Therefore, we performed the experiments with a greater range of datasets (approx. 350 sites) in order to capture the complexity under the actual environment and to maximize the real performance and accuracy of our predictive model.

Tables 1-4 are amongst other experiments we performed, describing how the "best" predicted U comp, V comp, velocity and direction values of the four data points were selected, after we performed the experiments using all the possible combinations of different ranges of historical datasets (time periods and number of days) as described above. The actual recorded values were listed in the leftmost column for assessing accuracy of predictions. We selected the best prediction of the four values based on the lowest RMSEs among the ranges (borderlined in red). From column two onwards, different configurations of (*p*,*d*,*q*) are tested with the level of `neighboring pixel = 1` and the range that provided the best prediction is listed in the table caption. In the same fashion, Tables 5-8, the RMSEs of Tables 1-4, were calculated. The four data points consisted of the following:

Data points with SRID = 4326	Latitude	Longitude
Grid#1	12.9229064	100.1194929
Grid#2	12.253999	100.1206079
Grid#3	12.2540281	100.1389894
Grid#4	12.254056	100.1573709

The map visualization of the data points is shown in Figure 4. For ease and convenience in interpreting results of Tables 1-4, we color-coded the results with different cell colors using an *accuracy scale*, with green being the most accurate and red being the least accurate while gray representing the undetermined. Similarly and in Tables 5-8, we color-coded the numerical values of the predicted results using a *validity scheme*, where the most valid predicted results are based on obtaining values from all four sites shown in Figure 4. In other words, black = validity achieved with 4 data points, green = 3 data points, purple = 2 data points, red = 1 data point, and no color = 0 data point. In Table 9, all the performed experiments were combined and compared to with previous approaches [7] using the greater range of datasets (~350 sites). Figure 5 shows the screenshot of the actual implementation of our forecasting system for the Gulf of Thailand, visualized on the map, including both direction (arrows) and velocity (colors).

In Tables 1-4, the "Real" column describes the true recorded values of the date of 2014-04-01 04:00 from the four coordinates given above, and the other columns are values predicted by our grid-based spatial ARIMA method.

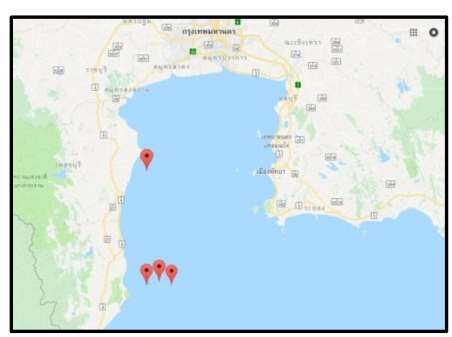

Figure 4. The selected four data points in our experiments

Table 1. Best U comp prediction, with the last 7 hrs within the last 10 days of data

Real	1,0,1	1,0,2	1,0,3	1,0,4	1,0,5	1,1,1	1,1,2	1,1,3	1,1,4	1,1,5
0.832	0.6365	0.5348				-0.0061				
4.252	2.2294					2.5266	2.7775	1.3284		
-4.428	1.5501	1.593			0.3228	1.5195	1.4773	1.4822	-0.6842	-0.0248
-0.205	1.8492			0.4829		1.6645	1.6963	1.2574	0.2014	0.2996
Real	2,0,1	2,0,2	2,0,3	2,0,4	2,0,5	2,1,1	2,1,2	2,1,3	2,1,4	2,1,5
0.832	0.7487	1.0880			0.9081	0.0182				
4.252	2.1349	4.1388				2.4508	2.7449			
-4.428	1.6826	0.3602				1.4230	-0.0481	-1.9748		0.3627
-0.205	0.4070	-1.8336		-0.1929		1.4196	0.1370			
Real	3,0,1	3,0,2	3,0,3	3,0,4	3,0,5	3,1,1	3,1,2	3,1,3	3,1,4	3,1,5
0.832	0.8923	0.5132	0.0807	0.2821		0.0381	0.3373			
4.252	3.1423					7.9026				
-4.428	1.6263	0.2520	-1.1654		-0.7529	1.1762	-0.3188	2.5374	-2.4852	-1.7593
-0.205	-0.6455	0.2824	-0.8668	-0.5805		-0.6069	0.7628			
Real	4,0,1	4,0,2	4,0,3	4,0,4	4,0,5	4,1,1	4,1,2	4,1,3	4,1,4	4,1,5
0.832	0.9977	0.8627	0.1040	1.8933		0.1658				
4.252	3.4451	6.2698	2.8311			8.2083				
-4.428	0.0354	-2.0829	-1.1121	2.3181	2.5157	0.0482	-1.6341	-2.2695		-1.7531
-0.205	-2.3049	-0.2923	-0.9412	-0.8398	-0.8921	-1.2548				
Real	5,0,1	5,0,2	5,0,3	5,0,4	5,0,5	5,1,1	5,1,2	5,1,3	5,1,4	5,1,5
0.832				1.1219	2.3957	0.1213		-0.2391		
4.252	4.0275	0.8544	5.1259	0.0352	0.2543	4.6881	5.4211			
-4.428	0.1846		-0.8571		2.0129	0.0632	-1.6154	1.8766		-2.7939
-0.205	-1.3676	-0.9282	-1.0101	-1.6360	-0.8016	-1.2725				

Table 2. Best V comp prediction, with the last 7 hrs within the last 7 days of data

Real	1,0,1	1,0,2	1,0,3	1,0,4	1,0,5	1,1,1	1,1,2	1,1,3	1,1,4	1,1,5
-21.466	-1.744	8.115	-23.100		-27.839	-4.782				
-28.081	-6.663	-4.015				-0.391	-9.448		-23.726	-10.949
-32.392	-7.273		-27.265	-22.355	-24.483	-9.115	-10.285	-30.507	-44.331	-34.773
-26.654	-7.015	-3.848	-26.867	-41.271	-26.352	-9.306	-10.181	-30.607	-38.861	-31.884
Real	2,0,1	2,0,2	2,0,3	2,0,4	2,0,5	2,1,1	2,1,2	2,1,3	2,1,4	2,1,5
-21.466	-14.055		-15.204	-2.020	-27.881	-17.034				
-28.081	-6.188	-7.233	4.331			-22.045	-1.465	-12.918		
-32.392	-16.628			-16.346	833.070	-18.621	-30.428		-36.660	
-26.654	-15.651	17.582		-18.347	-27.055	-17.354	-31.084	-28.691	-39.041	-35.210
Real	3,0,1	3,0,2	3,0,3	3,0,4	3,0,5	3,1,1	3,1,2	3,1,3	3,1,4	3,1,5
-21.466	-17.158		-2.099	-33.875	-33.323	-33.135	-38.896	-20.461		
-28.081	-5.380	-4.917	-3.272			-14.392	-1.173	-3.418	95.949	
-32.392	-23.543	-23.302		-13.459		-42.060	-38.508	655.608		
-26.654	-23.450	-62.379		-14.621		-43.135		-46.603		
Real	4,0,1	4,0,2	4,0,3	4,0,4	4,0,5	4,1,1	4,1,2	4,1,3	4,1,4	4,1,5
-21.466	-11.733	-4.202	-0.292			-16.120	-20.549			
-28.081	-1.790	-4.869	0.981			-14.288	-1.323		-22.540	
-32.392	-25.558	-11.670		-12.785	-19.684	-37.844	-42.119		2592.632	
-26.654	-25.524	-12.646		-14.027			-42.700	-32.227		
Real	5,0,1	5,0,2	5,0,3	5,0,4	5,0,5	5,1,1	5,1,2	5,1,3	5,1,4	5,1,5
-21.466	-22.674	-4.251	-0.002	-21.097						
-28.081	-3.995	-10.687	-0.414			-10.918				
-32.392		-26.787	-16.847		-12.786					
-26.654	-129.272	-140.173		-16.783	-27.918					

Table 3. Best direction prediction, with the last 7 hrs within the last 7 days of data

Real	1,0,1	1,0,2	1,0,3	1,0,4	1,0,5	1,1,1	1,1,2	1,1,3	1,1,4	1,1,5
177.8									188.97	
171.4	164.52	159.70				154.35	163.84	157.51	173.14	
187.8	159.48					150.55	156.79	157.14		
180.4	159.68					144.42	163.36	161.41		163.94
Real	2,0,1	2,0,2	2,0,3	2,0,4	2,0,5	2,1,1	2,1,2	2,1,3	2,1,4	2,1,5
177.8	157.42						209.53			191.78
171.4	156.80	153.15				155.54	163.39	121.54	122.60	
187.8	179.16	177.25				155.20	155.69	159.73		
180.4	200.91	178.41				155.79	159.13			
Real	3,0,1	3,0,2	3,0,3	3,0,4	3,0,5	3,1,1	3,1,2	3,1,3	3,1,4	3,1,5
177.8						210.19	183.74	178.30	176.70	195.93
171.4						111.67				
187.8	178.18	181.38				153.52		148.55		
180.4	199.23	209.42				162.65	162.73			
Real	4,0,1	4,0,2	4,0,3	4,0,4	4,0,5	4,1,1	4,1,2	4,1,3	4,1,4	4,1,5
177.8	171.76	202.58	694.94			211.03	194.47	174.75	207.06	
171.4	166.95		160.80			148.98	159.92	147.77		
187.8	172.64	178.58				156.41				
180.4	186.96					168.23	92.78			
Real	5,0,1	5,0,2	5,0,3	5,0,4	5,0,5	5,1,1	5,1,2	5,1,3	5,1,4	5,1,5
177.8				200.44		197.06	202.22	202.69	200.18	
171.4	160.84	164.73				142.89	159.75	134.55		
187.8	174.20	154.00	184.74		-463.08	158.52	121.55	126.60		
180.4	193.37	166.57	197.01	172.55	184.65	173.35	173.60			

Table 4. Best velocity prediction, with the last 7 hrs within the last 15 days of data

Real	1,0,1	1,0,2	1,0,3	1,0,4	1,0,5	1,1,1	1,1,2	1,1,3	1,1,4	1,1,5
21.482	7.870	11.631	13.377	19.828	18.487	9.979		10.767		
28.401	12.814	12.707						24.365	22.671	
32.693	11.450	16.257	15.570	22.520	20.200	12.994			21.315	21.160
26.655	12.739	17.488	17.801	26.142	22.897	13.320		26.764	24.481	23.774
Real	2,0,1	2,0,2	2,0,3	2,0,4	2,0,5	2,1,1	2,1,2	2,1,3	2,1,4	2,1,5
21.482	7.988	12.768	19.989	18.694	18.421	11.526	16.431	16.722		18.731
28.401	12.675	12.723	16.041				18.581	21.792	21.774	
32.693	13.254	15.844	22.790	18.777	18.758	17.602	18.622	19.882	21.142	23.329
26.655	13.257	17.694	25.090	22.162	21.454	18.852	20.440	21.350	19.390	25.545
Real	3,0,1	3,0,2	3,0,3	3,0,4	3,0,5	3,1,1	3,1,2	3,1,3	3,1,4	3,1,5
21.482	11.876	16.161			18.541	11.204	16.422	16.893	23.189	18.485
28.401	16.401		20.833			20.081	20.549	21.794		19.478
32.693	17.389	18.194	19.412	19.576	22.723	17.611	18.666	20.071	22.073	22.771
26.655	18.619	19.836	20.775	19.156	24.962	18.583	20.241	21.607	23.676	25.554
Real	4,0,1	4,0,2	4,0,3	4,0,4	4,0,5	4,1,1	4,1,2	4,1,3	4,1,4	4,1,5
21.482	11.518	16.145	16.778		18.237	15.954	18.135	18.503	17.971	18.240
28.401	19.216	19.659	20.823	22.191		20.624	21.179	20.970		
32.693	17.382	18.279	19.606	22.216	21.658	22.385	24.098	25.720	23.818	23.455
26.655	18.348	19.681	20.993	23.338	24.725	24.512	26.772	28.215	25.477	25.541
Real	5,0,1	5,0,2	5,0,3	5,0,4	5,0,5	5,1,1	5,1,2	5,1,3	5,1,4	5,1,5
21.482	16.099	17.958	18.328	17.434	18.111	16.807	18.976	17.359	18.670	
28.401	19.668	20.187	21.461	22.200		20.479	20.855		21.728	
32.693	22.100	23.756	25.216	23.302	19.646	24.026	25.635	25.759	20.045	22.357
26.655	24.213	26.373	27.704	24.838	24.900	26.114	27.967	28.113	28.566	

Table 5. Corresponding RMSEs of Table 1

1,0,1	1,0,2	1,0,3	1,0,4	1,0,5	1,1,1	1,1,2	1,1,3	1,1,4	1,1,5
3.319928	4.262743		0.68797	4.750834	3.261413	3.681566	3.89942	2.662842	3.133893
2,0,1	2,0,2	2,0,3	2,0,4	2,0,5	2,1,1	2,1,2	2,1,3	2,1,4	2,1,5
3.248184	2.532684		0.012067	0.076134	3.192964	2.681543	2.453163		4.790708
3,0,1	3,0,2	3,0,3	3,0,4	3,0,5	3,1,1	3,1,2	3,1,3	3,1,4	3,1,5
3.085595	2.722815	1.970371	0.470844	3.67506	3.373647	2.45403	6.965399	1.942791	2.668733
4,0,1	4,0,2	4,0,3	4,0,4	4,0,5	4,1,1	4,1,2	4,1,3	4,1,4	4,1,5
2.500514	1.547557	1.87656	3.95979	4.933906	3.470494	2.793901	2.158508		2.67486
5,0,1	5,0,2	5,0,3	5,0,4	5,0,5	5,1,1	5,1,2	5,1,3	5,1,4	5,1,5
2.749447	2.456289	2.172778	2.576363	3.881602	2.34549	2.15375	4.521926		1.634062

Table 6. Corresponding RMSEs of Table 2

1,0,1	1,0,2	1,0,3	1,0,4	1,0,5	1,1,1	1,1,2	1,1,3	1,1,4	1,1,5
21.58921	25.65373	3.109461	12.53823	5.866518	21.72462	19.2118	3.096488	10.1734	10.433
2,0,1	2,0,2	2,0,3	2,0,4	2,0,5	2,1,1	2,1,2	2,1,3	2,1,4	2,1,5
15.03155	34.57924	23.3423	15.32558	499.6885	9.113215	15.61935	10.81785	9.264453	8.556198
3,0,1	3,0,2	3,0,3	3,0,4	3,0,5	3,1,1	3,1,2	3,1,3	3,1,4	3,1,5
12.47487	25.13604	22.25524	14.80161	11.85689	13.12106	18.84365	344.3659	124.0304	
4,0,1	4,0,2	4,0,3	4,0,4	4,0,5	4,1,1	4,1,2	4,1,3	4,1,4	4,1,5
14.43916	19.12125	25.42574	16.49067	12.70835	9.102101	16.34721	5.573263	1856.177	
5,0,1	5,0,2	5,0,3	5,0,4	5,0,5	5,1,1	5,1,2	5,1,3	5,1,4	5,1,5
60.86062	58.13133	22.11963	6.984873	13.89268	17.16301				

Table 7. Corresponding RMSEs of Table 3

1,0,1	1,0,2	1,0,3	1,0,4	1,0,5	1,1,1	1,1,2	1,1,3	1,1,4	1,1,5
20.64558	11.70015				31.48078	20.88961	22.31387	130.3273	16.46071
2,0,1	2,0,2	2,0,3	2,0,4	2,0,5	2,1,1	2,1,2	2,1,3	2,1,4	2,1,5
16.75983	12.22305				25.30119	25.26904	40.45782	48.80285	13.97545
3,0,1	3,0,2	3,0,3	3,0,4	3,0,5	3,1,1	3,1,2	3,1,3	3,1,4	3,1,5
14.95325	21.01494				39.07298	13.18333	27.75728	1.096716	18.12848
4,0,1	4,0,2	4,0,3	4,0,4	4,0,5	4,1,1	4,1,2	4,1,3	4,1,4	4,1,5
9.07044	18.69534	365.7504			26.17372	51.92024	109.2959	29.26125	
5,0,1	5,0,2	5,0,3	5,0,4	5,0,5	5,1,1	5,1,2	5,1,3	5,1,4	5,1,5
12.44548	21.43314	11.94654	16.94416	460.2542	22.86124	35.94196	43.67815	22.37872	

Table 8. Corresponding RMSEs of Table 4

1,0,1	1,0,2	1,0,3	1,0,4	1,0,5	1,1,1	1,1,2	1,1,3	1,1,4	1,1,5
16.37965	13.20535	12.07323	5.958037	7.728154	15.25568		6.611038	7.461217	8.405572
2,0,1	2,0,2	2,0,3	2,0,4	2,0,5	2,1,1	2,1,2	2,1,3	2,1,4	2,1,5
15.7065	13.09519	7.992525	8.595086	8.767497	11.36879	9.467817	8.040781	8.758271	5.671202
3,0,1	3,0,2	3,0,3	3,0,4	3,0,5	3,1,1	3,1,2	3,1,3	3,1,4	3,1,5
11.56589	9.74759	9.455931	21.482	6.080623	10.81063	9.016102	7.897835	6.443738	6.860348
4,0,1	4,0,2	4,0,3	4,0,4	4,0,5	4,1,1	4,1,2	4,1,3	4,1,4	4,1,5
11.03461	9.504031	8.409419	7.287767	6.733352	7.104521	5.857793	5.365601	5.552384	5.688899
5,0,1	5,0,2	5,0,3	5,0,4	5,0,5	5,1,1	5,1,2	5,1,3	5,1,4	5,1,5
7.473543	6.32139	5.366525	6.048311	7.845843	6.325106	5.366067	4.732974	7.349352	10.33634

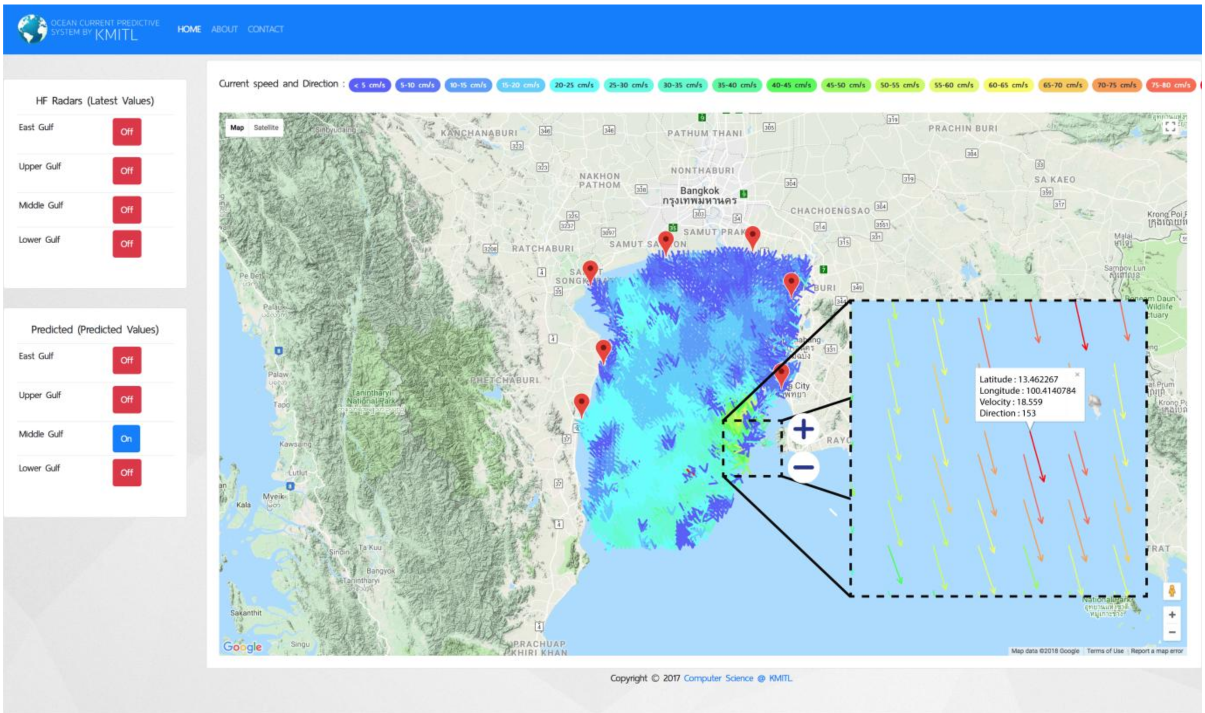

Figure 5. The screenshot of our ocean surface current prediction system for the entire Gulf of Thailand, visualized on the map.

The tables show predicted results, with the cell's color-coding schemes explained in the previous paragraph.

Tables 5-8 show the RMSE values from the predicted results. Each column shows the RMSE value derived from the predicted results compared to the actual values. The colors of the cell's background and the numerical values are as described above—to make it easier for readers to assess how close the predicted results are to the actual recorded values.

When we compared the results from our grid-based spatial ARIMA model with other methods, including baseline, KNN, traditional ARIMA, LSTM, the results showed that our approach outperformed other methods for V component prediction. In addition, we found that selecting the right range of data set had a direct effect on the accuracy of the prediction (see the red rectangular borders in Tables 1-8). We further elaborate on the selections and results below.

Particularly, for U comp prediction, a combination of (*p*,*q*,*d*) as (4,0,2) respectively, using 10 days of dataset with 7 hours prior produced the best prediction of RMSE = 1.547556, while for V comp prediction, a combination of (*p*,*q*,*d*) as (2,0,1), using 7 days of dataset with 7 hours prior, generated the best prediction of RMSE = 9.113215. Predicting velocity values is at the highest accuracy of RMSE = 5.365601 with (*p*,*q*,*d*) of (4,1,3) using 15 days of dataset with preceding 7 hrs., whereas predicting direction is at the highest accuracy of RMSE = 9.07044, with (*p*,*q*,*d*) of (4,0,1) using 7 days of dataset with preceding 7 hrs.

Nonetheless, we also found that, in some cases, when choosing the right configuration of (*p*,*d*,*q*) and historical range of datasets in practice, the *validity* of the estimated values can sometimes be less important than the accuracy aspect. For example, in the case of V comp prediction, although a combination of (*p*,*q*,*d*) as (2,0,1), using 7 days of dataset with 7 hours prior suggested the best prediction in terms of both RMSE (9.113215) and validity (all four estimated values), the second best of RMSE (2.322) and validity (only three out of four can be calculated) actually gave a better prediction when we added more sites (343 sites).

We also tested the model with higher degrees of *p* and *q* to see if they were affected in terms of the accuracy level and computational time. The result showed that even though the degrees of *p* and *q* were increased, the accuracy levels were not affected improved significantly.

Table 9. An averaged RMSE of each method

Method	RMSE (cm/s)	
	U Component	**V Component**
Baseline [7]	7.02	20.43
LSTM [7]	5.85	21.2
ARIMA [7]	4.33	11.81
kNN [7]	4.41	7.58
Our method	10.79	6.23

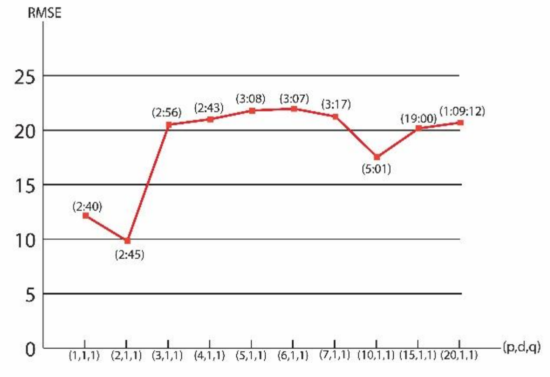

Figure 6. Prediction performance with different values of *p*

Figure 6 shows a sample result of increasing *p* (up to 20 degrees), which unfortunately did not improve accuracy performance. However, the increase also took a substantial amount of computational time to process.

V. CONCLUSION AND FUTURE WORK

In this research, the grid-based spatial ARIMA model is proposed and tested based on traditional ARIMA and dataset range selection criteria to analyze and forecast direction and velocity values of the ocean surface currents in the Gulf of Thailand. To evaluate the performance, the prediction results from our method were compared with other methods, including baseline, kNN, ARIMA, LSTM. The results showed that our spatio-temporal data mining approach generated quite a high level of accuracy for V comp prediction. The method and techniques proposed in our approach can further be adopted and used for other GIS applications in which spatio-temporal forecast values are involved in analyzing ocean current patterns, such as oil-spill trajectory prediction. Further adoption can maximize the positive impacts of our proposed prediction method. Such adoption could be facilitated by the diffusion of innovations theory [8], a communication framework for innovation and technology adoption.

The experimental results from this research showed the importance of selecting predictive data and the use of various parameters. In future development, more potential/relevant parameters should be added, such as increasing the number of points around in space for modeling. Further, more related data sets used, such as in the same period of the previous year or the same season, could also be added to improve the prediction of our proposed model. Finally, we believe that the proposed model can generate better predictions to further guide crisis communication in real time via mass communication and social media for evacuation efforts to save lives in the event of coastal natural disasters.

References

[1] D. E. Barrick, M. W. Evan and B. L. Weber, “Ocean surface currents mapped by radar,” Science, 198 (4313), 1997, 138-144.

[2] M. Marta-Almeida, M. Ruiz-Villarreal, J. Pereira, P. Otero, M. Cirano, X. Zhang and R. Hetland, “Efficient tools for marine operational forecast and oil spill tracking,” Marine Pollution Bulletin, vol. 71, no. 1-2, 2013, 139-151.

[3] T. Cheng and J. Wang, “Applications of spatio-temporal data mining and knowledge discovery (STDMKD) for forest fire prevention,” School of Geography and Urban Planning Sun Yat-sun University, Guangzhou, P R China, 2008.

[4] S. Shekhar et al., “Spatiotemporal Data Mining: A Computational Perspective,” ISPRS Int. J. Geo-Inf., 2015.

[5] N. Mamoulis, H. Cao, G. Kollios, M. Hadjieleftheriou, Y. Tao, and D. W. Cheung, “Mining, indexing, and querying historical spatiotemporal data,” In Proceedings of the ACM SIGKDD KDD’04, Seattle, Washington, USA, August 22-25, ACM, 2004, 236–245.

[6] A. Aunthong, “Autoregressive integrated moving average model”, Social Research Institute, Chiang Mai University, 2007.

[7] A. Jirakittayakorn, T. Kormongkolkul, P. Vateekul, K. Jitkajornwanich and S. Lawawirojwong, "Temporal kNN for short-term ocean current prediction based on HF radar observations," In Proceedings of the 2017 14th International Joint Conference on Computer Science and Software Engineering (JCSSE), Nakhon Si Thammarat, 2017, pp. 1-6.

[8] K. F. Kee, “Adoption and diffusion,” in International encyclopedia of organizational communication, C. Scott, and L. Lewis, Eds., Hoboken, NJ: Wiley-Blackwell, 2017, pp. 41-54….

[9] Coastal Observing Research and Development Center, HFRNet National Network: WERA Radial LonLatUV (LLUV) file format, Retrieved Feb. 27, 2019 from: https://cordc.ucsd.edu/projects/mapping/documents/HFRNet_WERA_LonLatUV_RDL.pdf