Stony Brook University

Constrained Variational Analysis Integrating Vertical and Temporal Correlations

A Dissertation Presented

by

Jun Huang

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Applied Math and Statistics

Stony Brook University

December 2012

Stony Brook University

The Graduate School

Jun Huang

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Wei Zhu – Dissertation Advisor
Professor & Deputy Chair, AMS

Xiangmin Jiao - Chairperson of Defense
Associate Professor, AMS

Haipeng Xing – Internal Committee Member
Associate Professor, AMS

Minghua Zhang – External Committee Member
Professor & Dean, SOMAS, Stony Brook University

This dissertation is accepted by the Graduate School

Charles Taber
Interim Dean of the Graduate School

Abstract of the Dissertation

Constrained Variational Analysis Integrating Vertical and Temporal Correlations

by

Jun Huang

Doctor of Philosophy

in

Applied Math and Statistics

Stony Brook University

2012

Understanding climate change is an increasingly urgent issue of our society. Existing global climate models are less accurate particularly in predicting severe weather and abrupt climate change, which will invariably cause dramatic losses in life and property. One of the four major impediments identified by the National Research Council is the rudimentary, stationary and often unrealistic statistical methods/models employed in climate modeling.

We endeavor to improve the stochastic modeling of atmospheric data by incorporating the spatial and temporal correlations with observations in the constrained variational analysis modeling (referred to CVA model hereinafter) approach pioneered by Zhang and Lin (1997) and further developed by Zhang et al. (2001). Thus far, we have successfully incorporated some spatial correlations into our model, especially the correlations of state variables across different vertical levels. Furthermore, we have incorporated temporal correlations via an AR(1) time series model.

The newly enhanced constrained variational analysis model is a significant improvement over the traditional methods in its core idea of (1) enforcing physical consistency through variational constraints, (2) stochastic modeling of the random errors, and (3) utilizing heterogeneous data through multiple nesting. Our novel integration of statistical methods and physical principles has given birth to a modern and superior climate model featuring better uncertainty quantifications and more accurate predictions as demonstrated by our final results.

Table of Contents

List of Figures ..... **vi**

Acknowledgments ..... **x**

Chapter 1

Introduction **1**

Chapter 2

Constrained Variational Analysis Integrating Vertical Correlations and Three Constraints **6**

2.1 The input variables and data sets **6**
2.2 Basic budget equations **7**
2.3 Model description **10**
2.4 Algorithm description **11**
2.5 Figures and results **13**

Chapter 3

Constrained Variational Analysis Integrating Temporal Correlations and Three Constraints **28**

3.1 Model description **28**
3.2 Algorithm description **32**
3.3 Figures and results **33**

Chapter 4

Constrained Variational Analysis Integrating Vertical Correlations and More Physical Budgets (190 Budgets) **48**

4.1 Model description **48**
4.2 Algorithm description **49**
4.3 Figures and preliminary results **50**

Chapter 5

Conclusion and Future Work **55**

5.1 Conclusion **55**
5.2 Future Work **56**

References **59**

Appendix **63**

A. Estimate of Q **63**

List of Figures

Figure 2.1: The area map of Darwin, Australia. The six radiosonde (dark-blue pentagons) stations observe the 3d-data set for temperature, dew-point temperature (for mixing ratio of water vapor or relative humidity), and two components of wind velocity. 17

Figure 2.2: Schematic diagram for the structure of the 3d-data set, where there are a total of *S* stations horizontally and *K*-pressure levels vertically. At each station and every level, and every three hours, we observe the *u*, *v*, *q*, and *s* variables. The measurements in Darwin cover from 19 January to 12 February, 2006. In this study, we have $S = 6$, $K = 45$, and $T = 201$. ..…….…….....…….....18

Figure 2.3: Flow chart of the algorithm for the new vertically correlated case. 19

Figure 2.4: Final results for u, v, q and s from vertical correlated case at station 1, shown in (a)-(d), respectively. ……..……..20

Figure 2.5: Adjustments of u, v, q and s from vertical correlated case at station 1, shown in (a)-(d), respectively. …….…...21

Figure 2.6:(a) and (b) for terms of column integrated moisture budgets before and after the vertical correlated variational method, respectively; (c) and (d) for terms of the column integrated heat budgets before and after vertical correlated variational method, respectively. …….22

Figure 2.7: The correlation structure of the adjustment for state variables *u*, *v*, *q*, and *s*, shown in (a)-(d), respectively. 23

Figure 2.8: Differenced vertical advection for moisture (a), Q_2 (b), differenced vertical advection for heat (c) and Q_1 (d) between vertical correlated case and the original case (Zhang & Lin 1997). ……..24

Figure 2.9: Correlation structures of difference values of (a) vertical advection for moisture, (b) vertical advection for static energy, (c) Q_2 and (d) Q_1 between vertical correlated case and the original case. …….25

Figure 2.10: Vertical advection for moisture of (a) original 3d-data versus (b) final results from new variational method; the apparent moisture sink of (c) original 3d-data versus (d) final results from new variational method. …….....26

Figure 2.11: Vertical advection for static energy of (a) original 3d-data versus (b) final results from new variational method; the apparent heat source of (c) original 3d-data versus (d) final results from new variational method. …….....27

Figure 3.1: Flow chart of the algorithm for the temporal correlated case. ……..39

Figure 3.2: Final results for u, v, q and s from temporal correlated case with AR(1) feature at station 1, shown in (a)-(d), respectively. ……..…40

Figure 3.3: Adjustments of u, v, q and s from temporal correlated case with AR(1) feature at station 1, shown in (a)-(d), respectively. ……..41

Figure 3.4: Curve of the residual sum of squares RSS v.s α . ……42

Figure 3.5: (a) and (b) for the terms of column integrated moisture budgets before and after temporal correlated variational method with AR(1) feature, respectively; (c) and (d) for terms of the column integrated heat budgets before and after temporal correlated variational method with AR(1) feature, respectively. ……..43

Figure 3.6: The autocorrelation structures for adjustments of u (a), v (b), q (c), and s (d) from temporal correlated constrained variational method with AR(1) feature. ……..……..…44

Figure 3.7: The autocorrelation structures of differenced vertical advection for moisture (a), vertical advection for static energy (b), Q_2 (c) and Q_1 (d) between temporal correlated case and the original case (Zhang & Lin, 1997). ……....45

Figure 3.8: Vertical advection for moisture of (a) original 3d-data versus (b) final results from the new CVA method with AR(1) feature; the apparent moisture sink of (c) original 3d-data versus (d) final results from the new CVA method with AR(1) feature. ...46

Figure 3.9: Vertical advection for static energy of (a) original 3d-data versus. (b) final results from the new CVA method with AR(1) feature; the apparent heat source of (c) original 3d-data versus (d) final results from the new CVA method with AR(1) feature. …….47

Figure 4.1: Final results of u, v, q, s and ϕ for vertical correlated case with 190 budgets at station 1, shown in (a)-(e), respectively. ……..……..…….52

Figure 4.2: Time series plots of the adjustments for u, v, q, s and ϕ from vertical correlated case with 190 budgets at station 1, shown in (a)-(e), respectively. ……..…..…53

Figure 4.3: Comparison of the adjustment for s from the original uncorrelated case (Zhang & Lin, 1997) with five budgets (a, c) versus the new case with 190 budgets (hydro-static balance incorporated) (b, d). ……54

Figure A.1: The long time series of X across all stations for pressure level k. ……...64

To My Great Parents

Acknowledgments

First, I wish to express my most sincere thanks to my advisor, Dr. Wei Zhu and collaborator Dr. Minghua Zhang, for their guidance, support and encouragement throughout these years, which is full of the happiness from great progress and frustration to the tough research problems. The thesis would have not been accomplished without their time and precious advice.

Special thanks also go to my committee members Dr. Haipeng Xing and Dr. Xiangmin Jiao. The temporal correlated constrained variational method will not succeed without the instruction of Dr. Xing. His long experiences in time series analysis have substantially benefited my current research and predictably my future career. Dr. Xiangmin Jiao gave me valuable advice regarding the algorithm design when I was doing the vertical correlated case. The way of thinking he taught me finally fosters the successfully convergent method.

I am also very grateful to all my friends in the meteorological area for helping me fully understanding the physical background of my project and solve any Linux and super computer related issues. They are Dr. Wuyin Lin, Dr. Ping Liu from SoMAS, Stony Brook University; Dr. Jian Li and Dr. Xiaocong Wang from meteorological administration; Dr. Bizheng Wang from China academy of sciences. Sincere thanks again go to Dr. Ping Liu for his long time work on revising my thesis and to my friend Guanan Yu for helping me with tricky skills of Word software when compiling this thesis.

I am also very grateful to all academic siblings in both statistical track and ITPA for those unforgettable discussions. I also want to thank the graduate program coordinator, Dr. David Green and the secretaries, Ms. Christine Rota and Ms. Gina Gartin for their kind help to process the tremendous paperworks through all these years.

Finally, I would like to express my deepest gratitude to my parents and my 92-year old grandma for their moral and financial support through these years, as well as to my girlfriend, Luqi Chen, for her support during this tough period. Their faith has encouraged me to move this far academically.

Chapter 1
Introduction

In the past twenty years, an unprecedented amount of atmospheric data has been collected from multiple measurement platforms at an increasing speed. There are, however, formidable challenges in using these data. First, the measurements from satellites and radars are highly heterogeneous in space, time, and resolution, and dependent on the specific design of instruments and the atmospheric variables they measure. Second, all data have more or less sampling errors that often render inconsistency to each other in quantifying the internal feedback processes of the climate system. Specifically in meteorology, global climate models, although overall following the universal physical conservations of energy, momentum, and mass, have many uncertainties in using parameterizations to describe sub-grid scale physical processes for example the phase change of water vapor that induces rainfall and fierce weather phenomena. These parameterizations bring larger uncertainties that are obstacles for the models to produce realistic climate simulations and projections.

The consequence induced by such uncertainties is serious. Errors of hurricane track predictions with a few hundred kilometers and poor predictions of severe weather associated extratropical cyclone positions (McMurdie and Mass, 2004; Zhang et al., 2002) can result in dramatic loss of properties and lives. For climate modeling, large uncertainties currently exist in the projection of future climate change, especially abrupt changes (Dessler and Parson, 2006; Cox, 2005; Alley et al., 2003; Drummond and Wilkinson, 2006). Improvement in both severe weather forecasting

and climate modeling requires accurate analysis of atmospheric data. In a report entitled “Abrupt Climate Change: Inevitable Surprises”, the Committee on Abrupt Climate Change from the National Research Council pointed out that “The abrupt changes of the past are not fully explained yet, and climate models typically underestimate the size, speed, and extent of those changes. Hence, future abrupt changes cannot be predicted with confidence, and climate surprises are to be expected”. In order to better modeling the past and predicting the future, the Committee laid out four major recommendations with Number 4 being “Improve Statistical Approaches” – because rudimentary statistical methods with unrealistic assumptions have thus far failed to model the complicated and dynamic climate system adequately and correctly.

The challenges call for a paradigm shift to significantly improve the utilization of statistical methods in climate modeling. The observed fields such as winds, water vapor, and mass have to be calibrated for physical coherence. Several different methods were used for this purpose: simple linear interpolation, optimum interpolation, and constraint variantional analysis (Zhang and Lin, 1997). In this thesis, we endeavor to integrate novel statistical error modeling with physics-based constrained variational analysis (Zhang and Lin, 1997) of atmospheric data.

Specifically, data from a sounding array will be used to calculate vertical velocity and advective tendencies for an atmospheric column. When measurements of surface pressure, sensible heat flux, latent heat flux, precipitation, wind stress, and radiative fluxes are available, the column-integrated budgets of mass, heat, moisture and momentum provide a validation on the accuracies of the state variables. However, due to the inevitable uncertainties in atmospheric measurements, physical terms such as wind divergence and tendencies typically do not satisfy the budgets which

they should follow. What we want to do is to force the state variables to satisfy the budgets after minimum adjustments. Previous studies (Zhang & Lin, 1997) reached this goal by using a constrained variational method under the non-correlated assumptions. In their study, they assumed that variables in the following cases are independent from each other: 1) if they are different control variables, 2) if they are same type of variable at different stations, levels or time steps. This assumption, though simple to implement, is not close to the reality, which is verified by our L-jung statistical test.

To make the novel method more realistic, we have several updated models constructed based on Zhang and Lin (1997) by incorporating the model with the vertical correlations across different pressure levels or with the temporal correlations across different time steps. Some problems have emerged with the new approach. The first is how to estimate the error structures more accurately. The independence assumption in previous study is not accurate enough to reflect the truth. Thus, how to add the error correlations, especially the spatial and the temporal correlations to the model is the most critical question we have to solve first. We then wish to explore some better results from our new model compared with the original one, especially the impact of the new assumption on the important physical variables such as the vertical advection, apparent moisture sink and apparent heat source. The results will be elaborated in Chapters 2, 3 and 4.

The second serious problem, as mentioned above, is related to the algorithm design. The original case adopted the fixed point iteration method that converges the results fast, because the weighting matrix is diagonal based on the independence assumption and it is simple enough to

make most of the iteration method convergent. However, the situation is quite different in our new case. With the vertical correlations being incorporated, the weighting matrix is no longer completely diagonal but block diagonal. The complexity of the matrix structure makes the original iteration method invalid. A new method needs to be explored to deal with this new condition by designing and testing numerous computing algorithms.

The third problem is about how to estimate the weighting matrix in the cost function. With the truth unknown, we treat the results from the original case (Zhang & Lin, 1997) as the truth to calculate the errors. The weighting matrix intuitively is the covariance matrix of the errors. The problem then becomes how to construct the time series to estimate the covariance.

The thesis has three objectives. One is to extend the structure of the original case by adding correlations of different types. Both statistical tests and meteorological knowledge have confirmed the existence of spatial and temporal correlations in the balloon sounding data. Efforts have been devoted to improve the method by considering the correlations in the weighting matrix. The new case with the vertical correlations is described in Chapter 2 and the same case with the time series feature is described in Chapter 3.

The second objective is to explore the improvements of the new results. This can be divided into two parts. First is to find out the new properties in the new final analysis; the other is to dig into the impact of the new assumptions on the important physical quantities, such as the vertical

advection, apparent moisture sink, apparent heat source, and etc. This will be deliberated in "figures and results" parts in both Chapter 2 and Chapter 3.

The third objective is to make the model as complete as possible by adding more physical constraints. Besides the column integrated budgets of mass, moisture and static energy, there are some more budgets to be balanced. Whether more constraints will be coherent with our new assumption and generate comparably better results is another issue worth considering. Chapter 4 gave a detailed example on how to add to more constraints to the vertical correlated case. In that case, column integrated budgets of momentums as well as hydro-static balance will be applied. The inclusion of hydro-static balance gave us a meaningful result, compared to the original case.

This thesis is organized as follows: Chapter 2 describes the constrained varitional analysis method with the vertical correlations integrated, including the budgets, method, alogirthm and final results. Similar procedure for the temporal correlated cases via AR(1) model is elaborated in Chapter 3. Chapter 4 gives an example on how more constraints will go along with our new assumption for the future exploration purpose. Chapter 5 contains the conclusion and future work.

Chapter 2

Constrained Variational Analysis Integrating Vertical Correlations and Three Constraints

2.1 The input variables and data sets

Basic input measurements for a CVA model can be generally categorized into two groups: the first is what needs to be adjusted and its members are in *x*-*p*-*t* directions (where *x* represents the geographical location in horizontal direction, *p* for pressure that decreases monotonically in the vertical, and *t* for time) so that they are referred to as 3d-data hereafter; the second group is used to obtain the terms of constraints and its members are integrals in the *p*-direction so that they are referred to as 2d-data hereafter.

Specifically the 3d-data set is a four dimensional array with a shape of 4×45×6×201, where the first dimension (4) represents the number of variables (they are temperature, relative humidity, southward and eastward components of wind); the second (45) is for pressure levels in the vertical; the third (6) for the weather stations and the last (201) for time steps. The temperature, dew-point temperature that can be used to derive relative humidity, and wind components are recorded by balloon-borne soundings at 45 pressure levels every three hours (one time step) during 19 January and 12 February, 2006. The six locations include one central station at the city of Darwin, Australia and five satellite stations (dark-blue pentagons in Figure 2.1).

The 3d-data set is recorded as shown in Figure 2.2. We use four vectors $\vec{u}$, $\vec{v}$, $\vec{q}$, $\vec{s}$ to denote the four variables at S stations and K levels (S = 6, and K = 45 in our design) as $\vec{x}^T = (x_{ik}) = (x_{11}, x_{12}, ...x_{1K}, x_{21}, ...x_{ik}, ..., x_{SK})$, where x can be any of the u, v, q, and s. Here u is the westerly wind; v is the northerly; q is the mixing ratio of water vapor, which is the mass of water vapor divided by that of total air in the same volume; and s is the dry static energy or potential temperature, a thermodynamic variable usually as a unique function of temperature.

The 2d-data set includes all terms on the right-hand side of the constraints introduced in Part 2.2 of Chapter 2. They are surface pressure, surface latent and sensible heat fluxes, wind stress, precipitation, net radiation at the surface and at the TOA, and the variability of column-integrated total cloud water content.

2.2 Basic budget equations

Numerical models for climate simulation and projection are formed on the laws of conservations and geo-fluid physics. Mathematically, they are a set of continuous partial differential equations representing geo-physical dynamics and satisfying energy and mass conservations of wind velocity, temperature and water-vapor mixing ratio. These equations are discretized at definite grids in space and time before being solved by modern computers. At each grid the atmosphere is simplified as a single column that may include some or all the radio-sonde stations in Figure 2.1, and thus the constrained variational analysis can be applied. Zhang et al. (2001) developed an example of such a single column. They used $\boldsymbol{X}_t$ to denote the atmospheric state variables (winds u and v, potential temperature s, and water vapor mixing ratio q) at time t at different

levels and stations. The essence of constrained variational analysis is to minimize the cost function of a set of physical equations described below.

1. Column-integrated budget of mass

$$\left\langle \nabla \cdot \vec{V} \right\rangle = -\frac{1}{g}\frac{dp_s}{dt} \tag{2.1}$$

where $\langle A \rangle$ represents mass-weighted vertical integral as $\langle A \rangle = \int_0^z \rho A dz$, ρ is air density, z is height, and p_s is the pressure at surface.

2. Column-integrated budget of water vapor

$$\frac{\partial \langle q \rangle}{\partial t} + \left\langle \nabla \cdot \vec{V} q \right\rangle = E_s - P_{rec} - \frac{\partial \langle q_l \rangle}{\partial t} \tag{2.2}$$

where E_s is the surface evaporation; P_{rec} is the total precipitation (In meteorology, precipitation, also known as one of the classes of hydrometeors which are atmospheric water phenomena, is any product of the condensation of atmospheric and water vapor that falls under gravity. The main forms of precipitation include drizzle, rain, sleet, snow, graupel and hail); q_l is the cloud liquid water content. The liquid water content (LWC) is a measure of the mass of water in cloud over a specified amount of dry air. It is typically measured as either mass per volume of air (g/m^3) or mass ratio of air (g/kg).

3. Column-integrated budget of static energy

$$\frac{\partial \langle s \rangle}{\partial t} + \left\langle \nabla \cdot \vec{V} s \right\rangle = R_{TOA} - R_{SRF} + LP_{rec} + SH + L\frac{\partial \langle q_l \rangle}{\partial t} \tag{2.3}$$

where R_{TOA} is the net downward radiative flux at the TOA (radiative flux, or radiative flux density, is the amount of energy moving in the form of photons or other elementary particles at a certain distance from the source per unit of area per second measured in $J \cdot m^{-2} \cdot s^{-1}$); R_{SRF} is the

net downward radiative flux at the surface; P_{REC} is the total precipitation; *SH* is the sensible heat flux (the energy exchanged by a thermodynamic system that has as its sole effect a change of temperature); L is the latent heat of condensation; and q_l is the cloud liquid water content.

4. Column-integrated budget of momentum

$$\frac{\partial \langle \vec{V} \rangle}{\partial t} + \langle \nabla \cdot \vec{V}\vec{V} \rangle + f\vec{k} \times \langle \vec{V} \rangle + \nabla \langle \phi \rangle = \tau_s \tag{2.4}$$

where τ_s is the surface wind stress (the shear stress exerted by the wind on the surface of large bodies of water – such as oceans, seas, estuaries and lakes. It is the force component parallel to the surface, per unit area, as applied by the wind on the water surface); f is the Coriolis parameter representing the earth rotation; and k is the unit vector in vertical direction.

The above four constraints are applied to each column in the vertical direction.

5. Hydro-static balance equation

$$\frac{\partial \phi}{\partial \ln p} = -R_d (s - \frac{\phi}{C_p}) \tag{2.5}$$

where ϕ is geopotential (the potential of the Earth's gravity field; for convenience it is often defined as minus the potential energy per unit mass, so that the gravity vector is obtained as the gradient of this potential, without the minus); R_d is the gas constant for dry air as a constant of 287.04 *J/kg/K*; and C_p is the heat capacity of dry air at constant pressure also as a constant of 1005.7 *J/kg/K*. Theoretically, this constraint is satisfied at each pressure level over each station.

The following set of equations introduces the computation of apparent moisture sink and heat source (referred to as Q_2, and Q_1 thereafter) based on the conservations of mass, moisture, statistic energy and momentum without any vertical integration. Note that they are not imposed

to the variational method in this thesis. The Q_2 and Q_1 are denoted in the budget equations for moisture and static energy.

$$\frac{\partial \omega}{\partial p} + \nabla \cdot \vec{V} = 0, \tag{2.6}$$

$$\frac{\partial q}{\partial t} + \vec{V} \cdot \nabla q + \omega \frac{\partial q}{\partial p} = E_{vap} - C_{onden} - \overline{\nabla \cdot \left(\vec{V}' q'\right)} - \frac{\overline{\partial(\omega' q')}}{\partial p} - \frac{\partial q_l}{\partial t} = -Q_2 \cdot C_p / L, \tag{2.7}$$

$$\frac{\partial s}{\partial t} + \vec{V} \cdot \nabla s + \omega \frac{\partial s}{\partial p} = Q_{rad} + L(C - E) - \overline{\nabla \cdot \left(\vec{V}' s'\right)} - \frac{\overline{\partial(\omega' s')}}{\partial p} + L \frac{\partial q_l}{\partial t} = Q_1 \cdot C_p, \tag{2.8}$$

$$\frac{\partial \vec{V}}{\partial t} + \vec{V} \cdot \nabla \vec{V} + \omega \frac{\partial \vec{V}}{\partial p} + f\vec{k} \times \vec{V} + \nabla \phi = -\overline{\nabla \cdot \left(\vec{V}' \vec{V}'\right)} - \frac{\overline{\partial(\omega' \vec{V}')}}{\partial p}. \tag{2.9}$$

2.3 Model description

Spatial correlations usually exist among all the four state variables, which have already been verified by the statistical Ljung box test for this dataset. Thus incorporation of the spatial correlations into the original uncorrelated model of Zhang and Lin (1997, referred to as original case hereafter) is a significant upgrade and more accurate. Given that the number in the vertical direction (45 levels) is much larger than that in the horizontal (6 stations), the vertical correlations are the main component of the spatial correlations and thus are first incorporated in the updated model. In this new model, we assume that 1) each state variable are correlated across all vertical levels at the same station and time step; 2) no correlation exists between different variables, or same variable at different stations, vertical levels and time steps.

Using the notations introduced in Part 2.1, we use the following statistical model to depict the error structure:

$$\forall\ i,\ \begin{pmatrix} x_{i1}^{*} \\ x_{i2}^{*} \\ \vdots \\ x_{iK}^{*} \end{pmatrix} = \begin{pmatrix} x_{i1}^{o} \\ x_{i2}^{o} \\ \vdots \\ x_{iK}^{o} \end{pmatrix} + \begin{pmatrix} \varepsilon_{xi1} \\ \varepsilon_{xi2} \\ \vdots \\ \varepsilon_{xiK} \end{pmatrix},\ \text{where } \vec{\varepsilon}_{xi} = \begin{pmatrix} \varepsilon_{xi1} \\ \varepsilon_{xi2} \\ \vdots \\ \varepsilon_{xiK} \end{pmatrix} \sim \mathrm{N}(\vec{0}, \sum_{xi}). \tag{2.10}$$

The analyzed products, *x** are derived by minimizing the cost function

$$I(t) = (\varepsilon_u, \varepsilon_v, \varepsilon_q, \varepsilon_s)^T Q^{-1} (\varepsilon_u, \varepsilon_v, \varepsilon_q, \varepsilon_s), \text{ where } \vec{\varepsilon}_x^{\,T} = (\vec{x}^{o} - \vec{x}^{*})^T, \tag{2.11}$$

which is weighted sum of square errors. Here *x* can be any of the *u*, *v*, *q*, and *s*. And *Q* is the covariance structure of errors.

Based on equation (2.10), we have

$$Q = \begin{pmatrix} \mathrm{cov}(\vec{\varepsilon}_u, \vec{\varepsilon}_u^{\,T}) & 0 & 0 & 0 \\ 0 & \mathrm{cov}(\vec{\varepsilon}_v, \vec{\varepsilon}_v^{\,T}) & 0 & 0 \\ 0 & 0 & \mathrm{cov}(\vec{\varepsilon}_s, \vec{\varepsilon}_s^{\,T}) & 0 \\ 0 & 0 & 0 & \mathrm{cov}(\vec{\varepsilon}_q, \vec{\varepsilon}_q^{\,T}) \end{pmatrix}, \tag{2.12}$$

where $\mathrm{cov}(\vec{\varepsilon}_X, \vec{\varepsilon}_X^{\,T}) = \begin{pmatrix} \Sigma_{x1} & 0 & \cdots & 0 \\ 0 & \Sigma_{x2} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \Sigma_{xS} \end{pmatrix}_{KS \times KS}$.

Since *Q* can be estimated from the data (Appendix A), it will be treated as known in the algorithm design.

2.4 Algorithm description

We use the Lagrange multiplier method to derive the final analysis. Three constraints (Equations (2.1)-(2.3)) in Part 2.2 above are imposed in this case.

The Lagrange function *J* is can be expressed as

$$J(\vec{X}) = I(\vec{X}) + \sum_{l=1}^{3} \lambda_l A_l(\vec{X}). \tag{2.13}$$

After taking partial derivatives to each variable, the variational (Lagrange) equations are

$$\nabla_{\vec{X}} J(\vec{X}) = Q^{-1}(\vec{X} - \vec{X}_o) + \sum_{l=1}^{3} \lambda_l \nabla_{\vec{X}} A_l(\vec{X}) = 0 \tag{2.14}$$

$$\nabla_{\lambda_l} J(\vec{X}) = A_l(\vec{X}) = 0.(l = 1,2,3) \tag{2.15}$$

We use Newton's iteration method to derive the solutions. Figure 2.3 is the flow chart for the procedure which is divided into five steps as below:

1) Treat $\lambda's$ in the same position as *X*.

Let $\vec{X}^{*T} = (\vec{X}, \lambda_1, \lambda_2, \lambda_3)$. Set 1.0 as the initial values for all $\lambda's$ and the original 3d-data as initial conditions of all state variables. The gradient can be calculated as:

$$\nabla_{\vec{X}^*} J(\vec{X}^*) = [\nabla_{\vec{X}} J(\vec{X}), \nabla_{\lambda_1} J(\vec{X}), \nabla_{\lambda_2} J(\vec{X}), \nabla_{\lambda_1} J(\vec{X})]^T. \tag{2.16}$$

Note that the length of the gradient is (4**I***K*+3).

2) Calculate the 2nd order derivative matrix ***H*** in terms of $\boldsymbol{X}^*$:

$$H_{ij} = \frac{\partial^2 J}{\partial X_i^* \partial X_j^*}, (i, j = 1, ..., (4IK + 3)). \tag{2.17}$$

3) Calculate the perturbation of X^{*T}, $\delta \vec{X}^{*T}$, from the linear system:

(2.18)

$$H \times \delta \vec{X}^{*T} = -\nabla_{X^*} J(X^*).$$

Thus, $\delta \vec{X}^{*T} = -H^{-1} \nabla_{X^*} J(X^*)$ (2.19)

4) Update $\boldsymbol{X}^*$ as: $\vec{X}^{*(n+1)} = \vec{X}^{*(n)} + \delta \vec{X}^*$, where n is the loop index. (2.20)

5) Repeat steps 1)-4) with the updated values of X^*.

6) Loop continues until some preset conditions are satisfied. For example, we terminate the loops when the largest squared element of gradient is smaller than 1.5×10^{-15}.

2.5 Figures and results

Final analyses of all state variables (refer to Figure 2.4 for station 1, others similar) after adjustments share the same pattern with the observations. All the residual plots (refer to Figure 2.5 for station 1, others similar) are within an acceptable range. However, the change varies in those figures: some are larger while some are smaller than that of the original case.

We next compare the results related to the budgets. The meaning of each curve is shown in the legend. The original data without the constraints are not balanced, which is clear in Figures 2.6 (a) and (c). The black line, as the summation of all the terms, does not overlaps the *x* axis for both moisture and heat budgets. This is caused by the inevitable instrumental errors that are included in the sounding data collected. Especially the wind velocities in the lower pressure levels are difficult to record accurately. After the variational method with constraints is applied, the black line is forced to be zero for both moisture and heat budgets, which is shown in plots (b) and (d). This is a significant improvement for our enough accurate data, because the column integrated budgets are zero in reality. After the constrained variational method is applied, even if

the truth can be never known due to systematic errors, we can have a much higher confidence to believe that the results are much closer to the true values than the one before the method. Additionally when precipitation is large, horizontal water vapor advection is known to be the dominant component of precipitation, and horizontal advection is a major part of energy transport for latent heat released from precipitation. This is in fact well demonstrated in plots (b) and (d), since the green line and the dotted line agree closely with each other in both plots. This is another advantage of the constrained variational method.

Figure 2.7 shows the correlation coefficients for adjustments between different vertical levels for state variables *u*, *v*, *q* and *s*. All four plots indicate that large correlations for the errors exit between different levels of all state variables, which confirm our original assumption. Generally, the adjustment for all four variables at each pressure layer has a positive correlation with the adjacent levels. The correlations tend to be smaller when the two levels become farther, and finally becomes negative when the two levels are departed far enough.

Figure 2.8 shows the difference of important physical variables between the our new model and the original case. This indeed gives us new information that is added in the results of vertical correlated case, as compared to the original case. Particularly the adjustment of the vertical advection for moisture and heat and Q_1 and Q_2 from the vertical correlated case is quite different from the uncorrlated case (Zhang and Lin, 1997). The incorporation of vertical correlation has apprent patterns for the values with different signs in the lower and higher levels, i.e., negative values in the lower level corresponding to positive values in the higher levels and vice versa. This change will be further demonstrated by correlation strctures in Figure 2.9.

Figures 2.9 shows the correlation structure of apparent moisture sink and heat source incorporated in our new model. In Figures 2.9(c) and (d), Q_2 and Q_1 in lower levels (lower panels) compensates well to those in higher levels (upper panels).

The correlation structure is similar for the vertical advection for both moisture and heat (Figures 2.9(a) and (b)). The results are interesting, because the vertical correlation, which is originally considered in 3d-data set in the case, is also significant in the error structures of physical variables, such as Q_1 and Q_2. Moreover, the structure of lower pressure level (around 600 hPa or below) compensates that of higher pressure levels (above 600 hPa). Such compensation promises balanced budgets of mass, moisture and static energy after vertical integration.

We next show the improved results after the new adjustments in the vertical advection of moisture and energy, Q_2, and Q_1. Figures 2.10 (a) and (b) shows the vertical advection for the moisture sink and apparent heat before and after the constrained variational method incorporated with the vertical correlation, respectively. Two peaks occur around January 21~23 in (a), while only one shows up in (b) with another less strong signal for the vertical advection of moisture across the levels between 1000 and 200 hPa. Given that precipitation is the main source for the vertical advection of moisture when the precipitation is of large amount, this result can be expected by noting that in Figure 2.6(b), the green curve, representing negative total precipitation, has only one peak around the same time as the peak vertical advection occurs. The other peak in (a) may be generated by the uncertainties in the variational analysis, which does

not occur in truth. Therefore, the results in (b) show great improvement over those in (a) by comparing with the precipitations. Similar improvement occurs in Q_2. For apparent moisture sink, the vertical advection is more dominant than the horizontal advection, and the structure of Q_2 is much more coherent with the precipitation distribution after the variational adjustment (Figure 2.10 d) than before (Figure 2.10 c).

Similar improvement is made from the viewpoint of energy. In Figure 2.11, because the vertical advection is dominant in Q_1 when precipitation is of large amount, the pair of (a) and (c) or (b) and (d) has similar patterns except for the sign. Besides, in the original 3d-data, the structure is not smooth and may contain large errors at upper levels where pressure is less than 200 hPa in (a) and (c). After new the adjustment by the incorporating vertical correlations in the constraint variational method, the errors are significantly reduced at those pressure levels, and the structure becomes rather smooth for the entire time period regardless of the pressure level.

To summarize, the new constrained variational analysis method with vertical correlations shows significant improvement for calibrating the sounding data in (1) dramatically minimizes the systematic errors by reinforcing balanced budgets, (2) significantly improves the important physical quantities such as the vertical advections, apparent moisture sink, and apparent heat source by removing the errors especially in the high pressure levels. This method also confirms the significant vertical correlations in the state variables, as well as demonstrates a compensation function of high pressure levels to the low pressure levels.

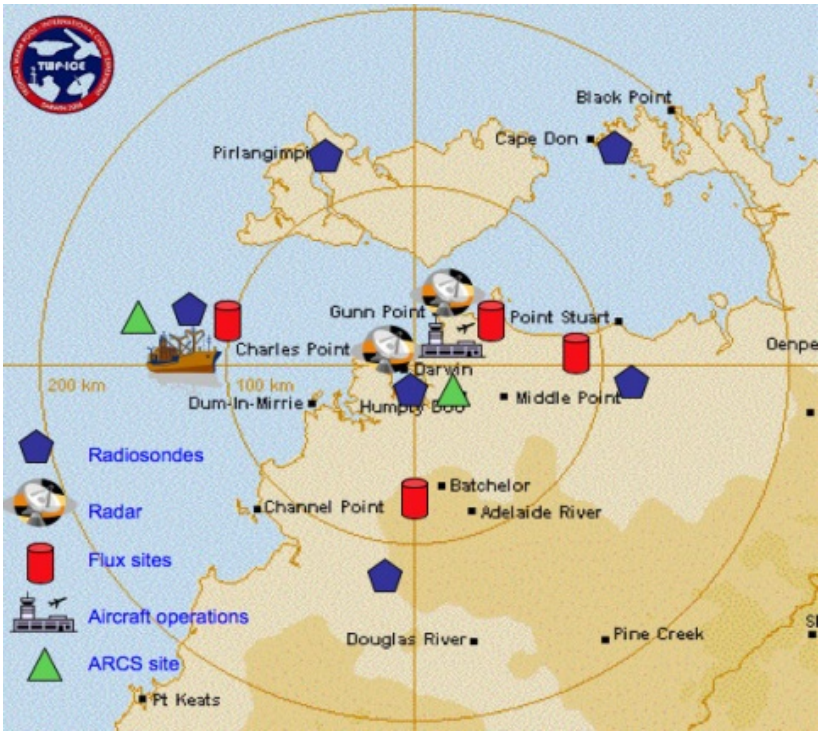


Figure 2.1: The area map of Darwin, Australia. The six radiosonde (dark-blue pentagons) stations observe the 3d-data set for temperature, dew-point temperature (for mixing ratio of water vapor or relative humidity), and two components of wind velocity.

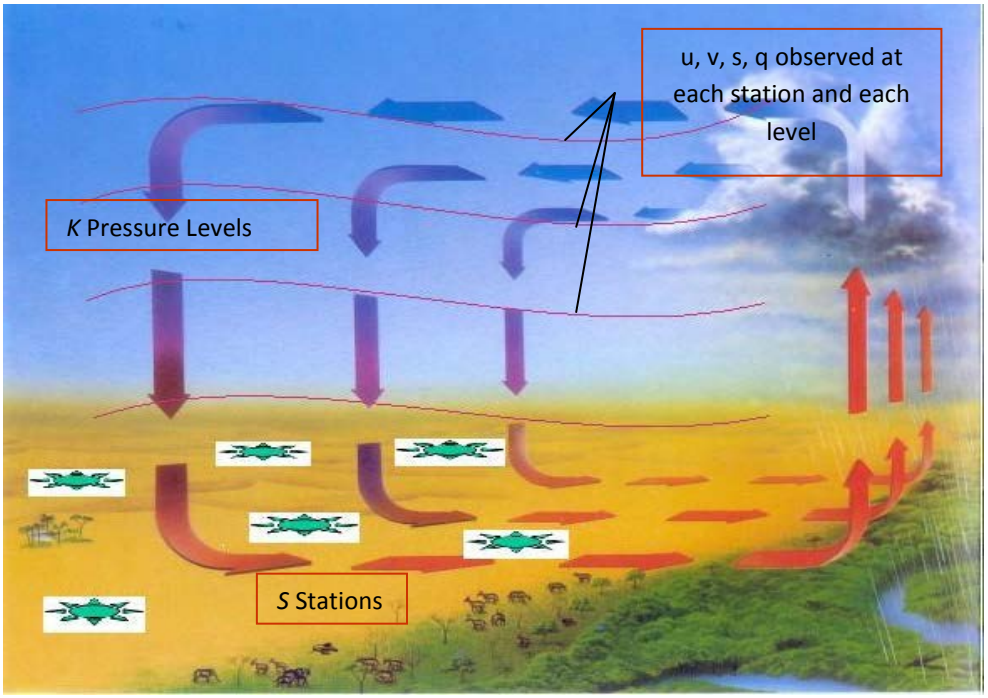


Figure 2.2: Schematic diagram for the structure of the 3d-data set, where there are a total of *S* stations horizontally and *K*-pressure levels vertically. At each station and every level, and every three hours, we observe the *u*, *v*, *q*, and *s* variables. The measurements in Darwin cover from 19 January to 12 February, 2006. In this study, we have *S* = 6, *K* = 45, and *T* = 201.

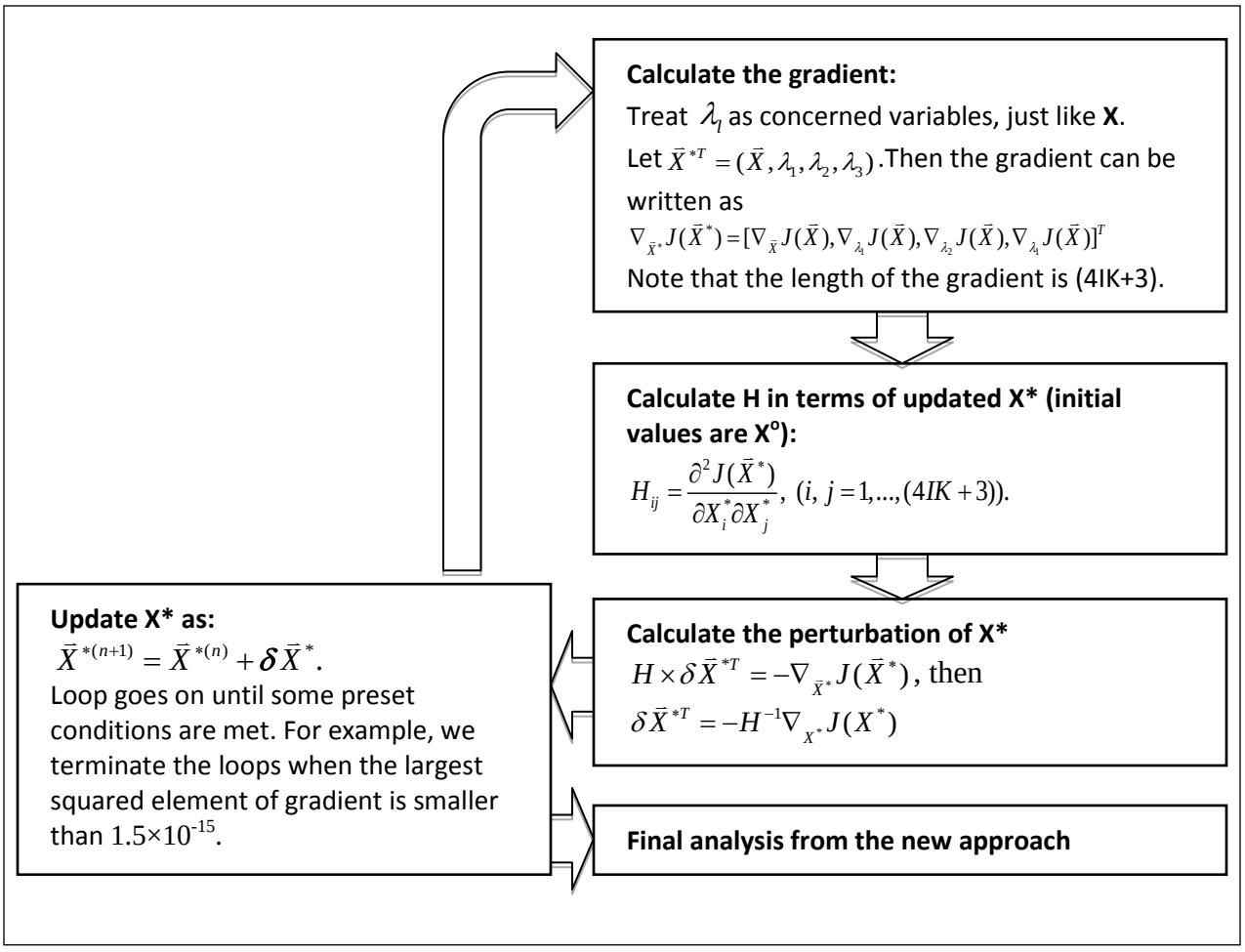


Figure 2.3: Flow chart of the algorithm for the new vertically correlated case.

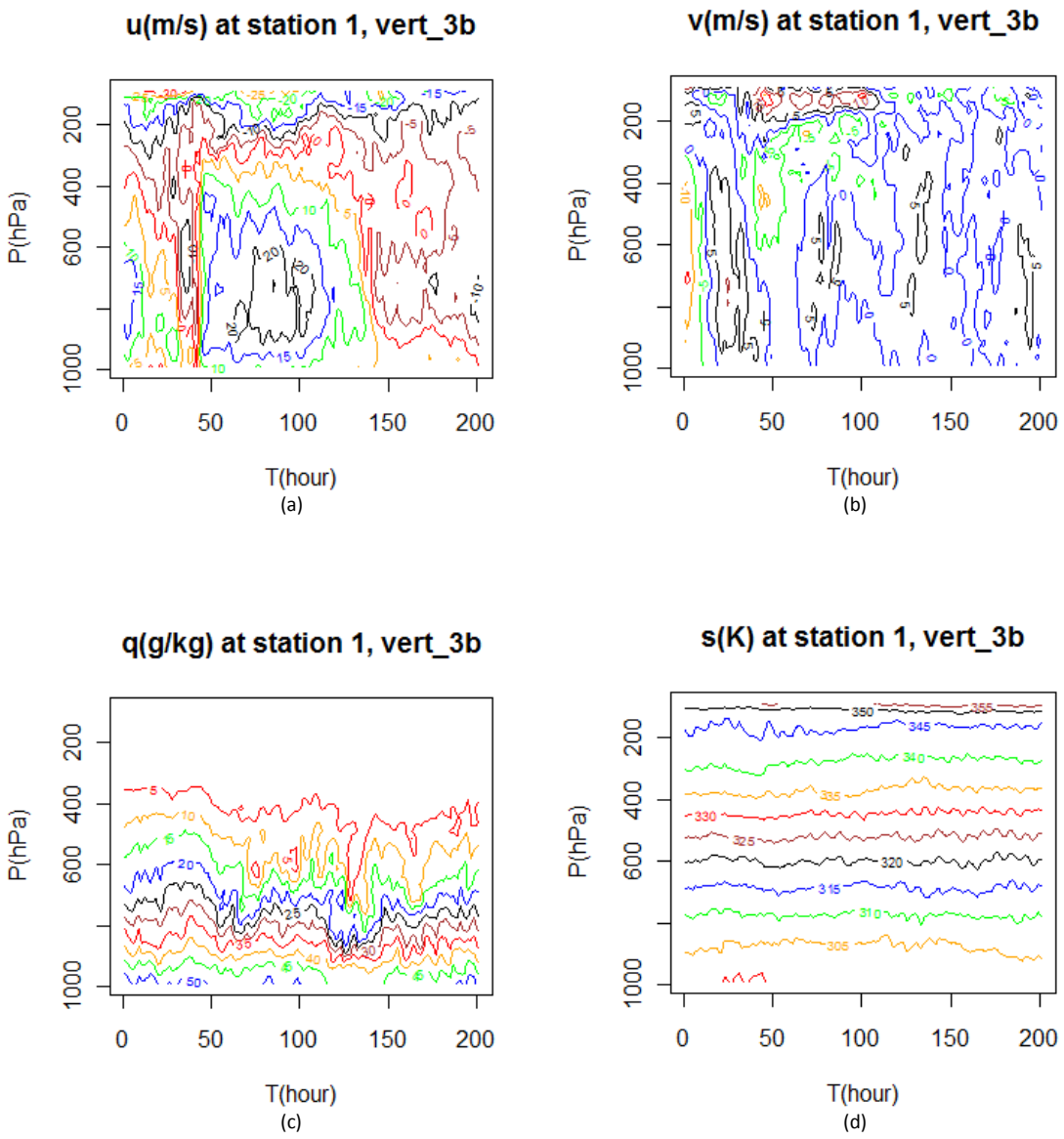


Figure 2.4: Final results for u, v, q and s from vertical correlated case at station 1, shown in (a)-(d), respectively.

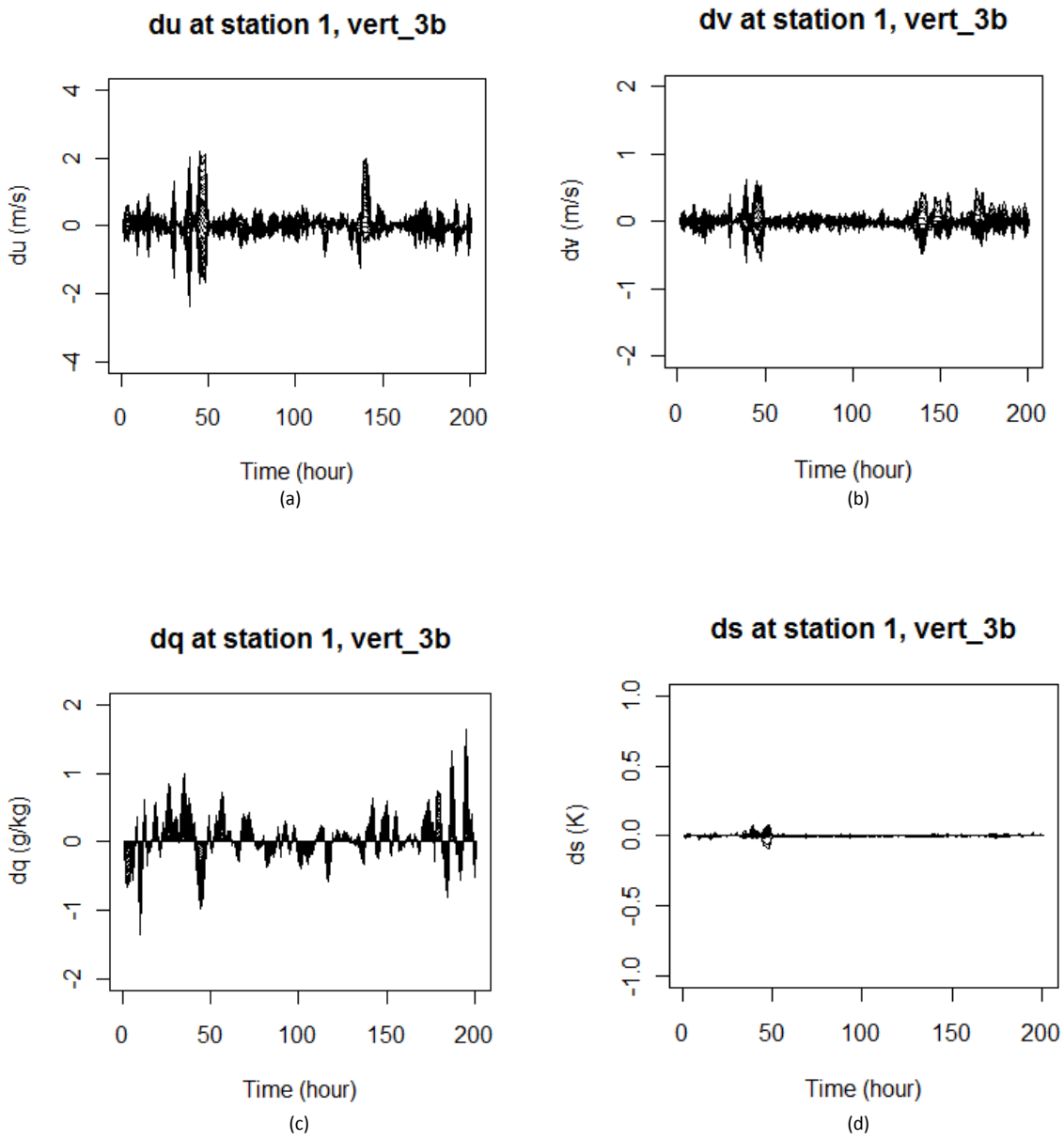


Figure 2.5: Adjustments of u, v, q and s from vertical correlated case station 1, shown in (a)-(d), respectively.

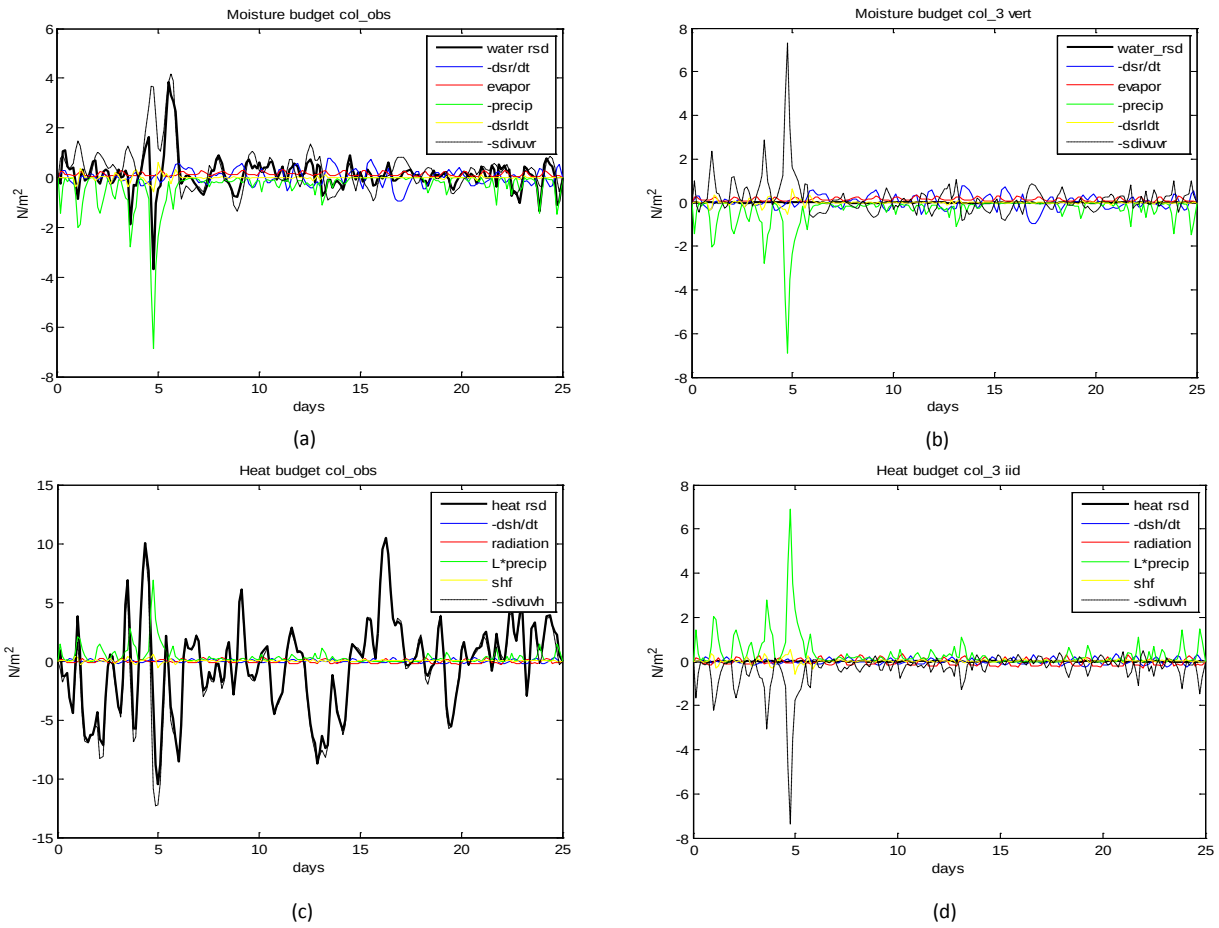


Figure 2.6: (a) and (b) for terms of column integrated moisture budgets before and after the vertical correlated variational method, respectively; (c) and (d) for terms of the column integrated heat budgets before and after vertical correlated variational method, respectively.

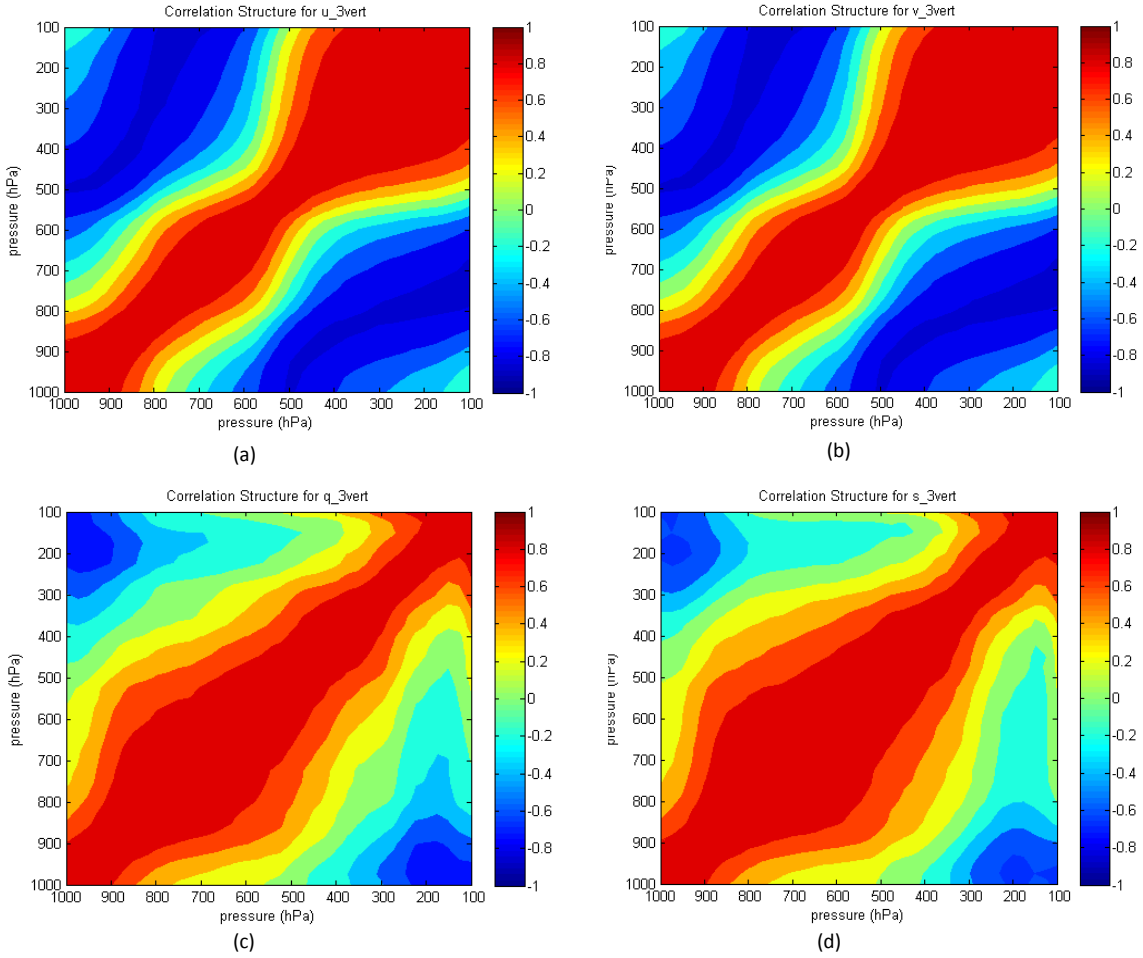


Figure 2.7: The correlation structure of the adjustment for state variables *u*, *v*, *q*, and *s*, shown in (a)-(d), respectively.

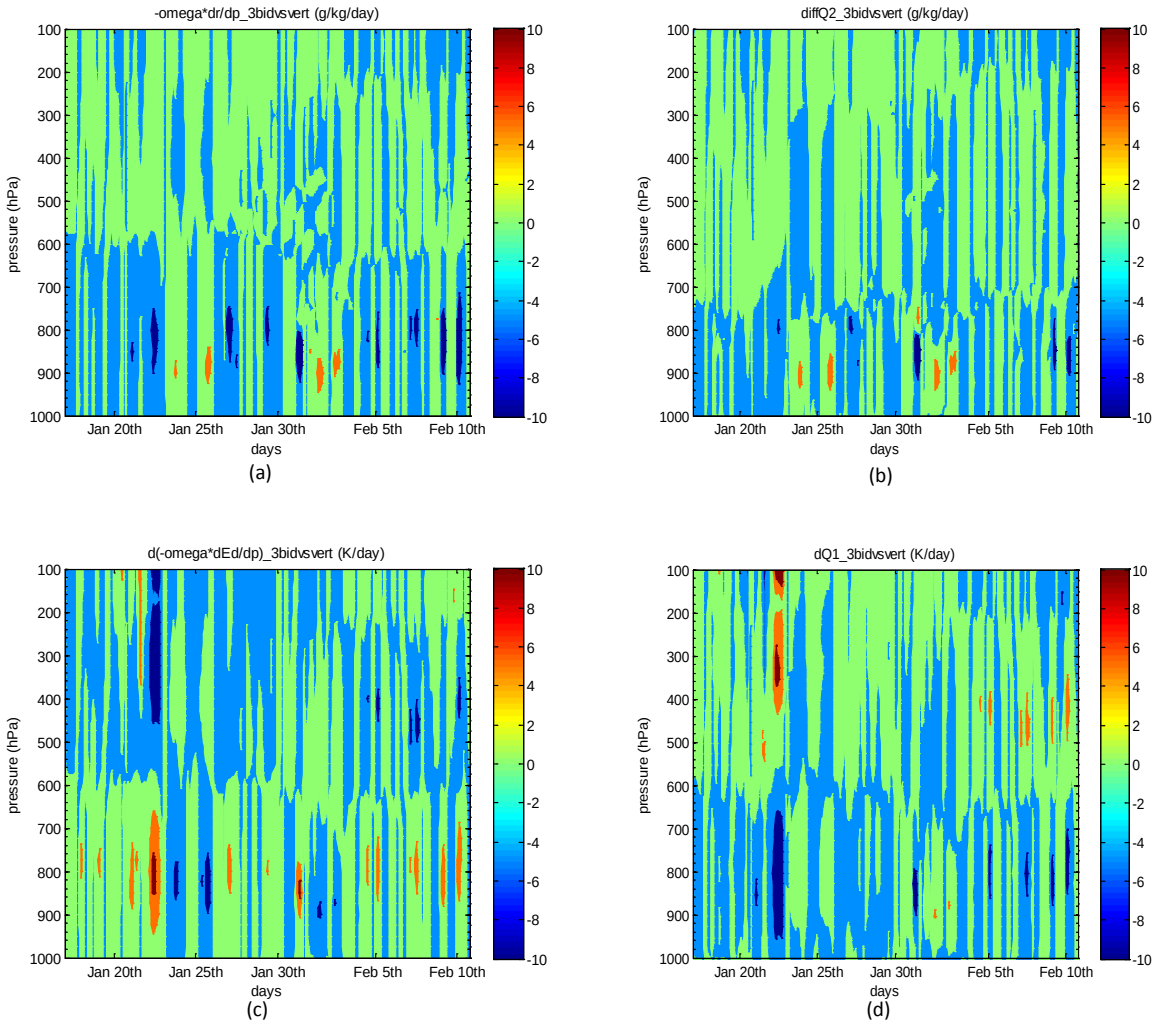


Figure 2.8: Differenced vertical advection for moisture (a), Q_2 (b), differenced vertical advection for heat (c) and Q_1 (d) between vertical correlated case and the original case(Zhang & Lin 1997).

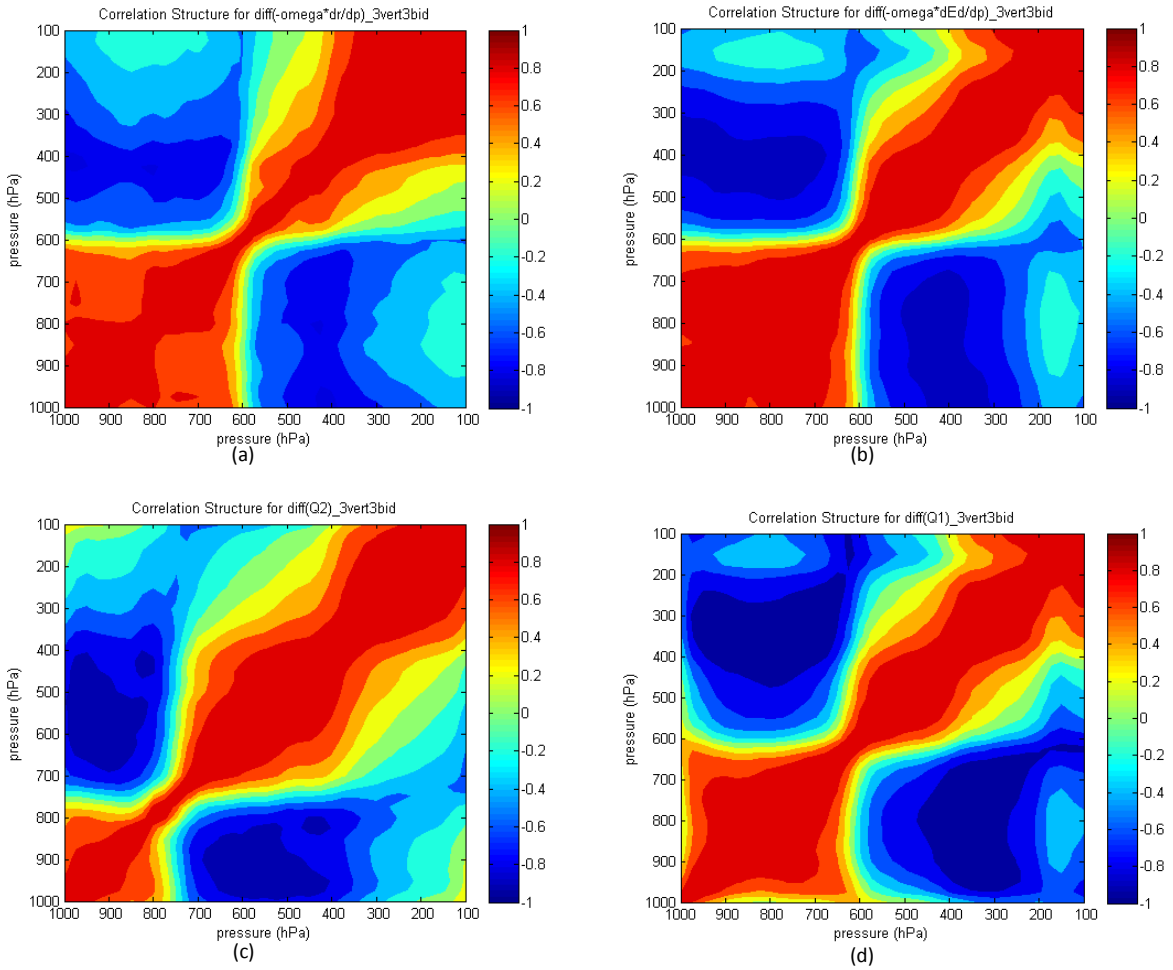


Figure 2.9: Correlation structures of difference values of (a) vertical advection for moisture, (b) vertical advection for static energy, (c) Q_2 and (d) Q_1 between vertical correlated case and the original case.

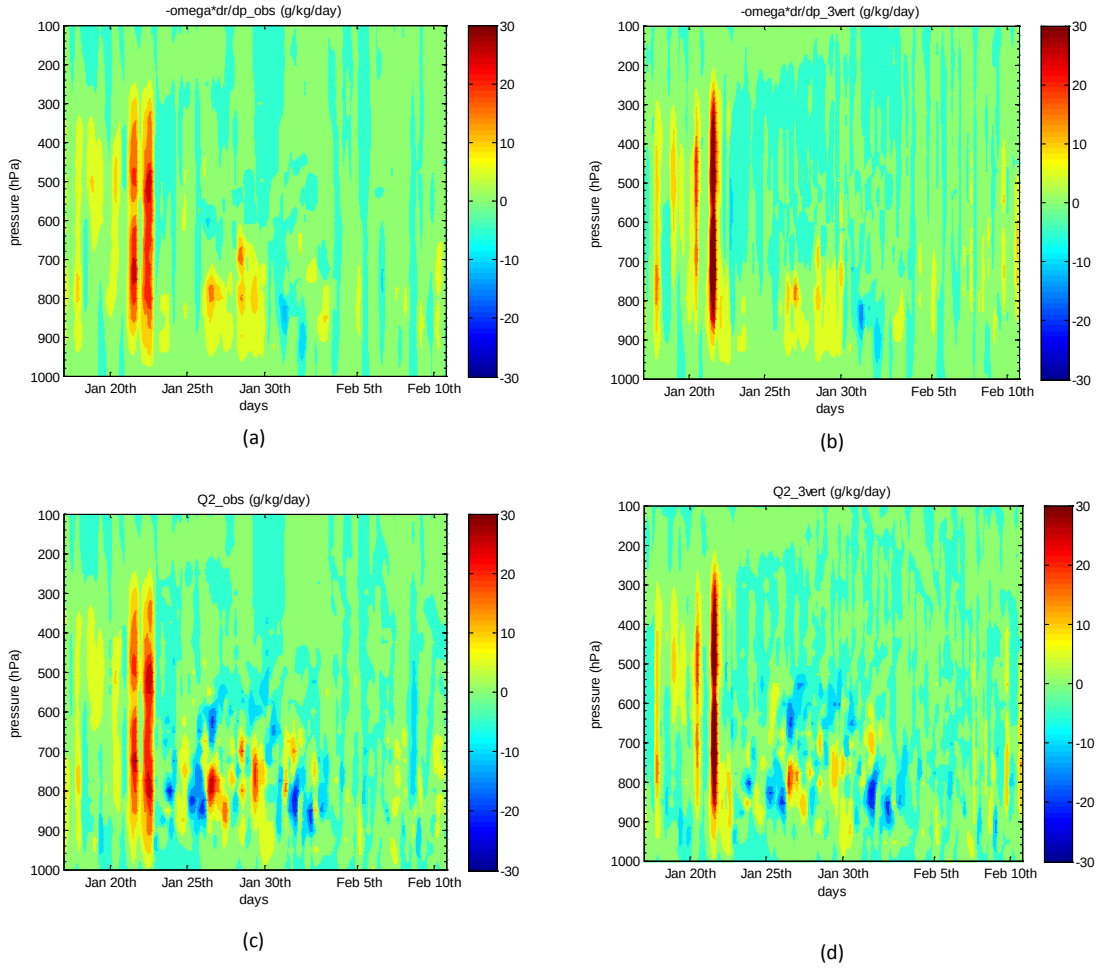


Figure 2.10: Vertical advection for moisture of (a) original 3d-data versus (b) final results from new variational method; the apparent moisture sink of (c) original 3d-data versus (d) final results from new variational method.

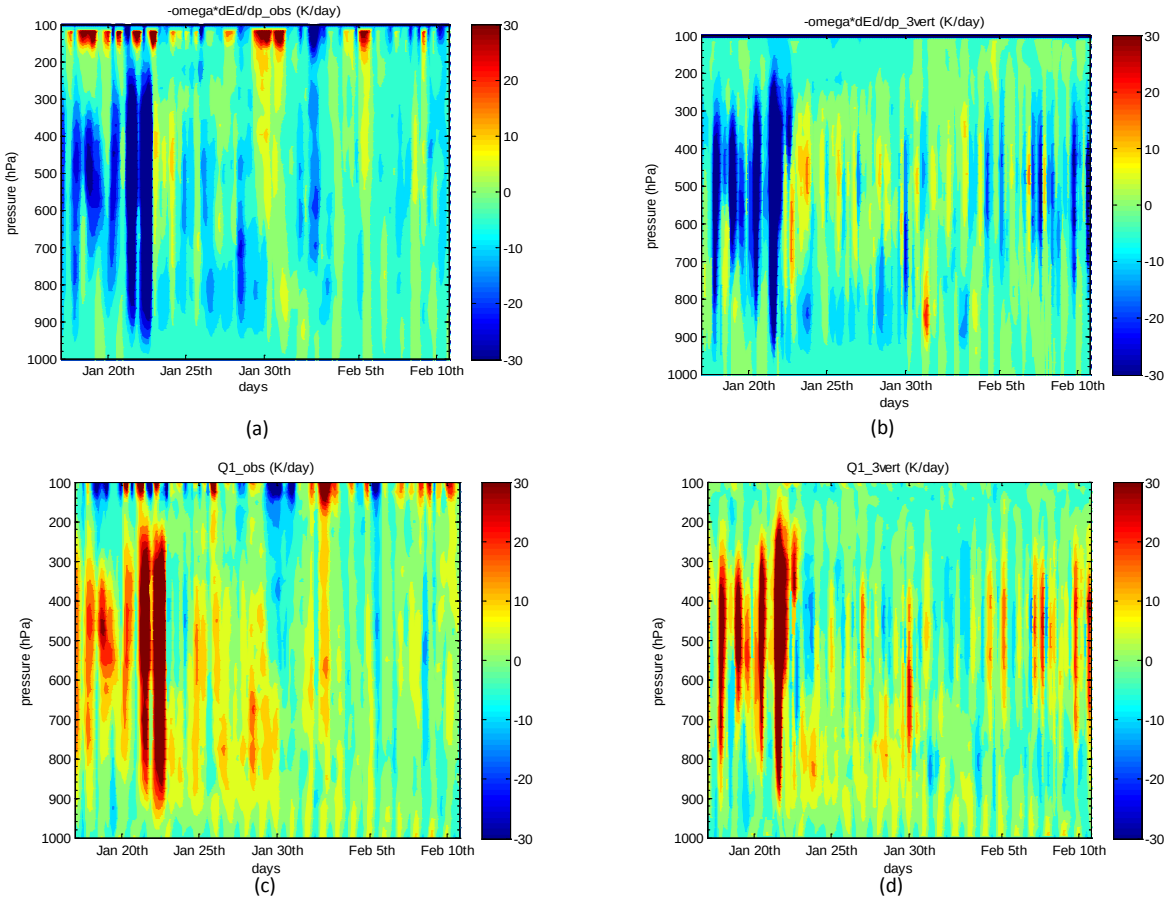


Figure 2.11: Vertical advection for static energy of (a) original 3d-data versus (b) final results from new variational method; the apparent heat source of (c) original 3d-data versus (d) final results from new variational method.

Chapter 3

Constrained Variational Analysis Integrating Temporal Correlations and Three Constraints

3.1 Model description

The data used here are identical to those in the vertical correlated case of Part 2.1, Chapter 2. The physical budgets are also the same with a detailed description presented in Part 2.2 of Chapter 2, except for the notations for the time dimension are changed slightly.

Let $x^o{}_t$ be the observed vector at time t and x_t the predicted vector (the analysis) determined by physical systems $L(x) = 0$, where L represents the operators for the partial differential equation (PDE) mentioned in Chapter 2. Thus we have

$$\boldsymbol{X}_t{}^T=(x_1, x_2, x_3,\ldots,x_T). \tag{3.1}$$

The vector $e_t:=x_t - x^o{}_t$ represents the difference between the observed and true values. The current literature assumes that the components of vector x_t are independent, based on which the x_t can be estimated by minimizing the residual sum of its squares (RSS)

$$C = \sum_{t=1}^{T}\sum_{j=0}^{\infty}(\vec{x}_t - \vec{x}_t^o)^T \Omega^{-1}(j)(\vec{x}_{t+j} - \vec{x}_{t+j}^o), \tag{3.2}$$

where Ω is the weighting matrix related to the error structure. Since the seasonal components of x_t and $x^o{}_t$ will be canceled out, the time series of x_t can be transferred to e_t without any seasonal effect. Specifically, let $e_t^{rv,p,s}$ represent the time series of state variables at pressure p and at the station s. Fitting an *AR(1)* model to yield

$$e_t = \alpha_{rv,i,k} e_{t-1} + \varepsilon_t, \text{ with } \{\varepsilon_t\} \sim N(0, \sigma_{rv,i,k}{}^2). \tag{3.3}$$

The autocovariance function is

$$\gamma_j = Cov(e_t, e_{t+j}) = \frac{\alpha_{rv,i,k}^j}{1-\alpha_{rv,i,k}^2} \sigma_{rv,i,k}^2 \tag{3.4}$$

which is derived as follows.

Multiplying e_t and e_{t-1} on both sides of (3.3), and then computing the covariance yield

$$\gamma_0 = \alpha\gamma_1 + \sigma^2 \tag{3.5}$$

$$\gamma_1 = \alpha\gamma_0 \tag{3.6}$$

From (3.5) and (3.6) we have

$$\gamma_0 = \frac{\sigma^2}{1-\alpha^2} \tag{3.7}$$

Multiplying e_{t-j} on both sides of (3.3) and calculating its covariance, we have

$$\gamma_j = \alpha\gamma_{j-1}, \text{ where } j = 1, 2, ... \tag{3.8}$$

Finally from equation (3.7) and (3.8), we have (3.4).

Hence the covariance structure of the observed data $e^{rv,i,k} = (e_1,\ldots,e_T)$ is expressed as

$$\Omega^{rv,i,k} = \frac{\sigma^2_{rv,i,k}}{1-\alpha^2_{rv,i,k}} \begin{pmatrix} 1 & \alpha_{rv,i,k} & \alpha^2_{rv,i,k} & \cdots & \alpha^{T-1}_{rv,i,k} \\ \alpha_{rv,i,k} & 1 & \alpha_{rv,i,k} & \cdots & \alpha^{T-2}_{rv,i,k} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \alpha^{T-1}_{rv,i,k} & \alpha^{T-2}_{rv,i,k} & \cdots & \cdots & 1 \end{pmatrix}. \tag{3.9}$$

As $\Omega^{rv,i,k}$ is always a symmetric positive definite matrix, it can be decomposed as $\Omega^{rv,i,k} = L \cdot L^T$ using the Cholesky's method, where L is a lower triangular matrix. Since L can be inversed easily, the inverse of $\Omega^{rv,i,k}$ can be expressed as

$$(\Omega^{rv,i,k})^{-1} = \frac{1}{\sigma^2_{rv,i,k}} \begin{pmatrix} 1 & -\alpha_{rv,i,k} & 0 & 0 & \cdots & 0 \\ -\alpha_{rv,i,k} & 1+\alpha_{rv,i,k}{}^2 & -\alpha_{rv,i,k} & \ddots & \ddots & \vdots \\ 0 & -\alpha_{rv,i,k} & 1+\alpha_{rv,i,k}{}^2 & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 1+\alpha_{rv,i,k}{}^2 & -\alpha_{rv,i,k} \\ 0 & \cdots & 0 & 0 & -\alpha_{rv,i,k} & 1 \end{pmatrix}_{T\times T}. \tag{3.10}$$

We next show that the cost function in this case has the form similar to the equation (3.2) by using the maximum likelihood method.

From (3.3), we have

$$e_t - \alpha e_{t-1} = \varepsilon_t. \tag{3.11}$$

Since ε_t follows $N(0,\sigma^2)$, the likelihood function is

$$M = \prod_{i=1}^{T} \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{\varepsilon_i^2}{2\sigma^2}), \tag{3.12}$$

After taking logarithm on both sides of equation (3.12), we have

$$\ln M = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{\sum_{i=1}^{T}\varepsilon_i^2}{2\sigma^2} \tag{3.13}$$

To maximize the likelihood function (3.12) is equivalent to minimize

$$\begin{aligned} Q &= \sum_{i=1}^{T}\varepsilon_i^2 \\ &= \sum_{i=1}^{T}(e_t - \alpha e_{t-1})^2 \\ &= \sum_{t=1}^{T}\sum_{j=0}^{\infty}(\vec{x}_t - \vec{x}_t^{\,o})^T \Omega^{-1}(j)(\vec{x}_{t+j} - \vec{x}_{t+j}^{\,o}) \\ &= \text{C, where } e_t = x_t - x_t^o \end{aligned} \tag{3.14}$$

This confirms the shape of cost function *C* is similar to equation (3.2).

Our goal, similar to that in the vertical correlated case, is to force the atmospheric state variables to satisfy conservations of mass, moisture and energy through a variational technique while to make minimum adjustment to the original sounding data by using the cost function *C*. We again use the Euler-Lagrange multiplier method. The Lagrangian function is

$$L = C + \lambda_1 A_1 + \lambda_2 A_2 + \lambda_3 A_3. \tag{3.15}$$

The derivative of *L* needs to be taken for each controlled variable and λ to find the optimal values that make all derivatives zero. As the budget functions are quadratic rather than linear, we have to select some iteration methods that make the process converge to the optimal values.

3.2 Algorithm description

Newton's iteration method is not suitable in this case. Because all the state variables are coupled vertically, horizontally and temporally, the inverse of a matrix with very large dimensions has to be calculated before updating the controlled variables at each step. This makes the convergence rather slow. To speed up the process, we adopt the algorithm of fixed point iteration.

Let $L = (A_{1t}, A_{2t}, A_{3t})$ be the three constraints at time step t, λ_{1t}, λ_{2t}, λ_{3t} be the corresponding multipliers and Q be the derivative matrix of the constraints to vector $\boldsymbol{X}_t$. The three constraints again are the column integrated budget of mass, moisture and static energy, as introduced in Part 2.2 of Chapter 2. Given that all constraints at a specific time step are independent with the state variables at other time steps, the derivative matrix Q has the form of

$$Q = \begin{pmatrix} \lambda_{11}\frac{\partial A_{11}}{\partial x_1} & \lambda_{21}\frac{\partial A_{21}}{\partial x_1} & \lambda_{31}\frac{\partial A_{31}}{\partial x_1} & 0 & 0 & 0 & \cdots & \cdots & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_{12}\frac{\partial A_{12}}{\partial x_2} & \lambda_{22}\frac{\partial A_{22}}{\partial x_2} & \lambda_{32}\frac{\partial A_{32}}{\partial x_2} & \cdots & \cdots & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & \cdots & \cdots & \lambda_{1T}\frac{\partial A_{1T}}{\partial x_T} & \lambda_{2T}\frac{\partial A_{2T}}{\partial x_T} & \lambda_{3T}\frac{\partial A_{3T}}{\partial x_T} \end{pmatrix}_{T\times 3T} \tag{3.16}$$

1) The variational equations (Euler-Lagrange equations) for the analyzed variables $\boldsymbol{X}_t^T$ are

$$2\Omega^{-1}(\vec{x} - \vec{x}^o) + Q = 0, \text{ or equivalenly, } \vec{x} = \vec{x}^o - \tfrac{1}{2}\cdot\Omega Q, \tag{3.17}$$

where Ω is calculated in equation (3.14).

2) By substituting *u*, *v*, *q*, and *s* with formula (3.17), constraint A_{it} ***(u,v,q,s)*** *= 0 (i = 1 , 2 , 3, t = 1, .., T)* becomes a function of $\lambda_{1t}, \lambda_{2t}, \lambda_{3t}$. The Taylor's expansion is then applied to each A_{it} for the 1st order of $\vec{\lambda}$ around $\vec{\lambda} = \vec{0}$, where $\vec{\lambda}^T = (\lambda_{11}, \lambda_{21}, ...\lambda_{3T})$. as

$$A_{it}(\vec{\lambda}) = A(\vec{0}) + \frac{\partial A}{\partial \vec{\lambda}^T}\Big|_{\vec{\lambda}=\vec{0}} \vec{\lambda} = 0. \tag{3.18}$$

3) The system (3.16) yields a linearized set of equations for $\vec{\lambda}$. We can numerically calculate the terms $\frac{\partial A}{\partial \vec{\lambda}^T}$ that can be used to solve $\vec{\lambda}$.

4) The α can be estimated from the updated value as $\alpha = \gamma_1/\gamma_0$, where γ_i is the sample autocovariance function of the *AR(1)* model.

Now the state variables can be updated with formula (3.17) and then the steps 1)-4) are repeated. This procedure continues recursively until some conditions are met. In this case, the condition is that the RSS becomes less than 1.5×10^{-10}. Results show that the convergence has already occurred within eight cycles of iteration for the whole measurement period.

3.3 Figures and results

Similarly to the results of the vertical correlated case in Chapter 2, the final analysis here also shares the same pattern with the observations, as shown in Figure 3.2 (all the plots are given in the appendix). Moreover, all the residuals plots are within an acceptable range at the central

station (Figure 3.3). For different stations, the results vary. Some are larger than that of the original uncorrelated case (Zhang and Lin, 1997) and some are smaller. Overall, improvement is clear.

If the α is set to zero, this CVA model is simplified to the original uncorrelated case (Zhang and Lin, 1997). We have selected different α and run the variational analysis, and then plot the residuals according to different α. The figure show that, in terms of sum of squared residuals, the temporal *AR(1)* model is much better than the uncorrelated case. The residual reaches minimum when α is around 0.69. Interestingly, the α is convergent to 0.69 even with more iterations. Therefore we can conclude that the weighted residual sum of square in the *AR(1)* model is much smaller than in the original case. We prove this conclusion mathematically below.

Let function *I(α)* be the cost function for the temporal correlated model. Note that all parameters except α, the coefficient of the *AR(1)* series model, are either given or can be estimated. The adjusted value of cost function is thus related only to α. Let *H* be the cost function of the uncorrelated case (Zhang and Lin, 1997). When $\alpha = 0$, the weighting matrix reduces to diagonal form that is for the uncorrelated case. Thus we have the following formula

$$H = I(0). \tag{3.19}$$

Since in the temporal correlated case $\alpha \in [0,1)$ and $\{0\} \in [0,1)$, we have

$$H = I(0) \le \min_{\alpha \in [0,1)} I(\alpha). \tag{3.20}$$

We next explain why the *I* does not reach minimum at $\alpha = 0$ for the uncorrelated case.

After taking the derivative of Lagrangian function *L* and setting it to zero, we get

$$\alpha(e_2^2 + e_3^2 + \cdots + e_{T-1}^2) - (e_1 e_2 + e_2 e_3 + \cdots + e_{T-1} e_T) = 0 \quad . \tag{3.21}$$

Thus

$$\alpha = \sum_{i=2}^{T} e_{i-1} e_i \Big/ \sum_{i=2}^{T-1} e_i^2 \,, \tag{3.22}$$

which is obviously not zero but a value between 0 and 1. This also justifies why our new model with temporal correlations has great advantages over the uncorrelated one.

The meaning of each line is shown in the legend. Similar to the vertical correlated case (refer to Chapter 2), the original measurements without the constraints are not balanced, as shown in Figures 3.3 (a) and (c), since the black line, which represents the total of all the budget terms, is not zero for both integrated moisture and water vapor budgets. The reason is anticipated due to the inevitable errors that exist in the original sounding data, especially the wind velocity at high levels (low pressures). On the contrary, in Figures 3.3 (b) and (d), the black lines lie exactly on the x-axis, which means that the budgets preselected become balanced again in the temporal correlated case. This balance suggests a high coherence with the theory that the column integrated balance should be zero. Based on this balance, we can predict that the final results are much closer to truth than those before adjustment and the perturbation can be an estimate of the data errors during measurement. In addition, it is known that when precipitation is large, the horizontal energy advection is a major component of the latent heat from precipitation, which is

clearly demonstrated in the plots (b) and (d), since the green and dotted lines are perfectly symmetric about the *x*-axis.

Figure 3.6 shows the error structure of the state variables. Plots are similar for all the four variables. Values are large on the diagonal and its adjacent lines, which verify that the errors have a very strong autoregressive feature.

We next show the auto-correlation error structure for physical variables, or the difference of vertical advection for moisture Q_2 and heat Q_1 of the *AR(1)* model compared to the uncorrelated case (Zhang and Lin, 1997).

Plots above have shown the new information that the *AR(1)* model has added to the physical variables in terms of correlation structure compared to the structure from original model. Figure 3.7 has clearly shown that the autoregressive feature occurs in the important physical terms, such as vertical advection for moisture and static energy, as well as the apparent moisture sink and heat source (Q_2 and Q_1, respectively), after using the *AR(1)* assumption in the original sounding data.

Figure 3.8 shows both vertical advection of moisture and apparent moisture sink before and after the new CVA method incorporated with the *AR(1)* feature, respectively. Similar to the CVA model with vertical correlations model (refer to Chapter 2), the two peaks around January 21st

and 23rd , 2006 in (a) and (c) have also been adjusted to one peak by the new model, as is shown in (b) and (d), respectively. This is due to the large precipitation during this period. As is shown in Figure 3.5(a), the green line, which depicts the precipitation, has only one large peak during the same period from January 21st. Since precipitation should dominate the vertical advection for moisture when it is of large amount, the analysis from the model is much improved. As the vertical advection of moisture is a much larger part of apparent moisture sink than the horizontal advection, the two peaks in (c) have also been adjusted to one peak in (d). Therefore, we conclude that the vertical advection for both moisture and Q_2 has shown a significantly improved consistency with the precipitation figure after being adjusted by the new CVA method.

For the static energy, the vertical advection is dominant when precipitation is large. This advection (Figures 3.9(a) and (c)) has a similar structure to the Q_1 (Figures 3.9(b) and (d)), except for the different sign. Impacted by the only one strong peak of precipitation, the peak has been weakened after the CVA method around January 21st. And also, large errors at high levels for the two physical variables (pressure below 140 hPa, refer to Figures 3.9 (a) and (c)) have been removed by the method (refer to Figures 3.9(b) and (d)), which shows a very large advantage of the final analysis.

The new constrained variational analysis method has successfully incorporated the temporal correlations among state variables via an *AR(1)* model. It improves the sounding data in (1) successfully providing a method to estimate the temporal correlation of true values by finding the optimal α through a new numerical iteration method; (2) dramatically minimizing the RSS

during the search of an optimal α; (3) significantly improving the important physical quantities, such as the vertical advections, apparent moisture sink, apparent heat source etc. by removing the errors especially at high pressure levels. The autoregressive temporal relationship among these variables has also been verified.

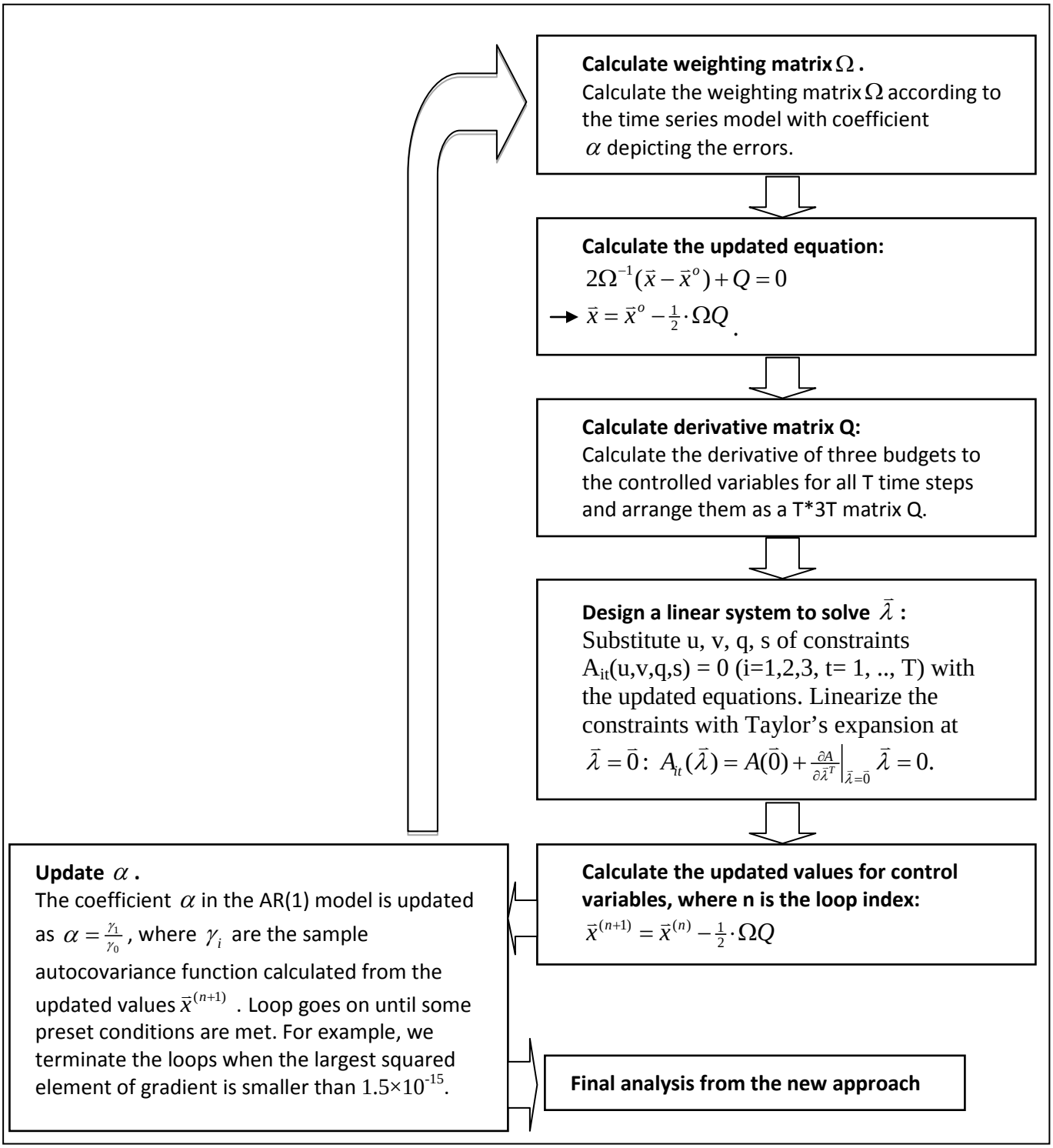


Figure 3.1: Flow chart of the algorithm for the temporal correlated case.

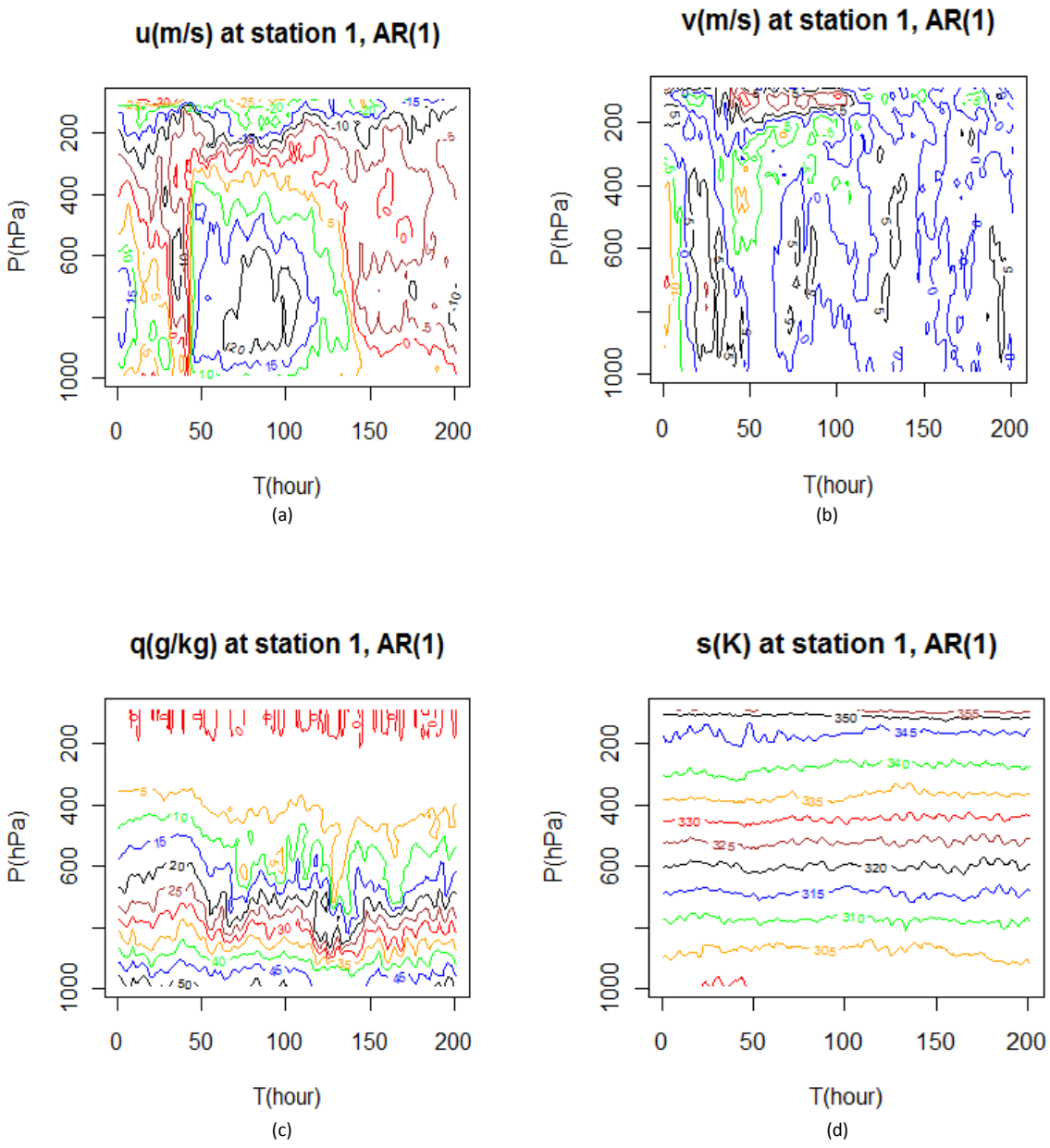


Figure 3.2: Final results for u, v, q and s from temporal correlated case with AR(1) feature at station 1, shown in (a)-(d), respectively.

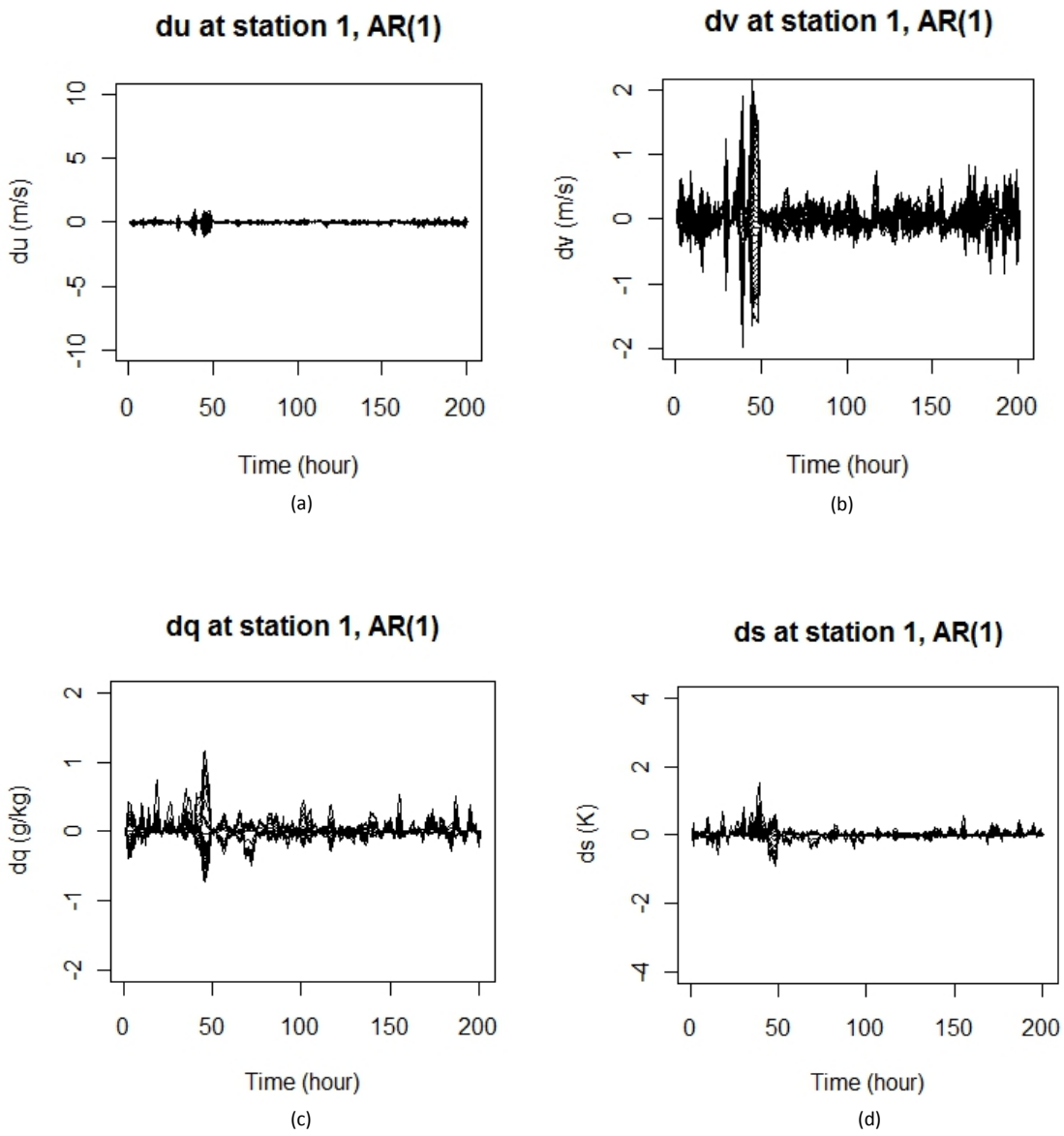


Figure 3.3: Adjustments of u, v, q and s from temporal correlated case with AR(1) feature at station 1, shown in (a)-(d), respectively.

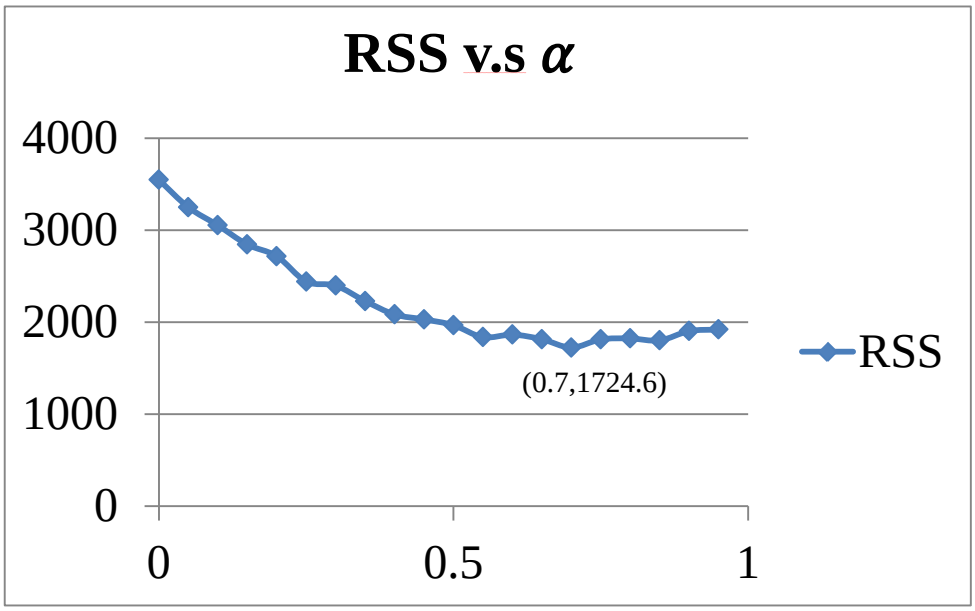


Figure 3.4: Curve of the residual sum of squares RSS v.s α.

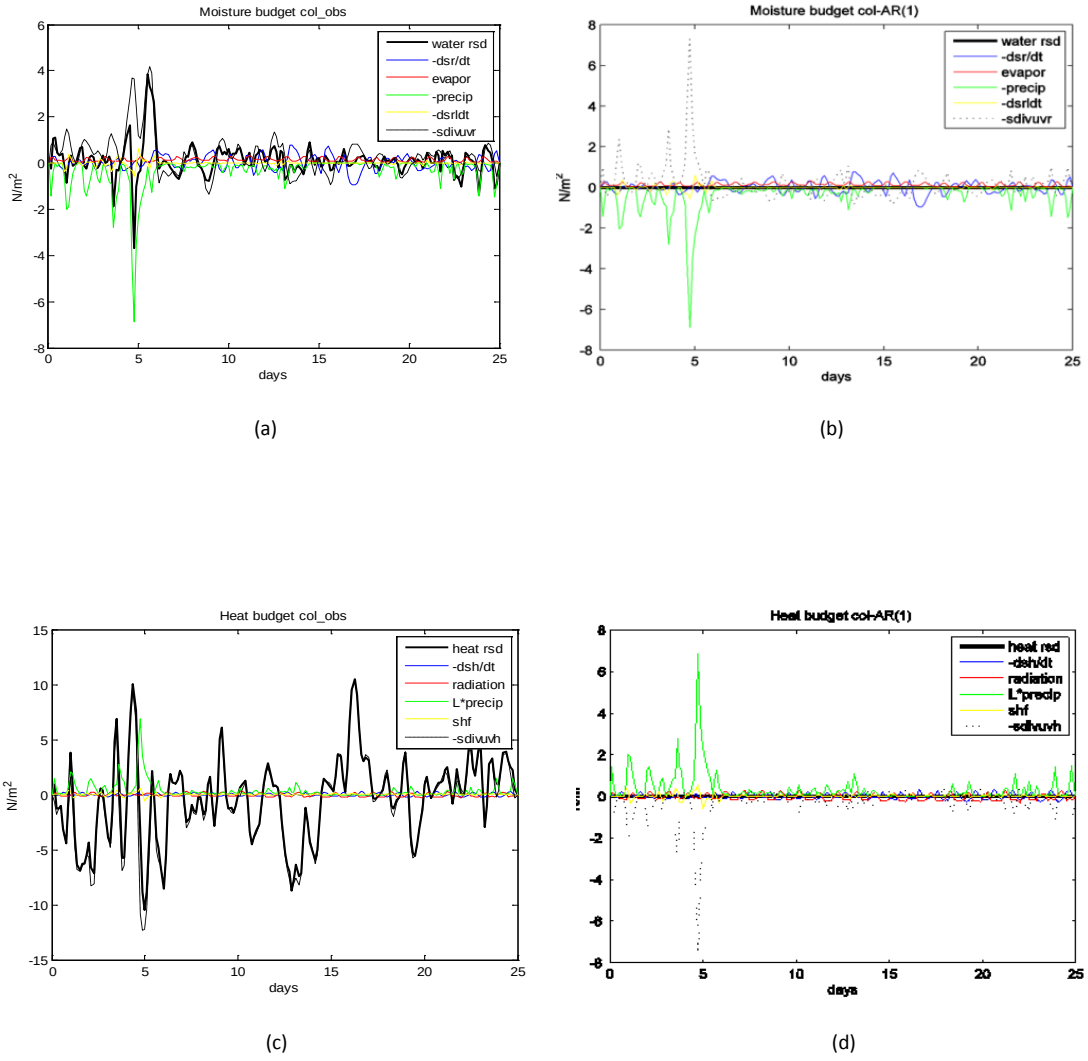


Figure 3.5: (a) and (b) for the terms of column integrated moisture budgets before and after temporal correlated variational method with AR(1) feature, respectively; (c) and (d) for terms of the column integrated heat budgets before and after temporal correlated variational method with AR(1) feature, respectively.

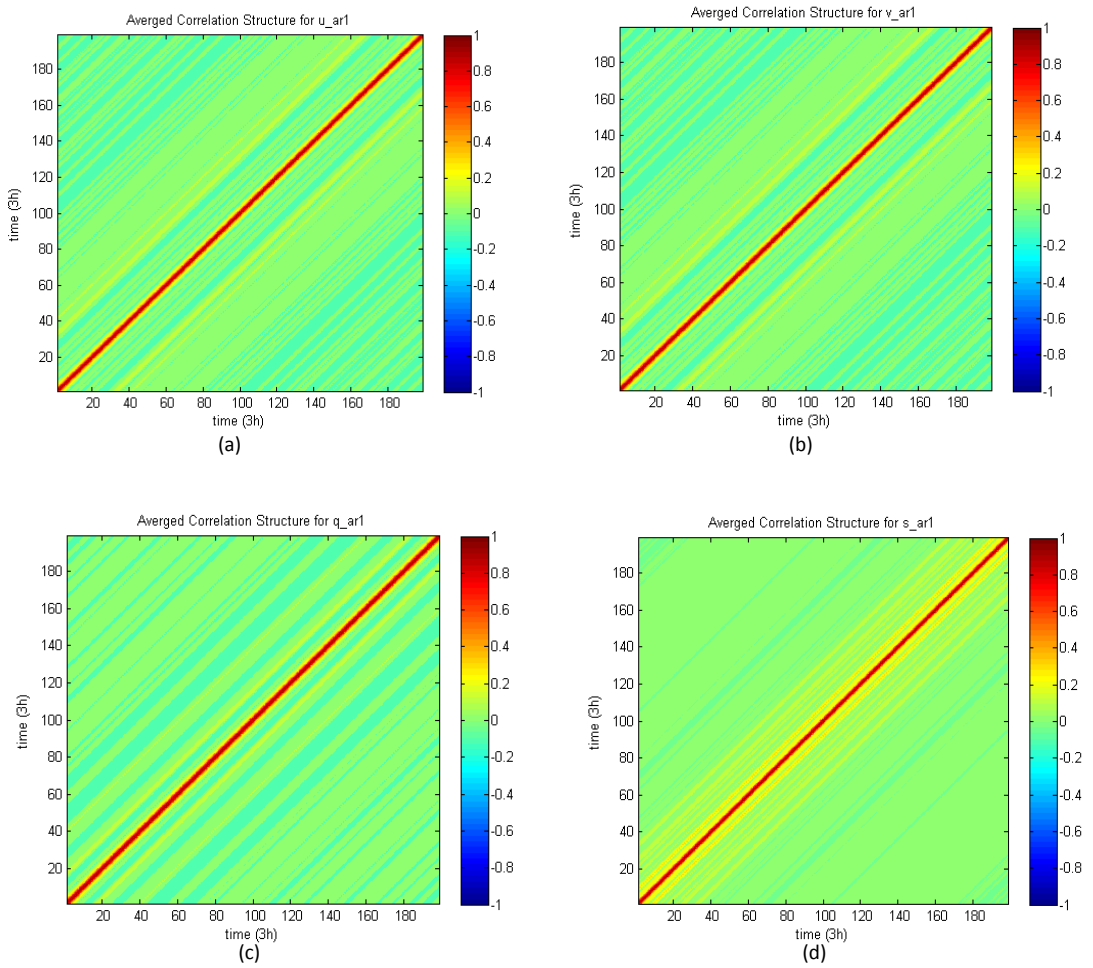


Figure 3.6: The autocorrelation structures for adjustments of u (a), v (b), q (c), and s (d) from temporal correlated constrained variational method with AR(1) feature.

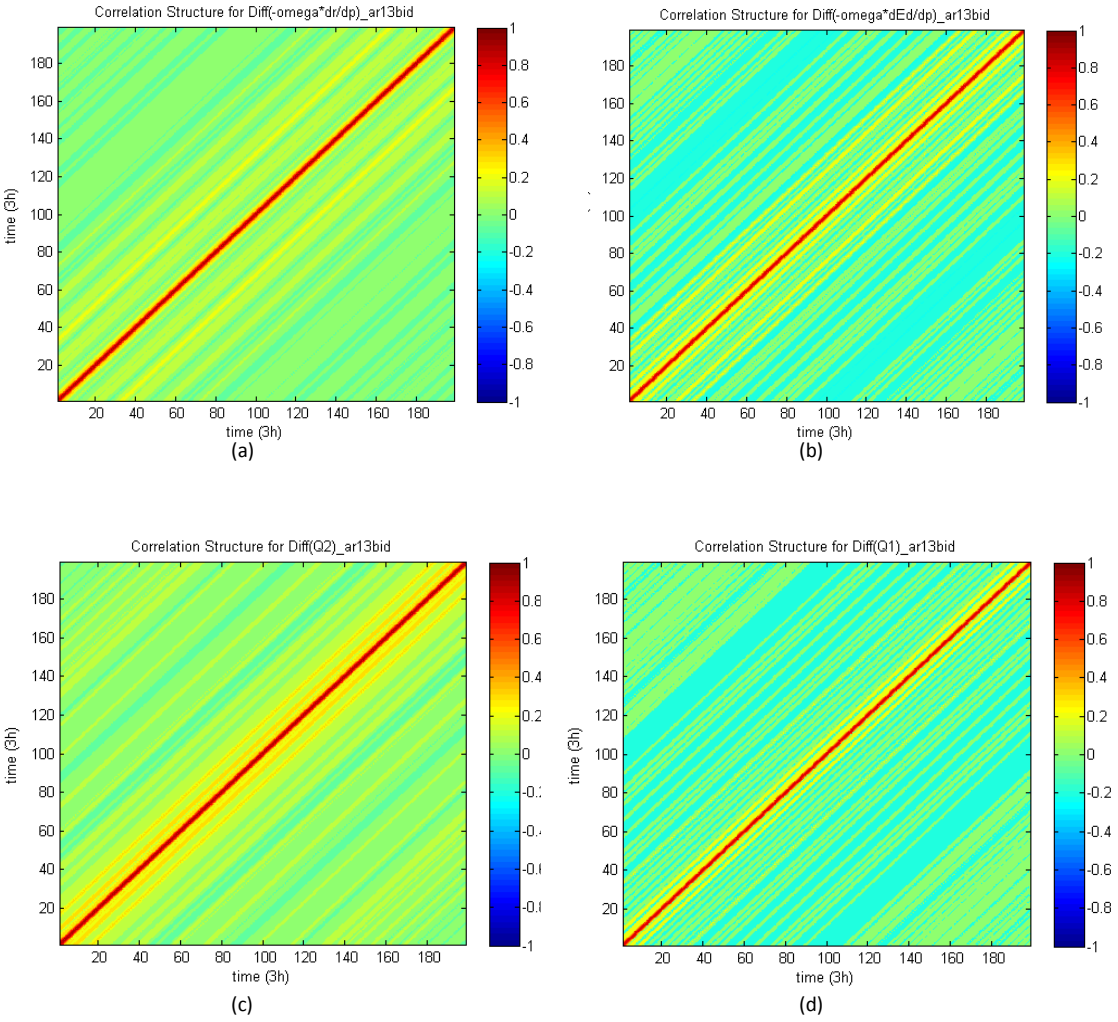


Figure 3.7: The autocorrelation structures of differenced vertical advection for moisture (a), vertical advection for static energy (b), Q_2 (c) and Q_1 (d) between temporal correlated case and the original case (Zhang & Lin, 1997).

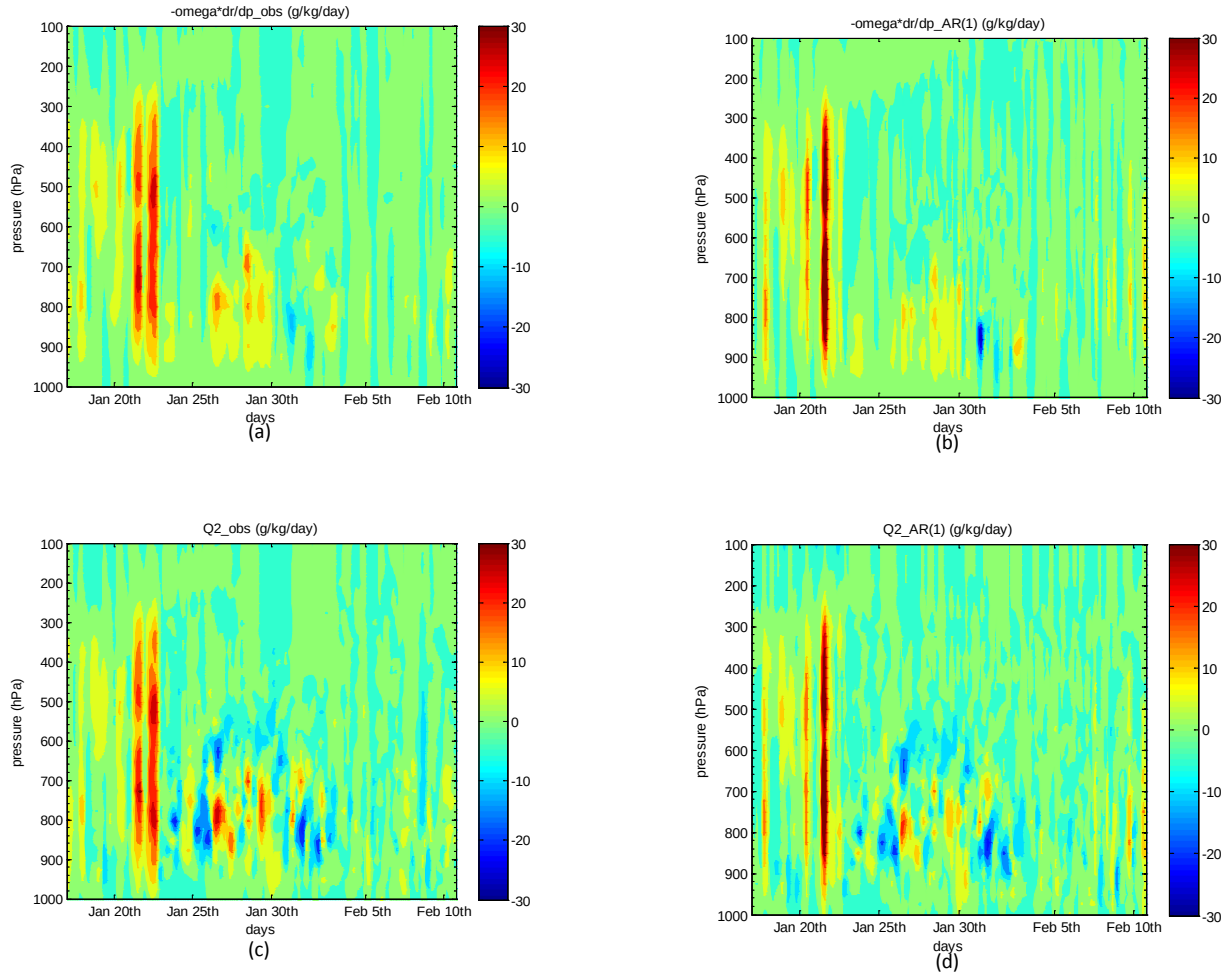


Figure 3.8: Vertical advection for moisture of (a) original 3d-data versus (b) final results from the new CVA method with AR(1) feature; the apparent moisture sink of (c) original 3d-data versus (d) final results from the new CVA method with AR(1) feature.

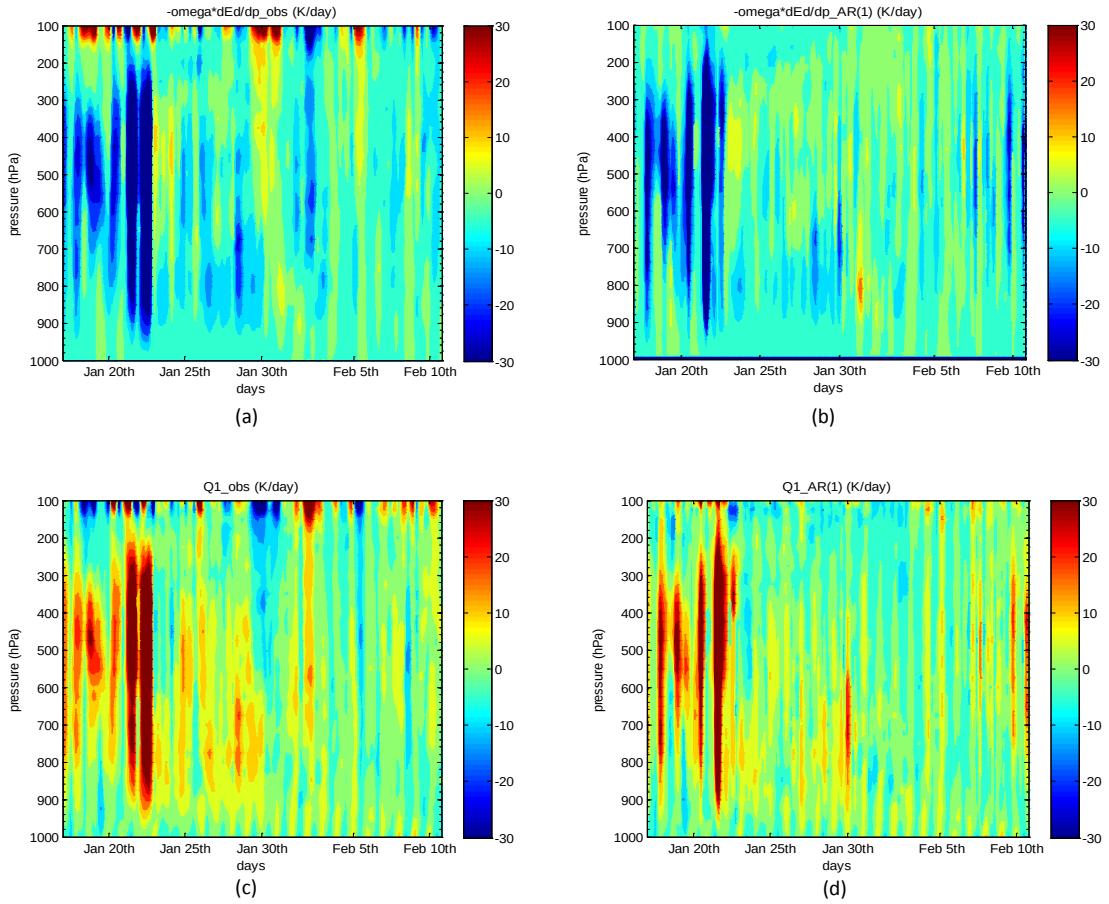


Figure 3.9: Vertical advection for static energy of (a) original 3d-data versus. (b) final results from the new CVA method with AR(1) feature; the apparent heat source of (c) original 3d-data versus (d) final results from the new CVA method with AR(1) feature.

Chapter 4

Constrained Variational Analysis Integrating Vertical Correlations and More Physical Budgets (190 Budgets)

4.1 Model description

The data set used here is exactly the same as in the previous two cases (refer to Part 2.1, Chapter 2 for detail). The three budgets incorporated for the previous two cases are the column integrated conservation of mass, moisture and static energy. To better adjust the sounding data, we add two more types of budgets, which are 1) the column integrated conservation of momentum in the *x* and *y* directions (refer to Eq. 2.4, Part 2.2 of Chapter 2); and 2) the hydro-static balance (refer to Eq. 2.5, Part 2.2 of Chapter 2). Note that the hydro-static balance is actually imposed on each of the 37 levels and 5 stations. Thus, 190 constraints have been adopted, including the five column integrated constraints. In this case, expect for vertical correlations, relationships between different stations, time steps or variables are neglected. Notations are generally similar to the previous two cases described in Chapters 2 and 3, except that we will introduce one more controlled variable, the geo-potential ϕ, which was derived from $\phi = gz$. Here $g = 9.8$ m/s^2, is gravitational acceleration; z is the geo-potential height.

We use five vectors $\vec{u},\ \vec{v},\ \vec{q},\ \vec{s},\ \vec{\phi}$ to denote all the variables at *S* stations and *K* levels (*S* = 6, *K* = 45) as $\vec{x}^T = (x_{ik}) = (x_{11}, x_{12}, ...x_{1K}, x_{21}, ...x_{ik}, ..., x_{SK})$, where *x* can be any of the *u*, *v*, *q*, *s*, ϕ. With only vertical correlations considered, the statistical model of the error structure is:

$$\forall\ i,\ \begin{pmatrix} x_{i1}^* \\ x_{i2}^* \\ \vdots \\ x_{iK}^* \end{pmatrix} = \begin{pmatrix} x_{i1}^o \\ x_{i2}^o \\ \vdots \\ x_{iK}^o \end{pmatrix} + \begin{pmatrix} \varepsilon_{xi1} \\ \varepsilon_{xi2} \\ \vdots \\ \varepsilon_{xiK} \end{pmatrix},\ \text{where } \vec{\varepsilon}_{xi} = \begin{pmatrix} \varepsilon_{xi1} \\ \varepsilon_{xi2} \\ \vdots \\ \varepsilon_{xiK} \end{pmatrix} \sim \mathrm{N}(\vec{0}, \Sigma_{xi}). \tag{4.1}$$

Our goal is to find the final analysis *x** by minimizing the cost function:

$$I(t) = (\varepsilon_u, \varepsilon_v, \varepsilon_q, \varepsilon_s, \varepsilon_\phi)^T Q^{-1} (\varepsilon_u, \varepsilon_v, \varepsilon_q, \varepsilon_s, \varepsilon_\phi) \tag{4.2}$$

where $\varepsilon_x^T = (x^o - x^*)^T$ and *x* can be any of the *u*, *v*, *q*, *s*, ϕ.

The weighting matrix *Q*, which is the variance-covariance matrix of the errors, is

$$Q = \begin{pmatrix} \mathrm{cov}(\vec{\varepsilon}_u, \vec{\varepsilon}_u^T) & 0 & 0 & 0 & 0 \\ 0 & \mathrm{cov}(\vec{\varepsilon}_v, \vec{\varepsilon}_v^T) & 0 & 0 & 0 \\ 0 & 0 & \mathrm{cov}(\vec{\varepsilon}_q, \vec{\varepsilon}_q^T) & 0 & 0 \\ 0 & 0 & 0 & \mathrm{cov}(\vec{\varepsilon}_s, \vec{\varepsilon}_s^T) & 0 \\ 0 & 0 & 0 & 0 & \mathrm{cov}(\vec{\varepsilon}_\phi, \vec{\varepsilon}_\phi^T) \end{pmatrix}, \text{ where}$$

$$\mathrm{cov}(\vec{\varepsilon}_X, \vec{\varepsilon}_X^T) = \begin{pmatrix} \Sigma_{x1} & 0 & \cdots & 0 \\ 0 & \Sigma_{x2} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \Sigma_{xS} \end{pmatrix}_{KS \times KS}. \tag{4.3}$$

4.2 Algorithm description

The Lagrange multiplier method is a classic method to derive the optimal solution. The cost function Eq 4.1 is to be minimized with 190 constraints being imposed.

Similar to the vertical correlated case with three budgets (refer to Chapter 2 for detail). The Lagrange function *J* is

$$J(\vec{X}) = I(\vec{X}) + \sum_{l=1}^{190} \lambda_l A_l(\vec{X}) \tag{4.4}$$

Making the gradient of *J* as zero to each variable yields:

$$\nabla_{\vec{X}} J(\vec{X}) = Q^{-1}(\vec{X} - \vec{X}_o) + \sum_{l=1}^{190} \lambda_l \nabla_{\vec{X}} A_l(\vec{X}) = 0 \tag{4.5}$$

$$\nabla_{\lambda_l} J(\vec{X}) = A_l(\vec{X}) = 0.(l = 1, 2, ..., 190) \tag{4.6}$$

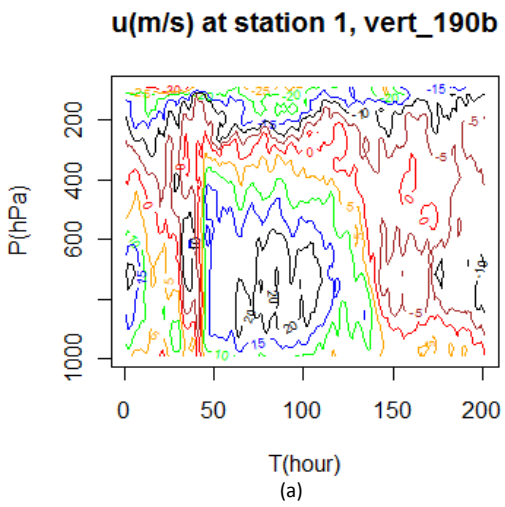
The robust Newton's interation memthod is adopted to search the final solution. For simplity, the whole algorithm is not expained again. The only difference between the case here and the one in Chapter 2 is the number of state variables and constraints. Please refer to Part 2.4 of Chapter 2 for detail.

4.3 Figures and preliminary results

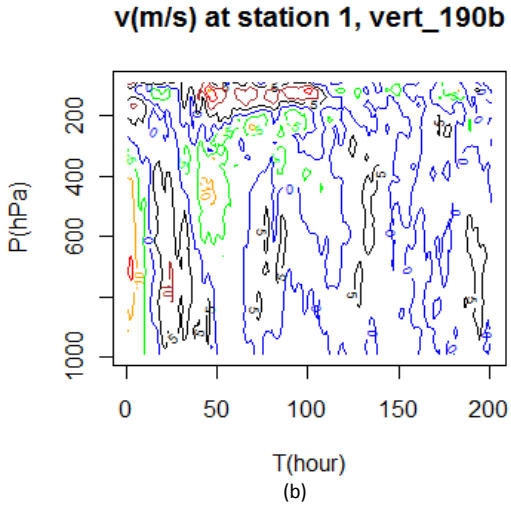
Figure 4.1 shows the final analysis of controlled variables *u*, *v*, *q*, *s*, ϕ at station 1. Other stations are similar. Like the vertical correlated case, the pattern of the final results of *u*, *v*, *q*, *s*, ϕ is quite similar to the original sounding data.

Figure 4.2 depicts the adjustment for all the five state variables, *u*, *v*, *q*, *s*, and ϕ at station 1 (Other stations are similar). The patterns of all the five figures are reasonable. Errors are symmetric about the *x*-axis for all variables and within an acceptable range compared to the order of magnitude of the variables.

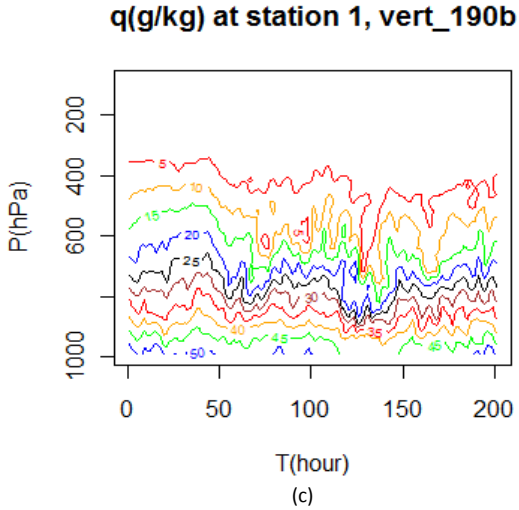
One great advantage of the model is the improvement of the adjusted *s*. One problem of the original method (Zhang and Lin, 1997) is that, when including the momentum budgets (Eq. 2.4), the pattern of adjustment for *s* will be one sided, which means for each station, the adjustments for all levels and time steps are either greater than zero or smaller than zero (Figures 4.3 (a) and (c)). The asymmetry of the adjustment about the *x*-axis indicates an ill pattern. How to reduce or even eliminate this phenomenon is an interesting topic. Since the Coriolis force term $f\vec{k}\times\langle\vec{V}\rangle$ and gradient $\nabla\langle\phi\rangle$ in the momentum budgets mainly shift the adjustment of *s*, adding more constraints to balance ϕ and *s* can be a feasible way. One of the important budgets, the hydro-static balance (Eq. 2.5, Chapter 2) is considered in this case. A detailed introduction to the budget is provided in Part 2.2, Chapter 2. Since it is not column integrated, the constraint is imposed on each layer and at each station. Figures 4.3(b) and (d) show the adjustment of *s* after combining the hydro-static balance and the momentum budgets. All adjustments are within an acceptable range and are symmetric about the *x*-axis, which makes the static energy *s* much more reasonable than in the original case. Other constraints will be analyzed in the near future.



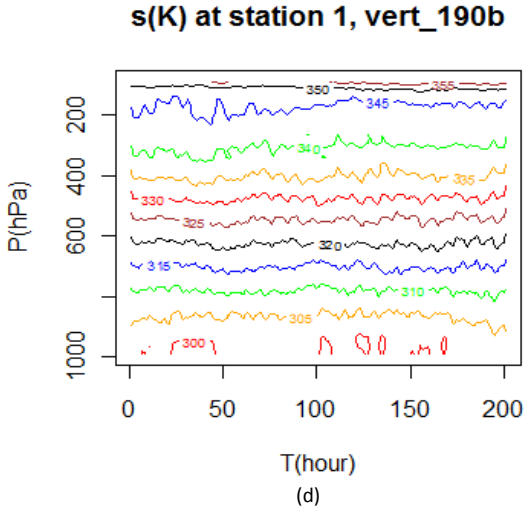
(a)



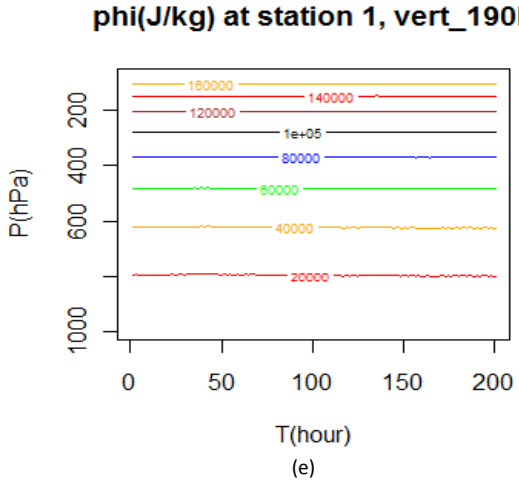
(b)



(c)



(d)



(e)

Figure 4.1: Final results of u, v, q, s and ϕ for vertical correlated case with 190 budgets at station 1, shown in (a)-(e), respectively.

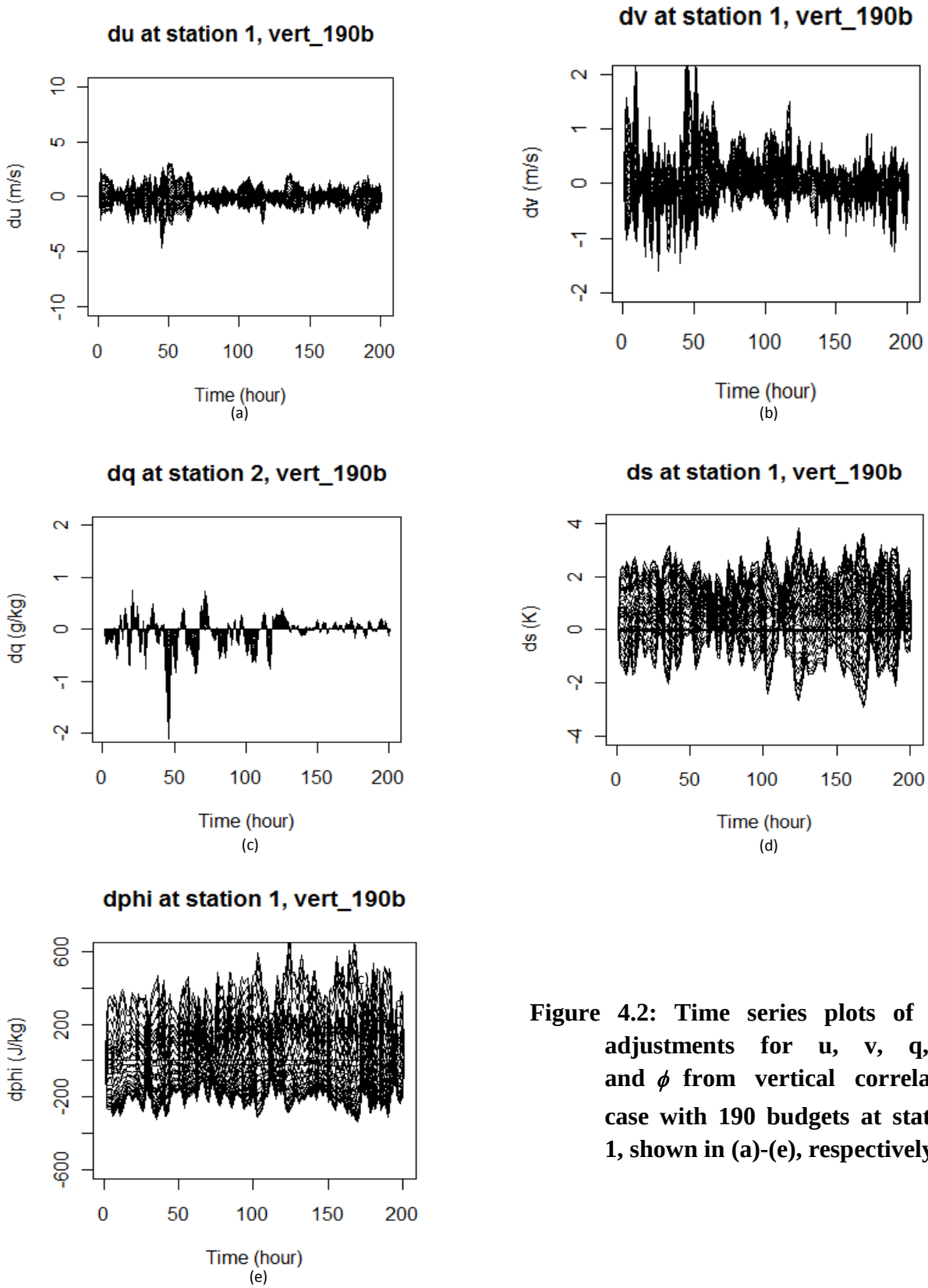


Figure 4.2: Time series plots of the adjustments for u, v, q, s and ϕ from vertical correlated case with 190 budgets at station 1, shown in (a)-(e), respectively.

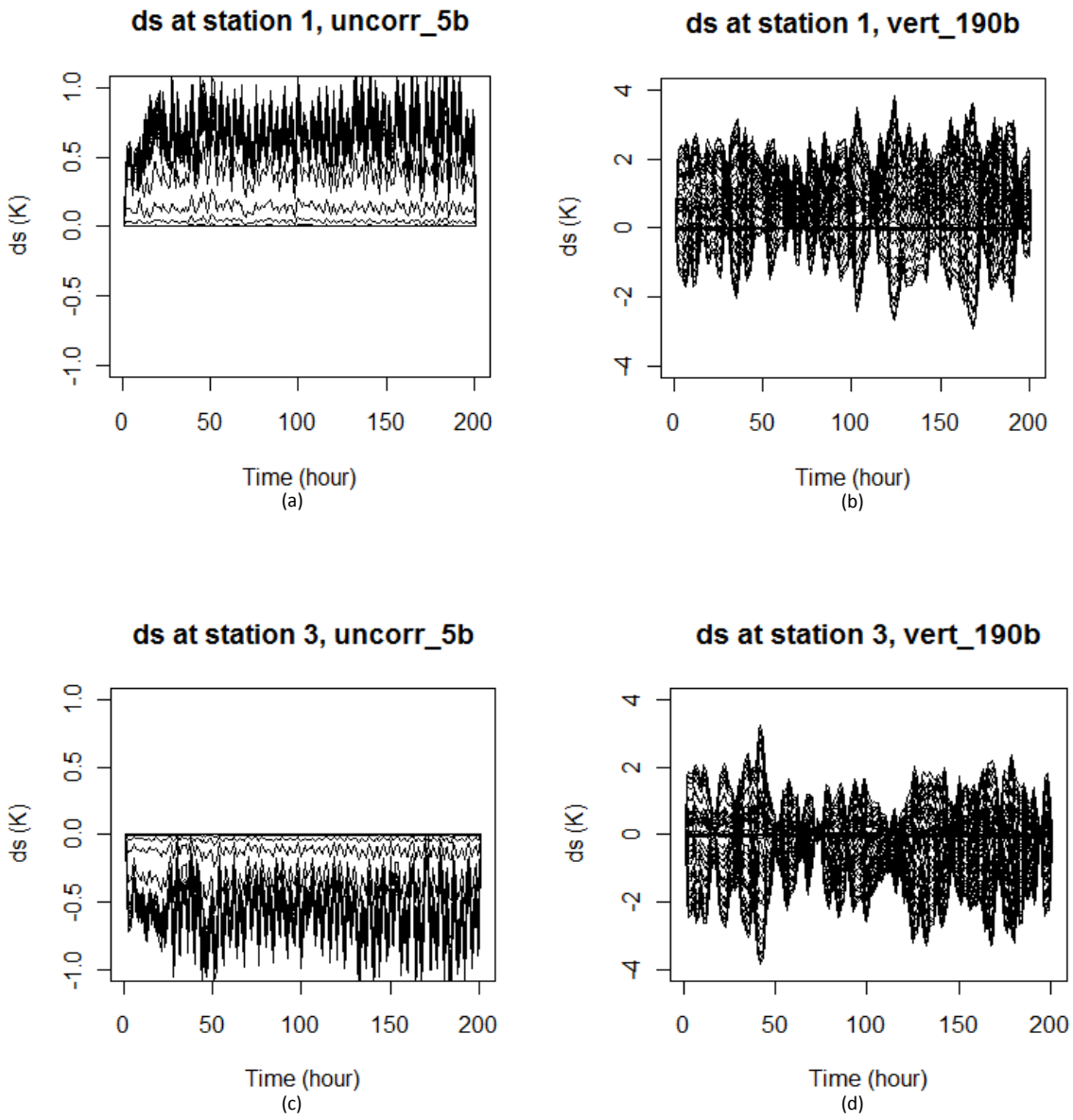


Figure 4.3: Comparison of the adjustment for s from the original uncorrelated case (Zhang & Lin, 1997) with five budgets (a, c) versus the new case with 190 budgets (hydrostatic balance incorporated) (b, d).

Chapter 5

Conclusion and Future Work

5.1 Conclusion

The thesis has elaborated in detail two newly constructed constrained variational analysis models with spatial correlations and temporal correlations and introduced briefly a new model with both momentum and hydrostatic balance equations integrated, based on the original model developed by Zhang and Lin (1997). From the extensive study, we can conclude that

- Both the vertical correlated case (Chapter 2) and the *AR(1)* temporal case (Chapter 3) for the CVA model have successfully incorporated correlations to the original model, which makes the model more accurate.

- For both new cases, the quantity and the adjustments for all controlled variables are reasonable. Meanwhile, error adjustments are confirmed to have the assumed type of correlations. That is, for the vertical correlated case, the vertical correlations have been verified in the final results and for the temporal correlated case with *AR(1)* feature, autoregressive feature have been confirmed in the final analysis.

- For both vertical correlated and temporal correlated cases, the important physical variables in meteorology, such as the vertical advection for moisture and static energy as well as Q_2 and Q_1, have been found to have similar type of correlations to what is assumed for the original sounding data in each case. These physical variables calculated

from the vertical correlated case have been verified to have strong vertical correlations, while those with the *AR(1)* feature case have been verified to contain strong auto-regressive feature. Also compared to the original sounding data, the four physical variables calculated from final analysis show considerably improvements, especially for the data at lower pressure levels (high altitude).

- The vertical correlated case with both momentum and hydro-static balance constraints (Chapter 4) has shown substantial improvements over both the original model and the vertical correlated case with only three budget constraints. Preliminary results have shown a much reasonable analysis over the original case with the momentum budgets. With the extra hydro-static balance incorporated, the ill phenomenon for static energy *s* in the original case has been reduced significantly. That is, the one sided adjustments of s in the original case have been replaced by the symmetric adjustments (about the *x*-axis) in the new case.

5.2 Future Work

The two new cases in the constrained variational analysis method for a single column have shown considerable improvements on the final analysis of important physical variables in meteorology in terms of the pattern of the variables structure and correlations. The first extension of current work is to apply the method to a larger area including more than one column. The hardest task for that is the algorithm design. Since the controlled variables of all columns will be coupled in the vertical correlated case, it will be unlikely to apply the Newton's iteration method

to up to millions of variables. Therefore, to search an innovative algorithm to deal with that large matrix is the first thing challenge.

A second extension is to include even more constraints for each case. Physical variables such as vertical advection (e.g. Figure 2.8(a)) should be zero in the clear sky when altitude is high enough. However, errors are still found in the figures after the balances of mass, moisture and static energy are incorporated. Applying more budgets may make the final analysis more accurate.

The third extension is to improve the statistical model for the error structure in the temporal correlated CVA method with *AR(1)* feature (Chapter 3). The incorporation of the AR(1) model is only for a model extension purpose. How to depict the errors better merits more investigation. One possible way is to apply the potential ARMA(p,q) model subject to a parameter selection based on the standards of AIC and BIC. The weighting matrix associated should be small; otherwise, new algorithm will be needed if the error structure is of a bad pattern.

The fourth extension is to upgrade the statistical model fundamentally by considering the scale and location factors. The statistical model can be modified by $\vec{x}^* = \vec{\beta}_0 + \vec{\beta}_1 \vec{x}^o + \vec{\varepsilon}$, with the same notations as in Chapter 2 and $\vec{\beta}_0$, $\vec{\beta}_1$ are two parameters that need to be estimated. In all three cases of this thesis, $\vec{\beta}_0$ and $\vec{\beta}_1$ are regarded as 0 and 1 respectively, which might not be the truth. This actually be indicated by the one-sided error structure of static energy s of the original case (refer to Chapter 4). What's more, it's better that $\vec{x}^o$ also need randomizing via the EIV model

(Error In Variable, refer to Casella and Berger, 2001), as $\vec{x}^o$ itself contains random factors when measured.

The fifth is to improve the cost function in the vertical correlated case (Chapter 2). The form of the cost function in Chapter 2 is derived in two steps: 1) estimating weighting matrix *Q*; 2) deriving the cost function via the maximum likelihood method. However, the two steps should actually be done simultaneously. Note that the determinant of *Q* will be included in the cost function if the *Q* is estimated during the maximum likelihood process. Though the new way will be closer to the reality, the implementation will be much more complicated. A comparison of the two ways needs further exploring.

References

Alley, R. B. et al., 2003: *Abrupt Climate Change,* Science 299 (5615): 2005-2010.

Chong and Testud, 1983: *Three-dimensional wind field analysis from dual-Doppler radar data. Part III: The boundary condition. An optimum determination based on a variational concept,* J. Climate Appl. Meteor, 22, 1227–1241.

Cram and Kaplan, 1985: *Variational Assimilation of VAS data into a mesoscale model: assimilation method and sensitivity experiments,* Mon. Wea. Rev., 113, 467–484.

Casella and Berger, 2001: *Statistical Inference (Second Edition)*. Duxbury Press.

Cox, John D., 2005: *Climate Crash: Abrupt Climate Change and What It Means for Our Future.* Washington DC: Joseph Henry Press.

Dessler, A.E. and E.A. Parson, 2006: *The Science and Politics of Global Climate Change: A Guide to the Debate*. Cambridge University Press, 0-521-53941-2.

Drummond, C. N. and B.H. Wilkinson, 2006: *Interannual Variability in Climate Data*, Journal of Geology, v. 114, p. 325-339.

Frank, 1979: *Individual time period analysis over the GATE ship array*. Mon. Wea. Rev., 107, 1600 –1616.

Heath, 2001: *Scientific Computing: An Introductory Survey (Second Edition).* McGraw-Hill Europe.

Ikawa, 1984a: *An Alternative Method of Solving Weak Constraint Problems and a Unified Expression of Weak and Strong Constraints in Variational Objective Analysis*, Papers in Meteorology and Geophysics, 35, No. 3, 71–79.

Ikawa, 1984b: *Generalization of Multivariate Optimum Interpolation Method and the Roles of Linear Constraint and Covariance Matrix in the Method, Part I: In Discrete Form,* Papers in Meteorology and Geophysics, 35, 81–102.

Ikawa, 1984c: *Generalization of Multivariate Optimum Interpolation Method and the Roles of Linear Constraint and Covariance Matrix in the Method, Part II: In Continuous Form,* Papers in Meteorology and Geophysics, 35, No. 4, 169–180.

Kuo and Anthes, 1984: *Mesoscale budgets of heat and moisture in a convective system over the central United States*. Mon. Wea. Rev., 112, 1482–1497. amples. J. Atmos. Sci., 53, 1123–1142.

Lewis, 1972: *The operational upper-air analysis using the variational method,* Tellus, 24, 514-530.

Lewis and Greyson, 1972: *The adjustment of surface wind and pressure by Sasaki's variational matching technique,* J. Appl. Meteor., 11, 586–597.

Lorenc, 1986: *Analysis methods for numerical weather prediction,* Quart. J. R. Met. Soc., 112, 1177–1194.

Lin and Johnson, 1996: *Kinematic and thermodynamic characteristics of the flow over the western Pacific warm pool during TOGA COARE.* J. Atmos. Sci., 53, 695–715.

Matheron, 1963: *Traite de G´eostatique appliqu'ee,* Vol. 1, 333 pages, and *Le Krigenge,* Vol 2, 173 pages, Memoires Bureau de Recherches Geologiques et Mini`eres, 19, Rue Vangirand, 75737, Paris, Cedex 15, France.

Matheron, 1970: *Random Functions and Their Applications in Geology,* in Geostatics (D. F. Merriam, editor), Plenum, New York, 79–87.

Matheron, 1981: *Splines and Krieging: Their Formal Equivalence,* Syracuse University Geology Contribution No. 8, Dept. of Geology, Syracuse University, Syracuse, New York, 13210.

O'Brien, 1970: *Alternative solutions to the classical vertical velocity problem.* J. Appl. Meteor, 9, 197–203.

Phillips, 1982: *On the Completeness of Multi-Variate Optimum Interpolation for Large-Scale Meteorological Analysis,* Monthly Weather Review, 110, 1329–1334.

Purser, 1984: *A New Approach to the Optimal Assimilation of Meteorological Data by Iterative Bayesian Analysis, 10th Conference on Weather Forecasting and Analysis, Clearwater Beach, Florida,* American Meteorological Society, Boston, pp. 102–105.

Bumgardner and Serafin, 1980: *Single and multiple Doppler-radar observations of tornadic storms,* Mon. Wea. Rev., 108, 1607–1625.

Sasaki, 1969: *Proposed inclusion of time-variation terms, observational and theoretical in numerical variational objective analysis,* J. Meteor. Soc. Japan, 47, 115–203.

Sasaki, 1970a: *Some basic formalisms in numerical variational analysis,* Mon. Wea. Rev., 98, 857–883.

Sasaki, 1970b: *Numerical variational analysis formulated under the constraints as determined by long-wave equations as a low-pass filter,* Mon. Wea. Rev., 98, 884–898.

Sasaki, 1970c: *Numerical variational analysis with weak constraints and application to the*

surface analysis of severe storm gust, Mon. Wea. Rev., 98, 899–910.

Sasaki, 1970d: *A theoretical interpretation of anisotropically weighted smoothing on the basis of numerical variational analysis*, Mon. Wea. Rev., 99, ndeg 9, 698–707.

Sasaki, 1971: *Low pass and band pass filters in numerical variational optimization*, J. Met. Soc. Japan, 49, 766–773.

Sasaki, 1973: *Mechanism of squall line formation as suggested from variational analysis of hourly surface observation*, preprint volume of the Eighth Conference on Severe Local Storms, October 15–17, AMS, Boston, MA.

Sasaki, 1955: *A fundamental study of the numerical prediction based on the variational principle*, J. Meteor. Soc. Japan, 33, 262–275.

Sasaki, 1958: *An objective analysis based on the variational method*, J. Meteor. Soc. Japan, 36, 77–88.

Semazzi and Navon, 1986: *A comparison of the bounded derivative and the normal mode initialization methods using real data*, Mon. Wea. Rev., 114, (11), 2106–2121.

Sheets, 1973a: *Analysis of hurricane data using the variational optimization approach with a dynamic constraint*, J. Appl. Meteor., 12, 963–976.

Sheets, 1973b: *Analysis of Hurricane Debbie modification results using the variational optimization approach*, Mon. Wea. Rev., 101, No. 9, 633–684.

Stephens, 1967: *Filtering responses of selected distance dependent weight functions*, Mon. Wea. Rev., 95 (1), 45–46.

Stephens, 1968: *Variational resolution of wind components*, Mon. Wea. Rev., 96, 229–231.

Stephens, 1965: *A variational approach to numerical weather analysis and prediction*, Ph. D. Dissertation, Texas A& M University, College-Station, TX, 77863, Rep. 3, 243 pages.

Stephens, 1965: *A variational approach to numerical weather analysis and prediction*, Ph. D. Dissertation, Texas A& M University, College-Station, TX, 77863, Rep. 3, 243 pages.

Stein, 1999: *Interpolation of Spatial Data---Some Theroy for Kriging*, published by Springer, 247 pages.

Thompson, Recker and Reed, 1979: *Structure and properties of synoptic-scale wave disturbances in the intertropical convergence zone of the eastern Atlantic.* J. Atmos. Sci., 36, 53–72.

Tseng, 1985: *Variational initialization and determination of weighting factors,* in Variational

Methods in Geosciences, (Y. K. Sasaki, ed.) Elsevier, 119–125.

Washington and Duquet, 1963: *An obphemsjective analysis of stratospheric data by Sasaki's method,* Dept. of Meteorology, Penn. State University, University Park, PA, 16802, 23 pages.

Wagner, 1971: *Variational analysis using observational and low pass filtering constraints*, M. S. Thesis, Dept. of Meteorology, The University of Oklahoma, Norman, OK, 73019, 39 pages.

Wahba, 1981b: *Some new techniques for variational objective analysis on the sphere using splines, Hough-functions, and sample spectral data,* in Proc. 7th Conference on Probability and Statistics in the Atmospheric Sciences, Monterey, CA, Nov. 26, 1981, 213–216.

Yanai and Johnson, 1993: *Impact of cumulus convection on thermodynamic fields. The Representation of Cumulus Convection in Numerical Models, Meteor. Monogr.,* No. 46, Amer. Meteor. Soc., 36– 62.

Zhang and Lin, 1997: *Constrained variational analysis of sounding data based on column-integrated budgets of mass, heat, moisture and momentum: approach and application to ARM measurements*, J. Atmos. Sci., 54, 1503-1524.

Zhang, F., C. Snyder, and R. Rotunno, 2002: *Mesoscale predictability of the 'surprise' snowstorm of 24-25 January 2000*. Monthly Weather Review, 130, 1617-1632.

Zhang, M. H., J. L. Lin, R. T. Cederwall, J. J. Yio, and S. C. Xie, 2001: *Objective analysis of the ARM IOP data: method and sensitivity*. Monthly Weather Review. 129, 295-311.

Zhang, M. H., and C. Bretherton, 2008: *Mechanism of low cloud feedbacks in idealized simulations with the single-column model of the NCAR CAM3*. J. Climate, Accepted.

Ziegler, 1978: *A dual Doppler variational objective analysis as applied to studies of convective storms*, M. S. Thesis, Dept. of Meteorology, University of Oklahoma, Norman, OK, 73019, 115 pages.

Appendix

A. Estimate of Q

As we consider only vertical correlation in this case, for each time step, we calculate the covariance from the combined time series across all five stations.

Take variable *X* as an example. To calculate $\mathrm{cov}(\varepsilon_{xk}, \varepsilon_{xk'})$, where ε_{xk} means the error terms of *x* at level *k*, we compute the covariance of the two time series *X* and *Y* as (Figure A.1):

$$\mathbf{X_k^{oT}} = (X^{o}_{k11},\ldots,X^{o}_{k1T},X^{o}_{k21},\ldots,X^{o}_{k31},\ldots,X^{o}_{kIT})$$

$$\mathbf{X_k^{*T}} = (X^{*}_{k11},\ldots,X^{*}_{k1T},X^{*}_{k21},\ldots,X^{*}_{k31},\ldots,X^{*}_{kIT})$$

where the superscript *o* represents the original sounding data and * means the results from original case (Zhang & Lin, 1997). Let $\varepsilon_{xk} = x_k^o - x_k^*$. ε_{xk} be the error time series of state variable *x* at the level *k*. The *X*, *k*, *i*, and *t* are the variable, level, station and time indices, respectively. By calculating the $\mathrm{cov}(\varepsilon_{xk}, \varepsilon_{xk'})$ numerically we can finally obtain the covariance matrix, or the weighting matrix *Q*.

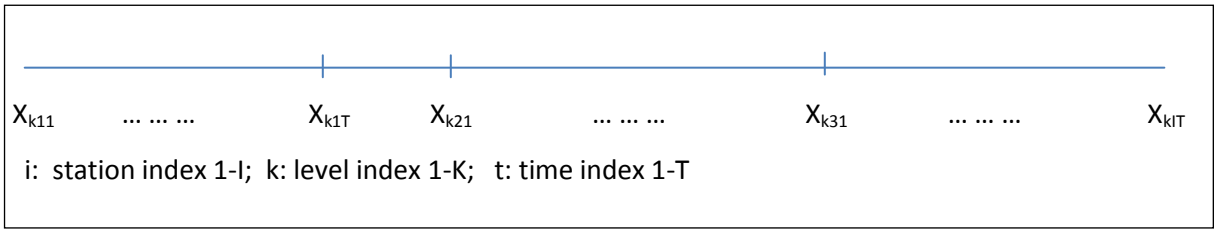


Figure A.1: The long time series of X across all stations for pressure level k