



شبكة المعلومات الجامعية  
التوثيق الإلكتروني والميكرو فيلم

# بسم الله الرحمن الرحيم



**HANAA ALY**



شبكة المعلومات الجامعية  
التوثيق الإلكتروني والميكروفيلم



# شبكة المعلومات الجامعية التوثيق الإلكتروني والميكروفيلم



**HANAA ALY**



شبكة المعلومات الجامعية  
التوثيق الإلكتروني والميكروفيلم

# جامعة عين شمس

## التوثيق الإلكتروني والميكروفيلم

### قسم

نقسم بالله العظيم أن المادة التي تم توثيقها وتسجيلها  
علي هذه الأقراص المدمجة قد أعدت دون أية تغيرات



### يجب أن

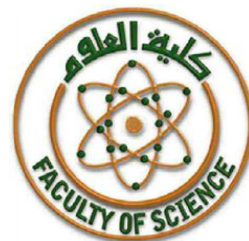
تحفظ هذه الأقراص المدمجة بعيدا عن الغبار



**HANAA ALY**



Ain Shams University  
Faculty of Science  
Department of Mathematics



**“Numerical Weather Prediction models and  
climate data interpretation using data mining  
techniques.”**

*A Thesis Submitted to Department of Mathematics –  
Faculty of Science – Ain Shams University in Partial  
Fulfillment of the Requirements of the Award of the Ph.D.  
Degree in Applied Mathematics.*

*By*

**Marwa Farouk Mohamed Ali**

*Supervised by*

*Prof. Dr.: Mohamed Ahmed Mohamed Hassan El-Arabawy  
Prof. of Applied Mathematics – Mathematics department –  
faculty of science – Ain shams university.*

*Prof. Dr. Mohamed Magdy Abd Al-Wahab  
Prof. of Meteorology – Astronomy and Meteorology  
department- faculty of science – Cairo University.*

*Dr. Somaya Abd AL-Hamed Ahmed Mohamed  
Lecturer of computer science – Institution of technology in 10<sup>th</sup>  
of Ramadan.*

**2021**

## **Acknowledgement**

First, I thank Allah the Almighty for giving me the strength and the ability to complete this thesis. Many thanks to my family who patiently encouraged and supported me to finish this work.

My Sincere thanks and real indebtedness are dedicated to my supervisors Prof. Dr. Mohamed Hassan El-Arabawy, Prof. Dr. Mohamed Magdy Abd Al-Wahab and Dr.Somaya M.Asklany, for their constructive guidance throughout the development of the work. The opportunity I had to share in their boundless vision, enthusiasm and advice is something for which I shall always be profoundly grateful, Dr. Zeinab Salah, for giving me feedback and for her support as research on this thesis; and Dr. Maha Harhash, for her support and patience. Without their guidance and advice this thesis would not exist. A special thanks should be presented to Prof.Dr.Fayez Nasif for his precious notes and guidance. It is pleasure to express my great thanks to Mathematics department, Faculty of Science, Ain Shams University for giving me this opportunity.

### **List of published papers**

- 1- Ali, M. F. M., Asklany, S. A., Abd El-wahab, M., & Hassan, M. A. ,Data Mining Algorithms for Weather Forecast Phenomena: Comparative Study. *IJCSNS*, 19(9), 76, (2019).
- 2- Ali, M. F. M., Asklany, S. A., Abd El-wahab, M., & Hassan, M. A., A comparison of Three Boundary Layer schemes for Numerical Weather Prediction. *Appl.Math.inf.Sci.14,No.6, 1-9*, (2020).

# Contents

Acknowledgement	i
List of published papers	ii
Table of contents	iii
List of figures	viii
List of tables	x
List of Abbreviations	xii
Abstract	xiii
<b>Chapter 1 : Basic concepts</b>	<b>1</b>
1.1 Atmospheric science and Meteorology	1
1.1.1 Types of scales of different Meteorological Phenomena	1
1.1.2 Atmospheric fundamental forces	3
1.1.3 Hydrostatic equation	5
1.2 Planetary Boundary Layer and Turbulent motion	6
1.2.1 Definition of Planetary Boundary Layer	6
1.2.2 PBL height	7
1.2.3 flux and convective flux	7
1.3 Turbulence and turbulent motion	9
1.3.1 PBL diurnal variation and turbulence	10
1.3.2 Mean and turbulent parts	13
1.3.3 Turbulent and Kinetic energy	13
1.3.4 Static stability and Convection	14

1.3.5 Dynamic stability and Kelvin-Helmholtz waves	14
1.3.6 Richardson number	15
1.3.7 Reynolds number	17
1.4 Air quality and air pollution	17
1.4.1 Air pollution	17
1.4.2 Air quality and ventilation index	19
1.5 Methods of prediction in Meteorology	19
1.5.1 Numerical weather prediction	19
1.5.2 Statistical modeling	22
1.6 Related work	23
<b>Chapter 2 : Data mining definitions and techniques</b>	<b>27</b>
2.1 Data mining definition and back ground	27
2.2 Data mining techniques	29
2.2.1 Classification	30
2.2.2 Prediction	30
2.2.3 Clustering	30
2.3 Some data mining algorithms	30
2.3.1 Decision tree	30
2.3.2 K- nearest neighbor algorithm	34
2.3.3 Naïve biased	36
2.4 Model evaluation techniques	36
2.4.1 Holdout method	37
2.4.2 Random subsampling	37
2.4.3 K cross validation	38



2.4.4 confusion matrix	38
2.5 Rapidminer	39
<b>Chapter 3 : Numerical modelling and Regional climate model (RegCM)</b>	<b>41</b>
3.1 Numerical modelling	41
3.1.1 Classification of Partial Differential Equations	41
3.1.2 Existence and uniqueness	42
3.1.3 Picard's theorem	42
3.1.4 Discretization	43
3.1.5 Finite difference method	43
3.1.6 Time finite difference	45
3.1.7 Three-level-schemes	45
3.1.8 Discretization of the advection equation using a centered difference both in time (leap frog) and space.	46
3.2 Regional climate models definition and survey	49
3.2.1 Regional climate models definition	49
3.2.2 Regional climate models survey	50
3.3 Model components	53
3.3.1 RegCM Model map projections, Horizontal and Vertical Grid	54
3.3.2 Model dynamics	56
3.3.3 Model physics	58
3.3.4 Model time step	59
3.3.5 Sources of model forecast errors	60
<b>Chapter 4 : PBL parameterization schemes in</b>	<b>63</b>

## **numerical modelling and their application**

4.1 A short overview of some PBL schemes	63
4.1.1 The Holtslag boundary layer scheme	64
4.1.2 The UW PBL scheme	66
4.1.3 Gfs scheme	69
4.2 Comparative study among some boundary layer schemes by means of a case study	75
4.2.1 1st case study	75
4.2.2 Site description	75
4.2.3 Data description	77
4.2.4 Model set up and configuration	77
4.2.5 Wind speed	78
4.2.6 PM10 concentration	80
4.2.7 Boundary layer height (PBLH)	81
4.2.8 The consistency of the model	83
4.2.9 Ventilation index	84
4.3 2 <sup>nd</sup> case study	85
4.3.1 Boundary layer height (PBLH)	86
4.3.2 Wind speed	87
4.3.3 Dust episode analysis	88
4.3.4 Ventilation index	89
<b>Chapter 5 : Data Mining Algorithms application on weather forecast phenomena</b>	<b>91</b>

5.1 Application of data mining techniques on actual data	91
5.1.1 Data collection	91
5.1.4 Data preparation	91
5.1.4 Data Mining	92
5.1.5 Experiments and results	93
5.1.6 The first experiment	93
5.1.7 The Second experiment	95
5.1.8 The Third experiment	97
5.2 Interpretation of model's output data by means of data mining algorithm	99
5.2.1 Experiment (1)	99
5.2.2 Experiment (2)	102
Results and discussion	106
Bibliography	109

## List of figures

Fig.1.1	Diurnal variation of PBL	11
Fig.2.1	The combined fields of Data mining science	28
Fig.3.1	The RegCM refined resolution of GCM, and the impacts of its predictions on many fields.	50
Fig.3.2	The horizontal grid (Arakawa B-grid) of the RegCM	54
Fig.3.3	$\sigma$ - levels illustration within the RegCM .	56
Fig.3.4	A flow chart summarizing the RegCM three components.	59
Fig.4.1	Egypt map	76
Fig.4.2	Model's domain map	76
Fig.4.3	The actual wind speed (m/s) vs. three schemes output wind speed	79
Fig.4.4	Comparison between the PM10 concentration output of the three PBL schemes and the real data.	81
Fig.4.5	Comparison between PBLH for both schemes and the reanalysis data.	82
Fig.4.6	Scattered plot relation between of PM10 concentration ( $\mu\text{g}/\text{m}^3$ ) and wind speed (m/s).	83
Fig.4.7	Diurnal variation of PM10 concentration ( $\mu\text{g}/\text{m}^3$ ) and wind speed (m/s).	84
Fig.4.8	2 <sup>nd</sup> domain, Kuwait map	85
Fig.4.9	The PBLH predicted data vs. the actual data	86
Fig.4.10	Wind speed resulted from the three schemes and actual data variation during the storm	87

Fig.4.11	The daily variation of PM10 concentration	88
Fig.4.12	Relation between ventilation index and boundary layer height for the Holtslag scheme	89
Fig.5.1	proposed method flow chart	92
Fig.5.2	1 <sup>st</sup> experiment decision tree diagram	94
Fig.5.3	The 1 <sup>st</sup> experiment confusion matrix evaluation	95
Fig.5.4	Displaying model in Rapidminer output	96
Fig.5.5	The 2 <sup>nd</sup> experiment decision tree diagram	96
Fig.5.6	The 2 <sup>nd</sup> experiment confusion matrix evaluation.	96
Fig.5.7	The 3 <sup>rd</sup> experiment decision tree diagram	98
Fig.5.8	Decision tree diagram	100
Fig.5.9	K-NN model in Rapidminer	100
Fig.5.10	Naïve biased model in Rapidminer	101
Fig.5.11	Confusion matrix of D.tree diagram	101
Fig.5.12	Confusion matrix of K-NN algorithm	101
Fig.5.13	Confusion matrix of Naïve-biased algorithm	102
Fig.5.14	The steps of building a decision tree model within the Rapidminer program	103
Fig.5.15	Decision tree diagram	103
Fig.5.16	Confusion matrix of D.tree diagram	104
Fig.5.17	Confusion matrix of K-NN algorithm	104
Fig.5.18	Confusion matrix of Naïve- biased	104

## **List of tables**

Table (3.1)	Advantages and disadvantage of the RegCMs	61
Table (4.1)	Schemes used in the model for different simulations of physical processes.	77
Table (4.2)	Statistical measures: mean bias, RMSE and MAPE measures for the validation of wind speed model output compared to the ERA-intrem data.	79
Table (4.3)	Statistics of PM10 concentration between RegCM4.7 PBL schemes and observations (real data) with the Performance indicators root mean squared error (RMSE), and mean absolute percentage error (MAPE).	81
Table (4.4)	Statistical measures: mean bias, RMSE and MAPE measures for the validation of the PBLH model output compared to the ERA-intrem data.	82
Table (4.5)	Relation between ventilation index, wind speed, wind direction and PBLH.	84
Table (4.6)	The errors in boundary layer height between the three schemes in El- Ahmady station.	86
Table (4.7)	Wind speed average error	87
Table (4.8)	PM10 concentration error against observed data	88
Table (4.9)	Ventilation index relation with wind and PBLH	89
Table (5.1)	Attributes description Of the 1 <sup>st</sup> experiment.	93

Table (5.2)	Statistical measures used for the model evaluation	94
Table (5.3)	Attributes description of the 2 <sup>nd</sup> experiment.	95
Table (5.4)	Statistical measures used for the model evaluation.	97
Table (5.5)	Attributes description of the third experiment.	97
Table (5.6)	Comparison between the three model performance using statistical measures.	98
Table (5.7)	The three models evaluation	102
Table (5.8)	Summarized results of the three models output using K-cross validation method.	105