CIE5050-09 Additional Graduation Work

A study on numerical problems of a coupled modelling of regional water balance and anthropogenic landcover change

by

Wenxing Zhang

to fulfill the requirement of CIE5050-09 Additional Graduation Work,
at Department of Water Management,
Delft University of Technology.

Student number:	4646231	
Submission date :	Sep 25, 2018	
Thesis supervisor:	Dr. Saket Pande,	TU Delft
	Dr. ir. Ruud van der Ent,	TU Delft

TU Delft

Contents

List of Figures

Abstract

This thesis describes the approaches applied to attempt to solve the numerical problems of the regional atmospheric model incorporated in the coupled modelling of regional water balance and anthropogenic land cover change in Amazon basin. For computational efficiency, the previous atmospheric model is evaluated at monthly scale. In order to cope with the numerical instability, nearest neighbouring averaging interpolation is iteratively performed to smooth the solutions as a transitional approach. Therefore a subsequent study is conducted to investigate the origin of the numerical instability and whether there are feasible measures to fix the numerical problem of the modelling.

Chapter 1 serves as an introduction, which briefly introduces the background and research question — Whether there is any possible remedy that can solve the numerical instability of the monthly-timestep regional water balance model and obtain convergent solutions? Chapter 2 contains 7 sections, each of which gives statement of the specific problem, the experimental method applied, the corresponding results and related discussions. It has been concluded from the series of experiments that — *a*. adding diffusion terms makes no sense; *b*. applying smaller fractions of wind helps alleviate instability but the applied monthly timestep length seems to make the model paradoxic and inherently not convergent; *c*. instability is not really relevant with the iterative method; *d*. after correcting a dormant error in previous research, the results gets no better; *e*. the model may be so oversimplified that cannot reflect the reality; f. wind and monstrous timestep length are the keys to the problem, especially the latter is more problematic. Chapter 3 summarizes the discussions and conclusions from chapter 2 and proposes several recommendations for future research such as trade-offs between model complexity and efficiency, a heuristic way of making wind endogenous and reconsideration of the model architecture.

Key words: Regional atmospheric model, Large temporal scale, Numerical instability

1 Introduction

A research on coupled modelling of regional water balance and anthropogenic land cover change was conducted by Ajar Sharma as part of his MSc thesis work for the purpose of simulating and predicting the deforestation in Amazon Basin in response to food price since Amazon rainforest is the lung of the earth and is currently suffering deforestation due to agriculture expansion. In his research, he developed a numerical atmospheric model that solves the governing continuity equation. The original motivation of the atmospheric model design is to be able to couple with the anthropogenic land cover change model for fast simulation, thus the timestep length of the model is set as one month which is quite large compared to temporal scales applied in other atmospheric models. However, the model is confronted with numerical instability problem. In order to reproduce reasonable results, Nearest Neighbouring Averaging interpolation is performed after every iteration for each time step to wipe out the oscillations and stabilize the modelled solution. But this method is merely a temporary solution as it lacks solid physical foundations. Apart from instability, the modified modelled solution is not conserved for water balance and is not to a certain degree convergent to the observations of ERA-interim reanalysis data.
As required by the course CIE5050-09 Additional Graduation Work, I select this topic as the content of my additional thesis. Because the coupled model focus on the coupling of water system and human system, it does not require a comprehensive and computationally expensive atmospheric model. Therefore, the tough spot of this project is instead of routinely applying a more conservative shorter timestep to control stability, we try to figure out whether it is possible to remedy our existing monthly-step model in a novel way. Therefore in my research, I dived into the model mechanisms and the codes to study and understand where the numerical instability comes from and attempted to try out possible methods that may resolve the problem. Chapter 2 introduces all the approaches that have been tested as well as corresponding results and discussions, such as adding diffusion term, modifying wind, applying sparse matrix solver, applying multiple time steps and etc. More specific problem-associated introduction is in the individual section in Chapter 2.
Despite continuous trial and error, the problem still remains unsolved since it has been concluded in Chapter 3 that monthly temporal scale seems to make the model inherently paradoxical and even though the solutions could be stabilized, convergence cannot be achieved. After conclusions, recommendations for future research possibilities are proposed that ask for reconsideration of the numerical model architecture.

2

Methodology and Results Discussion

This chapter articulates experimental methodologies to investigate the numerical instability of the regional water balance model and discusses corresponding results. It proceeds with the train of thought — First, a problem is raised, then an experimental study is conducted which leads to another problem to address. Section 2.1 discusses the impacts of adding diffusion terms at different levels on the modelling and briefly talks about the inherence of numerical diffusion in numerical simulation. Section 2.2 studies how the reduced fraction in magnitude of wind intensity influences the smoothness of the modelled moisture field. Section 2.3 tries a sparse matrix solver to solve the system of equations instead of Heun's algorithm applied in the previous work. Section 2.4 investigates the sensitivity of source term (precipitation and evaporation) on the simulation. Section 2.5 experiments time-invariant wind and spatial-independent wind to get further knowledge regarding spurious oscillation and numerical instability. Section 2.6 tries explicit scheme with multiple time steps at smaller time step size Δt. Section 2.7 conducts unit check and straightforward calculations on the observed data in order to grasp an first-order estimation of order of magnitude of wind speed to be applied in the modelling.

As we know from the previous work, the numerical solutions to the atmospheric model are oscillating, resulting in an unsmooth moisture field, which is irrational from the physical point of view. So an interpolation trick named "Nearest Neighboring Averaging" is applied in the formal research to eliminate this instability problem. This method indeed smooths the modelled moisture field and makes the solutions look more reasonable at least in terms of smoothness and magnitude, but it is not sufficiently scientifically rigorous due to the following facts. There are at least two crucial drawbacks of implementing this interpolation trick: Firstly, the solution after each loop of calculation is artificially bounded by default limits, in other words, the solutions might be unbounded in defect of those man-made limits. Solutions remaining bounded is a necessary condition for numerical stability. Beside, the conservation of mass is compromised due to the immediate cut off of exceedance and the absence of closure term. Hence it indicates that the interpolation trick may incorrectly deal with the numerical instability issue. Secondly, using Nearest Neighbouring Averaging interpolation to stabilize the results is merely a measure of expediency which has no strong theoretical foundations.

Anyway, the instability of the numerical recipe could arise from a number of aspects, for instance, the numerical scheme picked and whether the corresponding stability limit is

met, the missing of important elements in the governing equation, the magnitude of coefficients, variables and parameters (*e.g.* wind intensities, time step Δt, grid size ΔL), the spatial-variant and time-variant heterogeneous velocity field, the insufficient governing equations to build the modelling and etc.

2.1. Add diffusion terms

In this section, the impact of adding diffusion terms is studied as the starting point of a series of experiments. Diffusion is the movement of matter from a region of high concentration to a region of low concentration, driven by the concentration gradient (or pressure gradient)[4]. It is a physical spreading of substance depending on the concentration gradient whereas convection is merely a substance transport through bulk motion [50]. As the initial guess, diffusion term might be able to buffer the wiggles so it might reduce or even wipe out the spurious oscillation. Hence here comes the first research questions: Would the instability be alleviated, if the Nearest Neighbouring Averaging interpolation (NNA) were removed and the diffusion term were added to the simulation. If that were the case, in terms of the order of magnitude of diffusion coefficient, which value (or range) would be a reasonable pick? In order to study the impacts of diffusion terms, a set of coefficients at different levels are tested for two control groups. One control group is the original group plus diffusion terms and the other group is the RNNA (abbreviation for "remove nearest neighboring averaging") group adding diffusion terms.
First, the diffusion term is discretized with central difference in space, which is 2^{nd} order accurate in space discretization, of which the truncation error is $\tau_{\Delta x} = O(\Delta x^2)$. Below shows the discretized form of diffusion terms. The numerical solutions are evaluated at time level $t = n+1$. So it is an implicit (backward in time) differencing.

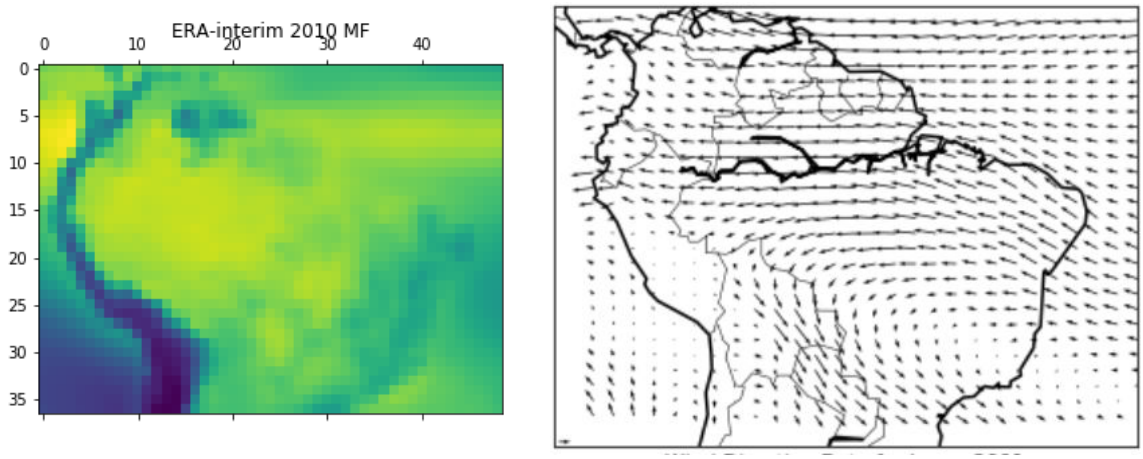
$$D\frac{\partial A^2}{\partial x^2} + D\frac{\partial A^2}{\partial y^2} \approx D\frac{A_{i+1,j} - 2A_{i,j} + A_{i-1,j}}{(\Delta x)^2} + D\frac{A_{i,j+1} - 2A_{i,j} + A_{i,j-1}}{(\Delta y)^2}$$

, where D is the diffusion coefficient. For simplicity, the coefficient is assumed to be homogeneous everywhere in the domain, while in reality the coefficients could be spatial and temporal dependent. $A_{i,j}$ is the atmospheric moisture at $cell_{i,j}$, $\Delta x = \Delta y = 110km$ are the grid size.
The diffusion coefficient of water vapor in the air is around $D_a = 2.3 \times 10^{-5} m^2/s$ [6]. So this value (converted from m^2/s to m^2/d) is taken as a starting point. Due to the built-in quite large temporal scale (1month) and relatively large spatial scale (1° by 1°, which is approximately $110km \times 110km$) , the actual diffusion coefficient at macro-scale level could be significantly greater than the initial value, so a set of diffusion coefficients of multiple order of magnitude are tested to study the impacts of diffusion terms and to examine whether those selected coefficients are within a physically reasonable range.

2.1.1. Control Group 1: Add diffusion terms based on original model

Moisture field modelled by ERA-interim

Figure 2.1a below is the atmospheric moisture field modelled by ERA-interim, the state-of-the-art global atmospheric reanalysis project conducted by the European Centre for Medium-Range Weather Forecasts (ECMWF). In comparison with the previous reanalysis



(a) Moisture field modelled by ERA-interim (b) Wind field for Jan 2001

Figure 2.1: Moisture field modelled by ERA-interim and wind pattern

ERA-40, ERA-interim has great improvements in terms of hydrological cycle, the stratospheric circulation, data assimilation in the observing system and etc, thus giving reliable near-real data due to its sophisticated modelling mechanisms [12]. Figure 2.1b shows the wind field over the study area for January 2001. The dominant direction of wind movement is due to the creation of low pressure zones due to the ITCZ. The wind pattern clearly indicates that there should be a wind-driven flow of moisture transporting water vapor from Atlantic ocean to Amazon basin. And when the dominant wind meets Andes Mountain, it diverts its direction. In the region of Andes, the atmospheric moisture decreases rapidly (marked in dark blue shadow) after the wind encounters the mountain, orographic precipitation explains this — the moisture is removed from the atmosphere since the moist air is forced to be uplifted by the rising terrain and the vapor condensates [34]. These two plots are indicative of a general idea of the moisture redistribution pattern of the Amazon basin: There is predominant convection of moisture from ocean to inland with continuous precipitation recycling as evaporation simultaneously along the flow path. Evaporation in the Amazon basin brings about a quarter of the rain that falls within the basin by estimation [16]. The orographic barriers named the Andes, which acts as an orographic control of the rainfall, casts a major precipitation shadow in the southwestern part. Hence the orographic precipitation determines the water vapor – precipitation interactions in the Andes region. However, this local topographic factor in Andes is not considered in the former research, so a next logical step in further research might make allowance for incorporating this important feature in the modelling or excluding the region from the modelling like Zemp's research because the evapotranspiration in Andes is mainly determined by temperature instead of by precipitation [49].

Original Model plus diffusion terms

Below are the results of the original control group, in which the nearest neighboring averaging interpolation is still being implemented after each iteration (50 iteration loops for each time step) until the solution of each cell is convergent to a fixed value at every time step. It should be noted that a small modification on the original code is applied here. The solution is bounded now with the upper and lower limits of $A_{max} = 0.06356m$ and $A_{min} = 0.00078m$

respectively, corresponding with the maximum and minimum value of moisture observations from ERA-interim, instead of primitive limits $A_{max} = 100000m$ and $A_{min} = 0$, which are apparently not so appropriate. As can be seen from the moisture field plot 2.3a, the initially modelled moisture field of Amazon basin is smooth but does not emerge the similar moisture pattern as that of ERA-interim. The blank gap on the right of the panel is induced by the interpolation algorithm applied in the original codes which generates a series of NaN. Nevertheless, the interpretation of the vast majority of the study domain is impeded little by this blank. Figure 2.3b depicts the modelled moisture numerical solutions and the ERA-interim solutions along the latitude $-7.11S^{\circ}$. Figure 2.3c and figure 2.4d are scatter plots and box plots of the precipitation and evaporation for R1, R2, R3 three region cases in previous research respectively for the year 2010. For further information on three selected case regions, please refer to Ajar Sharma's thesis work [2]. Each region contains 30 grids. For each region, the modelled solutions are validated with the observations (*i.e.* ERA-interim solutions) on a yearly basis. Figure 2.3c shows that the precipitation data for region 1 and region 2 are in a good match as points are scattered along the line and the root-mean-square-error ($RMSE$) is small. But the results of precipitation on region 3 is poorly modelled. In contrast, in terms of evaporation, an undesired poor linear relationship shows the modelled solutions for three regions are much less consistent with ERA-interim observed solutions.
Analysis on the original results raises several implicit anticipations for future researches relating to this topic:

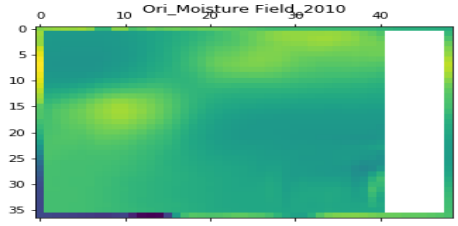
(i) First, a new atmospheric model should be able to reproduce a smooth moisture field with stable solutions, of which the pattern should emerge notable features similar as ERA-interim's.

(ii) Second, the validation of model solutions with the ERA-interim observations (the conservation of water) on a yearly basis should be improved, implying a better linear spreading of points along the straight line in the scatter plot is desired.

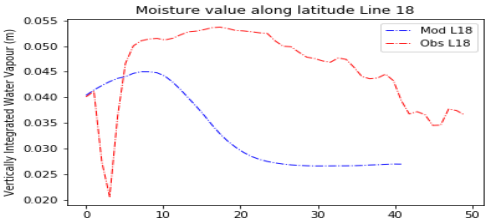
(iii) Last but not least, after modelled water balance being conserved on a yearly basis, the seasonal variation pattern should also be captured to a certain degree.

These anticipations ask for a simple but robust atmospheric model that is evaluated on coarse spatial resolution and on large monthly temporal step, which enables quick estimations and efficient predictions with a relative acceptable level of accuracy. However, these goals are virtually quite challenging. And the biggest challenge, as far as intuitive thinking, may lie in the temporal scale. Since most commonly used atmospheric models that solve a system of partial differential equations are evaluated from minute to hourly time step size for controlling numerical stability. Indeed there are some models evaluated at monthly scale but they are just simple models like bucket model of the entire study domain. Anyway, with or without a magic solution, it is always worthing trying since the charm of scientific research lies in exploring the unknown.

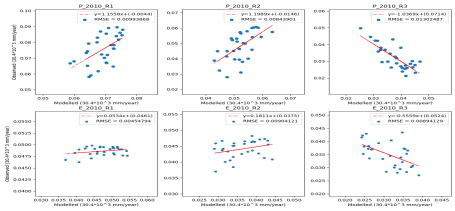
D = 0, original case:



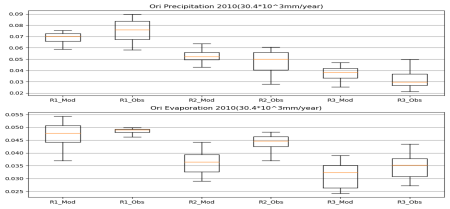
(a) Moisture field modelled in original model



(b) Moisture along mid line



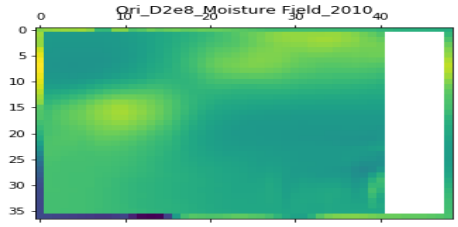
(c) Scatter plot of precipitation and evaporation



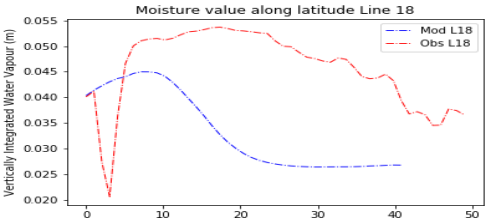
(d) Box plot of precipitation and evaporation

Figure 2.2: Original model's results

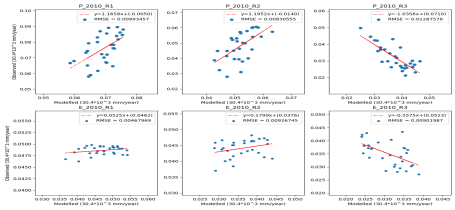
$D = 2 \times 10^8 m^2/d$:



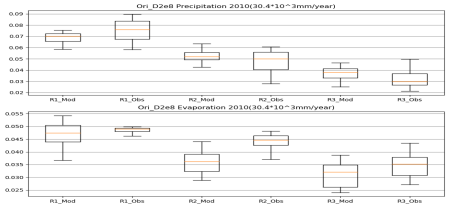
(a) Moisture field modelled in original model



(b) Moisture along mid line



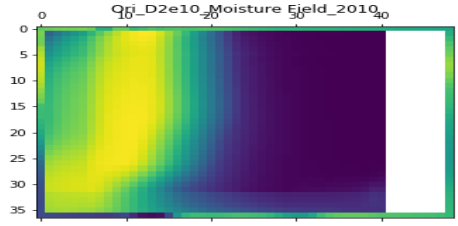
(c) Scatter plot of precipitation and evaporation



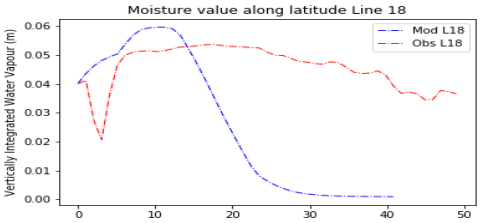
(d) Box plot of precipitation and evaporation

Figure 2.3: Results for Original model plus $D = 2 \times 10^8$ diffusion term

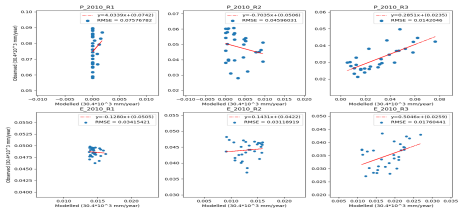
$D = 2 \times 10^{10} m^2/d$:



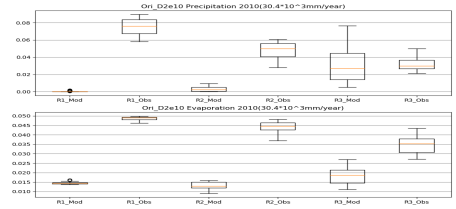
(a) Moisture field modelled in original model



(b) Moisture along mid line



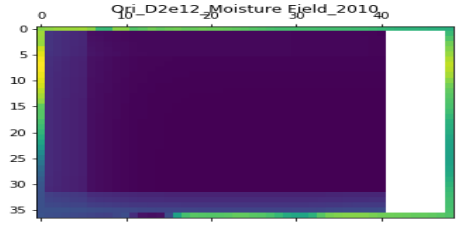
(c) Scatter plot of precipitation and evaporation



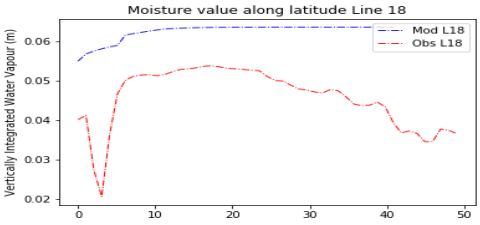
(d) Box plot of precipitation and evaporation

Figure 2.4: Results for Original model plus $D = 2 \times 10^{10}$ diffusion term

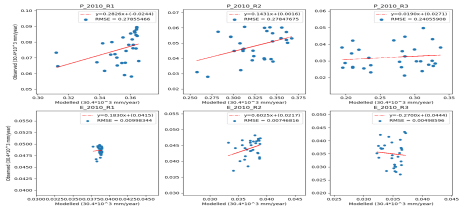
$D = 2 \times 10^{12} m^2/d$:



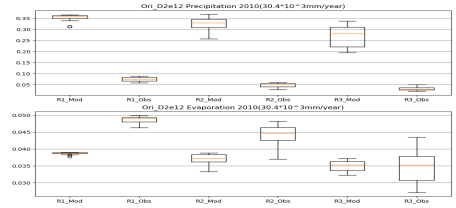
(a) Moisture field modelled in original model



(b) Moisture along mid line

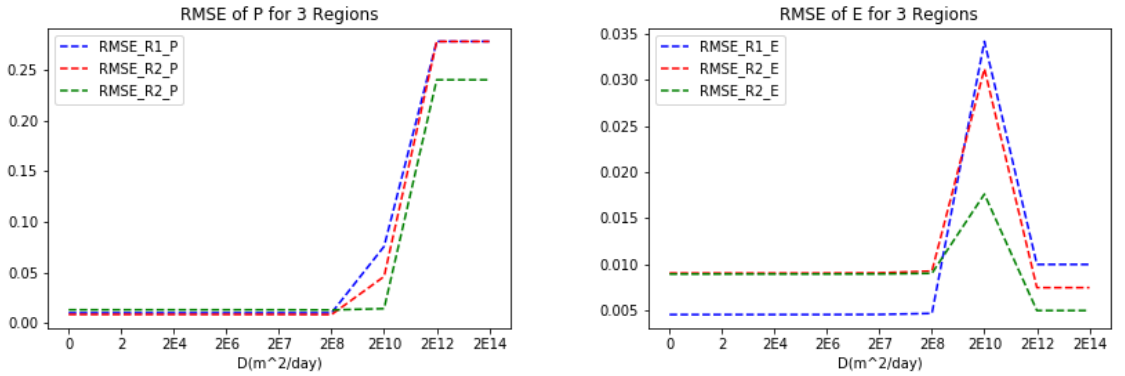


(c) Scatter plot of precipitation and evaporation



(d) Box plot of precipitation and evaporation

Figure 2.5: Results for Original model plus $D = 2 \times 10^{12}$ diffusion term



(a) RMSE of precipitation for three regions (b) RMSE of evaporation for three regions

Figure 2.6: RMSE for original cases with different diffusion coefficients

Comparing figure 2.2 and figure 2.3, it is not difficult to find that the diffusion term barely makes any influence with the given coefficient less than $2 \times 10^8 m^2/d$. From the figure 2.6, RMSE of precipitation and evaporation for three regions is greatly increased and the modelled moisture solutions begin to extend to the boundary limit once the diffusion coefficient is greater than $2 \times 10^8 m^2/d$, indicating the diffusion term has taken over advection term. However the modelled result no longer makes any sense from then on.

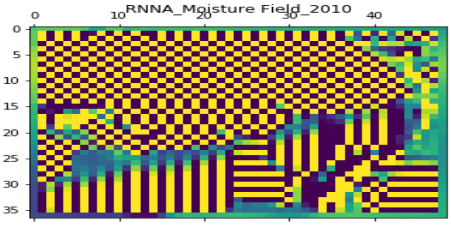
2.1.2. Control Group 2: Remove NNA and add diffusion term

Remove NNA and add diffusion terms

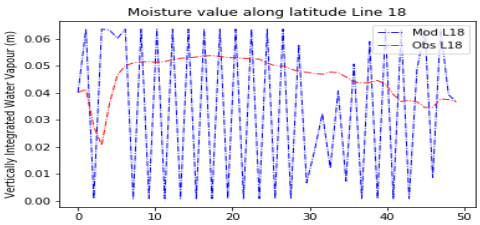
As stated above, a cascade of diffusion coefficients are tested for two control groups (original group and removing nearest neighboring averaging interpolation group). Figure 2.7 illustrates that if the Nearest Neighboring Averaging interpolation method is removed, the resulted moisture field is extremely unsmooth with abundant numerical oscillation, and the modelled solutions along the selected latitude is oscillating between the bounded limits in figure 2.7b. In the antecedent research, the instability is cut off by averaging interpolation, but this is not appropriate from physical point of view. Without this interpolation, the modelled moisture field becomes quite unstable, introducing two order of magnitude increased RMSE to the solution and making the modelled solutions totally invalid. Diffusion term is then incorporated into the modeling to see whether it may help alleviate the oscillation and probably improve the results. A bunch of coefficients are tested, but only the representative ones are selected and presented below.

Comparing the figure 2.8 and figure 2.9, it is found that the impact of diffusion term is quite minor if the magnitude of the given coefficient is less than 2×10^8. Only a minor reduction in RMSE of precipitation for three regions is detected. If the coefficient is greater than 2×10^8, even if the RMSE of evaporation is further reduced, the modelled results are no longer reasonable, clearly shown in the figure 2.10. However, one thing to note here is that a smaller RMSE does not necessarily means better results, mainly due to the following reasons: *a*. The scatter points could have a poorer linear relationship despite a smaller RMSE. *b*. The modelled moisture field could be even more unrealistically unstable. *c*. More attention should be put on the validation of precipitation since precipitation is directly converted from the modelled moisture, while evaporation is indirectly linked to the solution, in which uncertainties in computing may further induce errors to the solutions.

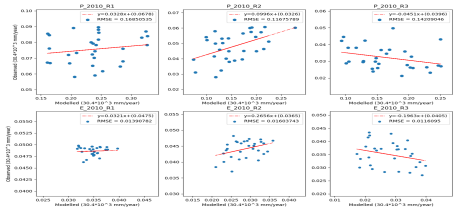
D = 0, RNNA case:



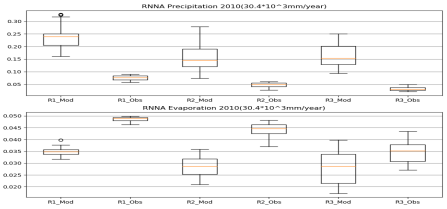
(a) Moisture field modelled in original model



(b) Moisture along mid line



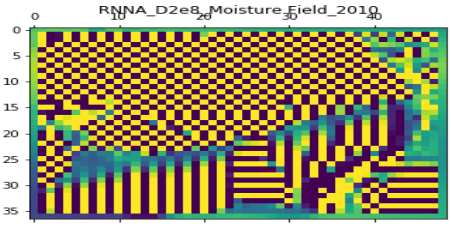
(c) Scatter plot of precipitation and evaporation



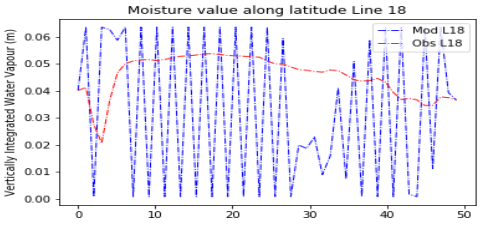
(d) Box plot of precipitation and evaporation

Figure 2.7: Results for RNNA model plus D = 0 diffusion term

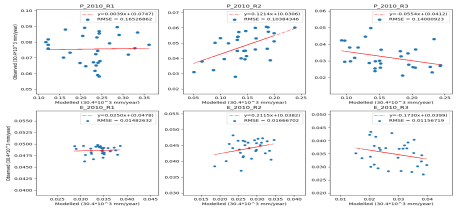
$D = 2 \times 10^8 m^2/d$:



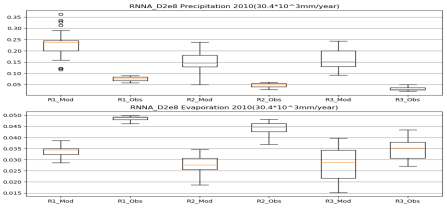
(a) Moisture field modelled in original model



(b) Moisture along mid line



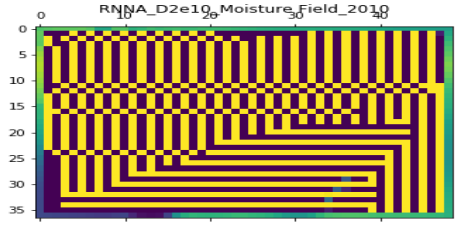
(c) Scatter plot of precipitation and evaporation



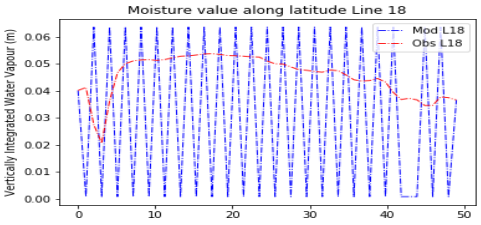
(d) Box plot of precipitation and evaporation

Figure 2.8: Results for RNNA model plus $D = 2 \times 10^8 m^2/d$ diffusion term

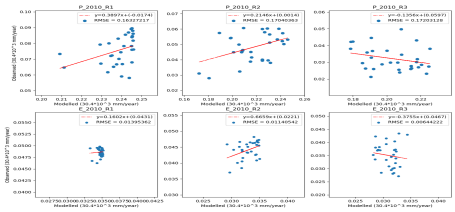
$D = 2 \times 10^{10} m^2/d$:



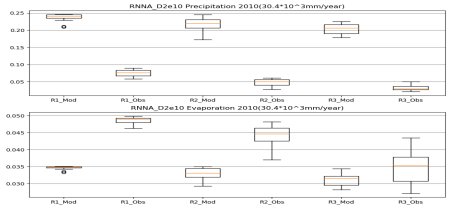
(a) Moisture field modelled in original model



(b) Moisture along mid line



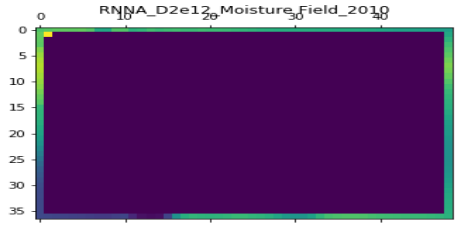
(c) Scatter plot of precipitation and evaporation



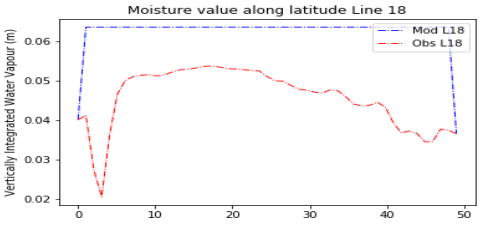
(d) Box plot of precipitation and evaporation

Figure 2.9: Results for RNNA model plus $D = 2 \times 10^{10} m^2/d$ diffusion term

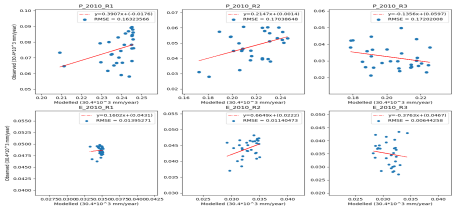
$D = 2 \times 10^{12} m^2/d$:



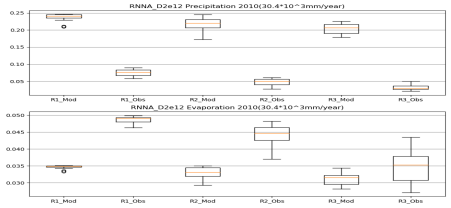
(a) Moisture field modelled in original model



(b) Moisture along mid line



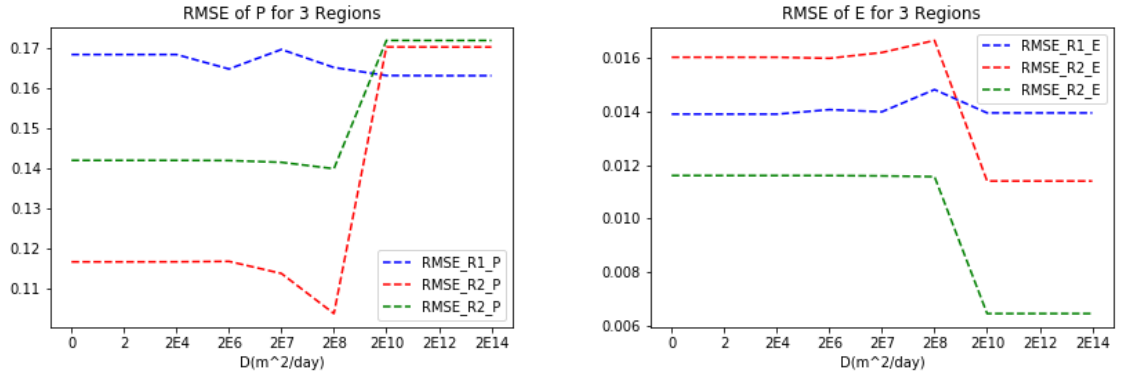
(c) Scatter plot of precipitation and evaporation



(d) Box plot of precipitation and evaporation

Figure 2.10: Results for Remove NNA plus $D = 2 \times 10^{12} m^2/d$ diffusion term

From the figure 2.11 below, it could be concluded that a diffusion coefficient around $2 \times 10^8 m^2/d$ would help improve the validation results slightly as the RMSE of modelled precipitation and ERA-interim's results is gradually reduced with increasing coefficient. However, a coefficient which is one order of magnitude above that threshold value will eventually ruin the simulation. Similar to the findings from control group 1, though diffusion term exerts almost no influence over the stability (smoothness), adding diffusion term with magnitude at certain degree does have some quite limited improvements on the modelled solutions and magnitude over threshold will destroy the reasonability and reliability of the simulation.



(a) RMSE of precipitation for three regions

(b) RMSE of evaporation for three regions

Figure 2.11: RMSE for RNNA cases with different diffusion coefficients

2.1.3. Result Discussion and numerical diffusion analysis

After summarizing the results of two control groups, several conclusions about adding diffusion terms into the continuity equation could be made. The diffusion does not add to numerical stability. The diffusion term's effect on improving the model is insignificant given a coefficient smaller than $2 \times 10^8 m^2/d$. The numerical solutions become unrealistic if a large diffusion coefficient is implemented. Diffusion coefficient of which range within $(0, 2 \times 10^8 m^2/d)$ should be applied is unsure, but values out of that range is definitely inappropriate. Considering the impact is negligible, I do not see any significance in incorporating diffusion terms. In fact, the numerical recipe used to discretize the governing continuity equation in the simulation — backward scheme in time and upwind scheme in space has already introduced an amount of numerical diffusion, which means the numerical recipe is more consistent with advection-diffusion equation other than advection equation. A simple 1-D case is introduced below to illustrate the existent numerical diffusion in the simulation. Basically all upwind schemes will always introduce some amount of numerical diffusion [50].

Considering a simple 1-D advection equation:

$$\frac{\partial c}{\partial t} + v\frac{\partial c}{\partial x} = 0$$

,which describes a wave propagating along the x-axis at a velocity v. In a 1-D domain, two directions are relevant with point i, either left (towards $-\infty$) or right (towards $+\infty$). If scalar

field c is positive, the traveling wave solution will propagate to the right, therefore the left direction of point i is called upwind and right side of i downwind. Vise versa. Provided the finite difference scheme for the spatial derivative $\frac{\partial c}{\partial x}$ has more points in upwind direction, the scheme is habitually referred to as an upwind-biased or simply an upwind scheme.
If we discretize the equation in FTFS scheme:

$$\frac{C_m^{n+1} - C_m^n}{\Delta t} + v\frac{C_{m+1}^n - C_m^n}{\Delta x} = 0$$

, where $v < 0$ meaning propagating to the left, so it is a upwind scheme in space differencing. Using taylor series to expand each term in the above equation and eliminating the remainder, the truncation error of this numerical recipe is calculated as

$$\tau_{\Delta x,\Delta t} = \frac{\partial c}{\partial t} + \frac{1}{2}\Delta t\frac{\partial^2 c}{\partial t^2} + v\frac{\partial c}{\partial x} - \frac{1}{2}v\Delta x\frac{\partial^2 c}{\partial x^2} + O(\Delta t^2, \Delta x^2)$$

Using modified equation approach [46] to convert the second term,

$$\because \frac{\partial c}{\partial t} = -v\frac{\partial c}{\partial x}, \therefore \frac{\partial^2 c}{\partial x \partial t} = -v\frac{\partial^2 c}{\partial x^2}, and \frac{\partial^2 c}{\partial^2 t} = -v\frac{\partial^2 c}{\partial t \partial x} \therefore \frac{\partial^2 c}{\partial t^2} = v^2\frac{\partial^2 c}{\partial x^2}$$

Hence the resulted truncation error is

$$\tau_{\Delta x,\Delta t} = \frac{\partial c}{\partial t} + v\frac{\partial c}{\partial x} + (\frac{1}{2}v^2\Delta t - \frac{1}{2}v\Delta x)\frac{\partial^2 c}{\partial x^2} + O(\Delta t^2, \Delta x^2)$$

, where the $\kappa = \frac{1}{2}v^2\Delta t - \frac{1}{2}v\Delta x$ is the numerical diffusion coefficient. It should be a positive value otherwise the system is undamped, hence $\left|\frac{v\Delta t}{\Delta x}\right| \leq 1$. Above example of 1-D advection case simply shows that due to the upwind scheme used in the modeling, numerical diffusion is already introduced into the simulation and it is inevitable. So if we add diffusion terms on the basis of the current scheme, we are kind of making an unnecessary move. Besides, we can not do better without knowing the actual magnitude and distribution of the diffusion. Increasing number of assumptions does not make the model better. Apart from upwind scheme, there are many other spatial differencing scheme, for instance the central differencing in space, will they make a better sense? Not really. Because through taking into consideration the dominant flow direction, the upwind-differencing scheme overcomes the weakness of the central-differencing scheme owing to that upwind scheme is primarily developed for strong convective flows with suppressed diffusion effects, especially for the case where Peclet number's absolute value is greater than 2 [45]. The moisture transporting in Amazon basin is a advection-dominant system, thus upwind scheme is suitable as the spatial differencing scheme. As for the time differencing scheme, implicit euler (backward in time) scheme is a good pick as it is unconditionally stable, though it does not mean arbitrary large timestep can be applied.
To sum up, adding diffusion terms brings insignificant improvements on the simulation results. Besides when specified diffusion coefficient exceeds the critical value 2×10^8, the model begins to collapse. Due to the inherent property of the upwind scheme, the numerical diffusion is inexorably extant in the modeling. Indeed more numerical recipes could be tried and tested to see the difference on the performance, but upwind scheme in space differencing should be put at primary place as it well suits the advection-dominant system.

2.2. Modify wind

The CFL condition is a necessary condition for solutions of a finite difference numerical recipe being convergent to a (non)linear hyperbolic PDE. Since published by Courant, Friedrichs and Lewy in 1920s, it has become one of the profoundest results in the history of computational numerical modelling [11]. The CFL condition stands on the concept of domain of dependence — A necessary condition of a convergent scheme is that the analytical domain of dependence is contained in the numerical domain of dependence. In terms of linear problems, the Lax equivalence theorem asserts that convergence is equivalent to stability. Hence the CFL condition can also be used to obtain a stability limit given a linear underlying PDE [27]. For an 1-D case, the CFL condition for an explicit scheme should be:

$$|\sigma| = \frac{|u|\,\Delta t}{\Delta x} \leq C$$

, where C (*i.e.* so-called Courant number) is a dimensionless constant determined only by the numerical recipe applied. For many schemes, typically explicit schemes, C is equal to 1, while larger C may be tolerated for implicit schemes. Apart from CFL condition, Von-Neumann stability analysis is at present the most commonly applied method to check the stability of finite difference schemes for PDEs. Von-Neumann stability analysis is based on the Fourier decomposition of numerical error which uses $\phi(x,t) = \widehat{\phi}(t)e^{ikx}$ to expand the equation. The necessary and sufficient condition for solutions to remain bounded (stability) is that the amplification factor:

$$|\Gamma| = \left| \frac{\widehat{c^{n+1}}}{\widehat{c^{n}}} \right| \leq 1$$

Von-Neumann method is especially suitable for linear PDEs with constant coefficients where the impacts of boundary conditions are ignored [3]. But Von-Neumann analysis to obtain stability limits might be troublesome or even impossible when we cope with nonlinear problems or PDEs with non-constant coefficients because the principle of superposition is not valid any longer for that case. To deal with these cases, one must linearize the problem by freezing the nonlinear terms or non-constant coefficients of the PDEs. Consequently, Von Neumann stability is merely a necessary condition rather than a sufficient condition since the analysis itself is an incomplete one. Therefore, satisfying the Von Neumann stability condition does not necessarily mean a numerical recipe becomes (conditionally) stable, while violation of this condition guarantees instability [50].

It can be easily inferred from the CFL condition stated above that if the time step Δt and the grid size Δx are required to remain fixed and unalterable, then the wind intensity becomes the determining factor of whether the stability limit is met. Apart from that, intuitively, there might be an upscaling effect in the modeling. When applied to larger situations, a system scales if it remains suitably efficient and practical, else it does not scale if it breaks down due to a quantity increase [21]. Defining the ad hoc requirements for scalability on key dimensions of a system is of great significance [14]. Unlike other atmospheric models that work on finer scale, the regional water balance model is evaluated at a considerably larger temporal-spatial scale without specially taking care of the scaling in quantity. Therefore potential issues may arise from this upscaling process especially in terms of the wind intensity: The figure 2.12 below shows how the previous wind intensity for each calculation cell is calculated — simply averaging wind intensity of the eight pixels at the boundary [2],

which is likely a misestimation. So further modifications on wind may be the key to solve the instability. Despite there could be some unknown functions to represent or endogenize the wind, the different fractions (magnitude) of wind intensity are studied as the starting point in this section.

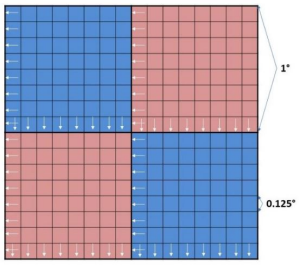


Figure 2.12: Illustration of wind intensity with spatial distribution in 1° and in 0.125°

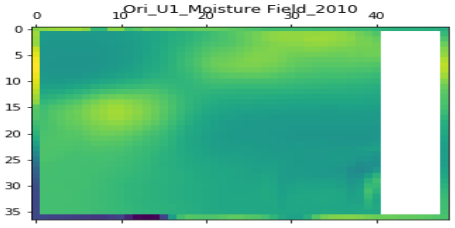
By the way, there are some special methods to deal with velocity field for controlling numerical stability that have been successfully implemented in other researches. As an example, ill-posedness in numerical simulation often results in unstable solution with spurious oscillations, of which the amplitude depends on the domain discretization. Besides generating physically unrealistic solutions, it may result in failure of numerical simulations. To eliminate this ill-posedness issue, a Two-fraction model, in which the wind consists of a fixed fraction and a dynamic pressure-gradient-driven component, is applied in a river morphodynamic modelling [9]. This methodology is not unusual or novel and similar method has been applied in many other studies *e.g.* atmospheric chemistry (ERCA) model [23]. Related discussion is in the recommendation chapter. More advanced and complicated techniques to adapt (or endogenize) velocity field should undoubtedly be given sufficient considerations in future researches, however, only fractions (magnitude) of wind intensity are tested in this section as the initial attempt. Similarly, two control groups (averaging interpolation reserved and removed) are set for comparison.

2.2.1. Control Group 1: Apply fraction of wind on the original model

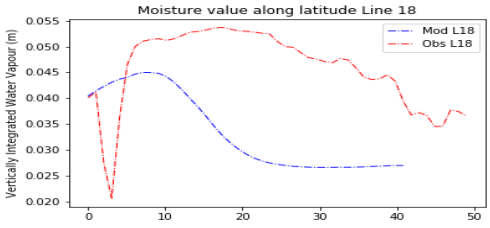
Wind fractions applied on original case are ranging from 0.05 to 1. Surprisingly it appears that the $U_{fraction} = 1$ case, *i.e.*, no modification on the wind case, has the best validation since $RMSE$ of precipitation for three regions is gradually increasing as the wind fraction is getting smaller. Scatter plot also has the best fit under original wind intensity . In terms of precipitation, modelled solutions and observations of region 1 and region 2 show desired linear relationship. Comparisons of the moisture field graphs and moisture along latitude graphs between four different wind fraction cases show that the smaller the fraction is, the "flatter" the modelled moisture field becomes.

However, it should be emphasized here that all the above analysis is insufficient and incomplete because the interpretation of the results could be physically pointless as the applied averaging interpolation method may greatly hamper the true expressions of the model. Besides, uncertainties in non-linear term calculations (*i.e.* P-A relationship and E-P conversion) could also contribute to a worse result validation. Since problems may attribute to many aspects, it is wise to do isolated study on individual component of the model. This thinking also gives stimulus to studies in follow-up sections.

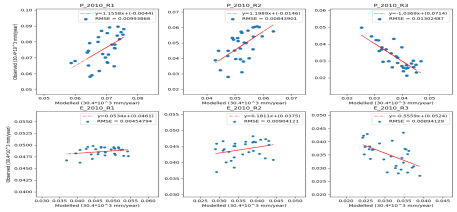
Ufraction = 1:



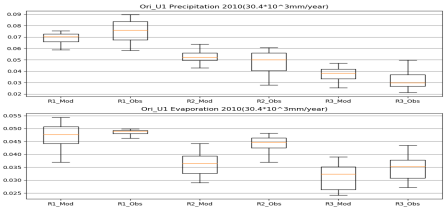
(a) Moisture field modelled in original model



(b) Moisture along mid line



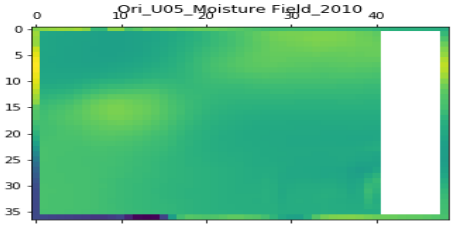
(c) Scatter plot of precipitation and evaporation



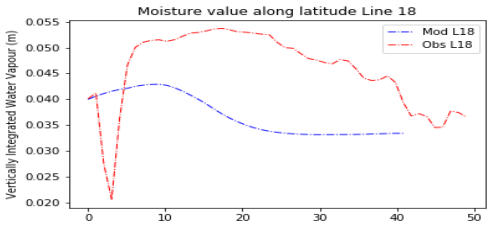
(d) Box plot of precipitation and evaporation

Figure 2.13: Results for ori model with Ufraction = 1

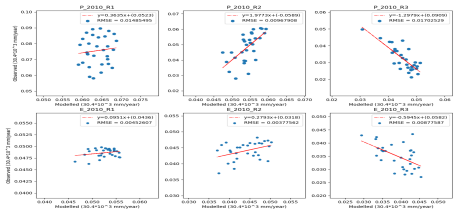
Ufraction = 0.5:



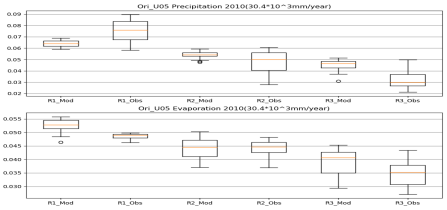
(a) Moisture field modelled in original model



(b) Moisture along mid line



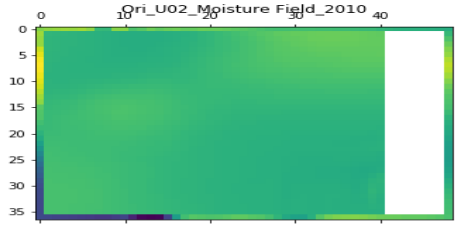
(c) Scatter plot of precipitation and evaporation



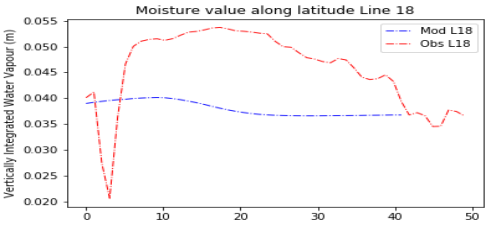
(d) Box plot of precipitation and evaporation

Figure 2.14: Results for ori model with Ufraction = 0.5

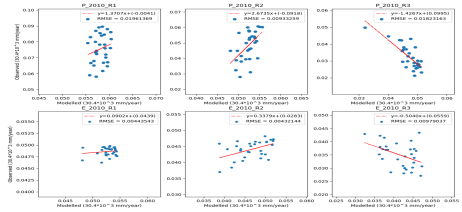
Ufraction = 0.2:



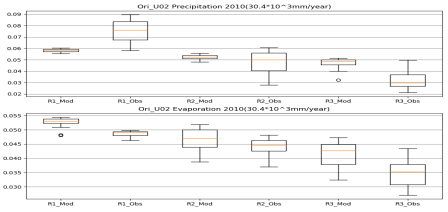
(a) Moisture field modelled in original model



(b) Moisture along mid line



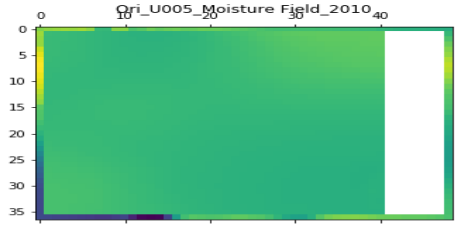
(c) Scatter plot of precipitation and evaporation



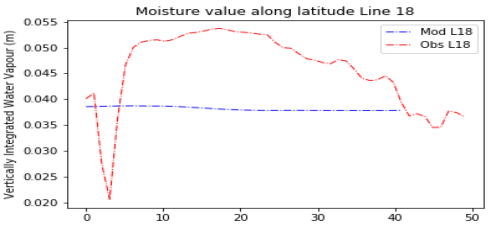
(d) Box plot of precipitation and evaporation

Figure 2.15: Results for ori model with Ufraction = 0.2

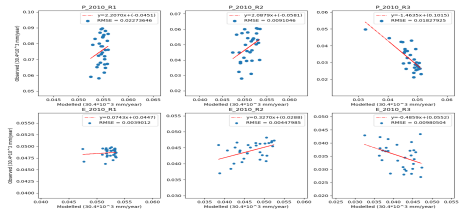
Ufraction = 0.05:



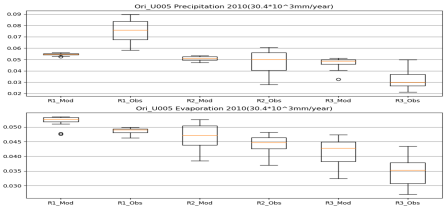
(a) Moisture field modelled in original model



(b) Moisture along mid line

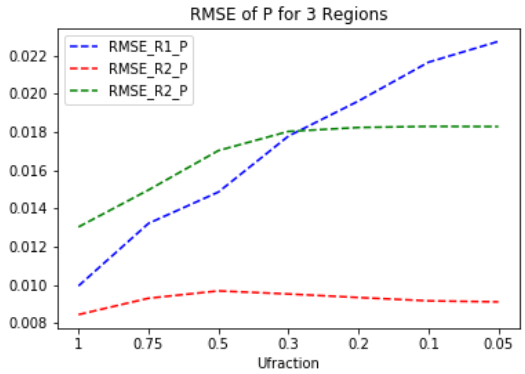


(c) Scatter plot of precipitation and evaporation

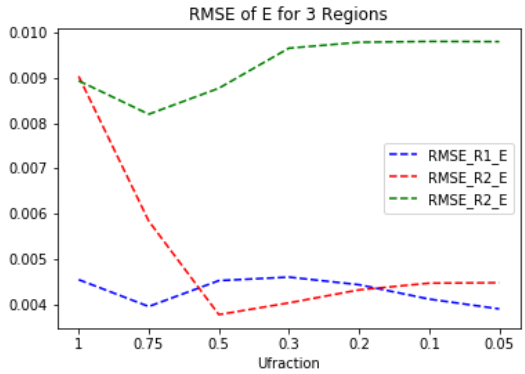


(d) Box plot of precipitation and evaporation

Figure 2.16: Results for ori model with Ufraction = 0.05



(a) RMSE of precipitation for three regions



(b) RMSE of evaporation for three regions

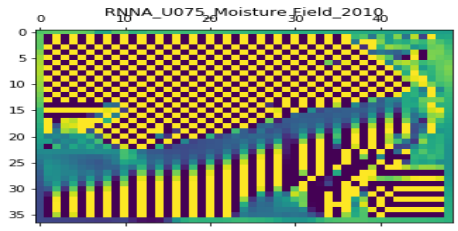
Figure 2.17: RMSE for original cases with different wind fractions

2.2.2. Control Group 2: Applying fraction of wind on RNNA model

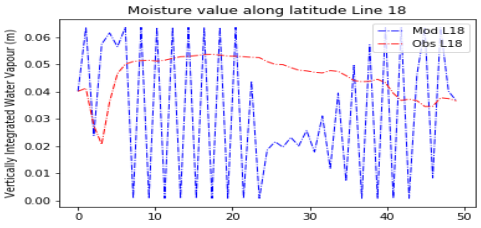
The control group 2 is given more concerns as it removes attached uncertainties of the applied averaging interpolation method. From the Courant number condition, for stability limit:$c = \frac{|u|\Delta t}{\Delta x} \leq 1$, given grid size ΔL and time step size Δt remaining fixed, applying lower wind intensity adds chances of meeting the stability criteria. A series of wind fraction coefficients are tested to examine whether this method would help solve numerical instability. A major difference to tell between $U_{fraction} = 0.75$ and $U_{fraction} = 0.5$ cases is that: Comparing figure 2.18a and figure 2.19a, the oscillation in a relative large area near the study domain central disappears in $U_{fraction} = 0.5$ case. In figure 2.19b, the wiggles on the western part along the specified latitude is replaced by a smooth curve, indicating an attenuation in the spurious oscillation in comparison with figure 2.18b. With wind fraction getting smaller, spurious oscillation on the modelled moisture field is further reduced, resulting in an increasingly smooth field. For the $U_{fraction} = 0.05$ case, it is interesting to find in the figure 2.21a that the water vapor field emerges the contour of the South America Continent. This may be explained by the reason that because of less moisture exchanged between cells due to small wind intensity, the modelled moisture pattern gets closer to the initial input pattern of moisture field where there is a distinct differentiation between the ocean and the continent.

The changes in $RMSE$ of precipitation and evaporation for three regions are shown in the figure 2.22. The averaged $RMSE$ of the precipitation decreases as wind fraction gets smaller until the point $U_{fraction} = 0.5$. In contrary, the $RMSE$ of the evaporation is increasing with reduction in $U_{fraction}$. One thing to be noted here is that the $RMSE$ (box plot) is merely an incomplete indicator and so is the scatter plot. Reducing in $RMSE$ does not necessarily means the outcome is better because if outcome is evaluated in terms of scatter plots, the linear relationship is getting negative (worse) when the wind fraction is further reduced from 0.75 on. Similarly, a better linear relationship in scatter plot does not really indicate a better simulation since errors could be arbitrarily random and coincidental as a result of complex interactions between multiple intertwined factors in the modelling. Therefore, more focus should be put on the spurious oscillation and the smoothness of the moisture field under different wind intensity conditions for stability is a prerequisite to accuracy.

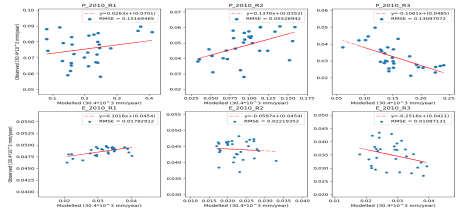
Ufraction = 0.75:



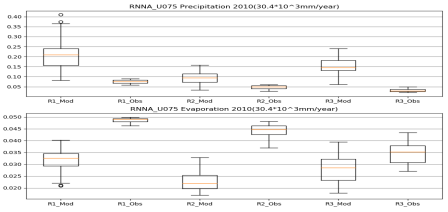
(a) Moisture field modelled in original model



(b) Moisture along mid line



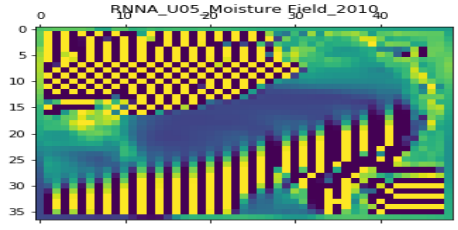
(c) Scatter plot of precipitation and evaporation



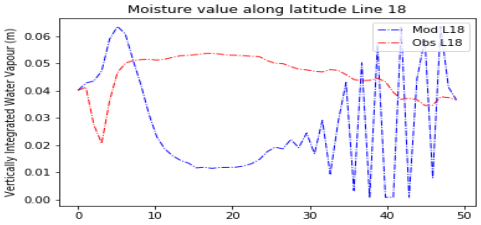
(d) Box plot of precipitation and evaporation

Figure 2.18: Results for ori model with Ufraction = 0.75

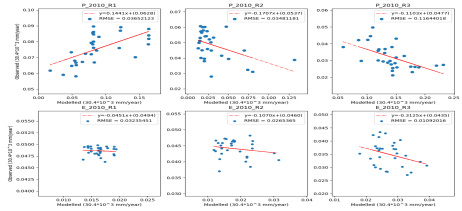
Ufraction = 0.5:



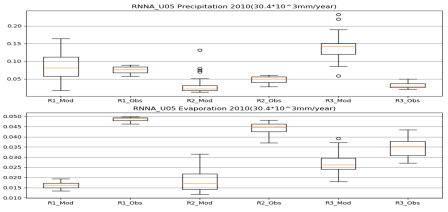
(a) Moisture field modelled in original model



(b) Moisture along mid line



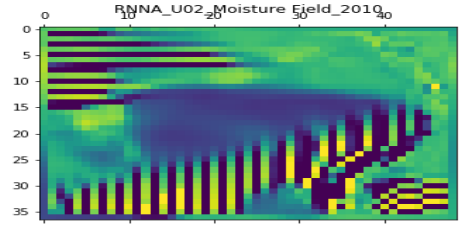
(c) Scatter plot of precipitation and evaporation



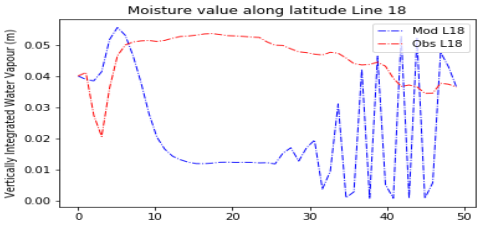
(d) Box plot of precipitation and evaporation

Figure 2.19: Results for ori model with Ufraction = 0.5

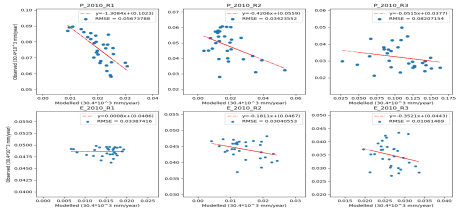
Ufraction = 0.2:



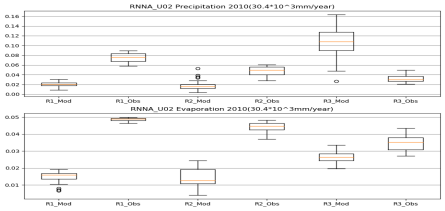
(a) Moisture field modelled in original model



(b) Moisture along mid line



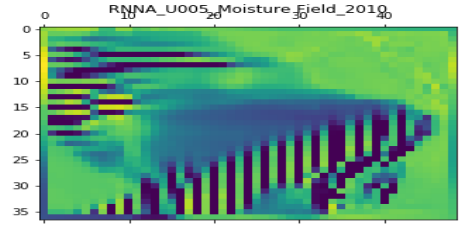
(c) Scatter plot of precipitation and evaporation



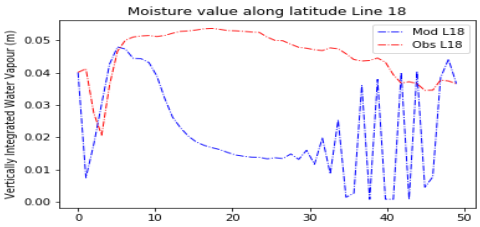
(d) Box plot of precipitation and evaporation

Figure 2.20: Results for ori model with Ufraction = 0.2

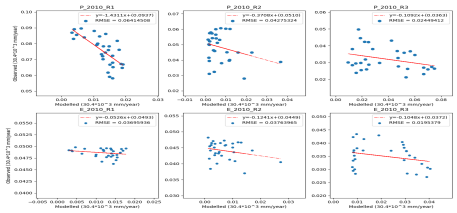
Ufraction = 0.05:



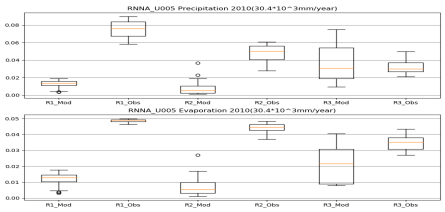
(a) Moisture field modelled in original model



(b) Moisture along mid line

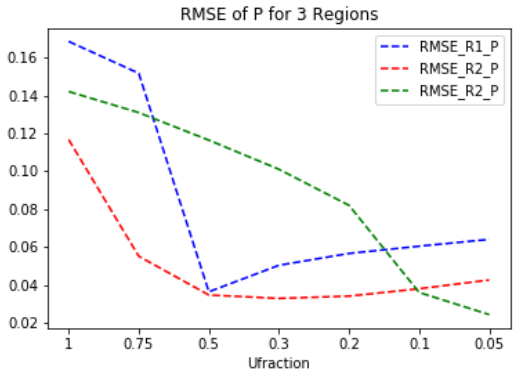


(c) Scatter plot of precipitation and evaporation

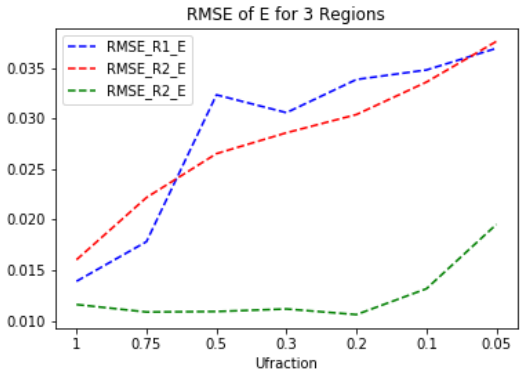


(d) Box plot of precipitation and evaporation

Figure 2.21: Results for ori model with Ufraction = 0.005



(a) RMSE of precipitation for three regions



(b) RMSE of evaporation for three regions

Figure 2.22: RMSE for original cases with different wind fractions

2.2.3. Result Discussion

Applying a smaller wind fraction helps alleviate the spurious oscillation in majority part of the study domain to a certain degree. However, the remaining part still suffers instability problem even if a minor wind fraction is chosen ($U_{fraction}$ <= 0.05). For cases of smaller wind fraction, the modelled moisture field tends to emerge the continent contour. Applied wind intensity indeed needs modifications due to upscaling. However, despite some improvements, it fails to root out the numerical instability since there is always existence of spurious oscillation at the northwest and southern part of the modelled field in spite of quite small applied fractions. Furthermore, through inspirations from other studies, there are smarter methods of handling the velocity field to control numerical stability. For example, wind is composed of fixed time-averaged component and dynamic fluctuation component, thus it is possible to link wind with pressure and friction to make it endogenous.
It seems a paradox to apply monthly time step in the modelling for that — According to CFL condition, for numerical stability, the information at certain cell should not propagate further than adjacent cell at current time step. However, in practical the moisture can be transported much further away from the neighbouring cell within one month. Thereby, I must reiterate the possibility of the original model being conceptually wrong and the feasibility of applying smaller time step (*e.g.* hourly) to fundamentally solve numerical stability.
Putting temporal scale issue aside, in the governing continuity equation, the wind term is unquestionably the most elusory factor as the result of following two respects: One one hand, wind velocity field is heterogeneous; On the other hand, it varies with time marching. Both ends can become impediments in the numerical modeling. This leads to experiments in section 2.5 and section 2.6, in which homogeneous wind distribution, constant and gradually changing wind are tested. Besides, non-linear term (source term, *i.e.* precipitation and evaporation) could also be problematic. All these uncertainties amount up to the difficulties of thoroughly correcting the modelling. The entire system somehow reflects "Murphy's Law": every time something doesn't fit it and anything that can go wrong will go wrong [7]. Besides, jumping out of the continuity equation, I was wondering whether a simplified wind-related routing system consisting embedded water balance nodes could be more representative of the regional water balance for large temporal scale compared to physically-based finite difference numerical modelling of governing equation.

2.3. Sparse Matrix Solver

In previous research, in order to solve the continuity equation implicitly, Heun' method was used as the iteration algorithm. However Heun' method is a 2^{nd} order Runge-Kutta Method for solving initial value problems of ordinary differential equations (ODEs) [1]. To solve a PDE like governing equation of this modelling, usually Jacobi and Gaussian-Seidel iteration method should be applied. Jacobi method is an algorithm used to determine the solutions to a diagonally dominant system of linear equations. Every diagonal ingredient is solved first, then an approximate value is put in and the process is iterated until convergence [25]. Gaussian-Seidel method is another frequently-used iterative method, however, of which the convergence is only ensured when the matrix is either diagonally dominant, or symmetric and positive definite [44]. Previous coding is based on Heun's method, using 50 iterations (with embedded interpolation) per time step to get a convergent solution. To be more detailed, the code scans through the first cell to the last one, and for each cell a solution is iteratively calculated based on the values of the nearby cells for each time step. This process is repeated until the difference between the newly-calculated value and the solution at previous iteration is smaller than predefined threshold or it counts to 50 iteration loops. Nearest Neighboring Averaging is performed in every iteration to ensure convergence. As already known from previous sections' experiments, without embedded interpolation, there is no convergence. The iterator solver is coded by the former researcher to solve the system of equations. Actually, there is a ready-made package called scipy.sparse in Python to solve a sparse matrix system of linear equations efficiently. Hence, the sparse matrix solver package is also tested in this section.

2.3.1. Setup

In order to test the package and setup the model, a sparse matrix system of linear equations is built up first as an example. A 3 × 4 study domain marked in yellow is used for setting up the code and for validation purposes. Moreover, it must be noted that a matrix system of equations may have no solutions as long as the coefficient matrix is a singular matrix which is irreversible [22].

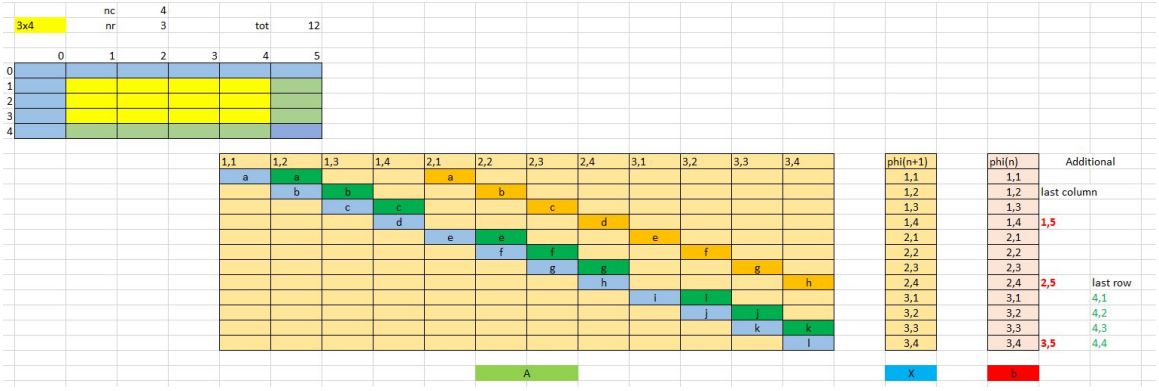


Figure 2.23: 3x4 test domain and its system of equations

The form of the sparse matrix system is

$$Ax = b$$

, where **A** is the coefficient matrix, x is the unknown matrix evaluated at time level $t = n+1$, and b is the known matrix at time level $t = n$.
The equation for a certain cell at specified location (i, j) and time level $t = n+1$ is then:

$$\left(1 + \frac{u_{i,j+1}^{n+1} - u_{i,j}^{n+1}}{\Delta x}\Delta t + \frac{v_{i+1,j}^{n+1} - v_{i,j}^{n+1}}{\Delta y}\Delta t - \frac{u_{i,j}^{n+1}\Delta t}{\Delta x} - \frac{v_{i,j}^{n+1}\Delta t}{\Delta y}\right)\phi_{i,j}^{n+1} + \frac{u_{i,j}^{n+1}\Delta t}{\Delta x}\phi_{i,j+1}^{n+1} + \frac{v_{i,j}^{n+1}\Delta t}{\Delta x}\phi_{i+1,j}^{n+1} = \phi_{i,j}^{n} + E\Delta t - P\Delta t$$

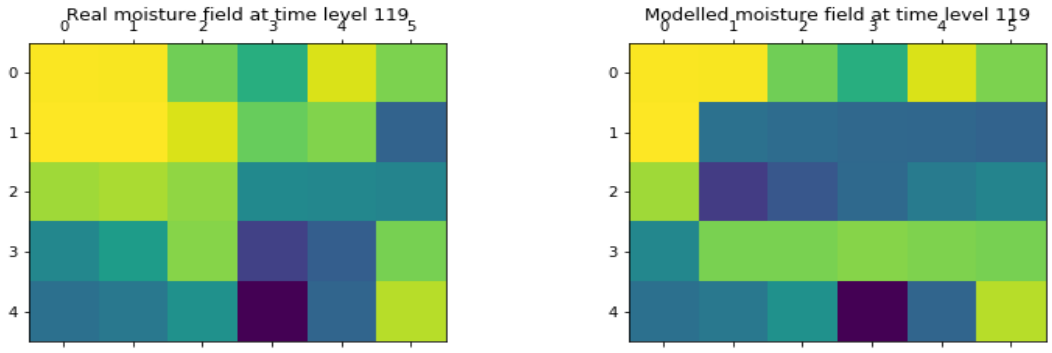
, where Δt is 1 (month).
1. First: Construct coefficient matrix A on the $l.h.s$ of the equation.
As can be seen from the illustrative figure 2.23 above, this coefficient matrix **A** is a 12×12 sparse matrix mostly filled with zeros. This matrix **A** we built to compute the finite difference solution is a sparse matrix with non-zero elements along three diagonals only. Storing such a sparse matrix as a full matrix means storing a lot of zeros. For larger models, with lots of cells, normally the computer memory could be run out of quickly. Hence it is much more efficient to only store the non-zero values of the matrix, together with their location (row and column number) in the matrix. Scipy.sparse package in python allows for fast manipulation of this type of coefficient matrix [24]. Our coefficient matrix **A** has three diagonals, and each of the diagonals contains the coefficients of the unknown at a specific location. The coefficients of the $\phi_{i,j}^{n+1}$ terms make up the diagonal 0 of matrix **A** , which is colored with light blue in the figure above. The yellow-marked coefficient of $\phi_{i,j+1}^{n+1}$ makes up the coefficient diagonal 1, which is marked in green. The coefficient diagonal nc is composed of coefficients of $\phi_{i+1,j}^{n+1}$.
2. Second: Construct known matrix b on the $r.h.s$ of the equation.
The rhs consists of the followings:

$$b = rhs0 - rhs1 - rhs2 + rhsPE$$

, where $rhs0$ is values at time level $t = n$, $rhs1$ is boundary conditions at right boundary, $rhs2$ is the bottom boundary conditions, and $rhsPE$ is the combined precipitation and evaporation term, which is explicitly obtained from solutions at time level $t = n$. The reason why the combined precipitation and evaporation term is made explicit here is for purpose of simplicity by avoiding uncertainties involved in the non-linear term calculation.
3. Third: Solve for the unknown matrix x
Using *spsolve(A, rhs)* function to solve for the unknowns at time level $t = n+1$.

2.3.2. Validation

The matrix **A** may have no inverse matrix, which means no solutions for the system of equations. A matrix will not be invertible if and only if determinant is not non zero. Besides, as stated above, due to heterogeneous coefficients in the coefficient matrix **A**, the calculated results could also be unreasonable. The sparse matrix solver is built upon a 3×4 study domain as a starting point. Prior to applying this sparse matrix solver to the 37×49 study domain, in order to test whether the algorithm is correctly programed, a validation test on the 3×4 domain is performed with a time span of 120 time steps. This validation test is more for testing the functionality of the code other than the accuracy of the results.

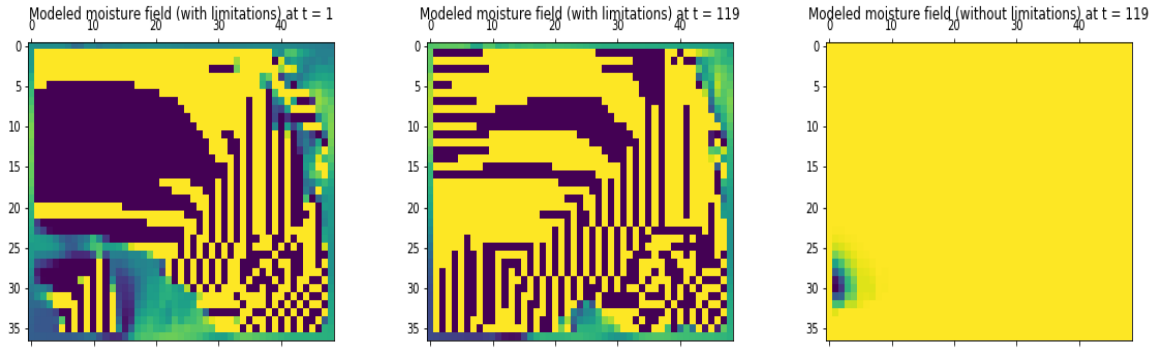


(a) Real moisture field at time level $t = 119$ (b) Modelled moisture field at time level $t = 119$

Figure 2.24: Validation of sparse matrix algorithm

As can be seen from the figures of moisture field 2.24 above, this 3×4 instance of sparse matrix system of equations does have solutions. Even though you can tell difference between the observations and modelled solutions of sparse matrix solver, the algorithm is proved to work on this example domain. Hence the next logic step is to apply this solver on the 37×49 study domain. But please note that this programming algorithm of sparse matrix solver guarantees neither certain solutions (it could have no solutions) nor reasonable solution (it could have ridiculous solutions *e.g.* negative, arbitrary, oscillating solutions).

2.3.3. Results

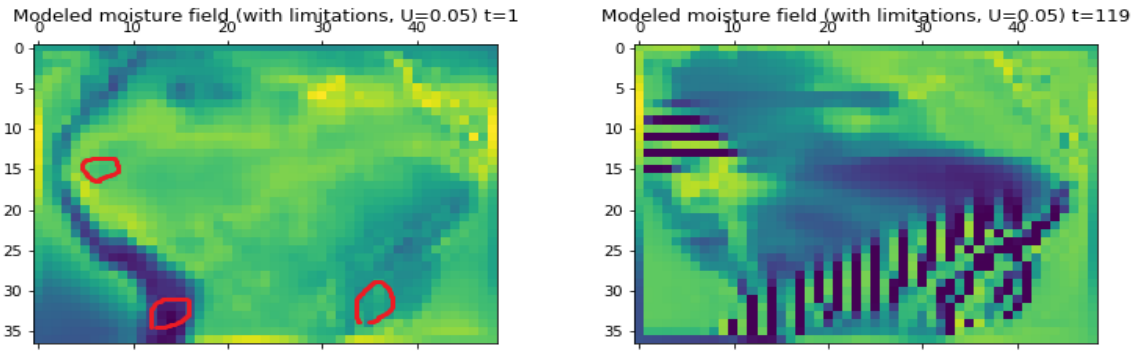


(a) Modelled moisture field at t = 1, with limits (b) Modelled moisture field at t = 119, with limits (c) Modelled moisture field at t = 119, without limits

Figure 2.25: Modelled moisture field by sparse matrix with and without boundary limits, $U_{frac} = 1$

In figure 2.25, when upper and lower limits (0.0636 and 0.00078, *resp.*) are applied to control the output solution of the sparse matrix solver, for calculated moisture after each iteration at every time step, the modelled moisture field is unstable with numerous oscillation. Besides the model fails in the first round of computing ($t = 1$) (figure 2.25a) and gets worse with time stepping (figure 2.25b). Moreover, if no man-made bounded constraint is predefined to limit the solution output, the simulation becomes even worse. As shown in figure 2.25c, the maximum and minimum solutions are $4.8890e^{216}$ and $-3.9481e^{216}$ respectively,

which are undoubtedly unrealistic ridiculous results.
What if a smaller wind fraction is applied? The figure 2.26 below shows, under $U_{frac} = 0.05$



(a) Modelled moisture field at t = 1, with limits (b) Modelled moisture field at t = 119, with limits

Figure 2.26: Modelled moisture field by sparse matrix with and without boundary limits, $U_{frac} = 0.05$

condition, the modelled moisture field at the first time step and the final time step. The model (with boundary limits) works well at early time steps. However after the beginning stage, the oscillation begins to occur originating from red circle in figure 2.26a. And with time stepping, the unsmooth field expands further gradually, resulting in the pattern in figure 2.26b. Compared to the result of original method without averaging interpolation, the result is a bit better. In contrast, if no limit is set for solutions, after around 15 time step, the model starts to fail (oscillations accumulate quickly and propagate), and eventually the modelled solution has 5728.1145 and -48509.7679 as maximum and minimum resultants *resp.*, of which the pattern looks similar as figure 2.25c. In general, when $U_{frac} = 1$, the model fails in the beginning, whereas a smaller wind fraction applied assures the model of delayed failing and a better moisture pattern. Next, making source term implicit is also tested. 50 iterations that consistently update solution for approaching convergence is added for each time step. Under $U_{frac} = 0.05$ case, the result of implicifying source term is not as good as explicifying them since it has as severe oscillation just like figure 2.21a.

2.3.4. Result Discussion

Sparse matrix solver solves the equation semi-implicitly. Under $U_{frac} = 0.05$ case, sparse matrix solver has similar results as original method without interpolation, but it excels in computing speed by 10×. Both two solvers suffer the problem of oscillation growth which originates from red circles after the initial phase and later propagates and expands further. Decreasing wind helps alleviate the oscillation to a certain extend but cannot eliminate it. This finding reiterates wind could be the key problem. Smaller fraction of wind can produce better results with emergence of some notable characteristics of the ERA-interim's. Therefore, previously used $U_{frac} = 1$ is such a big value that may overestimate the exchange limit of atmospheric moisture between two cells. That could explain why it relies on interpolation to average the huge jumpings between solutions. As a matter of fact, scanning through the evolution of moisture field of ERA-interim, the moisture distribution pattern does not change significantly as expected throughout time, only obvious seasonal variation is captured mainly in the south-eastern part of the continent.

2.4. Sensitivity of source terms

In the continuity equation, evaporation and precipitation are source (and sink) terms to the unit cell. The relationship between P and A (moisture), and the relationship between E and P are established as a series of nonlinear functions with parameter predefinition and threshold determination. Apart from that, the concept of maximum climatological water deficit is also used in the modeling process to include a residence storage in the soil to bond evaporation to precipitation [31]. The formulas are presented below:

$$E = min(A + \frac{P}{p_1}(1 - exp(-\frac{p_2}{P})), E_p)$$

$$A = -p_3 C + p_4, C_n = C_{n-1} + P_t - E_t$$

, where C_n is the current water deficit, P_t and E_t are the precipitation and evaporation in the current month, A is a carry-over factor relating to rootzone thickness and soil moisture, parameters p_1, p_2, p_3 and p_4 are dependent on surface vegetation type. These formulas describing the relationship between precipitation and evaporation have been used by Zemp, D.C in his research [49]. Note the above applied water deficit formula was the modified form of original formula, which is given below:

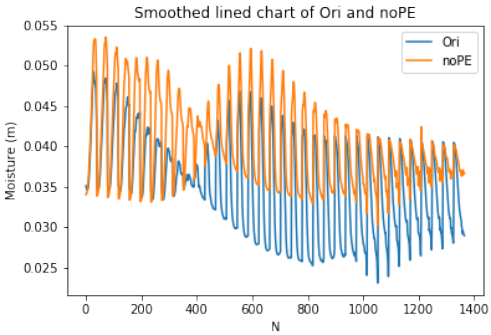
$$C_n = C_{n-1} + P_t - E_{fixed}, C_0 = C_{12}$$

, Where E_{fixed} is an approximation of evapotranspiration rate under favorable climatic condition and is fixed as $3.3 mm/d$. However, in Zemp's monthly-based model, moisture recycling network is built on the aggregated dynamic output of the 3-hourly-based $WAM-2layers$ model [48], thus the above formulas are not embedded in governing equation modelling. Consequently, endogenizing source terms with this intricate relationship might also give birth to numerical incorrectness due to superposition of errors and incorrect parameterization. Hence, in the sensitivity test, source terms are removed from modelling to investigate the impacts and sensitivity. Therefore it will leave the modelling to be simply a 2-D advection modelling within a time-variant heterogeneous wind field. It might also be noted, subsequently, more attention is put on the smoothness of modelled moisture field because getting stable solutions is a prerequisite for obtaining convergent solutions.

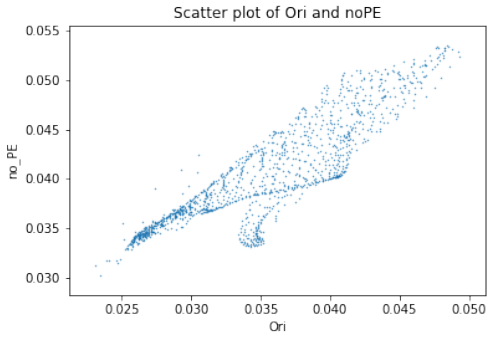
2.4.1. Results

Original solution with source terms is compared with modelled solutions without source terms. Though, logically speaking, source term should not be frozen and removed from modelling since it acts as an indispensable part in the moisture recycling, it is meaningful to perform isolate sensitivity tests to evaluate the built-in relationships. For comparison, solutions at final step for two cases with a 35×39 dimension are selected. The dimension of domain to compare is 35×39 instead of 37×49 because $Null$ values from the blank gap as well as boundaries are excluded.

F-test two-sample for variance is performed on two datasets. From table 2.1, as F value is greater than F critical, the null hypothesis that two subsets have same variance is rejected. Hence the variances of two datasets are significantly different. Nevertheless, the correlation coefficient $\rho_{x,y}$ between the two datasets is 0.88. And it can be found out from the smoothed lined chart figure 2.27a and scatter plots 2.27b that the interrelated relationship between the two datasets is evident. The solution for the case without source term is on



(a) Smoothed linechart of Ori and noPE cases



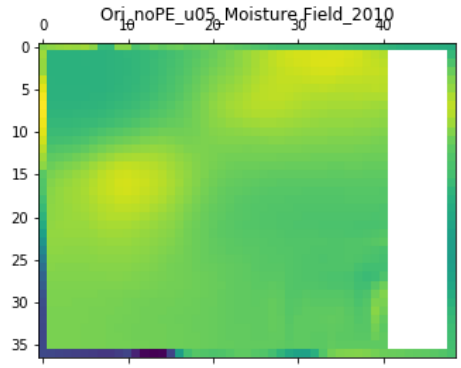
(b) Scattered plot of Ori and noPE cases

Figure 2.27: Comparison between the solutions of original case and freezing PE case

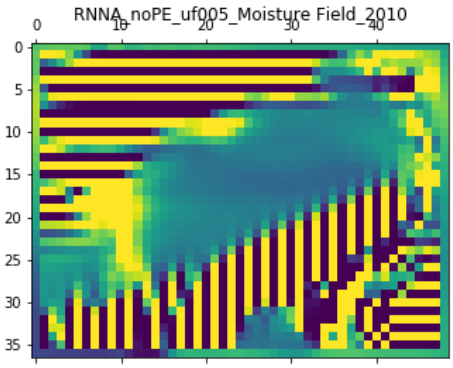
Table 2.1: F-test on original solution and freezing PE solution

F-Test	Mean	Variance	Observations	df	F	P(F<=f) one-tail	F critical one-tail
Ori	0.035012	3.6E-05	1365	1364	1.321208	1.41E-07	1.093195
noPE	0.040245	2.72E-05	1365	1364			

average greater than the original case solution, which can be explained by the removal of source term reducing the depletion of moisture storage in the unit cell.



(a) Ori moisture field (no P and E), uf = 0.5



(b) RNNA moisture field (no P and E), uf = 0.05

Figure 2.28: Results of modelled moisture field for two control groups (without P and E)

2.4.2. Result Discussion

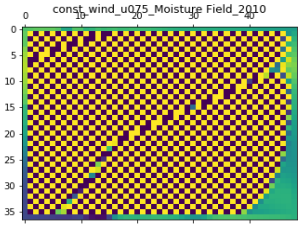
Comparing the figure 2.28b with figure 2.21a, If the NNA interpolation is removed from modelling, the modelled moisture field for the case with no source term is worse than that for the case with source term. The source (and sink) term — precipitation and evaporation are indispensable parts of moisture recycling and moisture redistribution. The established interrelationship between P and A and between E and P may not be problematic as expected, though there is still room for further work on the parameterization. However, the inaccurate calculated moisture A after every time step could induce an error in calculated source term thus probably producing error propagation in the water balance and fluxes. If not dealt with properly, the problem may get bigger causing failure in simulation.

2.5. Constant wind intensity

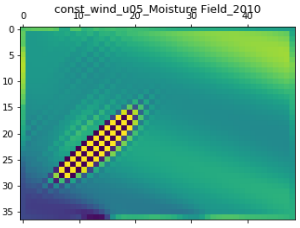
Wind field in previous simulation is a time-variant and heterogeneous vector field. It is normal to see in the study cases of graduate-level lectures of numerical modelling, to model advection-diffusion-reaction equation usually takes velocity as constant and homogeneous. However for most practical cases, the velocity field is varying with time and location. So here comes question: how to properly incorporate $u(x, y, t)$? Does it contribute or relate to the spurious oscillation? Furthermore, when previously man-made limit is set to control solutions, as a result of immediately cutting off exceedance without making it closure together with performing NNA interpolation after every iteration of computing, the water balance conservation is meanwhile compromised on. Yet it is commonly recognized that to numerically conserve conserved quantities is fundamentally significant. And to achieve this, usually a flux conserving scheme is required. A flux-conserving scheme makes cells out from grids rather than simply sampling grid points [15]. According to the flux-conserving advection algorithms in the book [28], there are a few feasible approaches to get higher order accuracy free from risk of spurious oscillations, *e.g.* Donor-cell advection, piecewise linear schemes, slope limiters, flux limiters and etc. Since the research is a reconnaissance, advanced techniques mentioned above are not tested but left as recommendations. But in this section, three forms of wind are tested for different wind fractions: a. global constant ($u(const.)$), b. time-invariant wind ($u(x, y)$), c. space-independent wind ($u(t)$). The purpose of carrying out these tests is to study the influence of spatial-temporal dependence of wind on the simulation to acquire more knowledge in numerical modeling.

2.5.1. Results

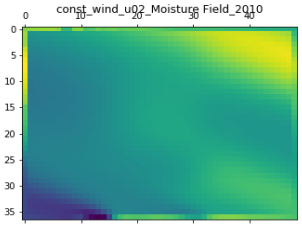
global constant wind: The wind u and v are assumed to be identical at anytime and anywhere. The mean of absolute value of all grids for 10 years, $-346644 m/d$ and $-132737 m/d$ for u and v *resp.* is taken as a starting point.



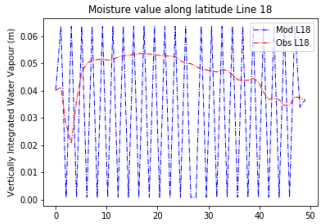
(a) global constant $U_{fraction} = 0.75$ moisture field



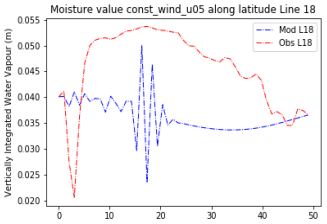
(b) global constant $U_{fraction} = 0.5$ moisture field



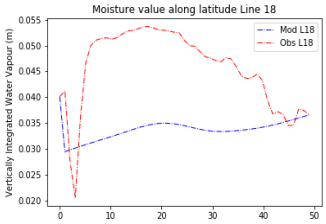
(c) global constant $U_{fraction} = 0.2$ moisture field



(d) Moisture along latitude $U_{fraction} = 0.75$



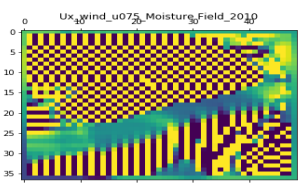
(e) Moisture along latitude $U_{fraction} = 0.5$



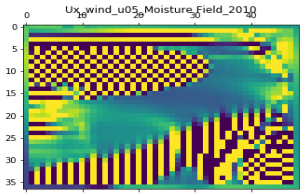
(f) Moisture along latitude $U_{fraction} = 0.2$

Figure 2.29: Results of global constant wind cases for three wind fractions

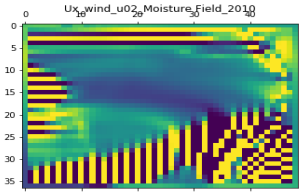
time-invariant wind: u and v at each grid are the mean of 12 months' value of that grid. So the wind intensity does not change with time. It is only heterogeneous.



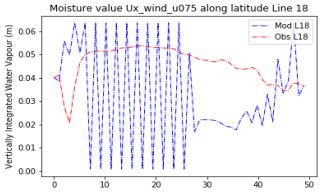
(a) time-invariant $U_{fraction} = 0.75$ moisture field



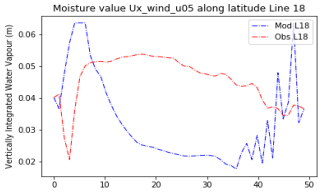
(b) time-invariant $U_{fraction} = 0.5$ moisture field



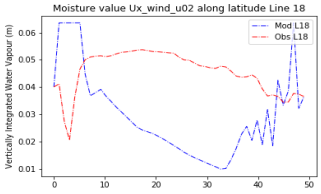
(c) time-invariant $U_{fraction} = 0.2$ moisture field



(d) Moisture along latitude $U_{fraction} = 0.75$



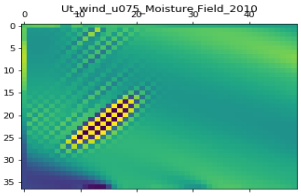
(e) Moisture along latitude $U_{fraction} = 0.5$



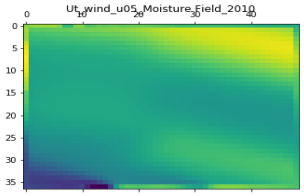
(f) Moisture along latitude $U_{fraction} = 0.2$

Figure 2.30: Results of time-invariant cases for three wind fractions

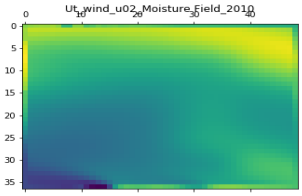
Space-independent wind: At certain time step, u (v) is homogeneous of which the value equals to the mean of u (v) of all grids at that time step. This global value changes as time stepping.



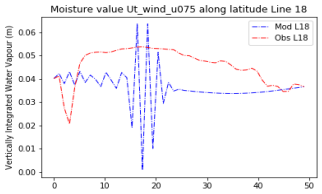
(a) space-independent $U_{fraction} = 0.75$ moisture field



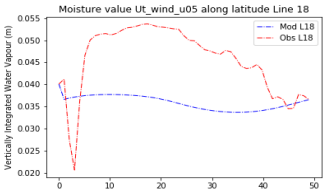
(b) space-independent $U_{fraction} = 0.5$ moisture field



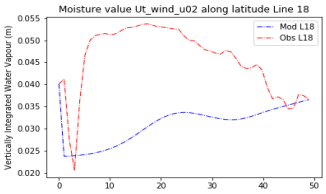
(c) space-independent $U_{fraction} = 0.2$ moisture field



(d) Moisture along latitude $U_{fraction} = 0.75$



(e) Moisture along latitude $U_{fraction} = 0.5$



(f) Moisture along latitude $U_{fraction} = 0.2$

Figure 2.31: Results of space-independent cases for three wind fractions

2.5.2. Result Discussion

Figure 2.29 illustrates what would happen if wind is a global constant, in which the homogeneous wind field is invariant throughout simulation. As long as the magnitude is relatively large ($U_{fraction}$ >= 0.5) (see figure 2.29b and figure 2.29e), spurious oscillation occurs. However, for $U_{fraction}$ = 0.5 case, despite occurrence of oscillation, the mass is still conserved as the solution is wiggling between the default limits (see figure 2.29e). Increasing the fraction up to a certain point, the water quantity is still conserved. Once the threshold is exceeded, the exceedance is immediately cut off by the limits, so the water balance is no longer conserved. When a smaller fraction of wind is taken as global constant, spurious oscillation and numerical instability no longer exist. This test indicates that the magnitude of wind determines whether stability limit is met in CFL condition.
Figure 2.30 shows the case where wind is not changing with time stepping and is only a function of location. Under this condition, the wiggles occur in the modelled moisture field right after the first iteration and gets exacerbated as simulation goes on, which could be explained by that the time-invariant heterogeneous wind field continuously exerts particular influences on moisture distribution and keeps intensifying the oscillation. The result is similar to "RNNA with small wind fractions" test studied in section 2.2 in that: *a*. Wiggles are detected after first computing iteration and are aggravated with the simulation going on. *b*. Smaller magnitude of wind can to certain extend alleviate the wiggles but cannot completely eliminate it. Note the major difference between this test and RNNA case test is: heterogeneous wind field is time-invariant in this test, whereas the wind field is inconstant in RNNA case test. Therefore it could be inferred from the comparison that the temporal dependency of wind is not literally responsible for the spurious oscillation in simulation.
Figure 2.31 shows the homogeneous wind field case in which the wind is assumed identical at every grid for current time step but changes with time. Given a large fraction of wind (see figure 2.31a and figure 2.31d), the oscillation exists due to the violation of CFL condition, while under small fraction of wind condition, the modelled moisture field is smooth and stable with no wiggles. This test reaffirms already-reached conclusion that the magnitude and perhaps spatial heterogeneity of wind field matter to the numerical stability issue.

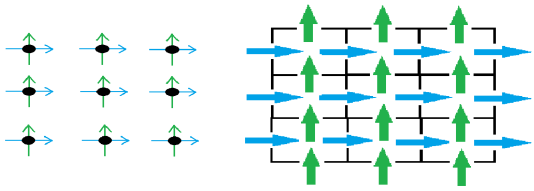


Figure 2.32: Grid and cell scheme comparisons

Figure 2.32 compares grid and cell. Artificial limits to cut off exceedance instead of making it closure and iterative averaging interpolation compromise on the conservation of mass, which also brings risk of spurious oscillation. Consequently, it is pivotal to apply a flux-conserving scheme, which creates cells from grids. Grid point then becomes a cell center encompassed by cell interfaces or walls that are established between centers. One commonly used scheme is piecewise constant advection algorithm. Besides, in order to prevent overshoots in algorithm, non-linear tools like flux limiters have to be applied to eliminate oscillation developments near a jump [28]. Hence, to apply an advanced flux-conserving scheme is recommended in future work.

2.6. Multiple time steps

As mentioned above in section 2.5, it is inferred that wind being time-variant is not (obviously) responsible for spurious oscillation in simulation given that oscillation still exists under constant wind condition. Meanwhile, as referred to in section 2.2, applying monthly-based temporal scale in modelling continuity equation seems inherently paradoxical. Therefore in this section, predefined time step size Δt of 1 month is downscaled to 0.1 month thus total time step is multiplied tenfold with wind component u and v dynamically interpolated. Besides, explicit scheme (forward euler in time differencing) is applied to save memory and reduce computing time.

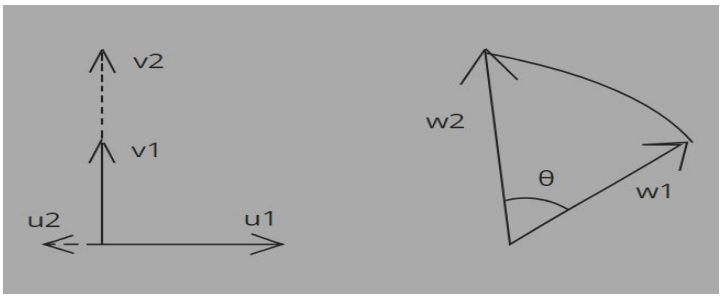
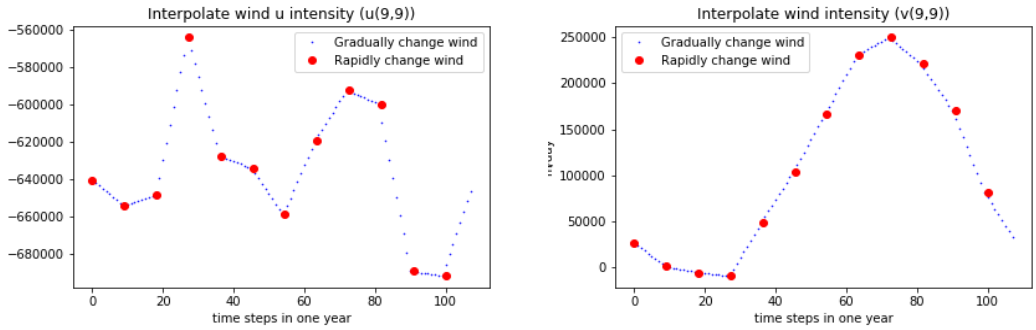


Figure 2.33: illustration of wind intensity interpolation

Due to lack of temporal dynamics of ERA-interim products, wind is interpolated based on monthly ERA-interim data. A more accurate method of interpolating would be considering vector w_1 gradually rotating by an angle of θ and extending to length of vector w_2, as reflected in right half of the figure 2.33. However, for simplicity purpose, u and v components are separately treated and interpolated linearly as shown in left of the figure.



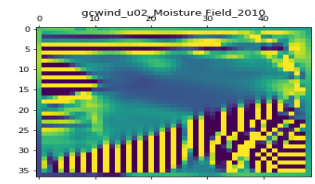
(a) wind component u interpolation at grid (9,9)

(b) wind component v interpolation at grid (9,9)

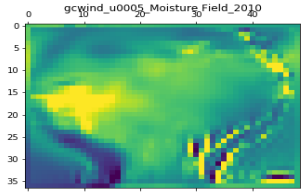
Figure 2.34: illustration of how wind intensity is interpolate for additional iteration steps

Figure 2.34a and figure 2.34b take grid (9,9) as an example—Based on two consecutive month, eight additional values are linearly interpolated in between. Hence total time step rises from 120 to 1080 for ten years simulation. Similarly, wind of different fractions are tested in order to meet the CFL condition. The results are selected and shown below.

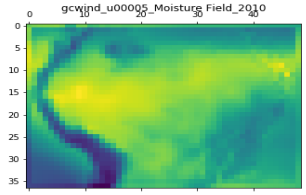
2.6.1. Result



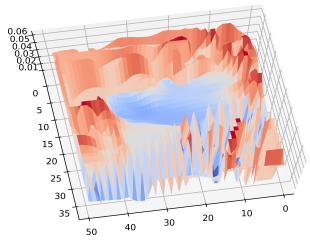
(a) Gradually changing wind $U_{fraction} = 0.2$ moisture field



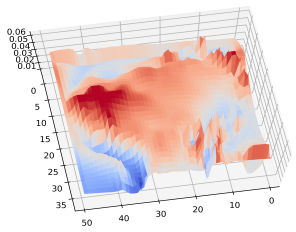
(b) Gradually changing wind $U_{fraction} = 0.005$ moisture field



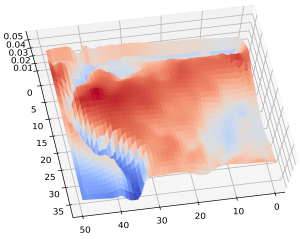
(c) Gradually changing wind $U_{fraction} = 0.0005$ field



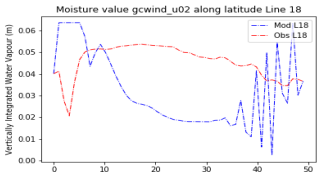
(d) $U_{fraction} = 0.2$ 3D moisture field



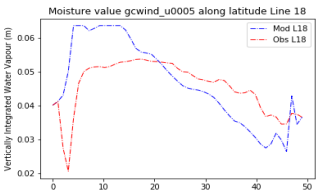
(e) $U_{fraction} = 0.005$ 3D moisture field



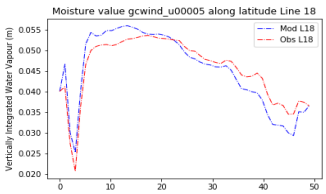
(f) $U_{fraction} = 0.0005$ 3D moisture field



(g) Mid latitude $U_{fraction} = 0.2$



(h) Mid latitude $U_{fraction} = 0.005$



(i) Mid latitude $U_{fraction} = 0.0005$

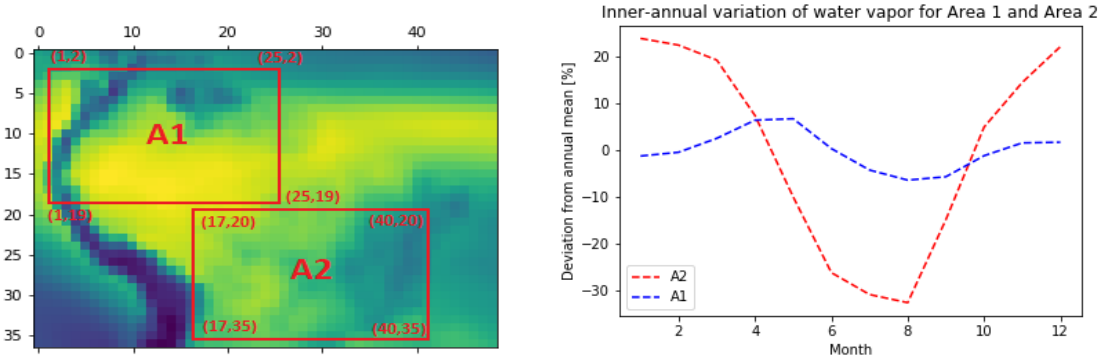
Figure 2.35: Results of gradually changing wind for three wind fractions

2.6.2. Result discussion

Through linearly interpolating wind intensity, reducing time step size Δt and extending total iterations, a multiple time step explicit model, which runs simulation of 10 years 10× faster, is built for different wind fractions, the results of which at final step are shown in the figure 2.35. Provided a tiny fraction is applied ($U_{fraction} <= 0.005$), spurious oscillation disappears and smooth moisture field is obtained (see figure 2.35b and figure 2.35c). Besides, comparing the moisture value along the mid latitude, the fitting between modelled solution and observation is getting better with wind fraction decreasing. Quite interesting! Because a tiny wind fraction under our explicit scheme means the exchange between grids is quite limited, therefore the modelled solution might not change dramatically throughout simulation, implying the final solution is approximating the initial input of ERA-interim, which explains why the modelled moisture field is smooth. So here come questions: *a.* Why does a tiny fraction of wind "work"? Is it a coincidence or the moisture content simply does not change significantly as expected and moisture transport between grids is just limited? How (much) does moisture in Amazon basin evolves with time practically? *b.* Is any possibility of technical error that might miscalculate the wind magnitude involved? Is it related to upscaling? What would then be wind's approximate order of magnitude?

2.7. Get the order of magnitude of wind intensity

Plotting the evolution of observed atmospheric moisture field of Amazon basin over ten years, it is not difficult to find that the seasonal as well as inter-annual variation of moisture field is at different levels for different regions. In order to illustrate this pattern, two sub ares that cover vast majority of the main study domain are taken (see in Fig 2.36a), of which the size are *resp.* 18×25 and 16×24. Firstly, the mean monthly averaged water vapor over the area is calculated from ten year time series. Then, the inner-annual variation of the moisture over the averaged annual mean is plotted to illustrate the averaged inner-annual variability of the moisture field. From figure 2.36b, it is obvious that compared to Area 1 which fluctuates within 10 percent, Area 2 has more significant variations of moisture throughout the year, varying from −33% to +23%.



(a) Sub domain selection

(b) Inner-annual variation of water vapor for Area1 and Area2

Figure 2.36: Moisture field inner-annual variation analysis

It indicates that throughout a year the moisture in northern part of study area does not change as much as in southern part of domain. Linking this conclusion with the conclusion from section 2.6, it might be inferred that wind possibly exerts different level of influence on different areas.

In order to get a general feeling of the approximate order of magnitude of the wind intensity to be applied in the modelling as well as to root out dormant errors in operating the model, unit check and direct calculations are performed.

First, unit check is performed. Considering 1-D advection-reaction case,

$$\frac{\partial A}{\partial t} = -\frac{\partial (Au)}{\partial x} + E - P = -A\frac{\partial u}{\partial x} - u\frac{\partial A}{\partial x} + E - P$$

$$\frac{[L]}{[T]} = [L]\frac{[L/T]}{L} + [L/T]\frac{[L]}{[L]} + [L/T]$$

If the time window is in second (s) and depth is in meter (m), then the unit would be:

$$\frac{m}{s} = m\frac{m/s}{m} + m/s\frac{m}{m} + m/s - m/s$$

However, if the time window is in month, and all the related parameters are monthly mean of daily means, then the data used in the previous simulation are expressed as below:

$$\frac{m}{mon} \neq m\frac{m/d}{m} + m/d\frac{m}{m} + m/d - m/d$$

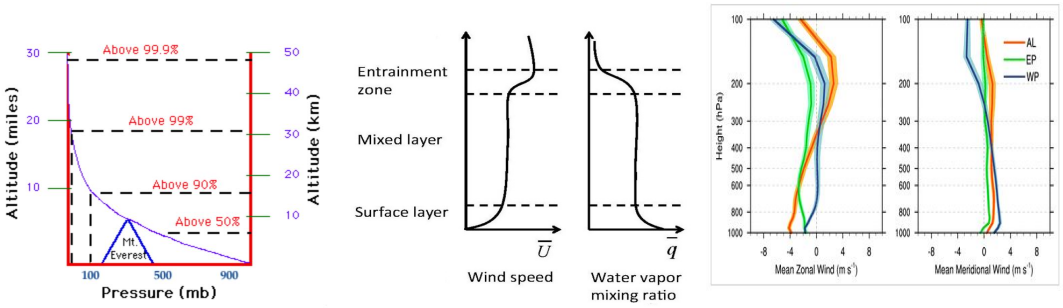
, where in the old simulation, as shown in the code, the water vapor input is converted from Kg/m^2 to m equivalent of water by multiplying 10^{-3}, the wind intensity is converted from m/s to m/d by multiplying 86400 and the unit of precipitation and evaporation flux is m/d. However there is no step in the code that converts it to a monthly value.
As can be seen from the unit check equation above, the unit on $l.h.s$ is m/Mon mismatching the unit m/d on $r.h.s$. Hence, a dormant technical error is spotted by the unit check. To fix this bug, all related parameters on $r.h.s$ of the equation — wind intensity u, precipitation and evaporation flux P and E are multiplied by a factor of 30.42 ($\frac{365}{12}$) to convert the unit from m/d to m/mon. So the equation is then evaluated on a monthly time step. However, multiplying the quantity by 30.42 will compromise the predefined P-A relationship in the old simulation since it is established based on the unit m/d. Hence these connections are also accordingly modified. Modifications of these built-in relationship notwithstanding, compared to previous $RNNA$ results in section 2.2, the final results are just as poor.
Previous experiments that attempt to solve the instability have been proven a failure. However, it has been concluded that the key to the question either lies in temporal scale or the wind component. If frontal breaking through the encirclement of the problem does not work, seeking for a chance at lateral might be a good choice. Therefore the next logic step is to get a general idea of the order of magnitude of wind intensity to be applied by performing data analysis and calculations. It is necessary to get a first-order estimation of wind's order of magnitude due to following reasons:
Firstly, currently the wind intensity incorporated in the modelling is the wind data at pressure level of 750 hPa. A formula proposed by the National oceanic and Atmospheric Administration (NOAA) suggests a conversion of atmospheric pressure in millibars (mb) to pressure altitude in feet (ft) [41] as below:

$$h = 145366.45 * [1 - (\frac{p}{1013.25})^{0.190284}]$$

Conversely, atmospheric pressure above sea level is calculated with below formula:

$$p = 101325 * (1 - 2.25577 * 10^{-5} * h)^{5.25588}$$

where, p is air pressure (Pa), h is altitude above sea level (m). After calculation, the altitude at the pressure level $750hPa$ is about 2.465.2 kilometer (8087.9 feet) above the sea level.
The figure 2.37a from a web library powered by University of Illinois [47] implies that atmospheric pressure decreases with increasing height. Due to more than 50% of the molecules in atmosphere are stored below the altitude of $5.5km$, the atmospheric pressure reduces around 50% (to *approx.* $500mb$) within the lowest $5.5km$. The pressure keeps decreasing with an increasingly slower rate when above $5.5km$.
Besides, influence of surface friction on atmospheric motion is decreased with altitude increasing to a certain level ($1-2km$) and the depth of this atmosphere in which friction plays a part in air motion is referred to as boundary layer [19]. In the boundary layer, there is usually a logarithmic-type wind vertical profile [38] as shown in figure 2.37b, in which the wind speed can be calculated by power-law empirical formula, log-law model and other modified formulas depending on data availability and specific conditions at different areas [13] [35]. However, between the ceiling of the boundary layer and the tropopause (around $17km$), the vertical wind profile in that range becomes a different story. The mean vertical zonal and meridional wind profiles on three major ocean basins of the Northern Hemisphere are shown in figure 2.37c, in which there are changes in wind speed magnitude as



(a) Pressure-Altitude relationship

(b) Vertical wind profile within boundary layer

(c) Vertical wind profile from altitude 0 to 15km in three ocean basins

Figure 2.37: Pressure-Altitude Relation and Vertical wind profile

well as directions from lower altitude to high altitude and those patterns vary with ocean basins [17]. Additionally, comparing vertical profiles of wind speed within the range from ground level to tropopause obtained from three independent research conducted in North America [5], Germany [36] and Maroc [20] respectively, it is not difficult to find that vertical wind profile above the boundary layer is determined by too many factors to be described and formulized generally. These conclusions reflect one bottleneck of the previous modelling. In previous work, the wind intensity data at $750hPa$ pressure level (around $2.5km$ altitude) from ERA-interim is taken as the transporter between grids. How could the wind intensity at one specified pressure level in the atmosphere be sufficiently able to represent the exchange between columns of vertically integrated moisture due to the complicated 3-D structure of atmosphere? Though I can understand this simple 2-D model is developed for quick modelling instead of using more complex multiple-layer 3-D model of different pressure levels (from $1000hPa$ to $125hPa$), I doubt the reliability of the model fundamentally. Since simplification is not simple, it is the ultimate sophistication. I have been trying to find some research on "general" wind formula so to use in the 2-D modelling, but nothing useful has been found. Consequently, if there was no directly relevant theory, it might be a good choice to do some statistics and calculations based on the data in possession. At least it could give us some clues or general idea on the reasonable order of magnitude of the wind intensity and maybe leave recommendations for future research possibilities.
Assumes there is a conversion factor K. The continuity equation is implicitly discretized into the following form, then the advection term is multiplied with a fraction K to be calculated. Data at time level $t = n+1$ is used to perform calculations.

$$\frac{\partial A}{\partial t} = -(A\frac{\partial u}{\partial x} + u\frac{\partial A}{\partial x}) - (A\frac{\partial v}{\partial y} + v\frac{\partial A}{\partial y}) + E - P$$

$$\frac{A^{n+1}-A^n}{\Delta t} = -(A^{n+1}\frac{u^{n+1}_{j+1,i}-u^{n+1}_{j,i}}{\Delta L} + u^{n+1}\frac{A^{n+1}_{j+1,i}-A^{n+1}_{j,i}}{\Delta L}) - (A^{n+1}\frac{v^{n+1}_{i+1,j}-v^{n+1}_{i,j}}{\Delta L} + v^{n+1}\frac{A^{n+1}_{i+1,j}-A^{n+1}_{i+1,j}}{\Delta L})K + E - P$$

Here there are a few considerations and assumptions to be made clear. Assumptions: Proper order of magnitude of the wind is the order of magnitude of wind intensity taken into the modelling multiplied by a fraction K, which is coequally applicable to zonal and meridional wind. Considerations: As this is a simple backward calculation, it is conceivable to

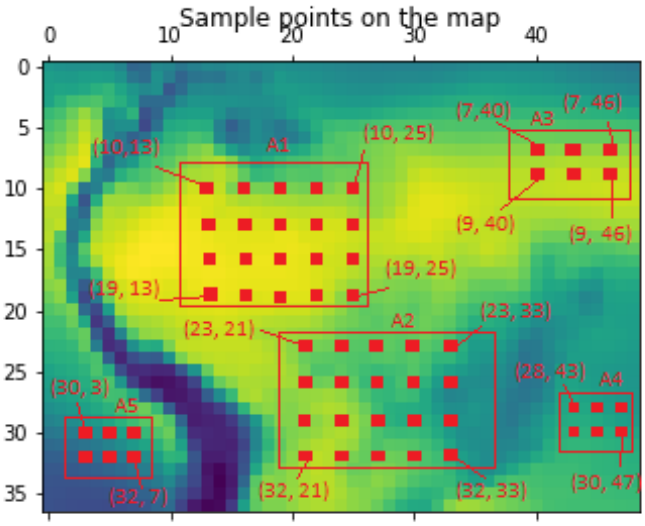
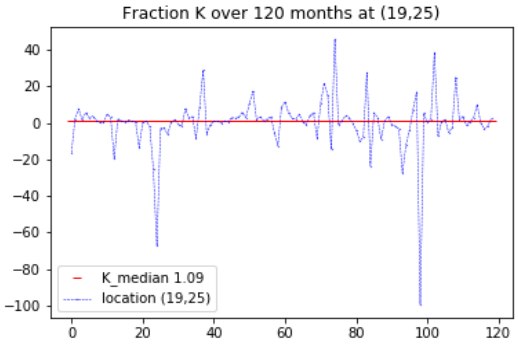


Figure 2.38: Sample points at different characteristic areas

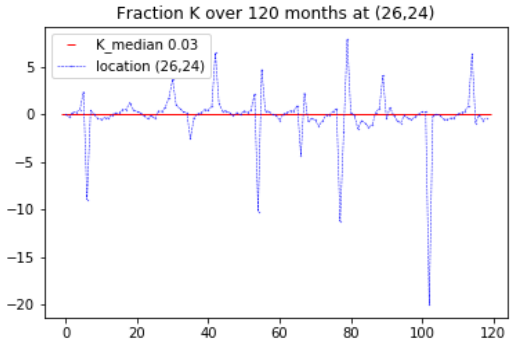
have unreasonable solutions in the results. So outliers of the calculated fraction K will be excluded. The calculation is performed on sample points, which are selected from characteristic regions (see in figure 2.38). In total, 58 sample points are chosen, 40 of which are chosen on land from two areas at northern part and southern part respectively, the remaining 18 points coming from areas in the ocean around the continent.

2.7.1. Results

Comparisons of results are performed both in temporal and spatial distribution. First, the temporal distribution is illustrated. As can be seen from the figure 2.39 below, taking two



(a) Calculated fraction K over 120 months at point (19, 25) in Area1



(b) Calculated fraction K over 120 months at point (26, 24) in Area2

Figure 2.39: Temporal variation of calculated fraction K at two selected sample points

sample point ((19,25) and (26,24)) as an example, the fraction K is varying with the time with changes in both magnitude and sign. Since error between modelled values and observations is inevitable, variation of K within a reasonable range is definite. However, there are quite a few calculated fractions K greatly deviating from the median. The bouncing pattern

of K over time suggests there is no reasonable approximate fraction range to be applied to one specified point.
Then comes the spatial distribution. Figure 2.40 describes the spatial distribution of the

A1	13	16	19	22	25		
10	-1.42	0.32	-0.03	0.46	-0.04	max	2.42
13	-2.39	-2.62	-1.57	0.90	-0.58	min	-2.62
16	1.94	2.42	-1.80	-1.29	-0.55	avg	-0.22
19	1.11	-0.37	1.32	-1.23	1.09	median	-0.21

(a) Calculated K median at Area1

A2	21	24	27	30	33		
23	0.96	0.08	0.59	-1.01	-0.07	max	0.96
26	0.03	0.03	0.14	-0.03	-0.14	min	-1.01
29	0.03	0.24	0.08	0.1	-0.05	avg	0.0595
32	-0.04	0.19	-0.16	0.18	0.04	median	0.035

(b) Calculated K median at Area2

A3	40	43	46		
7	0.29	0.01	0.32	max	0.32
9	-0.19	-0.44	0.04	min	-0.44
				avg	0.005
				median	0.025

(c) K median at Area3

A4	43	45	47		
28	0.37	0.62	0.86	max	0.86
30	0.61	0.8	0.63	min	0.37
				avg	0.64833
				median	0.625

(d) K median at Area4

A5	3	5	7		
30	-0.85	-0.76	-0.6	max	1.98
32	1.98	-0.61	-0.72	min	-0.85
				avg	-0.26
				median	-0.665

(e) K median at Area5

Figure 2.40: Backwardly calculated fraction K median at all five areas

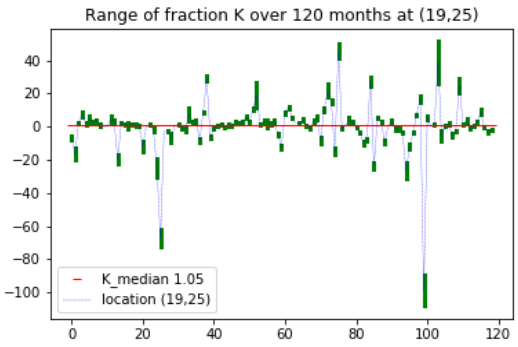
median of calculated fractions K over 120 months for five sample areas. Only for the small sample area Area4, which is in the ocean south-east to the study domain, the fraction shows kind of consistency for all internal sample points. However for other rest of areas, the fraction is varying significantly both in magnitude and sign from point to point within one area and averaged fraction differs in order of magnitude for different sample area.
It can be inferred preliminarily from above results that the wind magnitude to apply is much more complicated than expected. Not only it fluctuates over time but also it relates to the spatial configuration. However, it is not sufficiently proved because the previous backward calculation assumes both ends equal which leaves the possibility that error in fraction K is somehow amplified. Thereby to further verify above guesses, forward calculation on the data is conducted, which uses explicit discretization, presented in below formula, instead of implicit discretization. And tolerance of 50% deviation from the observation is applied in order to get the range of fraction. Data at time level $t = n$ is used to perform calculations. The final results are show in the figure 2.40.

$$A^{n+1} = \{-[(A^n \frac{u^n_{j+1,i} - u^n_{j,i}}{\Delta L} + u^n \frac{A^n_{j+1,i} - A^n_{j,i}}{\Delta L}) + (A^n_{j,i} \frac{v^n_{i+1,j} - v^n_{i,j}}{\Delta L} + v^n \frac{A^n_{i+1,j} - A^n_{i+1,j}}{\Delta L})]K + E - P\}\Delta t + A^n$$

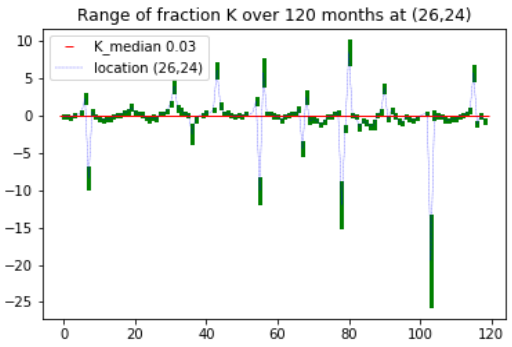
Comparing the below results (see in the next page) of the range of fraction K obtained from the forward calculation with the results of the value of fraction K from the backward calculation, it is obvious that the result K range is more or less the same thereby the variation in temporal distribution and the heterogeneity in spatial distribution of fraction K exists.

2.7.2. Result discussion

To sum up the findings in this section. After performing the unit check, it has been figured out that the wind speed applied in the previous modelling is $\frac{m}{d}$ on $r.h.s$ of the equation mismatching the $\frac{m}{mon}$ on $l.h.s$. However, the result gets no better after correction. Using wind data at one specified pressure level as the velocity scalar in advection term of the



(a) Calculated range of fraction K over 120 months at point (19, 25) in Area1



(b) Calculated range of fraction K over 120 months at point (26, 24) in Area2

Figure 2.41: Temporal variation of calculated range of fraction K at two selected sample points

A1	13	16	19	22	25		
10	-1.39	0.31	-0.08	0.44	-0.03	max	2.38
13	-2.47	-2.67	-1.64	0.94	-0.62	min	-2.67
16	1.94	2.38	-1.84	-1.26	-0.54	avg	-0.24
19	0.98	-0.50	1.39	-1.10	1.05	median	-0.29

(a) Forwardly calculated K median at Area1

A2	21	24	27	30	33		
23	0.92	0.03	0.64	-0.99	-0.05	max	0.92
26	0.03	0.03	0.13	0.00	-0.13	min	-0.99
29	0.04	0.26	0.09	0.12	-0.03	avg	0.06
32	-0.04	0.16	-0.16	0.14	0.03	median	0.03

(b) Forwardly calculated K median at Area2

A3	40	43	46		
7	0.26	-0.01	0.29	max	0.29
9	-0.21	-0.45	-0.03	min	-0.45
				avg	-0.025
				median	-0.02

(c) K median at Area3

A4	43	45	47		
28	0.38	0.69	0.88	max	0.88
30	0.53	0.78	0.67	min	0.38
				avg	0.655
				median	0.68

(d) K median at Area4

A5	3	5	7		
30	-0.84	-0.78	-0.61	max	1.93
32	1.93	-0.59	-0.74	min	-0.84
				avg	-0.27167
				median	-0.675

(e) K median at Area5

Figure 2.42: Forwardly calculated fraction K median at all five areas

original 1-layer-structured (soil water storage layer excluded) model oversimplifies the realistic atmospheric structure. For the purpose of having some clues on the approximate order of magnitude of the (range of) fraction to modify the wind input, backward and forward calculations were carried on. However, it has been found out that no consistent order of magnitude of the fraction can be applied, as it is time and space relevant, thereby applying one reduction factor on wind intensity for all grids is not a viable choice. It must be emphasized here that this section is more related to convergence problem instead of stability problem. Stability means error from any source *e.g* round-off, truncation, mistake and etc are not growing in the sequence of numerical procedure as the calculation stepping thereby the solution remains bounded. Consistency deals with the extent to which the finite difference discretization approximates the partial differential equation. Convergence addresses the degree to which the numerical solution to the numerical recipe approaches the true solution to the PDE given identical initial and boundary conditions. According to Lax's equivalence theorem, generally one can find a consistent, stable scheme is convergent [50]. It could be concluded from above discussions that, the model of original numerical scheme is inherently not convergent even though the solutions can be stabilized.

3

Conclusion and Recommendation

The purpose of this study was to investigate the numerical scheme problems of a coupled modelling of regional water balance and anthropogenic land cover change. The focus is put on regional water balance modelling. In previous chapter, various methods have been tested to see whether the model's solutions can be stabilized and made convergent to the observations. In this chapter, several conclusions have been drawn from result discussions in chapter 2. Besides, recommendations for future search are proposed which include limitations to consider, possible methods to use and feasible model architecture to apply.

3.1. Conclusion

Conclusion 1 In section 2.1, a study on impacts of adding diffusion term on stabilizing the results was done. Provided the diffusion coefficient is less than $2 \times 10^8 m^2/d$, the influence of it on numerical stability is negligible. Above that threshold, the modelled results become no longer reasonable. A diffusion term at reasonable range barely takes any effect on stability and accuracy. Besides, numerical diffusion is inherently introduced by the upwind scheme. Consequently, as far as our regional atmospheric modelling, adding diffusion term cannot address the numerical problem and there is no point in doing so.

Conclusion 2 It was investigated in section 2.2 that applying a smaller fraction of wind alleviates spurious oscillation to a certain degree. Smaller the fraction, more stable the solution, smoother the modelled moisture field. However, as the result of large temporal scale applied (1 month) and CFL condition requirement (information from a given grid or mesh element must propagate only to its immediate neighbours within timestep length), it seems that the original model is a paradox and inherently not convergent since moisture shall actually travel further than adjacent grid within one month.

Conclusion 3 Different from previous Heun's iteration method to implicitly solve the equation, a Python package of sparse matrix solver based on Gaussian-Seidel iteration method is applied in section 2.3 to solve the continuity equation semi-implicitly. This solver is quicker and has similar results as original solver in that the solution gets more stable and the modelled moisture field gets smoother given a smaller fraction of wind. However, it has the same problem stated above since the difference between the two only lies in the iteration algorithm but the underlying basics is the same.

Conclusion 4 Section 2.4 studies the sensitivity of source terms to see whether there is dormant errors in interrelationships between precipitation (P) and atmospheric moisture (A) and between evaporation (E) and precipitation (P). It is concluded that these parts pose no problems in the modelling though the parameterization might be further improved.

Conclusion 5 In section 2.5, the nonconstant as well as heterogeneity of the wind field is studied to learn more about numerical modelling. The numerical instability is not (primarily) attributed to the time-variance of wind but is relevant with the magnitude and perhaps spatial heterogeneity of wind. Besides, conservation of mass is compromised due to the direct cut off of exceedance, thus flux-conserving scheme is recommended for future research.

Conclusion 6 Multiple time steps with smaller timestep length is tested in section 2.6. Still, it is found out that a smaller magnitude means a smoother modelled moisture field, which could be explained by the solution is getting increasingly closer to the initial value input with magnitude of wind decreasing. However there is no meaning in achieving stability and smoothness in this way as it is not reflecting the true solution.

Conclusion 7 In previous coding, there is unit problem in data input. Though the quantity error in unit conversion is fixed, the modelled solution gets no improved. The original 1-layer-structured model which evaluates the advection at one specified pressure level could be too simple to represent the recycling transport network and reflect the reality. With the arbitrarily simplified setup, it seems not possible to have convergent solutions even though the solutions can be stabilized.

Conclusion 8 As we discussed before, the temporal scale and the wind component are the two keys to the numerical problem of the modelling. Even though endogenizing wind may solve the problem, the method of which is heuristically discussed in section 3.2.2, it is highly recommended to reconsider this regional water balance model from scratch since timestep length of one month is not preferred or appropriate in numerical models than in statistical simple bucket models. The spatial scale ($1° \times 1°$) is fine. But in terms of temporal scale, from minute-based to hourly-based timestep length is more commonly used in an atmospheric model at different levels of sophistication to control numerical stability, which is briefly introduced in recommendation section 3.2.1. To stabilize the solution, the stability limit has to be met, normally for Courant condition for 2-D case: $C = \frac{u_x \Delta t}{\Delta x} + \frac{u_y \Delta t}{\Delta y} \leq C_{max}$, it means the magnitude of velocity, the timestep length and length interval should satisfy this condition. For explicit scheme, $C_{max} \leq 1$. Although the implicit scheme is unconditionally stable, it means neither unconditionally accurate nor arbitrarily monstrous timestep length to be applied as a result of factors like integration error tolerance, boundary condition influence, solvability/unstability of non-linearity and etc. It is not uncommon to have these kind of confusion. Apart from stability, convergence is also quite essential. The original model is currently not capable of achieving both targets, thus it may ask for further reconsideration of the architecture of the modelling. Yet I am not denying the possibility of a feasible finite difference physically-based numerical atmospheric model evaluated at large temporal scale, but it definitely involves more research, more devotion and more in-depth knowledge in numerical modelling and computational fluid dynamics.

3.2. Recommendation

3.2.1. Atmospheric model position

It is of vital significance to choose the positioning of the atmospheric model depending on the purposes. According to the book "A climate modelling Primer" written by Kendal McGuffie [32], existing climate models are usually categorized into following four fundamental types:

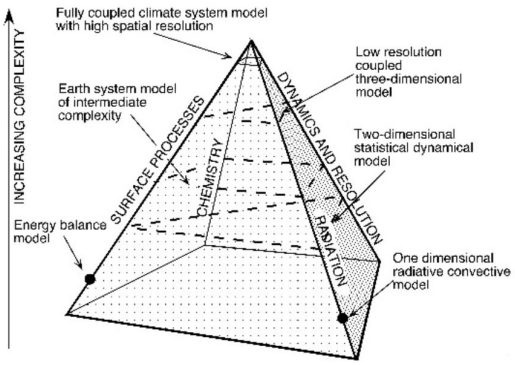
1. **Energy balance models (EBMs).** EBMs are 0- or 1-D models predicting the surface temperature based on energy balance [33].

2. **One-dimensional models.** For instance, a radiative–convective (RC) model computes globally-averaged temperature profile through explicitly modeling radiative processes with a predetermined lapse rate [29]. A single column model (SCM) is a single column extracted from a 3-D model without modelling horizontal energy transfers [40]. These 1-D models have no horizontal degrees of freedom and hence focus only on vertical processes.

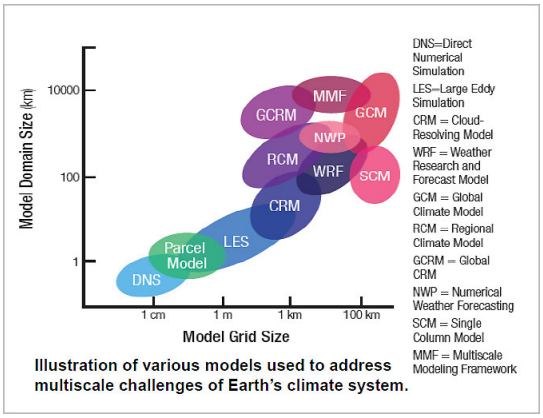
3. **Dimensionally constrained models.** This model generally represents either x and y dimensions or z dimension plus one horizontal dimension, the latter of which is initially more ordinary. For example, a statistical dynamical (SD) model combines the latitudinal dimension of a EMB with the vertical dimension of a RC model [30]. A dimensionally constrained model is usually considered as a preliminary attempt at climate modeling with intermediate complexity (EMIC) [10]. Rapid increase of computing capacity enables more people to run full three-dimensional resolution global circulation models (GCM). However, modern dimensionally constrained models have been born anew in recent years, especially for applications that involve human system and socio-economic change since they simplify physical dimensions specifically to incorporate human systems [32].

4. **Global circulation models (GCMs).** GCM are 3-D coupled ocean– atmosphere models. Normally these more advanced but computer-resources-consuming models simulate utmost processes to get a three dimensional picture of climate evolution with time. It is constructed by horizontal columns and vertical layers.

Figure 3.1a shows the hierarchy pyramid of climate models. As the pyramid is ascending, more progresses are integrated in the modelling hence complexity is increasing. Consequently, models climbing on top of the pyramid generally have finer spatial and temporal resolution which meanwhile means more computing power and more high resolution data for calibration and validation. In previous research on predicting the deforestation in Amazon basin, a coupled model of regional water balance and landuse change is developed, of which the atmospheric model part is problematic. This 2-D atmospheric model should be classified as the dimensionally constrained model, which is on around one-third of the height of the climate modeling pyramid. The model has only two horizontal dimensions but omits vertical exchanges in the column. Besides, the time window of the atmospheric model (one month) is pretty large, compared to GCM which is evaluated by the timestep size of minute to hour. Reducing the complexity is not only to cut consumption of computing resources because running a comprehensive simulation of tens of



(a) The climate modeling pyramid from the book "A Climate Modelling Primer"



(b) Climate model at various spatial scales

Figure 3.1: Climate model pyramid and climate model at multi-scale

years is very expensive, but also to handily incorporate anthropogenic land cover change since its consequence manifests at large time scale. Figure 3.1b illustrates climate models with various spatial scale. As the grid size is $1° \times 1°$ and the area of domain of interest is $4140km \times 5482km$, in terms of spatial scale, the model should be classified as regional climate model (RCM). However it is simpler than climate models since it only tries to grasp the essential characteristics — the water balance in the study region at large temporal and spatial scale.
Primitive equations are fundamental relationships governing atmospheric motion applied in most atmospheric models. They contains five main sets of balance equations [39]:

1. **Equations of motion (cons. of momentum):** How the zonal and meridional wind change with time. It depends on latitude, pressure gradient force, and friction. Detailed information can be found in next subsection. $\frac{DV}{Dt} = -2\Omega \times V - \rho^{-1}\nabla p + g + F$
2. **Continuity equation (cons. of mass):** $\frac{D\rho}{dt} + \rho\nabla \cdot v = 0$
3. **First law of thermodynamics (cons. of energy):** $\frac{DI}{Dt} = -\rho\frac{d\rho^{-1}}{dt} + Q$
4. **Moisture equation (cons. of moisture):** $\frac{D\rho}{Dt} = -p\nabla \cdot v + C - E$
5. **Equation of state:** It relates pressure, temperature and density $p = \rho RT$

Basically, all physical processes in atmospheric transporting are described by the above primitive equations. An ordinary three-dimensional global climate model (GCM) solves numerically these equations to simulate atmospheric circulation with finite time steps and grid boxes. It is evaluated in multiple-layer structure at finer temporal scale, which may even take months to run long time series simulations. "Essentially, all models are wrong, but some are useful" — Since models are simplification of reality with a series of assumptions, they can only somehow reflect the reality but it cannot perfectly represent the real world [8]. Besides, it is not uncommon to see totally different results from two advanced models. Coupled modelling asks for quick and simple solutions, meaning decrease in model complexity. Hence trade-offs between complexity, efficiency and accuracy have to be made. Nevertheless, the model should be backed up by solid physical foundations.

3.2.2. Endogenize wind

Another ultimate research question of this topic is to make every component (P, E, U, V) in the equation endogenous. Currently we already have linked the moisture A with the precipitation **P** and evaporation E in the continuity equation as following:

$$\frac{\partial A}{\partial t} = -(u\frac{\partial A}{\partial x} + A\frac{\partial u}{\partial x} + v\frac{\partial A}{\partial y} + A\frac{\partial v}{\partial y}) + E(A) - P(A)$$

, where zonal and meridional wind u and v *resp.* are taken straightforwardly from external ERA-interim database. However, it is viable to link wind component u and v with moisture A. Possible methods are recommended below:

First, link pressure P with moisture A. Pressure tendency is how pressure changes with time at a particular point. The pressure tendency at an altitude of z is due to the net mass change of the air above that altitude [18].

$$\frac{\partial p}{\partial t}\Big|_z = -g\int_z^\infty \frac{\partial(\rho u)}{\partial x}dz - g\int_z^\infty \frac{\partial(\rho v)}{\partial y}dz$$

$$d_P = -\rho g d_z \Rightarrow P = g\int_1^2 \rho d_z$$

Pressure in a column is due to the weight of the air above. Lower the mass of the air above, lower the pressure. Above equation states that the pressure at any level is directly proportional to the density of the atmospheric layer of thickness dz, or in other words, the weight of the slab of atmosphere of thickness dz. Taken level 1 as sea level and level 2 as the top of the atmosphere, the equation then simply states that pressure at sea level is just a function of the total weight of the atmospheric column. Based on this conclusion, we can then establish the conversion relationship between Pressure P and Moisture A by analyzing historic data statistically.

Second, link wind u and v with pressure P. It is based on the conservation of momentum. The acceleration of the wind ($l.h.s$) is equal to the imbalance between the body forces ($r.h.s$), including Coriolis, pressure gradient, gravity and friction:

$$\frac{DV}{Dt} = -2\Omega \times V - \frac{1}{\rho}\nabla p + g + F_r$$

Hence for horizontal west wind u, we have:

$$\frac{\partial u}{\partial t} = -(\underbrace{\frac{u\partial u}{\partial x} + \frac{v\partial u}{\partial y}}_{(1)} + \frac{w\partial u}{\partial z}) - \underbrace{\frac{1}{\rho}\frac{\partial P}{\partial x}}_{(2)} - fv - Friction$$

, where (1) is the advection of momentum, (2) is the pressure gradient and fv is rotation effect (due to Coriolis force), Frictional force term is relevant with speed of the air parcel as well as the roughness of the earth surface which could also be associated with surface vegetation type. Besides, the concept of wind consisting of a time-averaged component plus a fluctuation component together with the idea of mean advective flux component plus a parameterized turbulent flux component is also worth studying [37].

$$u(t) = < u > + u'(t)$$

, where $< u >$ is Time-averaged component composed of mean advective flux $< u >< n_i >$ and turbulent flux $< u' n_i' >$, $u'(t)$ is fluctuating component.

3.2.3. Possible model architecture

This section talks about some heuristic methods applied in other moisture system modelling as recommendation for future model architecture.

There are many other atmospheric motion models of intermediate complexity under simplified architecture. Compared to sophisticated GCM, they are usually one or several order of magnitude faster. However they are still to a certain extend computational expensive, thus not very applicable for our modelling motivation. We are looking for something simpler. Hence the trade-offs between computing efficiency and accuracy should be given deliberate considerations.

For the issue of the closure of water balance, WAM-2layers sheds some light on it which may serves as inspiration. The underlying water balance in WAM-2layers model is [42]:

$$\frac{\partial S_k}{\partial t} = \frac{\partial (S_k u)}{\partial x} + \frac{\partial (S_k v)}{\partial y} + E_k - P_k + \xi_k \pm F_v \left[L^3 T^{-1}\right]$$

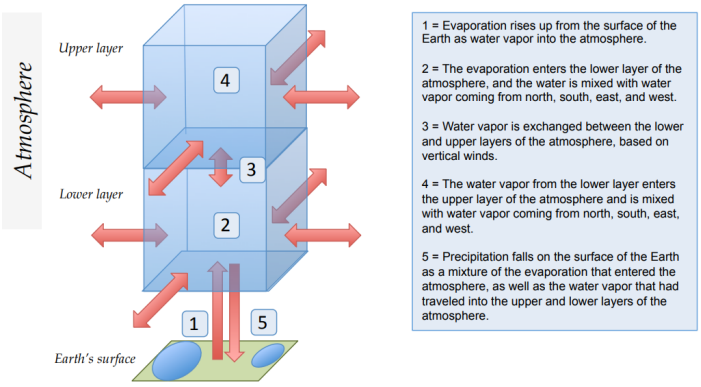


Figure 3.2: Overview of a Single Column of Atmosphere in the WAM-2layers

,where S_k is the moisture at layer k, E_k is evaporation entering layer k, P_k is evaporation removed from layer k, change in moisture $\frac{\Delta(Su)}{\Delta x}$ (take x dimension as an example) because of horizontal transport is modelled as the flux over western and eastern boundary of grid cell $F^-_{k,x} - F^+_{k,x}$, the total flux at layer k is calculated as vertical integral of flux from bottom pressure level to ceiling pressure level of the layer k. The model is divided as bottom and top two layers and the division pressure level is 81,283 Pa since as stated it best captures the division between sheared wind systems [42, 43]. ξ is residual term, F_v is vertical moisture transport flux between layers which serves as a closure term of the water balance relating to the residual term ξ. The initial time step is set as 0.25h, though a larger time step can be applied. A overview figure of single column [26] is attached above for further illustration.

Therefore it is recommended in that an atmospheric model of flux-conserving scheme evaluated at small timestep length (perhaps hourly) should be considered to be developed, which probably also makes allowance for making wind component (partially) endogenous. If monthly time step still has to be applied, then probably instead of using numerical modelling of the governing equation with finite difference method, it might be a solution to consider the problem from another perspective — a distributed-network water balance modelling which consists of separated nodes and routing between the nodes that relates to wind-associated moisture motion.

Bibliography

[1] S Abbasbandy, T Allah Viranloo, Oscar Lopez-Pouso, and Juan J Nieto. Numerical methods forfuzzy differential inclusions. *Computers & mathematics with applications*, 48(10-11):1633–1641, 2004.

[2] Sharma Ajar. Water-food nexus: modelling the inter-linkages between food prices, deforestation, and the water balance in northern south america. 2017.

[3] Larry C Andrews. *Elementary partial differential equations with boundary value problems*. Academic Press, 1986.

[4] Charles E Baukal Jr, Vladimir Gershtein, and Xianming Jimmy Li. *Computational fluid dynamics in industrial combustion*. CRC press, 2000.

[5] Gerald D Bell, Michael S Halpert, Chester F Ropelewski, Vernon E Kousky, Arthur V Douglas, Russell C Schnell, and Melvyn E Gelman. Climate assessment for 1998. *Bulletin of the American Meteorological Society*, 80(5s):S1–S48, 1999.

[6] R Byron Bird. Transport phenomena. *Applied Mechanics Reviews*, 55(1):R1–R4, 2002.

[7] Arthur Bloch. *Murphy's law*. Penguin, 2003.

[8] George EP Box and Norman R Draper. *Empirical model-building and response surfaces*. John Wiley & Sons, 1987.

[9] Víctor Chavarrías, Guglielmo Stecca, and Astrid Blom. Ill-posedness in modeling mixed sediment river morphodynamics. *Advances in Water Resources*, 114:219–235, 2018.

[10] Martin Claussen, L Mysak, A Weaver, Michel Crucifix, Thierry Fichefet, M-F Loutre, Shlomo Weber, Joseph Alcamo, Vladimir Alexeev, André Berger, et al. Earth system models of intermediate complexity: closing the gap in the spectrum of climate system models. *Climate dynamics*, 18(7):579–586, 2002.

[11] Richard Courant, Kurt Friedrichs, and Hans Lewy. On the partial difference equations of mathematical physics. *IBM journal of Research and Development*, 11(2):215–234, 1967.

[12] Dick P Dee, SM Uppala, AJ Simmons, Paul Berrisford, P Poli, S Kobayashi, U Andrae, MA Balmaseda, G Balsamo, d P Bauer, et al. The era-interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the royal meteorological society*, 137(656):553–597, 2011.

[13] Daniel R Drew, Janet F Barlow, and Siân E Lane. Observations of wind speed profiles over greater london, uk, using a doppler lidar. *Journal of Wind Engineering and Industrial Aerodynamics*, 121:98–105, 2013.

[14] Leticia Duboc, David S Rosenblum, and Tony Wicks. A framework for modelling and analysis of software systems scalability. In *Proceedings of the 28th international conference on Software engineering*, pages 949–952. ACM, 2006.

[15] CP Dullemond. Lecture numerical fluid dynamics.

[16] Elfatih AB Eltahir and Rafael L Bras. Precipitation recycling in the amazon basin. *Quarterly Journal of the Royal Meteorological Society*, 120(518):861–880, 1994.

[17] Peter M Finocchio and Sharanya J Majumdar. A statistical perspective on wind profiles and vertical wind shear in tropical cyclone environments of the northern hemisphere. *Monthly Weather Review*, 145(1):361–378, 2017.

[18] JM Fritsch and CF Chappell. Numerical prediction of convectively driven mesoscale pressure systems. part ii. mesoscale model. *Journal of the Atmospheric Sciences*, 37(8): 1734–1762, 1980.

[19] John Roy Garratt. The atmospheric boundary layer. *Earth-Science Reviews*, 37(1-2): 89–134, 1994.

[20] K Haustein, C Pérez, JM Baldasano, O Jorba, S Basart, RL Miller, Z Janjic, T Black, S Nickovic, Martin C Todd, et al. Atmospheric dust modeling from meso to global scales with the online nmmb/bsc-dust model-part 2: Experimental campaigns in northern africa. *Atmospheric Chemistry and Physics*, 2012.

[21] Mark D Hill. What is scalability? *ACM SIGARCH Computer Architecture News*, 18(4): 18–21, 1990.

[22] Roger A Horn, Roger A Horn, and Charles R Johnson. *Matrix analysis*. Cambridge university press, 1990.

[23] Daniel J Jacob, Robert M Yantosca, Isabelle Bey, Jennifer A Logan, Brendan D Field, Arlene M Fiore, Qinbin Li, Honguy Y Liu, Loretta J Mickley, and Martin G Schultz. Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation. *Journal of Geophysical Research: Atmospheres*, 106(D19):23073–23095, 2001.

[24] Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001–. URL `http://www.scipy.org/`. [Online; accessed <today>].

[25] IB Kalambi. A comparison of three iterative methods for the solution of linear equations. *Journal of Applied Sciences and Environmental Management*, 12(4), 2008.

[26] Patrick W Keys. *The precipitationshed: concepts, methods, and applications*. PhD thesis, Stockholm Resilience Centre, Stockholm University, 2016.

[27] Peter D Lax and Robert D Richtmyer. Survey of the stability of linear finite difference equations. *Communications on pure and applied mathematics*, 9(2):267–293, 1956.

[28] Randall J LeVeque. *Finite volume methods for hyperbolic problems*, volume 31. Cambridge university press, 2002.

[29] RM MacKay and MAK Khalil. Theory and development of a one dimensional time dependent radiative convective climate model. *Chemosphere*, 22(3-4):383–417, 1991.

[30] RM MacKay and MAK Khalil. Climate simulations using the gcrc 2-d zonally averaged statistical dynamical climate model. *Chemosphere*, 29(12):2651–2683, 1994.

[31] Yadvinder Malhi, Luiz EOC Aragão, David Galbraith, Chris Huntingford, Rosie Fisher, Przemyslaw Zelazowski, Stephen Sitch, Carol McSweeney, and Patrick Meir. Exploring the likelihood and mechanism of a climate-change-induced dieback of the amazon rainforest. *Proceedings of the National Academy of Sciences*, 106(49):20610–20615, 2009.

[32] Kendal McGuffie and Ann Henderson-Sellers. *A climate modelling primer*. John Wiley & Sons, 2005.

[33] Gerald R North, Robert F Cahalan, and James A Coakley Jr. Energy balance climate models. *Reviews of Geophysics*, 19(1):91–121, 1981.

[34] Gerard H Roe. Orographic precipitation. *Annu. Rev. Earth Planet. Sci.*, 33:645–671, 2005.

[35] Raoni Aquino Silva de Santana, Cléo Quaresma Dias-Júnior, Roseilson Souza do Vale, Júlio Tóta, and David Roy Fitzjarrald. Observing and modeling the vertical wind profile at multiple sites in and above the amazon rain forest canopy. *Advances in Meteorology*, 2017, 2017.

[36] Benjamin Schaaf and Frauke Feser. Is there added value of convection-permitting regional climate model simulations for storms over the german bight and northern germany? *Meteorology Hydrology and Water Management. Research and Operational Applications*, 6, 2018.

[37] John H Seinfeld and Spyros N Pandis. *Atmospheric chemistry and physics: from air pollution to climate change*. John Wiley & Sons, 2012.

[38] ZR Shu, QS Li, YC He, and PW Chan. Vertical wind profiles for typhoon, monsoon and thunderstorm winds. *Journal of Wind Engineering and Industrial Aerodynamics*, 168: 190–199, 2017.

[39] Joseph Smagorinsky. General circulation experiments with the primitive equations: I. the basic experiment. *Monthly weather review*, 91(3):99–164, 1963.

[40] Adam H Sobel and Christopher S Bretherton. Modeling tropical precipitation in a single column. *Journal of climate*, 13(24):4378–4392, 2000.

[41] Larry L Stowe, Alexander M Ignatov, and Ramdas R Singh. Development, validation, and potential enhancements to the second-generation operational aerosol product at the national environmental satellite, data, and information service of the national oceanic and atmospheric administration. *Journal of Geophysical Research: Atmospheres*, 102(D14):16923–16934, 1997.

[42] Rudi Johannes Van der Ent. A new view on the hydrological cycle over continents. 2014.

[43] RV van der Ent, OA Tuinenburg, Hans Richard Knoche, Harald Kunstmann, and HHG Savenije. Should we use a simple or complex model for moisture recycling and atmospheric moisture tracking? 2013.

[44] Charles F Van Loan. Matrix computations (johns hopkins studies in mathematical sciences), 1996.

[45] HK Versteeg. Malalasekera (1995) an introduction to computational fluid dynamics-the finite volume method.

[46] RF Warming and BJ Hyett. The modified equation approach to the stability and accuracy analysis of finite-difference methods. *Journal of computational physics*, 14(2): 159–179, 1974.

[47] R Wilhelmson, M Folk, M Ramamurthy, B Schatz, N Yeager, R Crutcher, and M Winslett. Horizon: A digital library project for earth and space data serving the public.

[48] DC Zemp, C-F Schleussner, HMJ Barbosa, RJ Van der Ent, Jonathan F Donges, J Heinke, G Sampaio, and A Rammig. On the importance of cascading moisture recycling in south america. *Atmospheric Chemistry and Physics*, 14(23):13337–13359, 2014.

[49] Delphine Clara Zemp, Carl-Friedrich Schleussner, Henrique MJ Barbosa, Marina Hirota, Vincent Montade, Gilvan Sampaio, Arie Staal, Lan Wang-Erlandsson, and Anja Rammig. Self-amplified amazon forest loss due to vegetation-atmosphere feedbacks. *Nature communications*, 8:14681, 2017.

[50] M. Zijlema. Computational modelling flow and transport. 2015.