博士論文

Reconstruction of Historical Weather with Data Assimilation Using Old Diaries

(古日記データ同化による歴史天候の復元)

Neluwala Gamethige Panduka Bandara Neluwala

ネルワラ ガメティゲ パンデュカ バンダラ ネルワラ

Reconstruction of Historical Weather with Data Assimilation Using Old Diaries

(古日記データ同化による歴史天候の復元)

Neluwala Gamethige Panduka Bandara Neluwala

ネルワラ ガメティゲ パンデュカ バンダラ ネルワラ

A dissertation submitted in partial completion
of the requirements for the degree of

Doctor of Philosophy

Department of Civil Engineering
Graduate School of Engineering
The University of Tokyo
東京大学大学院 工学系研究科 社会基盤学専攻

August 2018

Acknowledgements

First and foremost, I would like to express my deepest appreciation to my supervisor Assoc. Professor Kei Yoshimura for accepting me and guiding me to success. I consider it is an excellent opportunity to be a student of you. Your, guidance, leadership and valuable advice helped me to improve in academic and also in personal life within the last three years.

My special thanks go to my co-supervisors Prof. T. Oki, Assoc. Prof. J. Hirano, Prof. A. Kawasaki and Assoc. Prof. Y. Sekimoto for giving their valuable comments and suggestions to improve my research.

I am much grateful for Prof. Toshio Koike for the support and guidance in applying for PhD and all the support from the River lab members during my master's program and Prof. Weerakoon, Prof. Pathirana and all the staff members in the University of Peradeniya, Sri Lanka for guiding me to study in Japan.

I would like to appreciate the valuable comments from Assoc. Prof. Hyungjun Kim, Assoc. Prof. Yukiko Hirabayashi, Assoc. Prof. Dai Yamazaki. I should especially thank the support of Dr. M. Ichino and Dr. K. Masuda, Assoc. Prof. J. Hirano for giving valuable advice and spending their time for research discussions.

I like to specially mention Dr. Salem Ibrahim for the encouragement and as my closet friend in Oki lab, and I should mention Dr. Nobuyuki Utsumi the nicest person I met in Oki lab, Ms. Khin Thet Swe for your unique ideas and interesting time we had, Ms. Ayako Kurasawa for your unconditional affection to all the students, Ms. Azariah Amador for keeping the joy and love in the 5^{th} floor.

The sweet memories in Oki Laboratory will be in my mind forever. My special thanks to Ms. Misako Hatono for helping all the students without any limitation, Mr. Hasnain Aslam for your diligent work, Ms. Xiaoya Zhang for sharing the pizza, Ms. Isabel for showing the confidence, Mr. Ishitsuka and all the Students in Oki Lab for being with me all the time. Moreover, the 5^{th}-floor room would be memorable forever.

I should thank Ms. Inna Syafarina who was with me always listening to my long stories and sharing the experience. Dr. Hibino for your kind support to all the student, Dr. Onuma for sharing the long nights in the lab. Dr. Nitta for managing all the servers. Ms. Sunao Noguchi for the support as our secretary.

I want to especially thanks Ms. Ayoyama Junko for your kind support and motivation as a Staff member in the Civil Engineering Department Office. Ms. Masumi Tonegawa and all the staff in Foreign student office (FSO), Also, I should mention kind support from International student center and health center of the university when I seek help. I would like to acknowledge the funding received from the Japanese Ministry of Education, Culture, Sports, Science, and Technology (MEXT) to conduct my doctoral research. I should also thanks Minimini house agent for support to find three houses during my stay.

Moreover, Ms. Ayuko Akaike, UT Civil Alumina of our department for being with us in all events. All the members of The International Student Association (ISACE) and on the other hand, I should remember all of my friends and relatives in Sri Lanka for keeping contact with me all the time even though I am away for a longer time.

I like to specially mention Ven.Yalagamuwe Dhammissara Anunayaka Thero for organizing valuable activities in the Sakamuni International Buddhist Center which helped to keep my mind relax and free.

My deepest appreciation always goes to all the members of Sri Lankan Student Association and Sri Lankan community in Japan for helping me to enjoy my life in Japan with fullest potential by being as members of a single family while taking care of each other's in every challenging situation.

I want to express my gratitude for my Mother, Father and brother for your unconditional love and affection which always keep me happy and motivate.

Finally, I would like to convey my heartfelt gratitude and appreciation for each and every citizen in Japan, who helped me to add such an unforgettable experience to my life and invaluable ethics and values to my personality.

Abstract

Climate can control not only the personal lifestyle but also other living beings. It is crucial to investigate historical climate to understand the current and future climates. Information about daily weather can give a better understanding of past life on earth. Long-term weather influences crop calendar as well as the development of civilizations. Unfortunately, existing, reconstructed daily weather data are limited after the 1850s due to the availability of instrumental data. The climate data prior to that are derived from proxy materials (e.g., tree-ring width, ice core isotopes, etc.) which are either in an annual or decadal scale. However, there are many historical documents which contain information about whether such as personal diaries. In Japan, around 20 diaries in average during the 16^{th} - 19^{th} centuries have been collected and converted into a digitized form. As such, diary data exist in many other countries. This study aims to reconstruct historical daily weather during the 18^{th} and 19^{th} centuries using personal daily diaries which have qualitative weather descriptions such as 'cloudy' or 'sunny' by incorporating this information to a Climate model using a data assimilation scheme.

To reproduce climate, a numerical weather prediction model can be used. In this study, Global Spectral Model by Scripps Experimental Climate Prediction Center based on Global Seasonal forecast system in National Centres for Environmental Prediction's is used. This model was used as the operational forecast model there until 2004 and as the basis for several model development projects. However, these models are not perfect, and the results can be improved if observations available in the past can be incorporated to model results. Data Assimilation is useful to get the best estimate from a model and observations considering the model errors and the observation uncertainties. There are attempts to reconstruct past climate using other proxies such as Tree ring, Coral, Ice core, and sediment. The merit of these climate reconstructions is they cover several thousand years sometimes beyond last millennium. However, limitation of these proxy reconstructions is they are either annual or seasonal and not available in all the regions. On the other hand, the personal diary information provides more frequent information allowing to reconstruct climate in high resolution using online data assimilation techniques. In this study, we used Local Ensemble Kalman Filter which uses an ensemble forecast to calculate error covariances. Moreover it has localization ability that can-do assimilation grid wise to each state vector parallelly considering all the observation in the local area which makes the computation more efficient. Chapter 2 discuss the assimilation system in detail and characteristics of diary data. Japan has a digitized database of old personal diaries from the 17^{th} century. There are around 20 diaries in the 19^{th} century. Even though

personal diaries have valuable information about daily weather, they are limited to qualitative information such as descriptions like ‘sunny’ and ‘cloudy’ and it was a challenge to convert them to usable quantitative format to be used in the climate model. This qualitative information was converted to probabilistic representative quantitative values of Total column cloud content (TCC) and downward shortwave radiation (SR).

Chapter 3 further investigated the possibility of assimilating uncertain weather information. It was not clear about proper model settings and sensitivity of the number of observation and observation uncertainty for uncertain weather assimilation up to now. This chapter found solutions to them with several experiments. When TCC data is assimilated, the correlation in average over Japan improved to 0.47 from -0.01. In particular, the correlation of TCC improved to 0.64 from -0.13 at Choshi station. There are no significant contributions to other variables (i.e. correlation change in: Temperature 0.3 to 0.2, Precipitation -0.95 to 0.1 and Pressure 0.18 to 0.3). Experiments with a different number of observation stations showed improvement in the correlation coefficient and RMSE around the observations sites even with 18 number of stations. This indicates even the fewer number of weather records are available local improvement can be achieved over those regions. Further, the simulation using data from 418 stations improved the results of not only the exact areas near the stations but also in remote areas. For instance, correlation coefficients of TCC, Temperature, Precipitation and Specific humidity in a non-assimilated site (i.e. Choshi station) improved from -0.13 to 0.38 ,0.30 to 0.57, -0.10 to 0.53, -0.13 to 0.61 respectively. Simulations with different observations uncertainties were carried out to investigate the sensitivity to observation uncertainty and found that if a small observation uncertainty is given, assimilation scheme neglects the observations because ensemble spread is away from the observations. This was clear in results where an observation error with 1% achieved only 0.17 correlation while observations with an error of 50% correlation improved correlation coefficient to 0.42. Impact from the initial conditions was analyzed by doing simulations with perturbated simulations instead of initial conditions created from time shift method as in previous experiments. The correlation was better in time shift method (e.g., correlation decreased to 0.45 with perturbated initial condition method in comparisons to 0.64 in Choshi station using time shift method and RMSE increased to 39.9% in comparison to 32.6% in time shift method). Thus, it was decided to use time shift method to create ensembles.

Real weather diary data is much different from the regular observations or synthesized observations used in Chapter 3 because they do not have any numbers. Lack of boundary data such as SST and Sea-ice fraction are other main challenges for simulation of forecast model

during the 19th century. Currently, there is not any study which has overcome these challenges to assimilate qualitative description data. Hence in Chapter 4, we evaluated the impact from poor boundary condition. Assimilation system's skill was found to reduce mainly in precipitation when low-frequency Sea surface temperature and Sea-ice fraction data are used. Correlation of all station average in Precipitation in 1995 April reduced from 0.58 to 0.32 even though correlation in SR and TCC changed only slightly (i.e. 0.79 to 0.81 and 0.76 to 0.65 respectively). Another challenge is the sensitivity of assimilation time, diary data information is mostly available in daily scale and impact on assimilating at particular time step has not been investigated earlier. Separate experiments showed that assimilation results in morning and evening have only a slight difference. In spring the correlation coefficient of the average of all the station's changes from 0.54 to 0.43 in Precipitation, 0.66 to 0.72 in SR, 0.66 to 0.73 in TCC and 0.81 to 0.8 in Temperature when assimilation time changed to 3 pm from 9 am. Further, impact to model performance by assimilating only three weather classes data was evaluated in comparison to assimilating TCC from JMA observations with added 30% uncertainty, and it was found that the model could still capture the temporal variation even though correlation of TCC reduced to 0.47 to 0.57 and 0.32 from 0.54 in precipitation in comparison to direct TCC observation assimilation with 30% uncertainty at Choshi station.

Chapter 5 evaluate the skill of the model in assimilating document weather data. The main limitation was the lack of instrumental data in the past. Hence an alternative approach was followed by assimilating weather classes data derived from recent description data. All these experiments were carried out with real data keeping the consistency with 19th Century data quality. This is the first study to carry out such realistic experiments to investigate the performance of assimilating weather class data into a climate model. Several simulations were done in the recent period where observations data available for validation. Twentieth-century weather classes, data derived from JMA descriptions was utilized. SR assimilation could improve the correlation of TCC average in all the stations from 0.19 to 0.68 in spring while reducing RMSE by 8 %. Improvements in other seasons and fields such as precipitation could be achieved as well. Further, we investigated opportunities to improve the accuracy of the model by incorporating other information such as the absence of precipitation and found the correlation of precipitation in all the station average could be improved to 0.67 from 0.45 in spring. Monthly anomaly values over 1995-1999 showed good correlation in precipitation, TCC, and SR. By analyzing pressure fields, it could be shown that the model could capture the synoptic scale weather patterns such as extratropical cyclones (ExT cyclones). Bootstrap experiments were done using only half observations to check model performance when some

diaries are absence. Even though the model performance was reduced to some extent satisfying correlation could be achieved. Correlation of all the stations average in TCC was 0.57, in SR was 0.72 and in precipitation was 0.45.

Chapter 6 assimilate weather information from weather diaries in 19th century into the climate model for the first time in the historical data assimilation field using the settings and parameters identified from Chapter 2 and Chapter 3 with weather classes from other studies for 1830s and weather classes directly from derived from Historical Weather Data Base for 1860s as explained in Chapter 2. The model could capture weather types such as 'cloudy' and 'sunny' after data assimilation in 1830s similar to conditions of weather classes. Similar skill was observed in 1860s experiments. Due to the lack of instrumental data for daily comparison, monthly temperature from early instruments in Yokohama was used to check the model performance. The correlation coefficient in temperature without data assimilation and with assimilation were 0.96 and 0.94 respectively which are evenly high because the model can capture the seasonality. The model could produce ExT cyclones similar to 1995 when several diaries indicate rainy (weather class 3) during the spring period. By investigating precipitation anomaly from 1861 to 1864, 1861 May shown to be wet (19.0mm/month higher) and 1864 relatively dry year (13.5 mm/month less).

The last chapter makes the conclusions and discusses the recommendations for future studies. In this study, weather categories from qualitative description data could be converted to weather classes and assimilate into a general circulation model using a data assimilation scheme. The results showed that assimilation of weather classes using SR improved the correlation of non-assimilated variables as well, and it was revealed that the resulted atmospheric distributions could capture the actual synoptic weather events. To further expand this research, more information from diaries such as wind direction and snow information can be utilized. This study used only the diary data in Japan. In future, this methodology can be applied globally when more digitized diaries are available in different regions.

Publication Lists

Presentations on Conferences/Symposium (as a participant)

Panduka Neluwala, Kei Yoshimura, Junpei Hirano, Mika Ichino, 2017 Assimilating Various Fields from Historical Documents: Idealized Experiments, 水文・水資源学会 2017 年度総会・研究発表会, 北見, 2017/9/19-21 (Presentation)

P. Neluwala, K. Yoshimura, K. Toride, J. Hirano, M. Ichino, A. Okazaki Reconstruction of Historical Weather by Assimilating Old Weather Diary Data, PP31A-1264, AGU Fall Meeting 2017, New Orleans, 2017/12/11-15 (Poster).

Panduka Neluwala, Kei Yoshimura, Sensitivity analysis of historical weather documents assimilation for reconstructing past climate, fourth International Conference on Hydrology delivers Earth System Science to Society (HESSS4), Tokyo, 2017/05-16-19 (Presentation)

Panduka Neluwala, Kinya Toride, and Kei Yoshimura, Assimilation Experiment of Weather Diary Data Using an LETKF System, the 3rd RIKEN International Symposium on Data Assimilation, Kobe, 2017/2/28-2017/3/2 (Presentation)

Panduka Neluwala, Kei Yoshimura (2016) Reconstruction of historical weather patterns by data assimilation of old diaries and general circulation model World Archaeological Conference -8 (WAC8), Kyoto 2016/08/28-/09/02 (Presentation)

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1. Background

Information on past climate can help to study the influence of climate on agriculture, health and economy (Buntgen et al., 2011). To understand the changing atmospheric circulation due to the insolation and other factors investigation of past climate records is essential (Seppä and Birks, 2001) and to predict the future climate, it is essential to check the agreement between past data and the climate models (Lunt et al., 2013).

One of the main barriers to reconstruct long-term data series is lack reliable data in the past. As an alternative to instrumental data, climate proxies such as stable water isotope data, Atmosphere CO2 concentration, sediment depositions are used for climate reconstructions.

Tree ring data are shown to be a valuable proxy of precipitation and temperature. D'Arrigo et al., (2005) reconstructed December–February Nino-3 sea surface temperatures based on subtropical North American tree-ring records to examine characteristics of ENSO variability over the past six centuries (i.e. AD 1408–1978). Reconstructions of central European summer precipitation and temperature variability over the past 2500 years was done using tree ring data by Buntgen et al., (2011).

Past millennium climate anomaly in Europe was analyzed by Goosse et al., (2012) by assimilating temperature reconstructed using multiple proxies such as Tree ring, Coral, Ice core, sediment by of Mann et al., (2009) to Earth system model (LOVECLIM) (Goosse et al., 2010) using particle filter data assimilation method developed by (Leeuwen et al ., 2009).

The stable isotopic ratio of precipitation differs according to the scale of the precipitation, for example, tropical cyclones have a low value compared to that of normal summer

precipitation. Hence the isotopic analysis of ancient freshwater fossil carbonate shells, fossil mammal teeth or tree rings can be used to analyze past tropical cyclone activities (Lawrence, 1998). Steiger et al., (2017) showed water isotope data from ice cores could reconstruct historical climate in the 19th century by assimilating isotope-enabled atmospheric model ECHAM5-wiso. Okazaki et al., (2017) showed the possibility of global climate reconstruction by assimilating to Isotopic proxies (δ^{18}O in ice cores, corals, and tree-ring cellulose) to Isotope enable Global Circulation models (GCM).

The merit of about climate reconstructions is they expand to several thousand years sometimes beyond the last millennium. However, the limitation of these proxy reconstructions is they are either annual or seasonal and not available in all the regions.

1.2. Limitations in long-term annual or decadal reconstructions

Climate can influence to the crops, food production and economy. Flowering and harvesting period are observed to be changed with the climate. For instance, Cherry blossom day is changing over the last 50 years in Japan (Japan Meteorological Agency, 2016). On the other hand information about the flowering date can be used as a proxy for climate reconstructions. Aono et al., (1994) has calculated the March temperature in Kyoto from the 11th century using a statistical method with cherry blossom records. Temperature has a significant impact on the quality of wet rice during the ripening period. During the 2010 heat wave, in whole Japan except Hokkaido, quality of rice got deteriorated (MOEJ, 2012). To investigate this kind of phenomena, climate reconstruction with a higher temporal resolution would be better.

Precipitation frequency too can influence the vegetation (Zhang et al., 2013). Brien et al., (2013) has evaluated the impact of rainfall pattern on vegetation by changing the watering interval on different specie's seeds and evaluating the germination process. According to them, frequency influences biomass production and influence the competition between the crops. To investigate this kind of incidents a long-term higher frequency weather information is useful.

The two main limitation of traditional proxies are lack of availability, and their low temporal resolution can be overcome with document data which has monthly, daily or even sub-daily information (Mikami, 2008) in several areas where other proxies are not available.

1.3. Data availability and previous attempts

Old documents are recently utilized for historical climate constructions (Franke et al., 2017). However, the documents currently in used are limited to old instrumental records. On the other hand, there are qualitative data such as personal diaries. These records do not have numerical values but provide a lot of information about the historical period. The value of weather information in personal descriptions is essential today as well. Even today many people post in social media about the weather events. A recent study about Twitter messages showed the possibility of retrieving weather information from these qualitative data in the present (Hannak et al., 2010).

Lorrey et al., (2015) digitized the meteorological data from an instrumental diary belongs to New Zealand. That is one of the oldest instrumental observations in the mid-19th century (i.e. 1839-1844) in the southern hemisphere. It has qualitative instrumental observations and remarks about the weather such as the snow cover. The study has found lower winter temperature in winter and warmer summer than present which may be linked to influence from ENSO. Analysis of weather on these periods helps to understand the difficulties faced by the early settlers.

Monthly, seasonal and annual temperature reconstructions for Central Europe has been derived using documentary evidence since 1500 AD (Dobrovolný et al., 2010), they found long-term temperature decreasing trend over the last five centuries in three seasons winter, spring and summer. This example shows the importance of expanding similar investigations globally. Wang, (1992) has discussed sources of Chinese historical weather reports from 18th century Bell et al., (1978) has analyzed the historical data with weather information in Europe during the medieval period in Europe. Zhang et al. (2013) reconstructed seasonal change in Beijing 1867-1897 using a diary with weather records considering the number of rainy days. However, it is limited to seasonal reconstructions without a physical model. Lake freezing date records are available since the 15th century and has been used to reconstruct winter temperature for five centuries (Mikami, 2008).

Gathering of historical document data is ongoing for instance Yoshimura et al., (1993) have collected hundreds of Japanese personal data from the 17th century to 18th century and made a Historical Weather Data Base (HWDB) and those data are available online. Similar information is available in other countries and has been used for climate reconstructions.

Toride et al. (2017) have shown the possibility of reconstructing past weather using uncertain weather information found in documents with idealize experiments. In that study, they have converted total column cloud percentage (TCC) into three classes 10% (Sunny), 50% (partially cloudy) and 90%(Cloudy) using recent climate data to synthesize personal weather data. That study can be recognized as the first step towards assimilating uncertain weather classes into a climate model. However, there are several limitations in that study. It is limited to making weather classes based on numerical values from recent instrumental observations and reanalysis data. However real diaries are much complicated and have only qualitative descriptive data. Hence it is required to develop a method to convert descriptive information to usable numbers. Moreover, that study used recent SST and Sea Ice data even though in the 19th century SST data is available only in the monthly low-quality state. Moreover, they do not cover the impact of observation availability or observation uncertainty level. There is diverse information in a diary such as presence of precipitation and can be assimilated even though the above study only assimilates TCC. These limitations indicate that further studies are crucial before using the actual weather diary data. Ichino et al., (2001) has developed weather classes using historical weather descriptions patterns and recent Japan Metrological Agency (JMA) weather patterns. In this study we utilized the weather classed derived from the JMA weather description data in the recent period to evaluate the model performance and weather classes from HWDB to reconstruct past weather.

Only Europe and some other regions have instrumental observations before the mid-19th century (Lamb, 2005). Japan is a blind point in this pre-industrial period (Zaiki, 2006).On the other hand, Japan has much descriptive information in documents such as personal diaries. Hence, this study focused on Japan. The developed system can apply to the rest of the world as well.

This study is the first study to assimilate information from a description data set into the climate model and will be an immense contribution to historical weather reconstruction studies

1.4. Dissertation outline

Chapter 2 introduce the historical weather data availability and the process of converting these data into the simple weather classes such as sunny, cloudy and rainy. Furthermore, it describes how the simplified classes were transformed into numerical values by comparing with instrumental observations. Illustrate numerical models; Global circulation model is used as the climate model, and a data assimilation scheme is used to assimilate weather information to the climate model.

Chapter 3 introduce the model with idealize experiments. Sensitivity to the number of observations, number of ensembles, sensitivity to observation error, improvements to other variables from single variable assimilation and impact from initial conditions are also discussed.

Chapter 4 provides details about the experimental setup for past weather description data assimilation and test the influence of the boundary conditions , the influence of assimilation time and contribution from weather classes relative to exact values.

Chapter 5 validates the proposed model using JMA observations in the recent period. Simulations results after assimilating Solar radiation derived from descriptive information against the ground observations are compared. Assimilation of alternative variables such TCC is also investigated. Improvement by incorporating additional information such as the absence of precipitation is also analyzed.

Chapter 6 apply the model to the historical period during the 1830s and 1860s. Model results are compared with the 1995 experiments and observations and other proxy data. Importance of using weather documents over other regions are also shown.

Chapter 7 conclude finding of this study and recommend possible future directions.

Chapter 2

Material and Methods

Abstract:

To reproduce climate, a numerical weather prediction model can be used. In this study, Global Spectral Model by Scripps Experimental Climate Prediction Center based on Global Seasonal forecast system in National Centres for Environmental Prediction's is used. This model was used as the operational forecast model there until 2004 and as the basis for several model development projects. However, these models are not perfect and the results can be improved if observations available in the past can be incorporated to model results. Data Assimilation is useful to get the best estimate from a model and observations considering the model errors and the observation uncertainties. There are attempts to reconstruct past climate using other proxies such as Tree ring, Coral, Ice core, and sediment. The merit of these climate reconstructions is they cover several thousand years sometimes beyond last millennium. However, limitation of these proxy reconstructions is they are either annual or seasonal and not available in all the regions. On the other hand, the personal diary information provides more frequent information allowing to reconstruct climate in high resolution using online data assimilation techniques. In this study, we used Local Ensemble Kalman Filter which uses an ensemble forecast to calculate error covariances. And it has localization ability that can-do assimilation grid wise to each state vector parallelly considering all the observation in the local area which makes the computation more efficient. Chapter 2 discuss the assimilation system in detail and characteristics of diary data. Japan has a digitized database of old personal diaries from the 17th century. There are around 20 diaries in the 19th century. Even though personal diaries have valuable information about daily weather, they are limited to qualitative information such as descriptions like 'sunny' and 'cloudy', and it was a challenge to convert them to usable quantitative format to be used in the climate model. This qualitative information

was converted to probabilistic representative quantitative values of Total column cloud content (TCC) and downward shortwave radiation (SR).

2.1. Introduction

Weather information can be found on personal diary records in several countries. For example, Mikami et al., (1988) discussed about Japanese diary data. Similarly, Zhang et al., (2013) using Chinese diary data and Bernhardt et al., (2015) using the United States have reconstructed regional weather patterns with information from diary data. A sample of a Japanese diary page is shown in Fig. 2.1. The descriptions in diaries provide information about precipitation, wind and cloudiness. Most of the documents have only keywords such as 'sunny', 'cloudy', 'rainy' etc. as shown in Fig. 2.2. Several diaries have information in detail with time and weather intensity such as 'heavy rain', 'early morning rain' etc.

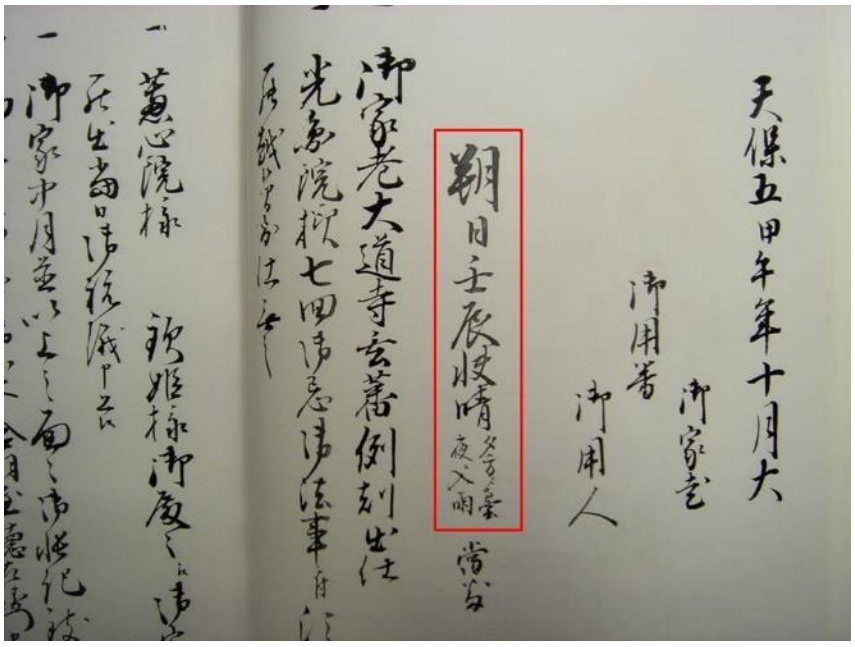


Fig. 2.1: An example of daily weather descriptions in old diaries: "Hirosaki-han Edo" Diary ("Weather diary records since the 18th century," 2017)

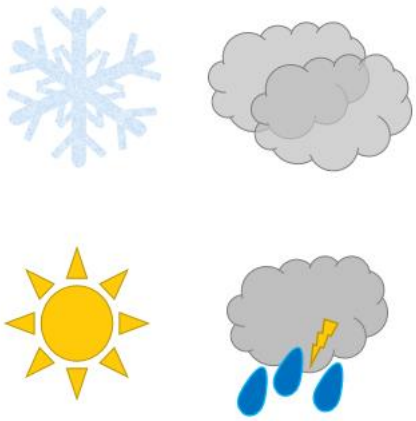


Fig. 2.2: Information from weather descriptions

This study used the weather information from the HWDB, which has 50 diaries with weather information during the 19th century, 23 diaries during the 18th century and 11 diaries during17th century. However, a single diary does not cover the whole period or continues. For instance, there are only 21 diary records in 1860. The weather data is available online (Yoshimura, 2006) at https://tk2-202-10627.vs.sakura.ne.jp/. The website shows 'good weather', 'bad weather' of all the weather diaries in Japan on each day in maps or tables. Here the good weather means the state of weather that is closest towards a sunny day and bad weather mean state furthers from sunny. In someday the weather can be sunny throughout the period. In this case, both these may have similar information. However, in many days weather can change during the day. If the weather was cloudy in the morning and become sunny in the afternoon, good weather would indicate sunny and bad weather will indicate cloudy. However, as it is difficult to extract data from the website the raw digitized weather was collected through personal communications and will be explained in the next section.

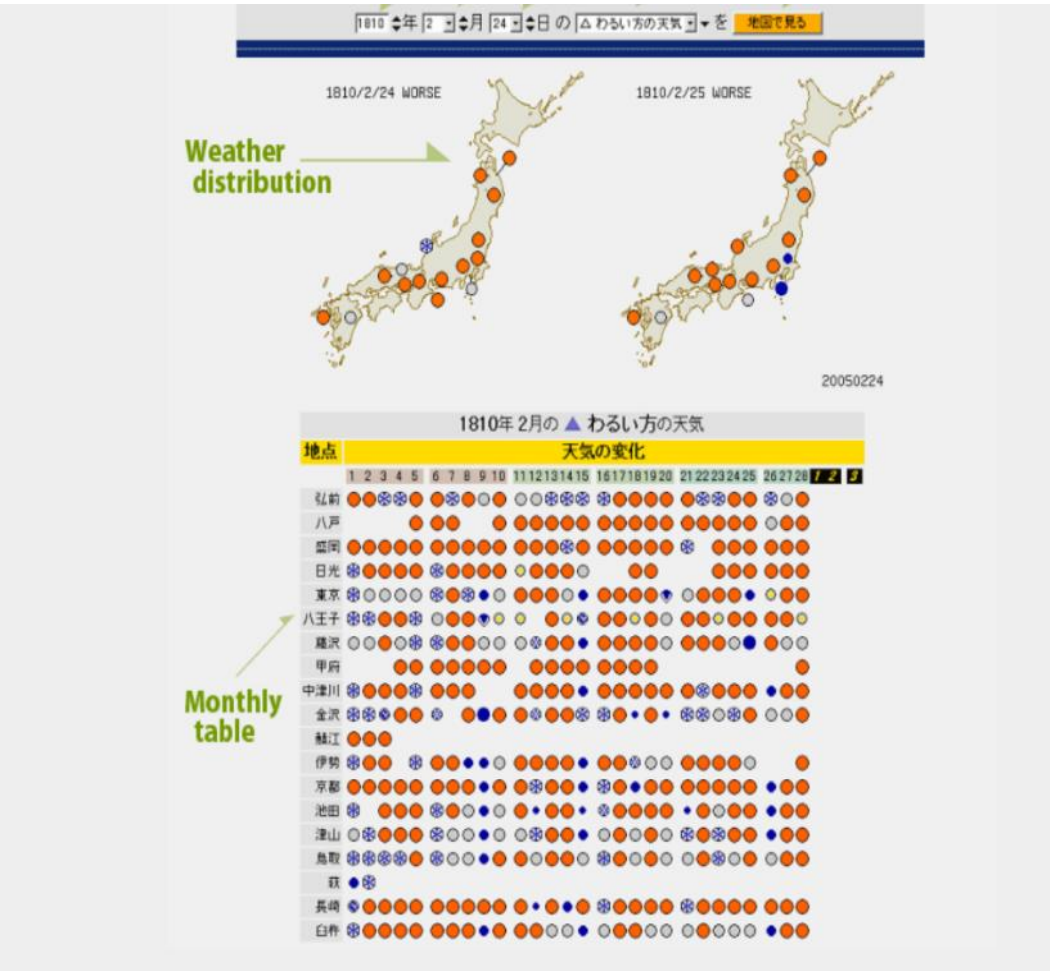


Fig. 2.3: Historical Weather Database on The Web(歴史天候データベース・オン・ザ・ウェブ)

2.2. Data

2.2.1. Weather classes

The information from the diaries cannot be used directly as they have a lot of information in various styles. HWDB has information about wind, thunder and warmth. However, to reduce the complexity, only the information representing sunshine and precipitation was utilized as this is the first attempt to use such data in a climate model. Even though information about these two variables provides a lot of information as the given categories as Table 2-1, converting these categories into a usable format is a challenging task. If we have a digital range we can follow the Toride et al. (2017) as explained in the introduction, however, because these data are qualitative which only words, first it was necessary to convert the weather categories into a usable format (simple weather classes). Converting to weather classes can be done by

simplifying either category information in Table 2-1 or reading the text descriptions. Ichino et al., (2007) has converted the text descriptions in diaries into three weather classes (i.e. 1,2 and 3) considering sunshine and precipitation information in the diaries. However, the weather classes data produced by them is available only for a limited period and locations, a separate algorithm was prepared to convert whether categories in Table 2-1 to weather classes in this study while keeping the consistency with classes by Ichino et al., (2007).

Table 2-1 Key information types from the digitized weather diary database

TH code	Meaning in Japanese	Meaning
0	記 述なし 判読不能(虫 食いその他)	No data
1	快 晴, 晴, 吉, 能	Clear, Sunny
2	薄 晴, 薄 曇 り	Slightly Cloudy
3	曇, 陰	Cloudy
4	に わか雨,時 雨(雷 鳴,雷 電はなし)	Shower
5	雷 雨	Thunder Storm
6	大 雷雨, 雷 大雨	Heavy Thunderstorm
7	小 雨, 細 雨	Light rain
8	雨, 降	Rain
9	大 雨, 甚 雨	Heavy Rain
A	あ られ, 雹, 霙	Hail
B	に わか雪	Snow Shower
C	小 雪	Light Snow
D	雪, 大 雪	Snow, Heavy Snow

In early 19th century, Japan had a strong famine during the 1830s in the Edo period. To see the possibility of reconstruction of weather in 1830, using the weather classes data of Ichino et al., (2007) an experiment was done in this period. However, there is no instrumental observation to evaluate model performance in the past. Hence several experiments were carried out in 1860 where few instrumental observations are available. 1860s diary data was converted

to classes in this study. The conversion was done to keep the consistency with the classes of Ichino et al., (2007) as follows.

Class 1 - Sunny and no precipitation

Class 2- Change from sunny to cloudy or sunny day with little precipitation or cloudy day

Class 3- Cloudy and Precipitation

The Fig. 2.4 shows the weather classes in Yokohama station during 1863 from Ichino et al., (2007) and the derived weather classes from the HDWB in this study. According to the results, only a few days have a mismatch due to the different interpretation of nighttime rainfall in few days.

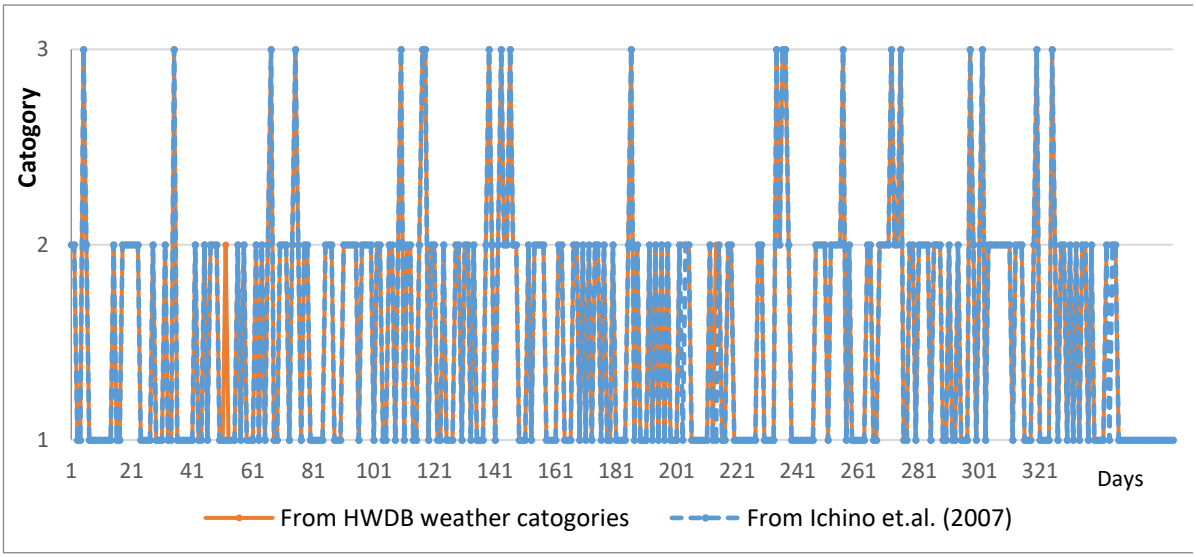


Fig. 2.4: Weather Classes at Yokohama in 1863

This study aims to use information in historical details into climate models. However, due to lack of validation data in the past, we focused in recent descriptions data of JMA as Ichino et al., (2007) has converted the JMA description data into weather classes keeping the consistency with the diary weather information. Hence, we considered they are similar and used to evaluate the model performance before applying the historical weather data. We utilized 18 stations (see Fig. 2.5) data in Japan.

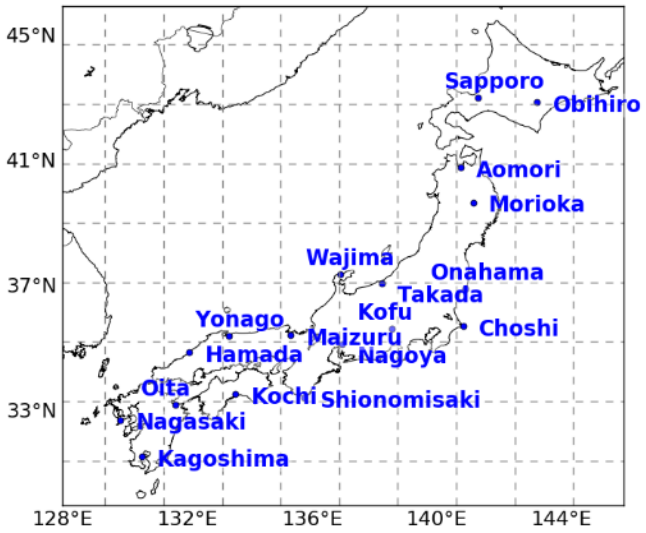


Fig. 2.5: Available of weather classes data in 1995, data from 18 stations were used considering the availability of diary data in the past.

2.2.2. Converting to numerical values

2.2.2.1. Using weather classes

To assimilate the weather classes, it is essential to convert them to physical variables. Weather classes are found to be correlated to daily downward shortwave solar radiation at the surface (SR) Ichino et al., (2007). On the other hand, the Previous idealized study has shown the potential to assimilate TCC classes (Toride et al., 2017) in numerical models. Toride et al., (2017) converted TCC in JMA data to weather classes to check the contribution from weather classes using following simple approach

Less than 20% –> 10% (Sunny)

- 20-80% -> 50% (Partially Cloudy)
- More than80% -> 90% (Cloudy)

However, there was no evaluation for the suitability of above classes, and it limits to convert weather classes from TCC. Hence the correlation between the weather classes and SR and TCC were evaluated and empirical relationships were developed between the variables and weather classes using the recent JMA observations in these stations.

Ichino et al., (2007) introduced an empirical relationship between K_T (Daily clearness index) and weather classes in each month. Values are calculated from in $Q_d = Q_s K_T$ Eq. 2-1

$Q_d = Q_s K_T$ Eq. 2-1 using 20 year's (i.e. 1979-1998) JMA SR data (Q_s) and Top atmosphere solar radiation (Q_d).

$$Q_d = Q_s K_T \quad \text{Eq. 2-1}$$

These Tokyo station's K_T values were used for some basic experiments over Japan. However, as climatology differs from one location to another (see Fig. 2.6) K_T values were calculated using a similar approach for each of the 18 stations using JMA data and weather classes data in 1995-1999.

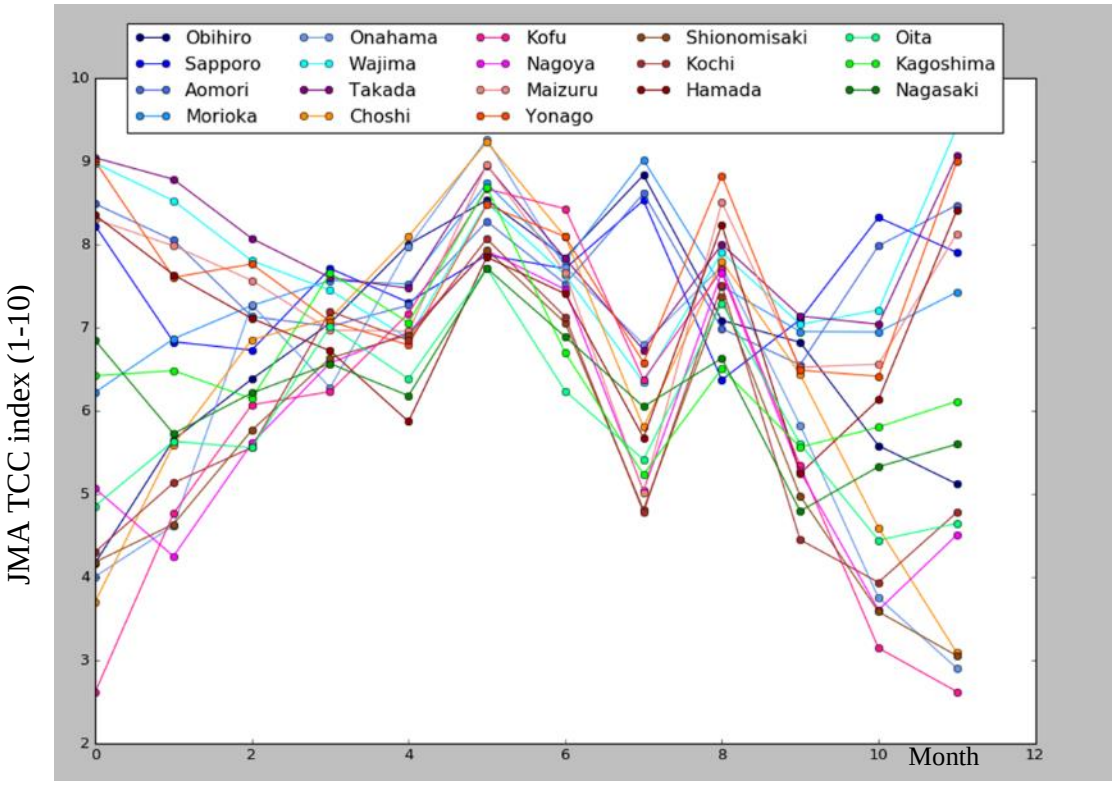
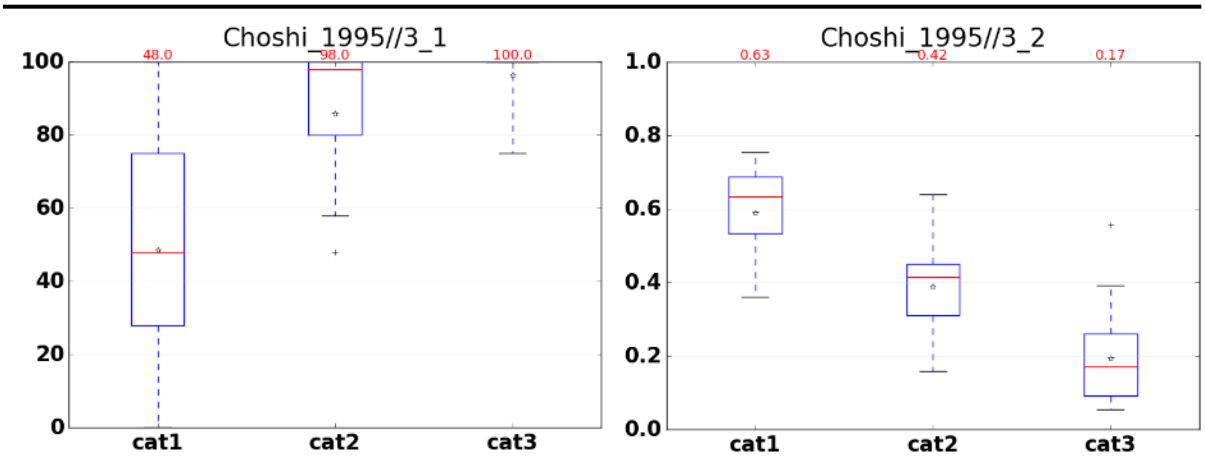


Fig. 2.6: Monthly JMA Total Cloud cover variations in 1995 over different stations

Similarly, for each month TCC values for each weather class was calculated using the weather class' data and JMA daily average cloud cover data for the 5-year period.



(a) (b)

Fig. 2.7: Empirical relationships between (a): weather classes (1, 2 and 3) and TCC (b) Weather classes (1,2 and 3) and KT, the red line represents the median and star represent the mean in each box plot

Box and whisker diagram in Fig. 2.7 (a) show the calculated monthly SR and TCC values at Choshi observation station for each weather classes. A noticeable trend in SR for each weather class can be identified as stated by (Ichino et al., 2001) and SR values have smaller error variance compared to TCC. On the other hand, TCC ranges have a higher overlap (for instance weather classes 2 and 3 has a very high overlap and large error variance). To represent these errors, for SR 50 w/m^2 value random error is added for solar radiation as they have similar variance. For TCC considering their variance 10%, 20%, 30% random error is added for TCC values of each category respectively.

These calculated TCC and Kc values for each weather class would be a probabilistic representative value, and the actual value lies in the range. Hence we added a random normalized error considering the standard deviation of these values.

2.2.3. Observation data for validation

To evaluate the model, JMA observations are considered. However, the model has a larger grid size around 200 km even though JMA is point observations. On the other hand, the model may produce average results for the whole area causing to average different climatology and topography as well. Due to this model may have a bias from the JMA observations and may deviate from the point observations. To evaluate the impact, a comparison between Tokyo JMA observation station and Chiba observation station data were done and as shown in Fig. 2.8. Even though they are only around 100 km away and fall into the same grid, a noticeable difference can be seen. Chiba station has more precipitation events, and only two precipitation

events are overlapped, in those two events (01/14,01/21) too Chiba has a considerably higher amount of precipitation and around 3-degree low temperature.

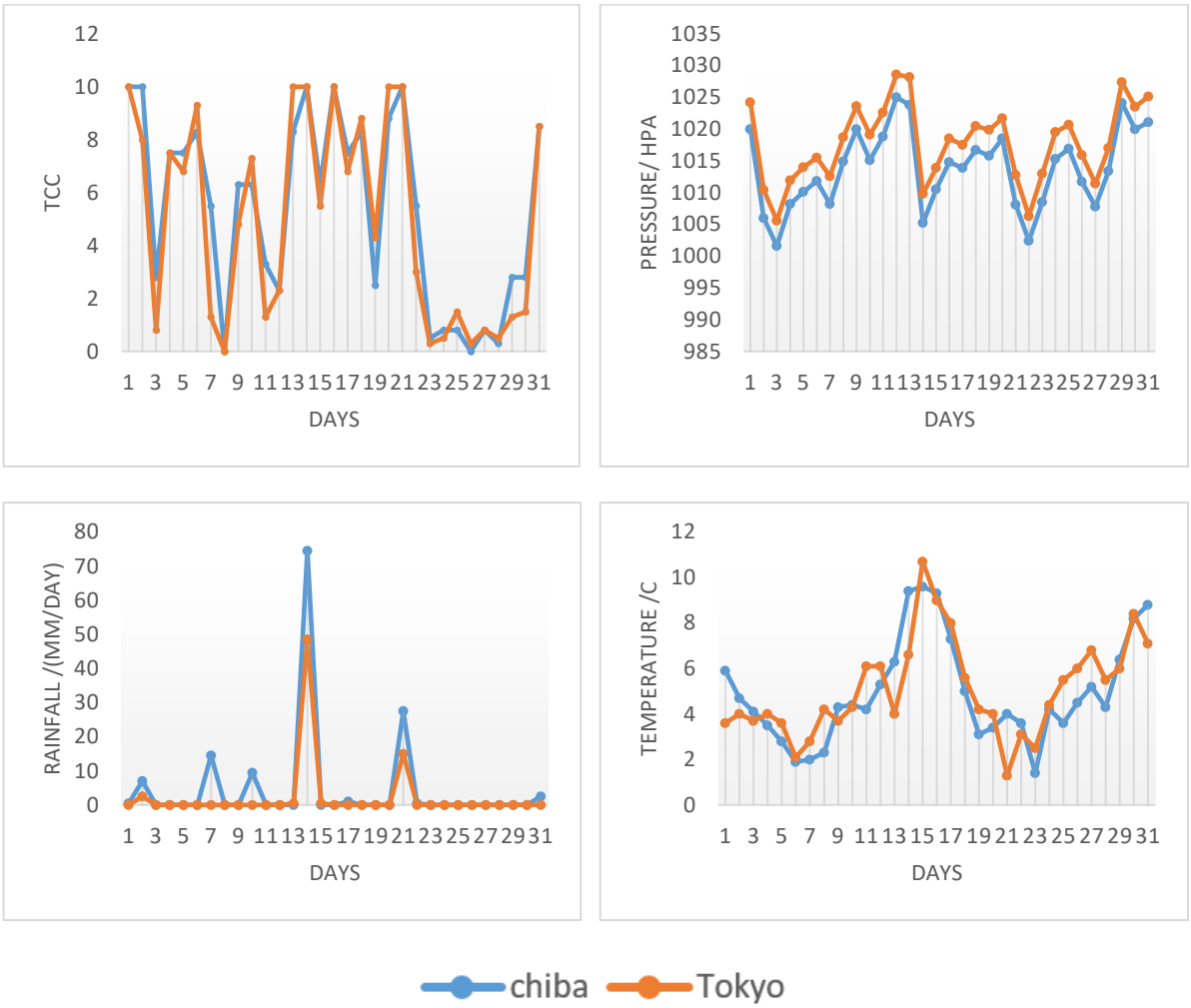


Fig. 2.8: Comparison of JMA observations at Tokyo observation station and Chiba observation stations.

To check the impact of resolution further, JMA results were compared with a reanalysis dataset which has a similar characteristic with the proposed model. A data set produced by applying spectral nudging to National Centers for Environmental Prediction–Department of Energy (NCEP–DOE) Reanalysis-2 "NCEP_Reanalysis 2 data" (n.d.), Kanamitsu et al., (2002a) will be referred as NCEP data here onwards was used for this purpose. Fig. 2.9 shows the results, here a significant bias in the pressure can be seen in NCEP data which may be due to the averaging topography in the grid. Temperature also has a positive bias even though temporal correlation is very high. JMA does no directly measure TCC directly instead it provides Cloud cover in 1-10 scale, to compare the two types, NCEP TCC percentage also brought to a 1-10 scale by dividing 10. According to the figure, it has a good correlation with

NCEP data thus JMA cloud observations are used as TCC observations in this study. The pressure between JMA and NCEP has a considerable bias through the period indicating that model results too has a similar bias. To investigate further the influence of resolution, it is required to compare the more grids which have various local climatologies.

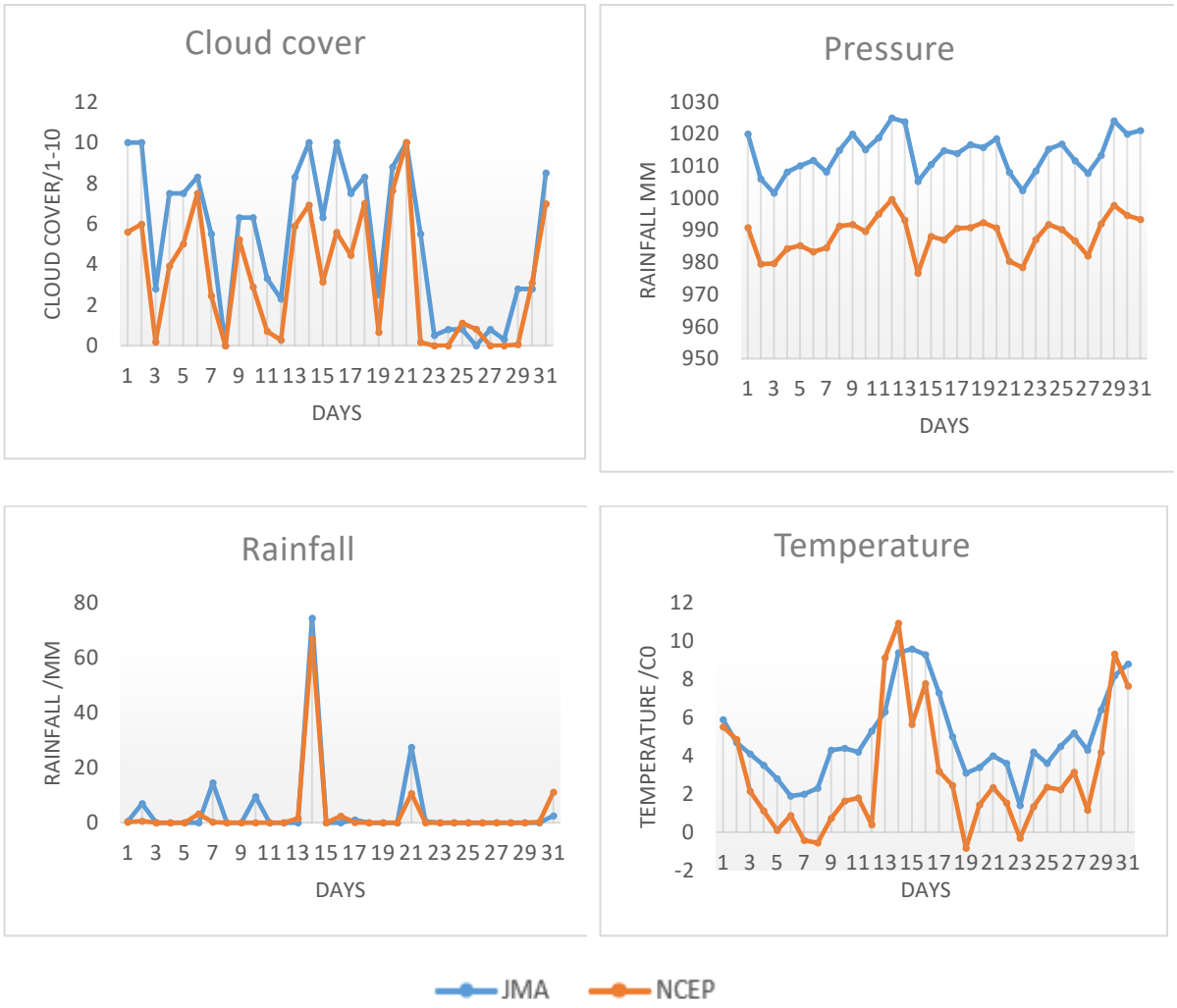


Fig. 2.9: Comparison of JMA observations vs NCEP reanalysis observations at Choshi station

2.3. Weather Forecast model

2.3.1. Importance of a physical model

In historical climate reconstruction, statistical and empirical methods are used. (Ge et al., 2005) used an empirical method based on field experiments to reconstruct precipitation using memo documents to the emperor which consists of 104,996 records of moisture penetration depth after precipitation and snow depth information. However, from those data, it would be

possible to estimate other variables if a numerical weather prediction model base on physics could be used. In a numerical weather prediction model once a variable is improved other deterministic variables would also improve. (McGregor et al., 2013) used a coupled general circulation model (CGCM) was used to investigate changes in El Niño–Southern Oscillation variance over the past six centuries. Compo et al., (2006) used NCEP global medium-range forecast model used in 1998 (Kanamitsu et al., 1991) to check the possibility of developing 100-Year Reanalysis Using Only Surface Pressure Data.

2.3.2. Numerical Weather prediction models

To represent the physical process in planetary atmosphere, numerical models are used based on Navier–Stokes equations. Atmosphere Global Circulation models (AGCMs) are developed in two different approaches called grid pin models and spectral models. In grid point model variables are assigned above the intersection of each grid in a regular matrix whereas in the spectral model's known as Global Spectral Models (GSM) variables are represented by a periodic function which is calculated as the sum of spectral harmonics. The spectral representation of variables has several advantages such as the exact calculation of space derivatives, no pole problems, and no instability arising from aliasing in an ideal situation (Orszag et al., 1970). Hence in this study global spectral model is used. Several researchers used Global Spectral Model (GSM) developed by NCEP to investigate data assimilation techniques. Szunyogh et al., (2007) used NCEP Global Forecast System (NCEP-GFS) GSM that was in operational use at the beginning of 2001 to investigate performance of LEKF, Whitaker et al., (2008) used NCEP-GFS GSM which was operation in March 2004 to investigate the performance of ensemble data assimilation including LETKF.

2.3.3. GSM

In this study, GSM by Scripps Experimental Climate Prediction Center (ECPC) based on Global Seasonal forecast system (SFM) in National Centers for Environmental Prediction's (NCEP) which was an operational seasonal forecast system by NCEP (Kanamitsu et al., 2002b) is used. This model was used as the operational forecast model at NCEP until 2004, and as the basis for several model development projects (Saha et al., 2006). Yoshimura et al., (2014) assimilate information from isotopes by carrying out observation system simulation experiment (OSSE) where synthetic data is created mimicking the actual data from satellite or ground observations. This study demonstrated how only isotope data assimilation could improve the isotope fields and atmospheric dynamics (temperature, wind speed, humidity, and

surface pressure) into the above GSM. Toride et al., (2017) used the same GSM to evaluate the performance of climate model with uncertain weather information. Considering the similarity of this study, the same GSM was utilized in this study.

2.3.3.1. Physics packages

The model's physics packages include the following schemes;

- Longwave radiation scheme of Chou and Suarez (1994),
- Shortwave radiation scheme of Chou (1992),
- Relaxed Arakawa–Schubert convective parameterization (Moorthi and Suarez, 1992),
- Non-local vertical diffusion (Hong and Pan, 1998),
- Mountain drag (Alpert et al., 1988),
- Shallow convection (Tiedtke, 1983),
- Noah land surface scheme (Ek et al., 2003).

The commonly use T62 grid system (about 200 km horizontally) and vertically 28 sigma layers was applied in this study. The time step of the model was set to 20-30 minutes depending on the situation.

2.4. Data Assimilation

2.4.1. Introduction

After the model forecast, if there are observations, information from the observations can be incorporated to the guess and use as the initial condition for the next model run. Both the information from diaries and the model has an uncertainty. Data Assimilation is useful to get the best estimate from a model and observations. In this study, our focus is to obtain an accurate estimation with the available information and know physical laws between the variables (i.e. model). Considering the model errors and the observation errors with a Bayesian approach as shown in the schematic diagram (see Fig. 2.10). As in Eq. 2-2 Bayesian methods provide the probability to occur some event when there is some knowledge about it.

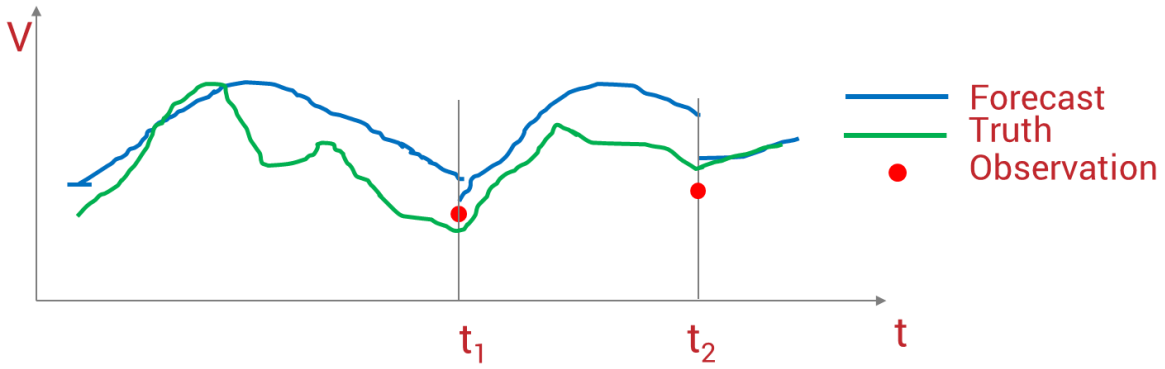


Fig. 2.10: Schematic diagram showing the process of data assimilation

$$P(x \mid y_0) = \frac{p(y_0 \mid x) P(x)}{p(y_0)} \qquad \text{Eq. 2-2}$$

Where $P(x)$ is the probability density function (PDF) of model state before the observation is known (prior) and $p(y_0)$ is the PDF of the observations and $p(x \mid y_0)$ is probability density function of model state if the observation occurs (Likelihood).

2.4.2. Kalman filter

A Kalman filter is one of the classical data assimilation method among various assimilation schemes. However, as this model is a linear model it is problematic to be used in a nonlinear forecasting models such as the GSM because for error covariance of the nonlinear models cannot be estimated beforehand unlikely in linear models on the other hand computational cost is significantly high due to the calculation of the time evolution of the error covariance (Tippett et al. 2003).

There are two kinds of approaches, stochastic and deterministic approaches in 1994 (Evensen, 1994) introduced Ensemble Kalman Filter (EnKF) a stochastic approach to handle the above limitations of the Kalman Filter by using ensembles to forecast error covariance and model estimate error covariance. This Enkf was successfully applied to a perfect model experiment using a low-resolution atmosphere model by Houtekamer and Mitchell (1998). However stochastic approach requires to add and perturbation to observation to update ensemble members who can add a bias to the estimate of analysis error covariance (Whitaker and Hamill, 2002).

On the other hand perturbed observations are not necessary for creating the ensembles when using Deterministic methods, such as the Ensemble Square Root Filter (EnSRF),

(Tippett et al., 2003). To increase the efficiency Bishop et al., (2001) introduced a deterministic method, ensemble Transform Kalman Filter (ETKF) which finds a transformation matrix which makes covariance calculation more efficient.

Below we provide a basic explanation to the ensemble Kalman filter method used in this study Following the notation in Whitaker and Hamill (2002), the EnKF update equations are

$$\mathbf{x}^a = \bar{\mathbf{x}}^b + \mathbf{K}(\mathbf{y}^0 - \mathbf{H}\mathbf{x}^b) \qquad \text{Eq. 2-3}$$

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^b$$

Where,

$\mathbf{x}^b$ - *m*-dimensional model background state vector

$\mathbf{x}^a$ - model analysis state vector

$\mathbf{y}^0$ - *p*-the dimensional vector of observed values

H – observational operator that converts the models state to the observation space

$\mathbf{P}^b$ is the *m* × *m*-dimensional background error covariance matrix

$\mathbf{P}^a$ - analysis error covariance

I - *m* × *m*-dimensional identity matrix, and the overbar denotes an ensemble mean

K - *m* × *p*-dimensional Kalman gain matrix

$$\mathbf{K} = \mathbf{P}^b\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{P}^b\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}$$

Where,

R -*p* × *p*-dimensional observational error covariance matrix.

In this study, the model state vectors consist of air temperature, specific humidity, surface air pressure, precipitation, total column cloud content and downward shortwave radiation. The model background state $\mathbf{x}^b$ is calculated by running the GSM to the analysis model state $\mathbf{x}^a$ from the previous time step.

Even though H can be a complex nonlinear operator here, it only extracts the variables from the $\mathbf{x}^b$ at observation locations. Next chapter explain about the $\mathbf{y}^0$ extraction from weather dairies and the observational error **R.**

2.4.3. Local Ensemble Transform Kalman Filter (LETKF)

Ott et al., (2004, 2002) introduced Local Ensemble Kalman Filter (LEKF), which has localization ability that can-do assimilation grid wise to each state vector parallel considering all the observation in the local area which makes the computation faster. Also, a lesser number of ensembles are required to this. In this study, we used the LETKF (Hunt et al., 2007), Harlim

et al., (2005) which applies the ETKF locally as in the LEKF. Final equations of LETKF are as below

$$X^a = \overline{x^f} + \partial X^f T$$

$$T = P^a (\partial Y)^T R^{-1} \left(y^0 - \overline{H(X^f)} \right) + \left[(m-1) P^a \right]^{1/2}$$

$$P^a = \left[(m-1) \mathrm{I} / \rho + (\partial Y)^T R^{-1} \partial Y \right]$$

Pa – analysis error covariance matrix m-ensembles N-dimensional

If the observational errors are uncorrelated and R is diagonal

20 ensembles enough (Miyoshi and Yamane, 2007)

ρ - Inflation parameter to keep filter divergence

H- Linear observation operator

R- Observational error covariance matrix

2.4.3.1. Localization and inflation techniques

Due to the localization, data assimilation can minimize the errors due to the random correlations among distant localization allowing to use lesser number of ensembles (Hamill et al., 2001). To reduce the influence of discontinuity, we use a Gaussian function as in Eq. 2-3 introduce by Miyoshi and Yamane (2007) to smoothen the weighting function.

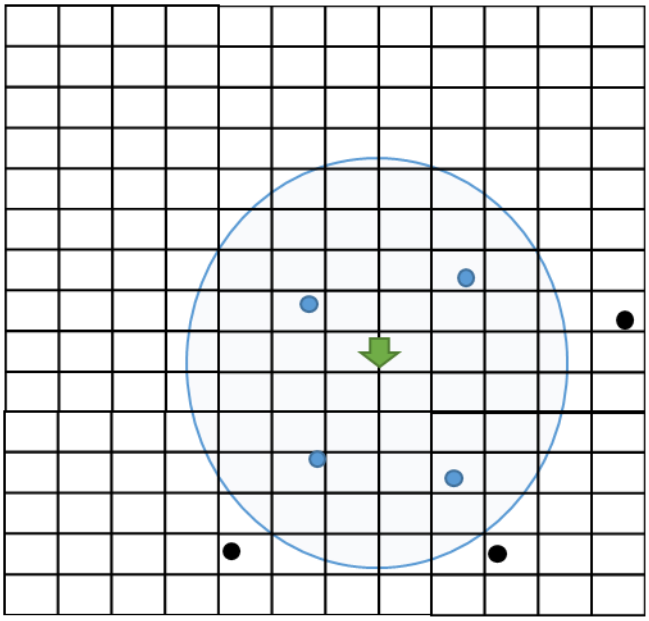


Fig. 2.11: Schematic diagram showing the process of localization

We use a Gaussian function introduce by Miyoshi and Yamane (2007)to smoothen the weighting function to minimize the discontinuity. 500km is used as the minimum physical distance

$$w(r) = \exp\left(-\frac{r^2}{2\rho^2} \right)$$ Eq. 2-4

Where,

r- Distance between the local patch center and observation

ρ - Minimum physical distance

Furthermore, the covariance inflation method was used to overcome a common limitation, an underestimation of the error variance (Anderson, 2009) in ensemble filters. In this study adaptive covariance inflation method by Miyoshi et al., (2011) was used, which estimates multiplicative inflation parameters adaptively.

2.5. Development of Data Assimilation system

Fig. 2.12 show a schematic diagram of the developed model in this study. In this study GSM explained in section 2.3.2 is used to forecast the climate and the data assimilation technique (i.e. LETKF) is used to incorporate the observation whenever the observations are available. Forecast model provides an ensemble guess from the initial ensemble condition from the previous time step and the assimilation scheme incorporate observations into the model guess ensembles and provide the ensemble analysis for the next initial condition of the model. If no observations are available, model guess will be directly used as the next initial condition. This procedure repeats until the end of the simulation in 6-hour cycles.

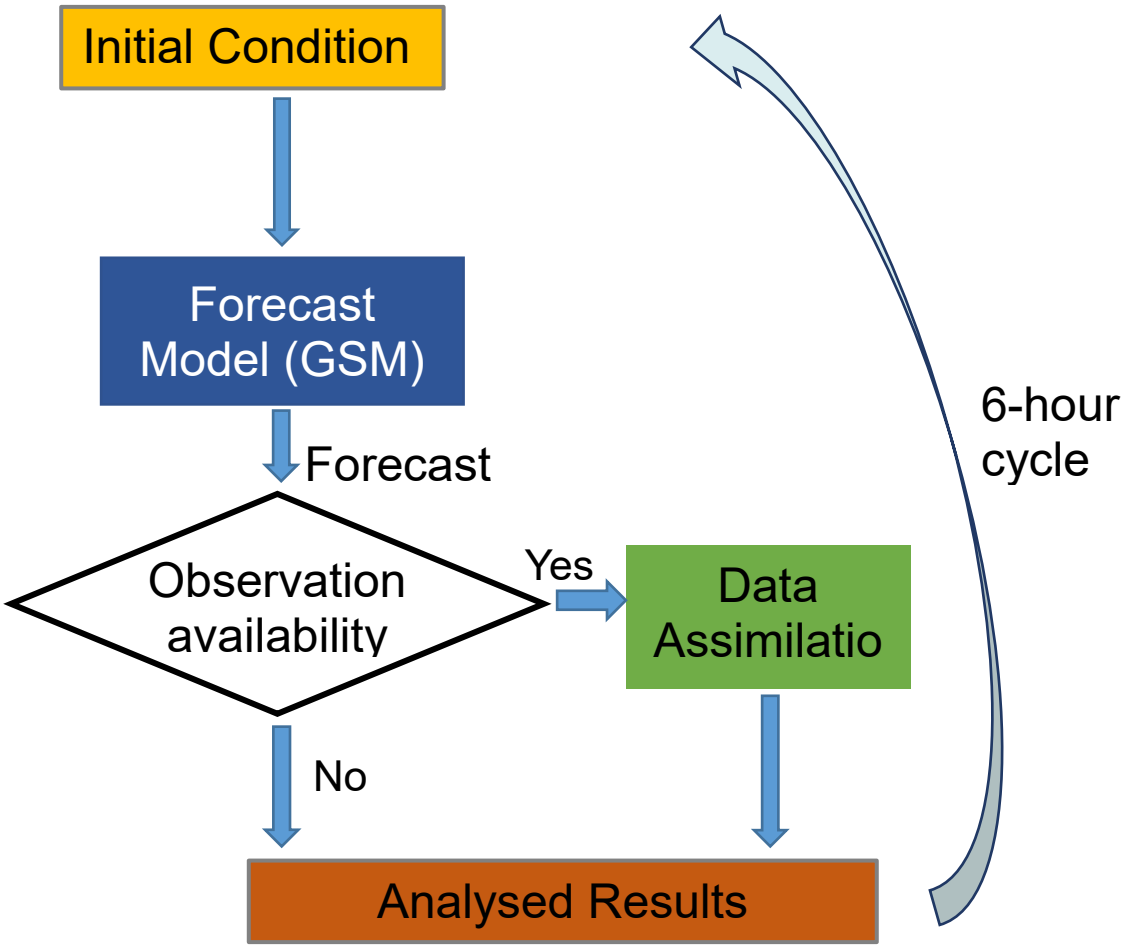


Fig. 2.12: Schematic diagram of the experimental setup

2.5.1. Initial condition

In order to start a model which is known as initializing the model, it is essential to specify specific variables (Reichler and Roads, 2003). As the assimilation system is an ensemble forecast, the forecast model runs parallelly multiple runs initiated with different initial conditions.

In most of the ensemble forecast studies ensembles are created by picking the initial condition from another timestamp (Yoshimura et al., 2014), (Toride et al., 2017). Even though the initial condition is not so critical in an atmosphere model due to the low memory about the initial condition in the atmosphere, another method to create ensembles was investigated to check model performance. Ensembles can be created by adding a perturbation to the variables of the actual data (Toth et al., 1993), ("Ensemble Prediction Systems," 2016). Both of these methods were investigated in section 3.6 and found to have no significant difference in model

behavior. Hence the commonly used Time shift method is used in this study as well to create ensembles.

However, before the 19th century, there isn't any daily dataset that has sufficient data to initialize the mode. Thus initial condition from a random recent year was selected, and one-year spin-up was carried out to bring the model to an equilibrium state. Then each ensemble was then initiated from a different day from the starting date following Yoshimura, et al., (2014) to create the ensemble forecast. The same procedure was used for the experiment in the resent period to keep the consistency.

2.5.2. Boundary conditions

GSM model requires SST and Sea-Ice fraction data as a boundary condition to the atmosphere model. In the present these products are available in high temporal resolution and spatial resolution. Lack of high quality in SST and Sea-Ice fraction data is a limitation to climate reconstructions (Reichler and Roads, 2003). Even though there are data sets on annual scale for thousands of years, higher temporal resolution data is limited. The daily SST reconstructions such as National Oceanic and Atmospheric Administration (NOAA) Optimum Interpolation Sea Surface Temperature (OISST) (Reynolds et al., 2007) are limited to 1981 and monthly SST data such as Hadley Center Global Sea-Ice fraction and Sea Surface Temperature (HadISST) (Rayner et al., 2003) is limited to 1870s. The annual SST and sea-ice fraction data products are not sufficient for daily reconstructions like this study. Recent studies of (Franke et al., 2017) has developed monthly SST data since 1600 by incorporating seasonally anomaly and ENSO signals from HadISST into the 1500 year annual SST reconstruction of Mann et al., (2009) and a sea-ice fraction data climatology from HadISST 1.1 data set before 1870. Franke et al., (2017) SST is the optimum SST and Sea-ice fraction data that we can find. Hence, we used SST data and Sea-ice fraction data of Franke et al., (2017) after interpolated into daily values as in Fig. 2.13. Those are the optimum SST data and sea-ice fraction available for the focus period of this study. In this study even though several experiments were carried out in 1995 for validation purposes, the boundary conditions were kept in similar quality to the interested period in the 19th century. We did a separate experiment to check the influence between weekly SST data from NOAA OISST and Monthly Franke et al., (2017) data to the model performance with same observational error and model settings by assimilating solar radiation similar to the other experiments and results are shown in section 4.3.

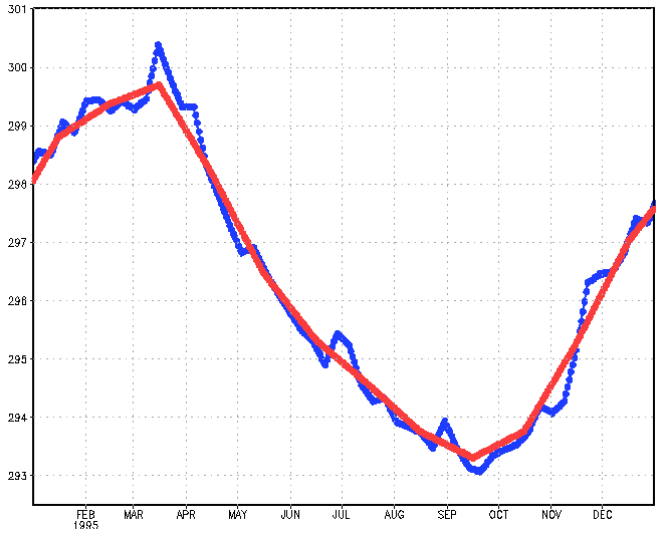


Fig. 2.13: SST data from NCEP-OI SST and Franke et al. (2017) over a random grid

2.6. Conclusion

Weather documents were shown to be an excellent alternative proxy in the absence of instrumental data. The data from the documents are not numbers, and the uncertainty of the documents was explained. The online database has data as 'good weather', and 'bad weather' and the raw digitized data consists of many weather categories as explained. It was shown that this information can be simplified to weather classes such as 'sunny', 'cloudy' and 'rainy'. The weather classes were successfully converted to numerical value considering the probabilistic spread. The consistency of converting the data was confirmed with continues weather class data from Ichino et al., (2007) at Yokohama station. Only one weather event was differently simplified which was due to neglecting nighttime precipitation by the proposed method. A limitation to validate the proposed system was lack of instrumental data, hence experiments in the recent period with description data of JMA was proposed. Impact of the model resolution was briefly examined by comparing two-point observation stations in the same grid which are apart around 100 km, and results showed a bias in some variables and a different number of precipitation events. Both stations have the same major weather events even though the magnitude is slightly different. The impact of resolution was further analyzed by comparing JMA point data observations with NCEP data which has a similar resolution with the proposed model and found to lies between the two points in temporal variation while several variables have a uniform bias throughout the period. Selection of Climate forecast model was discussed, and GSM was selected considering its ability to couple with a Data assimilation scheme and

extensively used in similar studies. For the data assimilation scheme LETKF was selected due to the low computational cost and higher performance and ability parallel computation. Preparation of initial conditions and boundary conditions is a challenge in historical climate reconstructions and alternative methods were discussed to overcome the limitations of data.

Chapter 3

Evaluation of Model Performance and Sensitivity Analysis of Environmental Settings

Abstract:

This chapter further investigated the possibility of assimilating uncertain weather information. It was not clear about proper model settings and sensitivity of the number of observation and observation uncertainty for uncertain weather assimilation up to now. This chapter found solutions to them with several experiments. When TCC data is assimilated correlation improved to 0.47 from -0.01 in average over Japan. In particular, the correlation of TCC improved to 0.64 from -0.13 at Choshi station. There are no significant contributions to other variables (i.e. correlation change in: Temperature 0.3 to 0.2, Precipitation -0.95 to 0.1 and Pressure 0.18 to 0.3). Experiments with a different number of observation stations showed improvement in the correlation coefficient and RMSE around the observations sites even with 18 number of stations. This indicates even the fewer number of weather records are available local improvement can be achieved over those regions. Further, the simulation using data from 418 stations improved the results of not only the exact areas near the stations but also in remote areas. For instance, correlation coefficients of TCC, Temperature, Precipitation and Specific humidity in a non-assimilated site (i.e. Choshi station) improved from -0.13 to 0.38 ,0.30 to 0.57, -0.10 to 0.53, -0.13 to 0.61 respectively. Simulations with different observations uncertainties were carried out to investigate the sensitivity to observation uncertainty and found that if a small observation uncertainty is given, assimilation scheme neglects the observations

because ensemble spread is away from the observations. This was clear in results where an observation error with 1% achieved only 0.17 correlation while observations with an error of 50% correlation improved correlation coefficient to 0.42. Impact from the initial conditions was analyzed by doing simulations with perturbated simulations instead of initial conditions created from time shift method as in previous experiments. The correlation was better in time shift method (e.g., correlation decreased to 0.45 with perturbated initial condition method in comparisons to 0.64 in Choshi station using time shift method and RMSE increased to 39.9% in comparison to 32.6% in time shift method). Thus, it was decided to use time shift method to create ensembles.

3.1. Introduction

Toride et al., (2017) developed a system to assimilate TCC into GSM using LETKF as explained in Chapter 2. In that study, idealized experiments were done by assimilating TCC data of 18 stations in Japan from NCEP data and weather classed derived from JMA Instrumental data. The NCEP data assimilation experiments were done with high uncertainty of 30% TCC, to evaluate the feasibility of assimilating uncertain data. They further did weather class assimilating derived from JMA data as explained in section 2.2.1 to investigate the possibility of assimilating weather classes. To assimilate real weather description data, it is essential to understand the model performance due to several factors such as observation uncertainty, number of observations, assimilation time and improvement in the surrounding area. Furthermore, the required number of ensembles for a unique assimilation system may differ from traditional experiments. This information is lacking in above idealistic study.

Hence, separate experiments were done to investigate the above points. Firstly, idealistic simulations similar to Toride et al., (2017) were done by assimilating NCEP data and weather classed data Idealize experiments were set up following observation system simulation (OSSE) experiments Miyoshi and Yamane, (2007), Yoshimura et al., (2014) and Toride et al., (2017).

3.2. Experimental settings and input data

NCEP data is used to create synthetic cloud observations, and 20 ensembles were used for the experiments with the boundary conditions and initial conditions from NCEP data as explained in Chapter 2.5. The period of experiments is from January 2006 to March 2006. TCC assimilation was done once a day.

3.3. Data Assimilation over Japan

In this experiment, the influence of assimilating uncertain weather information over Japan to represent Japanese weather diary data investigated. Here only 18 stations were considered as observations stations considering the availability of diary data. 30% of random error was added to the TCC values of NCEP data to incorporate the uncertainty similar to Toride et al., (2017).

Experiment results are shown in Fig. 3.1 and Fig. 3.2. According to the RMSE maps in Fig. 3.1, a considerable improvement can be seen in cloud cover all over Japan region represented by dark blue color. In temperature, a slight improvement can be seen in most if the area while no considerable change in China seaside, Aomori and Morioka areas. In precipitation and pressure, only southern side and the northern side has a slight improvement. In most of the areas, there is no considerable change.

Fig. 3.2 shows the model performance in January at Choshi station (model grid: latitude 35.238, longitude 140.625). There is a significant improvement in TCC; correlation coefficient increased to 0.64 from -0.13 correlations in control simulation. Slight improvement in other variables can also be observed. In pressure, Correlation coefficient increased to 0.33 from 0.12, and in temperature, it increased to 0.47 from 0.30, and in humidity, it improved to 0.44 from -0.12. U wind improved to 0.2 from 0.06. Precipitation correlation in the control simulation without data assimilation is very poor and has a negative correlation of -0.1. With data assimilation, it is improved 0.1. In overall poor TCC data assimilation could improve the model's TCC and several other variables.

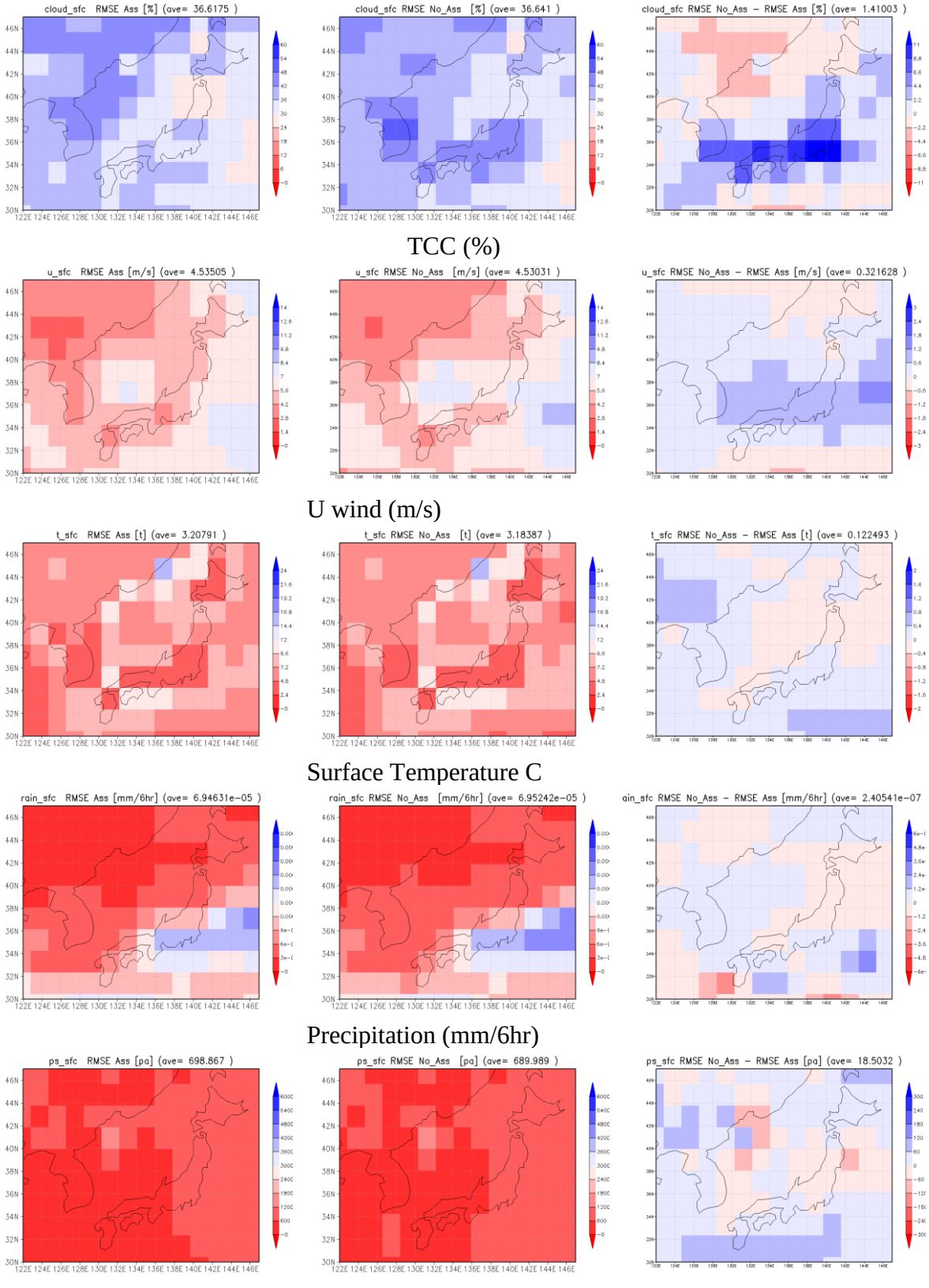
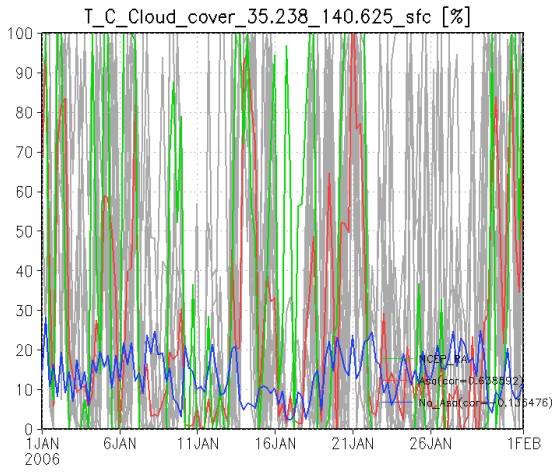
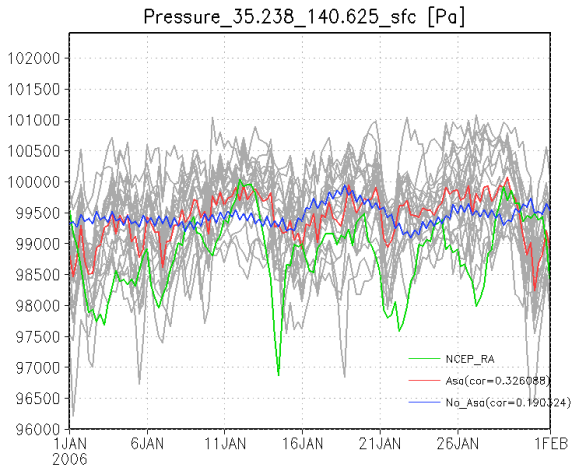


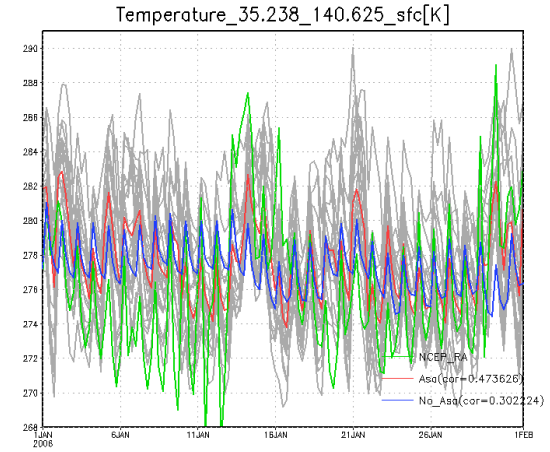
Fig. 3.1: RMSE of TCC over Japan; (a) data assimilation simulation, (b)- control simulation without data assimilation, (c) improvement in RMSE (%) (b-a)



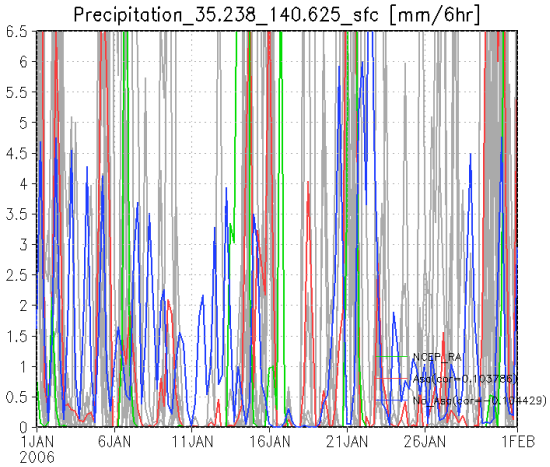
(a) TCC



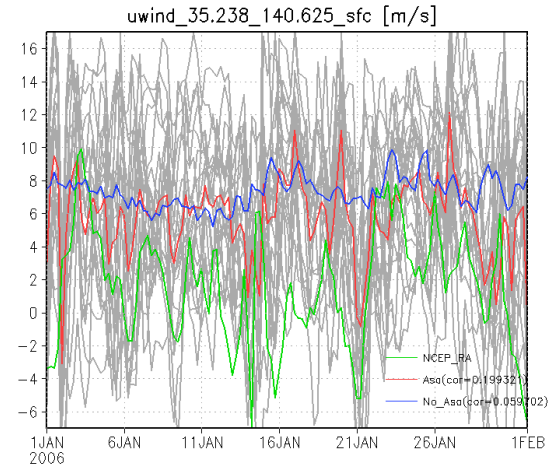
(b) Pressure /Pa



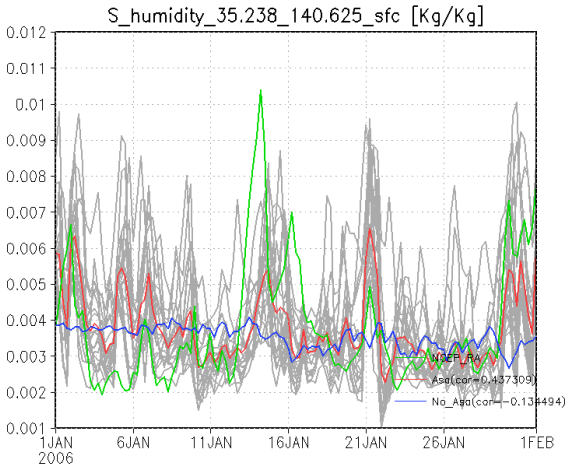
(c) Temperature/ K



Precipitation / (mm/day)



(e) U-Wind /(m/s)



(f) Humidity / (Kg/Kg)

-- Assimilation simulation, -- No Assimilated Simulation, --Observation

Fig. 3.2: RMSE of TCC at Choshi station

Table 3-1: Correlation coefficients for simulations at each station

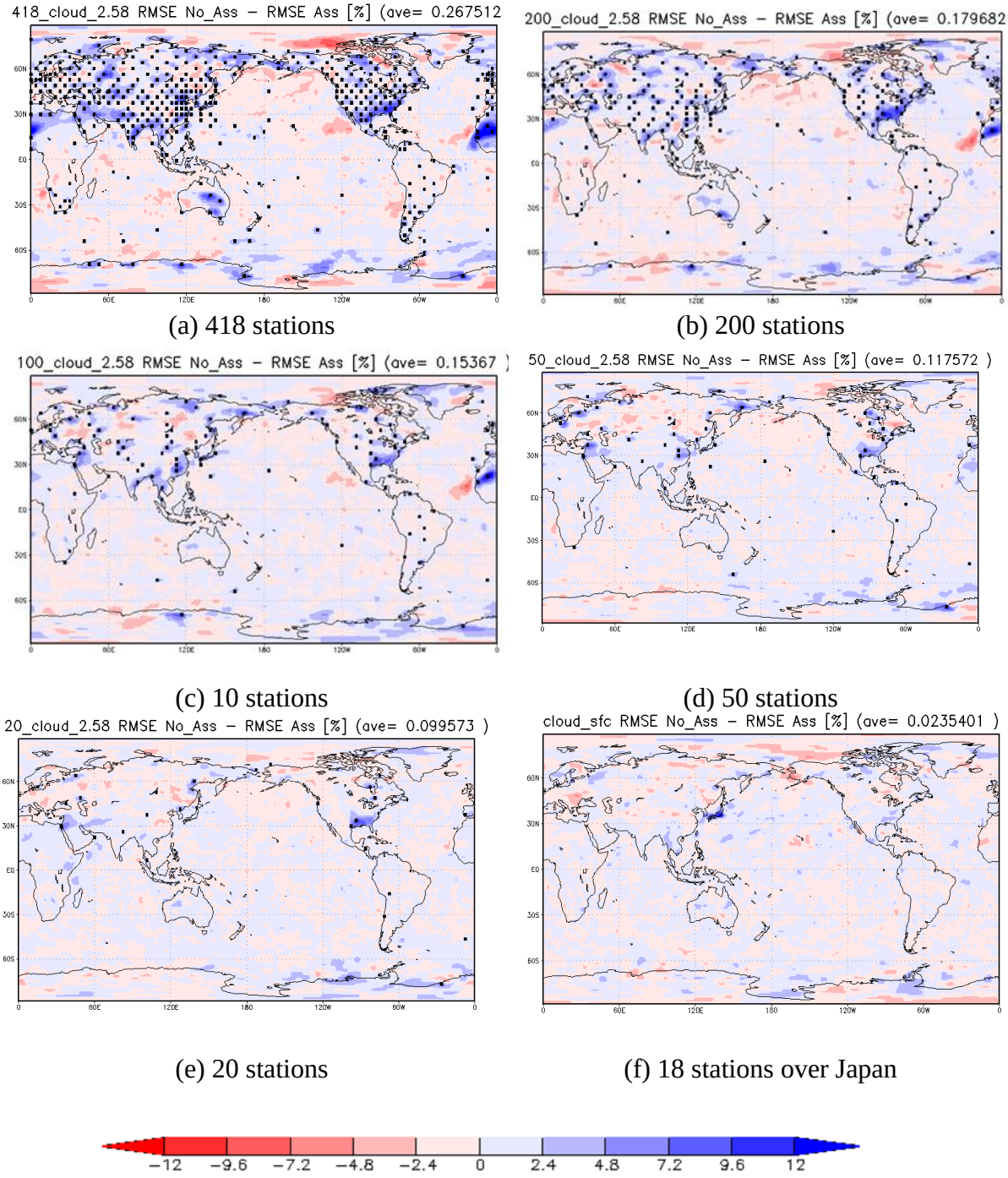
Station	DA	NA
	Correlation	
1	0.317282	-0.01946
2	0.318258	0.071102
3	0.545725	0.163876
4	0.48921	-0.02382
5	0.412977	0.087332
6	0.555097	0.005416
7	0.354535	-0.0829
8	0.501692	-0.02067
9	0.571858	-0.00704
10	0.434217	0.037434
11	0.638069	-0.13648
12	0.540657	0.022072
13	0.46876	-0.03124
14	0.638592	-0.13648
15	0.542312	0.022072
16	0.470525	-0.03124
17	0.311282	-0.31897
18	0.434656	0.188849
Average	0.474761	-0.01167

3.4. Importance of the diaries from other regions

Document data is available over several countries. In the 19th century, instrumentation was available mainly in Europe and few other countries as explained in Chapter 1.1. If the information from all over the world can be utilized with available instrumental observations, a better historical climate can be constructed. We did several experiments to check the performance with different observation numbers of 418, 200, 100, 50, 20 in addition to 18 station experiment over Japan. Synthetic TCC derived from NCEP reanalysis data was assimilated. 30% normalized random error is added to represent high uncertain data. In this study one single simulation was done for each number of experiments to save time. However, multiple simulations would improve the accuracy further removing impact due to random error added to true observations in idealistic simulations.

The global maps in Fig. 3.3 shows the RMSE difference in TCC between the assimilation experiment and no assimilation experiment which represent the skill of the assimilation from 2006 January to March. With an observation number is 418 per day considerable improvement could be achieved. Many areas around observations have an improvement over 10% TCC (i.e.

represented by dark blue). When the number of diaries decreases to 20, improvement limits to local stations. Fig. 3.3 (f) shows the experiment over Japan with 18 stations explained in the previous section. Even though the global change is minimal similar to 20 observation simulation as in Fig. 3.3 (e), a clear regional improvement can be seen over Japan. Around 10% TCC RMSE improvement can be seen over Japan as represented by blue color over Japan. However, the influence on other regions is negligibly small. While some areas have a slight improvement, some areas have worsened the accuracy which may be a result of random correlations.



(a) 418 stations (b) 200 stations

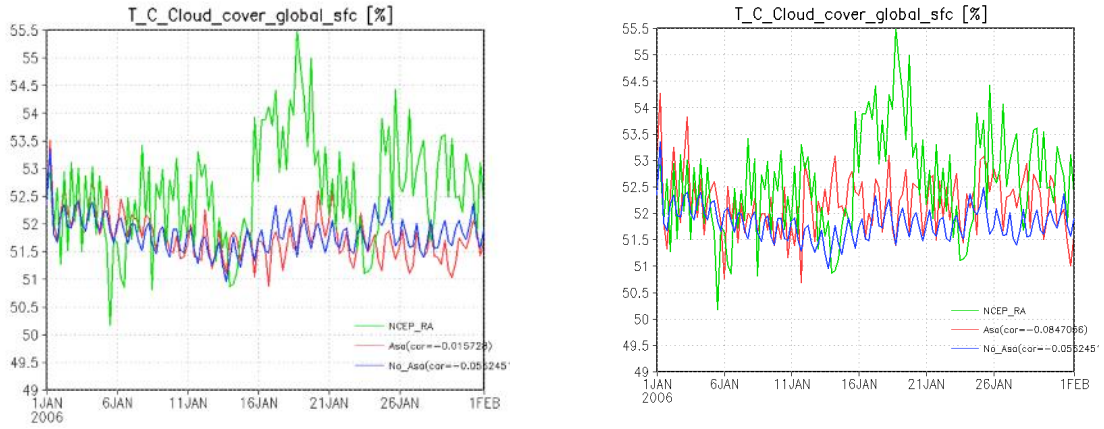
(c) 10 stations (d) 50 stations

(e) 20 stations (f) 18 stations over Japan

Fig. 3.3: Influence of number of observations to model performances, RMSE difference of the assimilation and No assimilation model runs

Fig. 3.4 shows simulations with 418 stations and 18 stations observations (i.e. max and min number of stations). In Fig. 3.4 (a) both Data assimilation simulation and control run has a similar variation which confirms Toride et al., (2017) that by only assimilating uncertain data in a particular region, cannot make a noticeable improvement in a global scale. Performance of non-assimilated variables is also having no improvements on a global scale. In Fig. 3.4. (b)

realistic variation in TCC can be seen, however still the correlation is not improved in the global average.



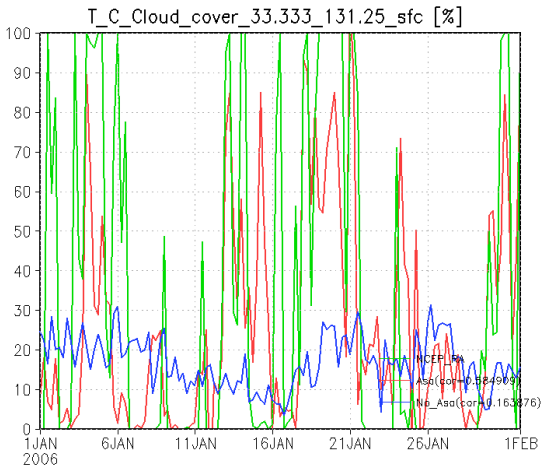
-- Assimilation simulation, -- No Assimilated Simulation, --Observation

(a) 18 stations in Japan (b) 418 stations

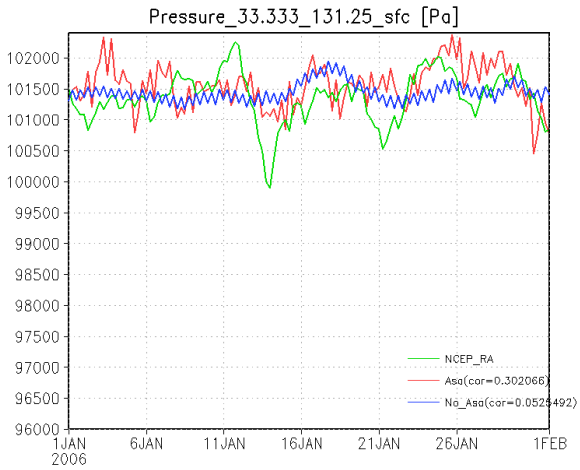
Fig. 3.4: Global mean of TCC (%) in (a) 18 station assimilation in Japan (b) 418 station Data assimilation simulation globally

In global observation data assimilation (i.e. with 418 stations) clear improvement in the assimilated variable and non-assimilated variables at the observation stations can be seen. Fig. 3.5 shows the performance at Oita station (model grid lat. 33.33 lon. 131.25). Summary of performance of 18 station assimilation and 418 stations are given in Table 3-2. Orange color represent the higher correlation among the two. It is clear that 418 has a better performance at Oita station.

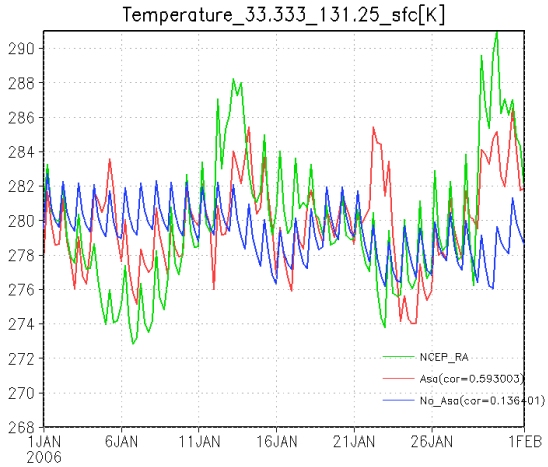
Furthermore, non-assimilated areas are improved. Fig. 3.6 shows the performance at Choshi station where observations were not assimilated. All the variables are improved well compared to the regional observation including this observation site as in figure Fig. 3.6. As expected, improvement in TCC is less than using this station as an observation site directly.



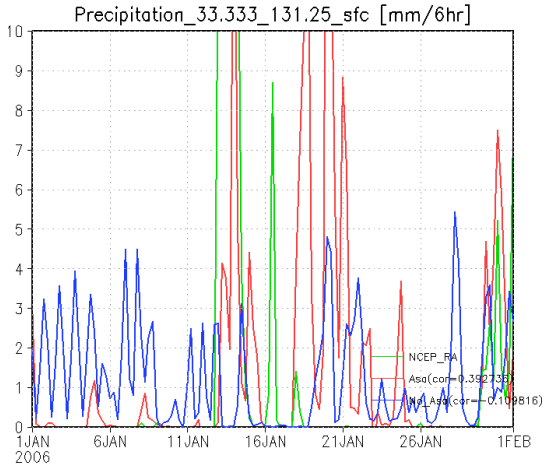
(a) TCC



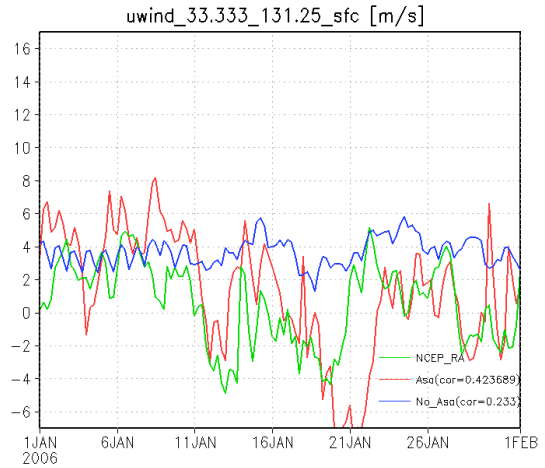
(b) Pressure /Pa



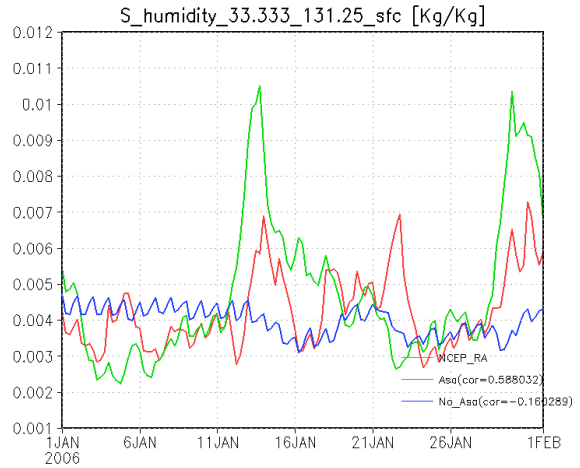
(c) Temperature/ K



Precipitation / (mm/day)



(e) U-Wind /(m/s)



(f) Humidity / (Kg/Kg)

Fig. 3.5: RMSE of TCC in 418 observation assimilation at Oita station

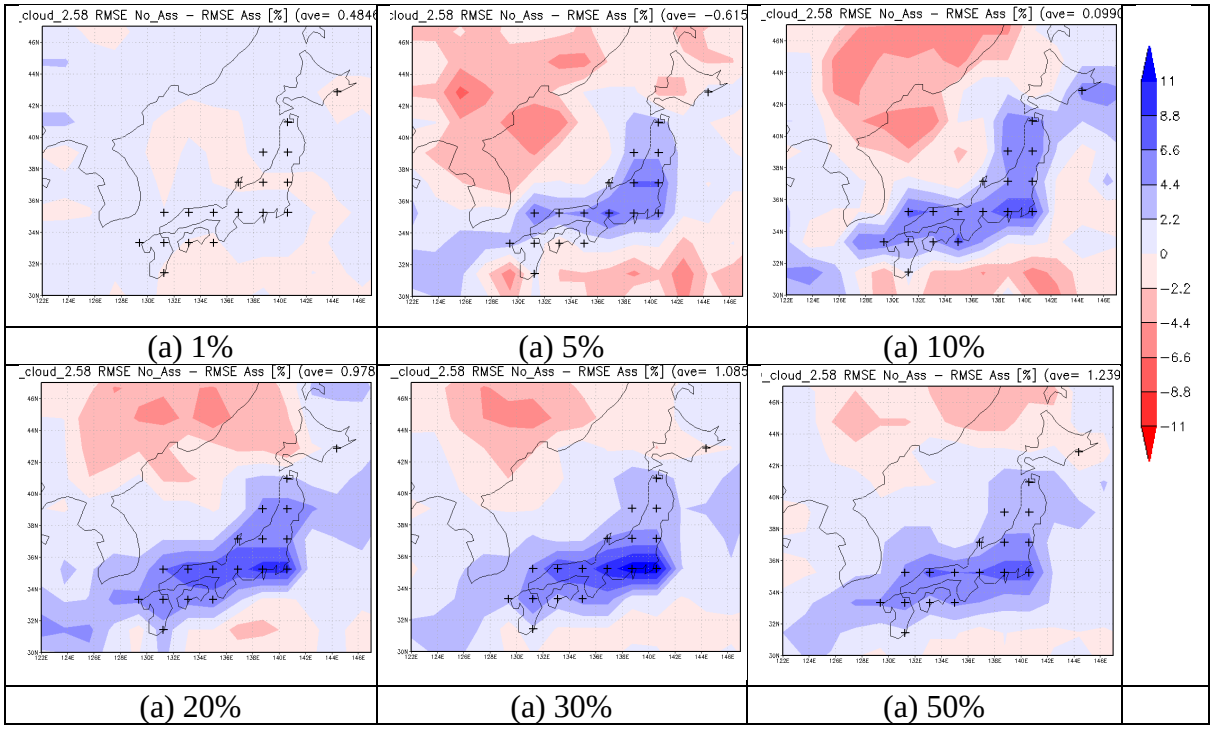
Table 3-2: Summary of the model performance statistics for simulations with different number of observations in Oita station.

No of Observations	TCC%	Precipitation $Kg/m^2/s *10^{-5}$	Humidity / (g/kg)	U / (m/s)	Pressure /Pa	Temperature/ K
RMSE						
No obs.	43.7	6.6	2.2	3.97	437	4.08
18 obs.	35.6	7.2	2.0	3.39	426	4.14
418 obs.	34.4	7.5	1.7	3.54	496	3.18
R						
No obs.	0.16	-0.1	-0.16	0.23	0.05	0.14
18 obs.	0.54	0.27	0.22	0.26	0.35	0.19
418 obs.	0.58	0.39	0.59	0.42	0.30	0.59

Fig. 3.6: RMSE of TCC in 418 observation assimilation at Choshi station

3.5. Observation uncertainty

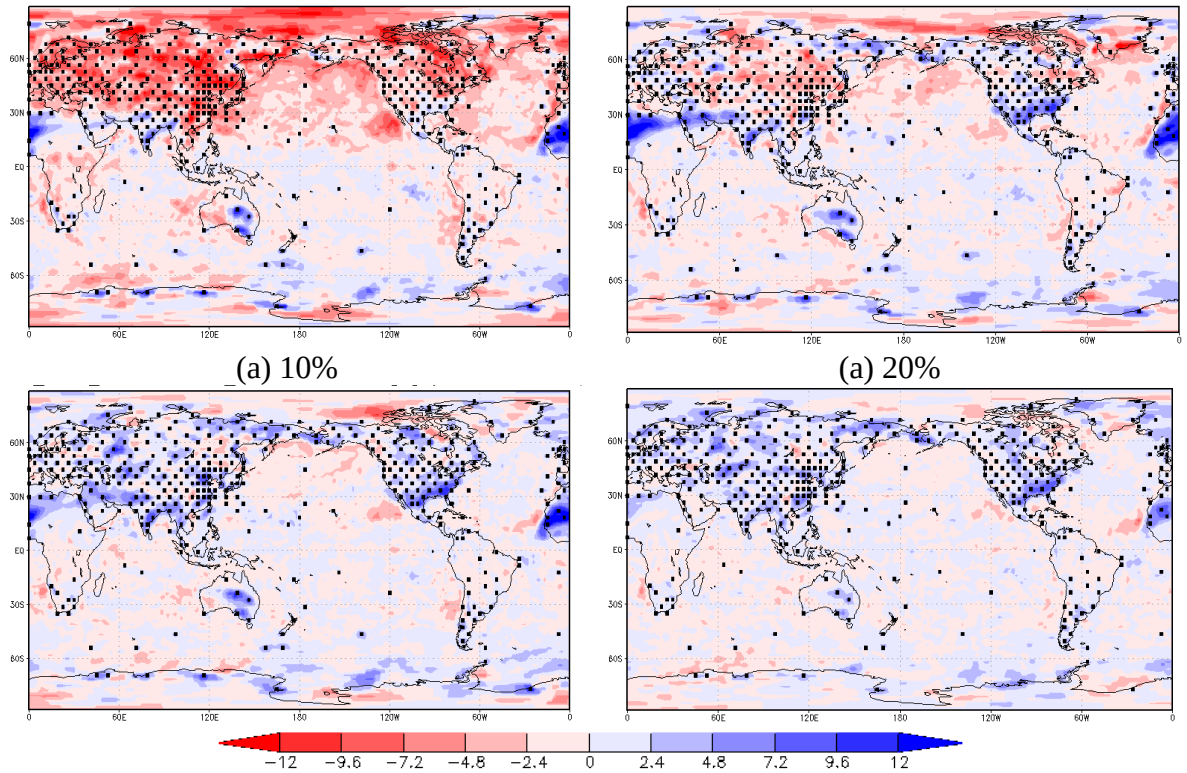
The weather documents have a very higher uncertainty which is represented by the observational error in the model. Hence several experiments were done with random errors of (1%, 5%, 10%,20%, 30% 50%) with a normal distribution. Two types of experiments were done with 18 stations over Japan and 418 stations over globally. Fig. 3.7 and Fig. 3.8 shows the improvement of RMSE in each observation error percentage in regional and global assimilations respectively. The results show that when the observation error is 1% (near to true observation), the results do not improve and when the observation error increases the results improve. Similar results can be seen in global experiments as well in Fig. 3.8. This is due to the model ensemble spread is away from the truth due to the model bias and errors causing the assimilating scheme to neglect the observations. This can be clearly seen in ensembles plots in Fig. 3.9. For example, in Fig. 3.9 (b) because the model ensemble spread has a very low TCC it neglects the observations which have a very high TCC, on the other hand in Fig. 3.9 (a) has a wide ensemble spread enabling it to capture the observation. So even the observations are very accurate if the model is not that accurate data assimilation system would not be very effective. However, in diary assimilation, it would not be an issue because observation uncertainty is much higher than model uncertainties. Table 3-3 shows a summary of the model simulation performance (i.e. RMSE and correlation coefficient) at Choshi station. Those results too explain the same phenomena.



(a) 1% (a) 5% (a) 10%

(a) 20% (a) 30% (a) 50%

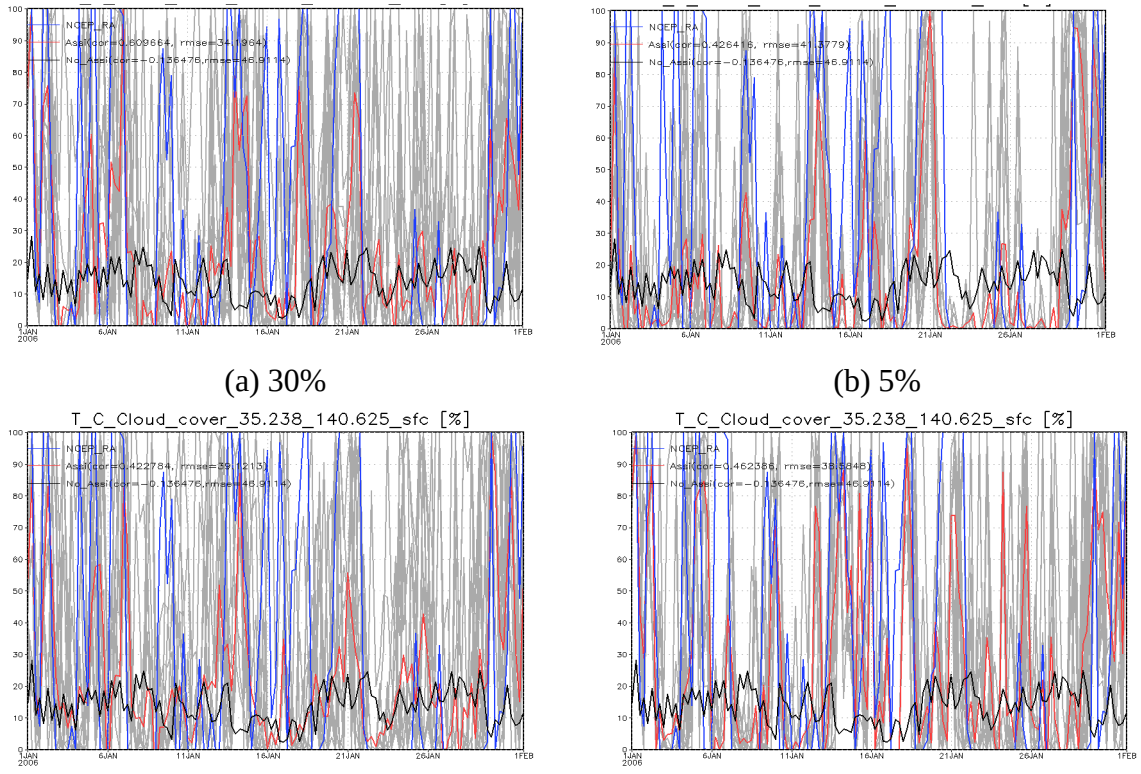
Fig. 3.7: Sensitivity of observation uncertainties, TCC RMSE % difference between assimilation and No assimilation simulations



_(a) 10% (a) 20%

(a) 30% (a) 50%

Fig. 3.8: Sensitivity of observation uncertainties, TCC RMSE % difference between assimilation and No assimilation simulations



(a) 30% (b) 5%

(c) 50% (d) 10%

-- Assimilation simulation, -- No Assimilated Simulation, --Observation

Fig. 3.9: Performance of the model with different observation uncertainties

Table 3-3: Summary of model simulation performance at Choshi station with difference observation uncertainties.

Error %	Point R.M.S.E TCC	Point C.C.
No Assimilation	46.9	-0.14
1	45.37	0.17
5	41.38	0.43
10	38.56	0.46
20	36.96	0.50
30	34.20	0.61
50	39.12	0.42

3.6. Initial conditions

To create the initial conditions for the ensembles, below two methods were investigated as explained in section 2.5.1.

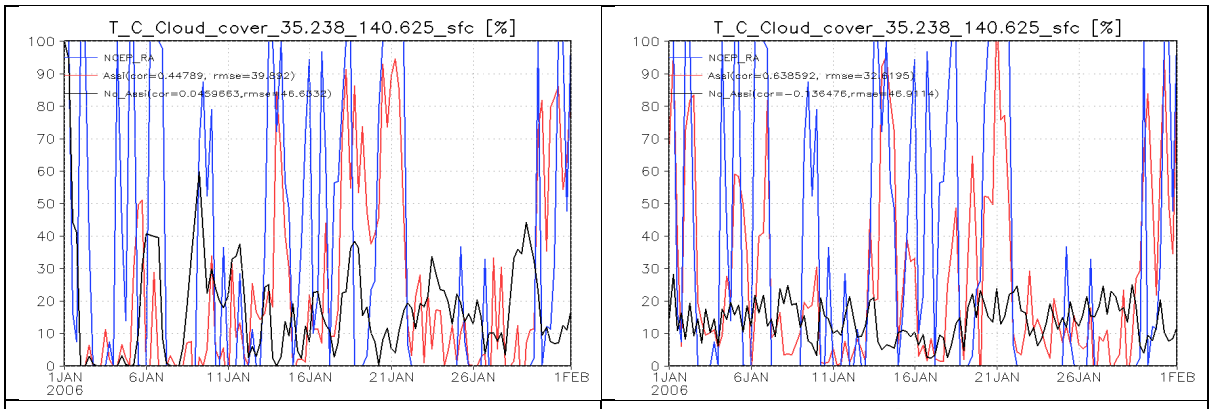
1- Perturbation method - Adding a perturbation to temperature and pressure

For this experiment, perturbation was added by adding a perturbation to temperature and pressure. 1 C temperature and 100 pa pressure variation were added to each ensemble. For example, for an assimilation scheme with 7 ensembles a temperature difference of -3, -2, -1, 0, 1, 2, 3 C and pressure difference of -300, -200, -100, 0, 100, 200, 300 Pa would be added respectively.

2- Timeshift method- Adding a temporal shift

For this experiment, ensembles are chosen from a different continuous period from the model starting date. For example, in this case for each ensemble data from different date since January 1st was used.

The results are shown in Fig. 3.10. The behavior of both assimilation simulations gave similar results, the perturbated simulation has a slightly lower performance with 0.45 correlation and 39.9 % TCC RMSE in comparison 0.64 correlation and 32.6 % TCC RMSE of Time Shift simulations. Similarly, control run without data assimilation also has slightly less performance in Perturbation method. Considering the model performance and usage in previous studies, timeshift method is used to create the initial conditions of the model.



(a) (b)

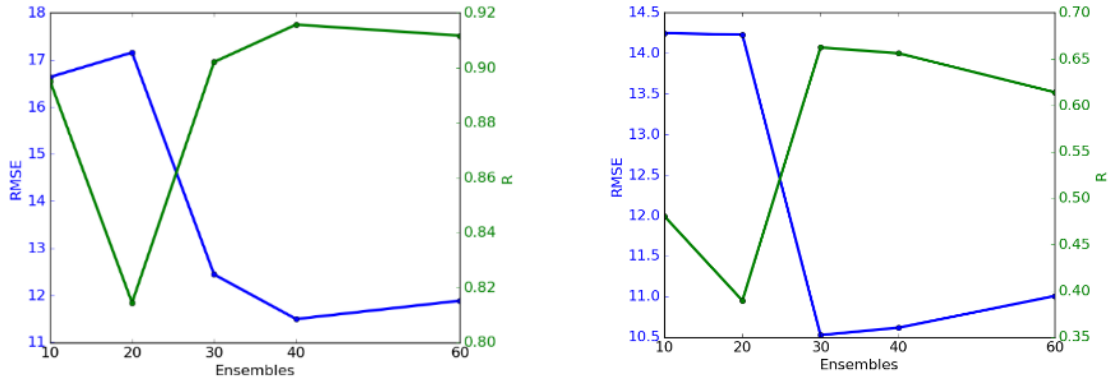
-- Assimilation simulation, -- No Assimilated Simulation, --Observation

Fig. 3.10: Performance of the model with different methods of ensemble initialization, (a) perturbation method, (b) Time shift method

3.7. Ensembles results

Fig. 3.10 show the RMSE and R change with the number of ensembles (a) SR and (b) TCC respectively RMSE decreases in SR and TCC when the number of ensembles increases. However, there is a slight increase in RMSE when the number of ensembles are increased to 60. However, it is not considerable compared to the reduction of improvements from 10 to 30 ensembles (i.e. 16.5 to 12.5 W/m^2). The correlation coefficient is also increased from 10 to 30.

Considering the computational cost 30 ensembles are used for regular experiments and 20 ensembles are used to some sensitivity tests which does not require higher accuracy.



(a) SR (b) TCC

Fig. 3.11: Performance of the model with different number of ensembles

3.8. Conclusion

This chapter checked the performance of the model with similar experiments as in previous literature with Idealistic experiments to evaluate the assimilation system. It was found that by assimilating TCC , correlation coefficient was increased to 0.64 from -0.13 in the control run. Also, the other variables like wind (Correlation 0.23 to 0.25), humidity (Correlation -0.13 to 0.44), temperature (Correlation 0.30 to 0.47) and precipitation (Correlation -0.1 to 0.1), Pressure (Correlation 0.19 to 0.32) was also improved proving that assimilating of a single variable can improve the other variables to some extent. The impact of the observation error was investigated with different values of observation error percentages. As expected even if the observation is very accurate if it stays outside the model ensemble spread the assimilation would not utilize those observations. Thus, the Correlation coefficient was improved up to 30% TCC error (0.61) and tends to decrease afterward at Choshi station. Furthermore, sensitivity to number of observations was investigated, and global RMSE was improved as expected with the number of observations. Sensitivity experiment indicated 30 number of ensembles would be suitable considering the accuracy and the computational cost.

Chapter 4

Preparation for Realistic Past Data Assimilation Experiment

Abstract:

Real weather diary data is entirely different from the regular observations or synthesized observations used in Chapter 3 because they do not have any numbers. Lack of boundary data such as SST and Sea-ice fraction are other main challenges for simulation of forecast model during the 19th century. Currently, there is not any study which has overcome these challenges to assimilate qualitative description data. Hence in Chapter 4, we evaluated the impact from poor boundary condition. Assimilation system's skill was found to reduce mainly in precipitation when low-frequency Sea surface temperature and Sea-ice fraction data are used. Correlation of all station average in Precipitation in 1995 April reduced from 0.58 to 0.32 even though correlation in SR and TCC changed only slightly (i.e. 0.79 to 0.81 and 0.76 to 0.65 respectively). Another challenge is the sensitivity of assimilation time, diary data information is mostly available in daily scale and impact on assimilating at particular time step has not been investigated earlier. Separate experiments showed that assimilation results in morning and evening have only a slight difference. In spring the correlation coefficient of the average of all the stations' changes from 0.54 to 0.43 in Precipitation, 0.66 to 0.72 in SR, 0.66 to 0.73 in TCC and 0.81 to 0.8 in Temperature when assimilation time changed to 3 pm from 9 am. Further, impact to model performance by assimilating only three weather classes data was evaluated in comparison to assimilating TCC from JMA observations with added 30% uncertainty, and it was found that the model could still capture the temporal variation even though correlation of TCC reduced to 0.47 to 0.57 and 0.32 from 0.54 in precipitation in comparison to direct TCC observation assimilation with 30% uncertainty at Choshi station.

4.1. Observation time

Information about the occurring time of weather incidents is only available in limited records. Records like 'heavy rain at 10 am' provides specific information about the time of the rain. Examples for records with time information in weather diaries are given in Table 4-1. However, the information about time is still difficult to utilize as the information about the time is not yet digitized into a usable format and many weather records do not have information about time. In this study information about the time was not taken into consideration, which is a limitation of the current study. Instead, all the information is assumed to occur at a specific time during the day.

Table 4-1: Examples of weather incidents with occurred time in personal diaries

	Description	**English Translation**
1	ヨニイリ / アメ / フル	early **nigh**t / rain / fall
2	ハレ / ケサ マデ / フウウ ノ / トコロ / ミコク マエ ヨリ / ヤミ / ハレ	sunny/ until morning / rain and wind / but / around **10 am**/ (rain and wind) stopped/ sunny (after that)
3	セイテン / ミナミカゼ / フク / ヨナカ / アメ / フル	fine / southerly wind / blown / at night / rain / fall

As this is an online assimilation system, it is computationally costly and challenging to include daily variables in the state vector. Hence the instantaneous values are either assumed as daily averages or calculated empirically from the daily average. In TCC assimilation, it is assumed that instantaneous value at 3 pm is equal to the daily average. This adds extra uncertainty to the observations. For the SR accumulated 6hourly SR was calculated empirically using the daily solar radiation values. A satisfactory relationship could be achieved between the daily average of SR and 3 pm SR. Following empirical equation was found using trial and error method. Here TOA is used to get the proper sub-annual variation. Fig. 4.1 shows the daily SR and SR at 3 pm.

$$hrSR = DSR \times 3.26 / Tx \qquad \text{Eq. 4-1}$$

$$Tx = \left(\frac{TOA}{TOAS} \right)^{.236} \times 1.1$$

Where;

hrSR - SR at 3 pm

DSR - Daily SR

TOA – Top atmosphere Download solar radiation

TOAS- Annual sum of TOA

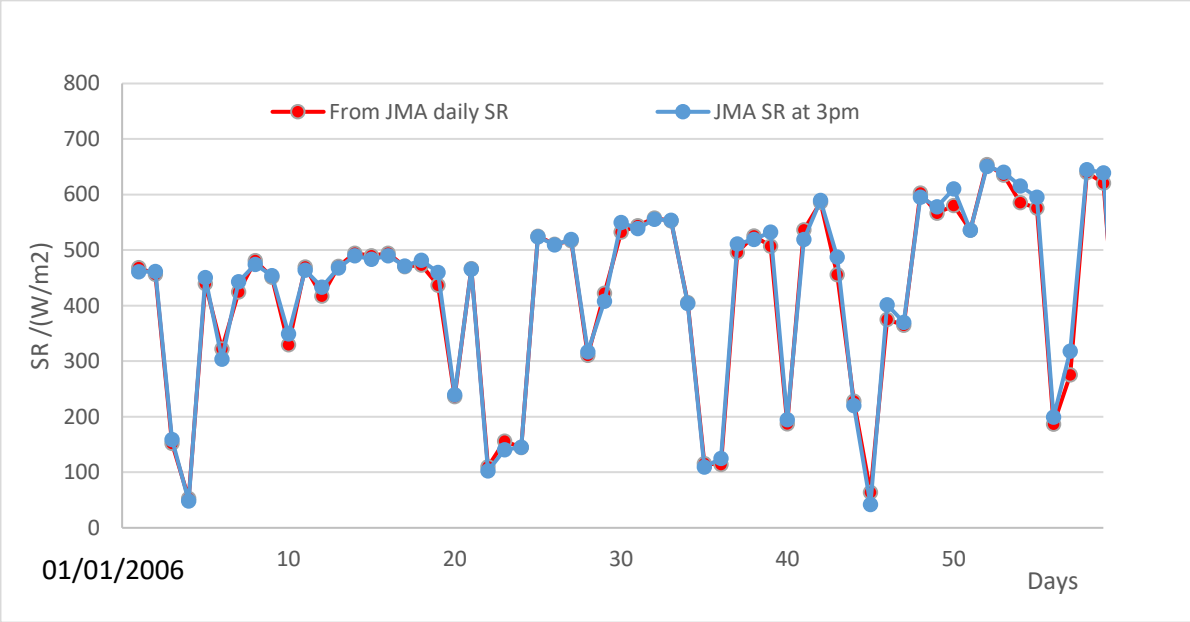


Fig. 4.1: 6hour accumulated SR at 3 pm from January to March 2006 calculated using daily values vs 3 pm JMA SR data

To check the influence of the data assimilation time, separate experiments were done. Assimilation of TCC was done instead of SR to remove strong impact from the diurnal cycle. TCC data from the weather class data were assimilated at 9.00 am and 3.00pm local time

Fig. 4.2 show the influence of assimilation time on the Cloud assimilation. Assimilating weather information in the evening (i.e. 3 pm) has a slightly higher performance than the morning (i.e. 9 am). Both simulations have a similar TCC RMSE values 19.4% and 19.2% respectively. However, Correlation coefficient in afternoon assimilation is higher by 0.1. Similar improvement could be seen in SR. However, correlation is slightly decreased in the precipitation even though the RMSE value is decreased. Assimilation at a particular time may shift the precipitation time causing mismatches in the model behavior. For instance, if the precipitation was in the afternoon, it may be assimilated in the morning because diary data do not always have occurrence time. This may be one reason the change performance in the

precipitation. However, proper time of assimilation may differ from season to season, and more investigation can be done for further investigation. As there is no other alternative at the moment due to the lack of details in diaries assimilation was done at 3.00 pm where more variables get improved.

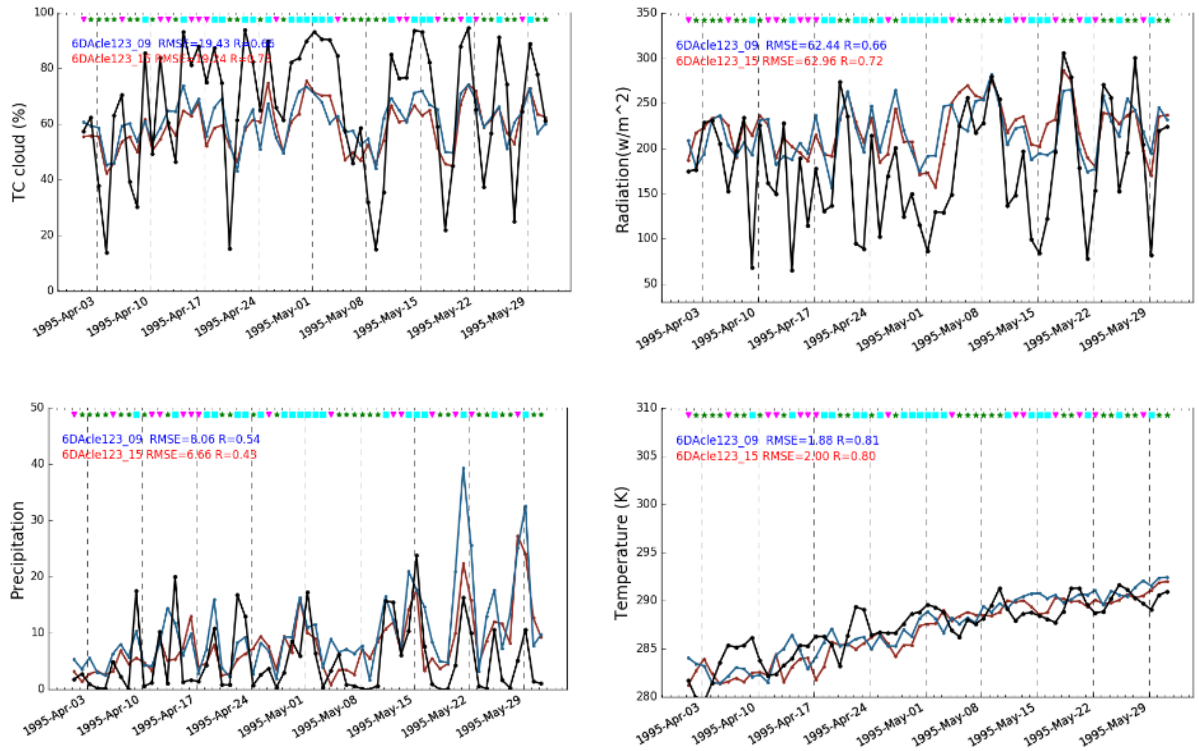


Fig. 4.2: Model performance in 1995 spring over Choshi observation station with the assimilation of TCC at two different times using three weather classes data.

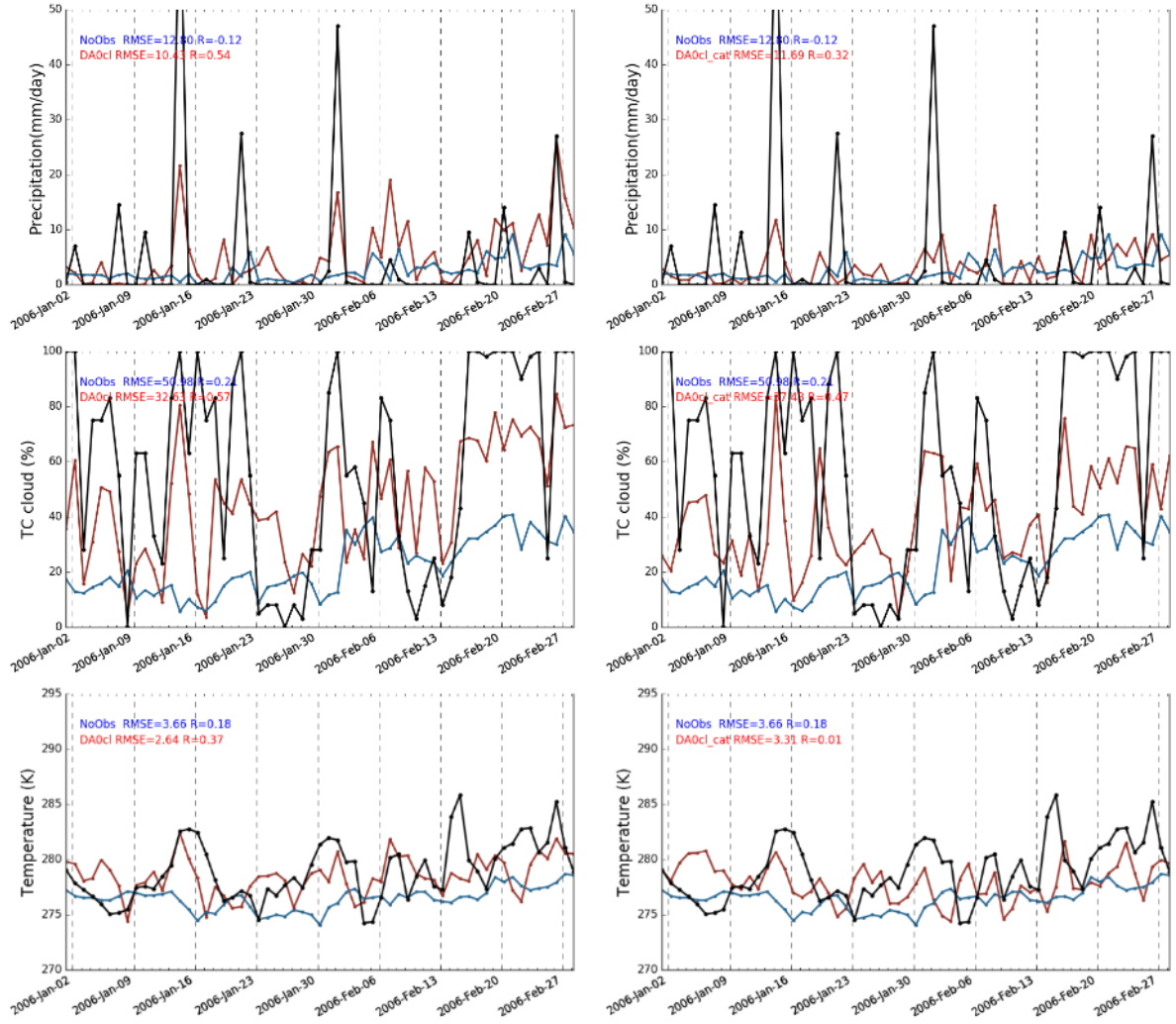
4.2. Impact of assimilating weather classes

In Chapter 3, experiments were done using NCEP data. However, dairy data are local observations and has different character it is to the grid wise NCEP results. Hence. JMA point observations were used to analyze model performance with the assimilation of uncertain weather data further. In the diaries, variations are not linear instead consists with only weather classes such as ‘sunny’ and ‘cloudy’ as explained in Chapter 2. Hence in this section assimilation results of weather classes and exact values are compared using below two experiments.

a- Using exact JMA TCC values

b- Using TC weather classes made according to (Toride et al., 2017)

According to the Fig. 4.3 when only weather classes are given due to the higher uncertainty model produce lower estimations even though it could still capture the temporal variation well. Model miss the precipitation event on February 27 due to underestimation of TCC and in both two major precipitation events on 15th January and 2nd February the precipitation amount decreases. The accuracy of precipitation reduced considerably. The correlation coefficient reduced to 0.01 from 0.37. All the station average statistics in Table 4-2 shows the reduction in all the stations averages as well. Hence it is clear that when the observations are limited to weather classes information, the accuracy reduced. However most importantly still it improves correlation and RMSE values than no assimilation simulation,



(a) TCC assimilation using direct JMA observations

(b) TCC assimilation using simple weather classes from JMA observations

Fig. 4.3: Performance with simple JMA weather classes

Table 4-2: Summary of the model performance statistics for simulations with JMA observations and simplified weather classes.

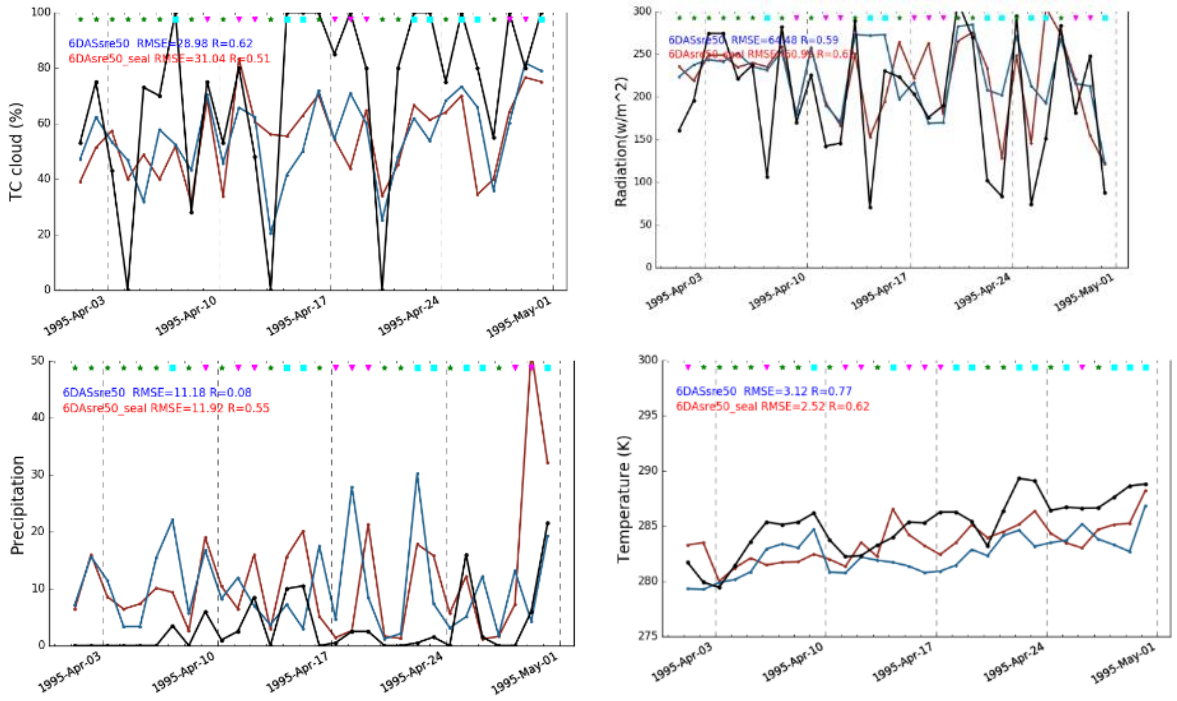
Observations	TCC	
	RMSE %	R
Average of all stations		
No observation	30.9	0.21
Weather classes ass.	20.7	0.58
Direct ass.	16.9	0.62
At Choshi station		
No observation	51.0	0.21
Weather classes ass.	37.4	0.47
Direct ass.	32.63	0.57

4.3. Boundary conditions

Model performance over monthly SST and weekly NCEP OI SST is shown in Fig 4.4. The blue line shows Control simulation without data assimilation, and the red line shows the assimilation simulation. The assimilated weather classes are shown on the top (i.e. blue square-Rainy, Pink Triangle –Cloudy, Green star – Sunny). The impact from different SST would be discussed in this section. There are several weather events in this period. Assimilation system's skill capturing those events are discussed below the change in the RMSE difference in an assimilated variable (i.e. SR) in Fig 4.4 is smaller (3.5 W/m^2 increase), and R change is negligible (0.03), and in Temperature, RMSE is worsened by 0.7 K, and R improved by 0.12. In TCC RMSE improved by 2% and R improve by 0.11. However, in precipitation, there is a considerable reduction in correlation (reduced from 0.55 to 0.08). This is visible in the precipitation graph. The blue line has an inferior skill compared to the red (with accurate SST and Sea-Ice fraction).

Idealize experiments with JMA TCC observations was also done. The correlation coefficient was increased to 0.6 from 0.2 at Choshi station, and daily RMSE was decreased from 51 TCC to 33 TCC. The remaining large RMSE was due to the bias of the model. Precipitation degradation is apparent in average of all the station.

Table 4-3 as well. This may be due to change in the weather pattern due to poor quality SST.



(a) (b)

-- Assimilation with monthly Sea-Ice & SST, – Assimilated with weekly SST & Sea-Ice, --Observation

- Rain - Cloudy - Sunny

Fig 4.4: Influence of SST temporal resolution on model performance. The red and blue lines indicate the simulations results using SST from NCEP-OI SST (weekly SST & Sea-Ice fraction) and Franke et al. (2017) (monthly SST & Sea-Ice fraction) respectively at Choshi station from April 1995 to May 1995.

Table 4-3: SST impact on the average of all the observation stations

	RMSE			
	TCC (%)	Radiation(w/m^2)	Temperature K	Precipitation (mm/day)
SST weekly	19.58	55.03	2.52	5.11
SST monthly	19.15	53.27	3.12	6.61
	R- correlation coefficient			
	TCC	Radiation	Temperature	**Precipitation**
SST weekly	0.67	0.79	0.62	**0.58**
SST monthly	0.65	0.81	0.77	**0.32**

In conclusion low-quality SST in 19th century significantly reduces the precipitation forecast.

4.4. Conclusion

Assimilating daily weather information at a particular time add further uncertainty and it was found that assimilating in the afternoon has slightly higher performance (i.e. the correlation coefficient increased to 0.75 from 0.65). Assimilating weather classed gave minor improvement as expected but still provide higher accuracy than no assimilation simulation. Their assimilation increased by 4.2% and correlation decreased by 0.04. It was found that poor SST and Sea-Ice fraction data lead to reduce the model performance especially in precipitation (average R reduce from 0.6 to 0.3). This gives the message that when the SST and SI products improve the future, historical reconstruction results would be improved further.

The simplified weather classes data too added some improvement even with the 30% error. TCC RMSE% average of all the stations was decreased to 33% from 51%. The correlation was improved by 0.26, and there was no improvement in temperature and correlation reduced than the control simulation. In precipitation there RMSE was improved by only 1.1 TCC % even though correlation was improved by 0.2. In conclusion assimilation of uncertain weather information (i.e. TCC classes) can improve the assimilating variable while and correlation of few other variables even though improvement is less than assimilating exact values. In the absence of accurate observation, these document data would be an excellent source of information to climate models.

This chapter has overcome the barriers against assimilating real weather description data for the first time in weather diary assimilation field. Those findings would be utilized for the weather description data assimilation in Chapter 5.

Chapter 5

Application of Proposed Data Assimilation System and Validation

Abstract:

This chapter evaluates the skill of the model in assimilating document weather data. The main limitation was the lack of instrumental data in the past. Hence an alternative approach was followed by assimilating weather classes data derived from recent description data. All these experiments were carried out with real data keeping the consistency with 19th Century data quality. This is the first study to carry out such realistic experiments to investigate the performance of assimilating weather class data into a climate model. Several simulations were done in the recent period where observations data available for validation. Twentieth-century weather classes, data derived from JMA descriptions was utilized. SR assimilation could improve the correlation of TCC average in all the stations from 0.19 to 0.68 in spring while reducing RMSE by 8 %. Improvements in other seasons and fields such as precipitation could be achieved as well. Further, we investigated opportunities to improve the accuracy of the model by incorporating other information such as the absence of precipitation and found the correlation of precipitation in all the station average could be improved to 0.67 from 0.45 in spring. Monthly anomaly values over 1995-1999 showed good correlation in precipitation, TCC, and SR. By analyzing pressure fields, it could be shown that the model could capture the synoptic scale weather patterns such as extratropical cyclones. Bootstrap experiments were done using only half observations to check model performance when some diaries are absence. Even though the model performance was reduced to some extent, satisfying correlation could be achieved. Correlation of all the stations average in TCC was 0.57, in SR was 0.72 and in precipitation was 0.45.

5.1. Experiments in 1995

5.1.1. Solar Radiation Assimilation

The Fig. 5.1 shows the regional average in TCC, SR, Precipitation, and Temperature. Here spring period is shown as it has a subtle variation in weather with several precipitation days and sunny days. The blue line shows Control simulation without data assimilation, and the red line shows the assimilation simulation. The assimilated weather classes are shown on the top (i.e. blue square- rainy, Pink Triangle – cloudy, Green star – Sunny) There are several weather events in this period. Assimilation system's skill capturing those events are discussed below.

There are two longer cloudy periods (three days of weather class 3) from 28th April to 4th May and 11th May to 18th May (three days with weather class 3 and two days with weather class 2) which are shown by purple dotted rectangles. The observed SR and TCC clearly show low SR and high TCC in this period. The model could nicely capture both these events. In both of these periods, precipitation present in the observations, the model was able to produce the precipitation in the first period. However, the model could not capture the precipitation peak in the second period as the SR and TCC on 1st May was not constrain well by the data assimilation.

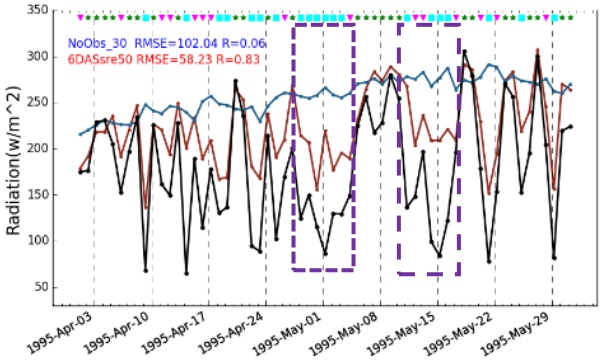
There is a longer sunny period from 5th May to 11th May (continues weather class 1). The observation shows very low cloudiness and very high radiation during this period without any precipitation events. The control run could not capture this variation and in the Model simulation with data assimilation could undoubtedly produce high solar radiation and lower cloudiness with negligible precipitation events.

Other than in above weather events there are nine precipitation events in this period which has more than 10 mm/day precipitation (04/09, 04/12, 04/14, 04/19, 04/22, 04/23, 05/21, 05/25, 05/29, shown by purple downward arrows). During all these days the weather class is three except April 12th which has weather class 2. The model could not capture only three precipitation events including the event on April 12th (04/09, 04/12, 04/19).

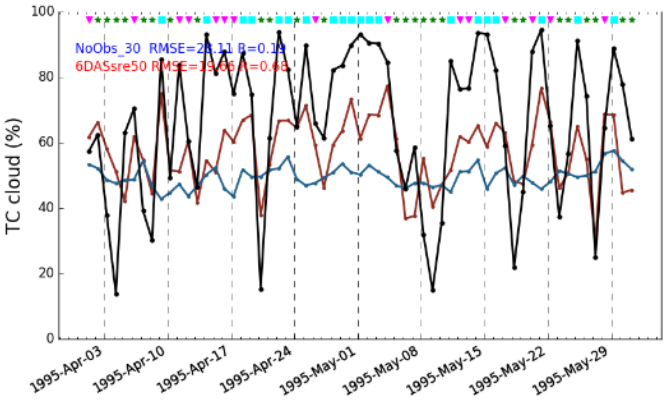
In overall weather class assimilation could constrain the SR of the model throughout the period capturing all the local events. The RMSE value was decreased to 58.2 W/m^2 from 102.0 W/m^2 which is almost half of the control run. The control simulation without assimilating SR has a marginal correlation value of 0.06 which is very low compared to the 0.83 correlation in the simulation with data assimilation. Non-assimilated variables also improved. In TCC, the

model's correlation improved from 0.19 to 0.68 and RMSE reduced to 20% from 28%. However, the model has underestimation in all the cloud peaks which may be due to the large size of the grid compared to single observations sites. In precipitation model capture many precipitation events but few unrealistic precipitation events occurred as in event near April 17th. Overestimation in precipitation in most of the precipitation events occurred which may be due to the impact of lowering the radiation and increasing the TCC than the model's usual range caused from assimilating from point observations which may have a bias between the large grid of the model and the observations or shift in synoptic scale event location. This is further investigated in section 5.1.3. Due to the overestimation, the RMSE value was only slightly improved even though the correlation coefficient increased to 0.45 from negligible correlation in the control run. On the other hand, in temperature even though several changes in daily temperature could be captured (green arrows) some variations could not be captured (orange arrows). The control simulation does not have a noticeable daily variation. Hence the RMSE is less as it lies along the mean of the observations. Both simulations have a good long-term seasonal correlation as the model has the skill to produce seasonal variations without data assimilation.

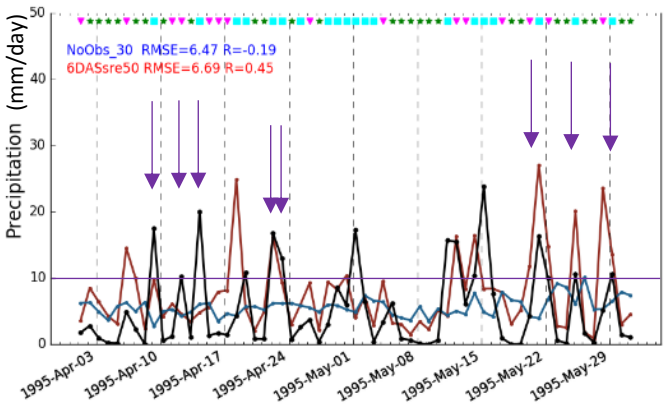
In summary, the simulation could capture most of the observed incidents. The missed event may be due to the observation uncertainty and poor boundary data.



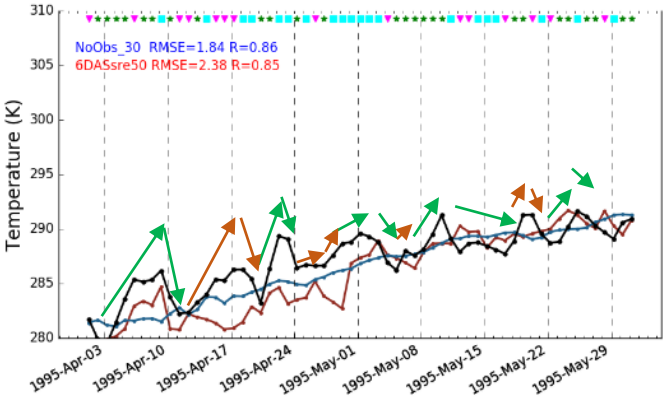
(a) SR



(b) TCC



(c) Precipitation



(d) Temperature

-- No Assimilation, – Assimilated, --Observation - Rainy - Cloudy - Sunny

Fig. 5.1: Model performance in 1995 spring over all the station average with the assimilation of 3 classes (i.e. Rainy- Class 3, Cloudy- Class 2, Sunny Class -1) of weather data.

5.1.2. Performance in local observation stations

Model performance is evaluated over several observation stations. Fig. 5.2 shows the weather class assimilation results in Choshi station during the same period as in the previous section. The data assimilation simulation could capture the temporal variation quite well by decreasing the RMSE from 111 w/m^2 to 76 w/m^2 relative to the control run without data assimilation. On the other hand, the control simulation has a very low correlation compared to the correlation of 0.57 in the SR. Similarly, the other variables too have a similar performance as in the average results.

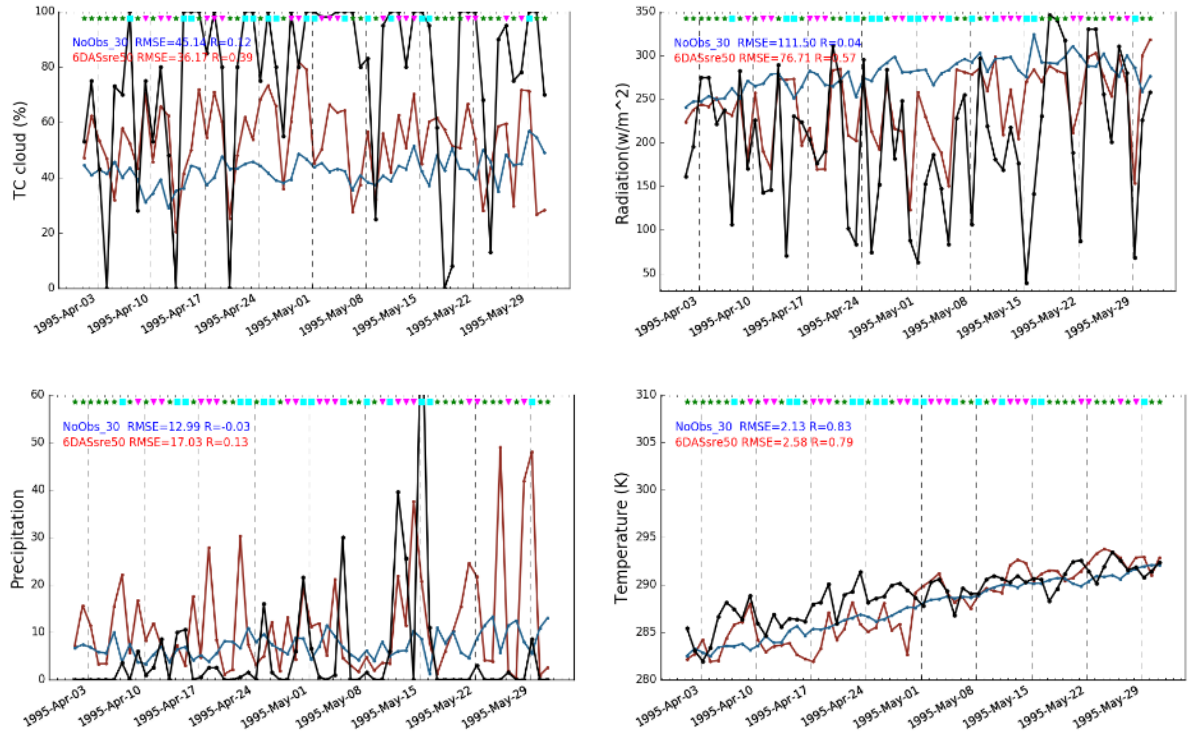


Fig. 5.2: Model performance in 1995 spring over Choshi observation station with the assimilation of 3 classes of weather data.

5.1.2.1. Performance in different regions

Japan was divided into four areas as shown in below figure.

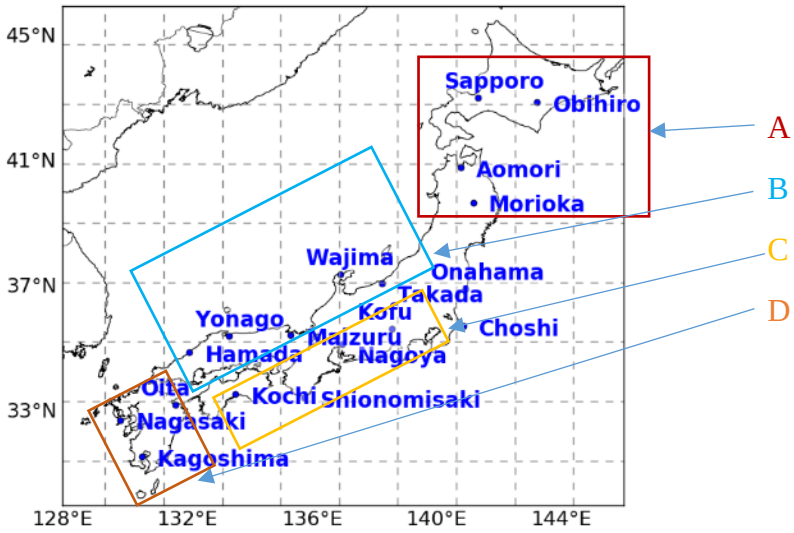


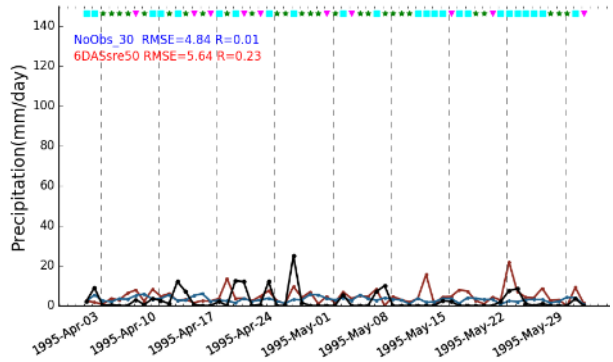
Fig. 5.3: Regions

Table 5-1: Stations covered in each region and improvement to the correlation coefficient.

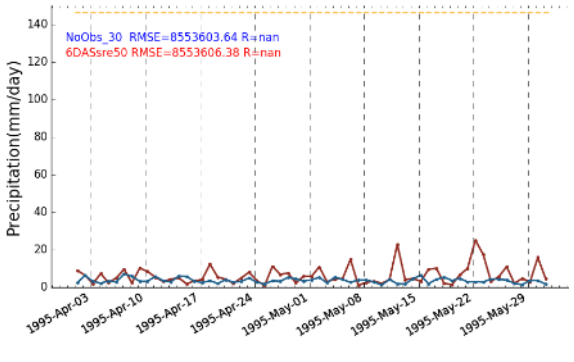
Region	Station and improvement in Correlation coefficient
A	Sapporo (0.22), Obihiro (NA), Aomori (0.33), Morioka (0.17)
B	Yonago (0.5), Maizuru (0.25), Wajima (0.27), Takada (0.11)
C	Kofu (0.43), Choshi (0.16), Nagoya (0.49), Shinomisaki (0.43)
F	Kagoshima (0.57), Nagasaki (0.2), Oita (0.23)

Below sections show the results in region wise. Considering the figures Fig. 5.4- Fig. 5.7 and table 1, the region A has less intense precipitation events, which is due to less frequency of extratropical cyclones (ExT cyclones) in region A during this period. This area has a lower correlation which may be due to the weather in this area is not governed by synoptic scale events like other regions and has fewer observation stations relative to the west. In region B too, stations in the east side have a lesser correlation.

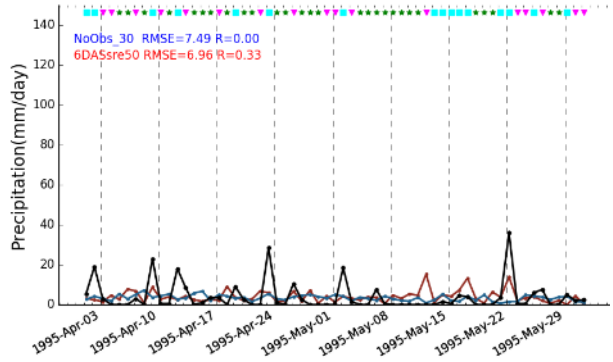
Region C and D have a higher correlation which is due to the higher influence of synoptic scale events such as ExT cyclones over that area and a higher number of observation stations.



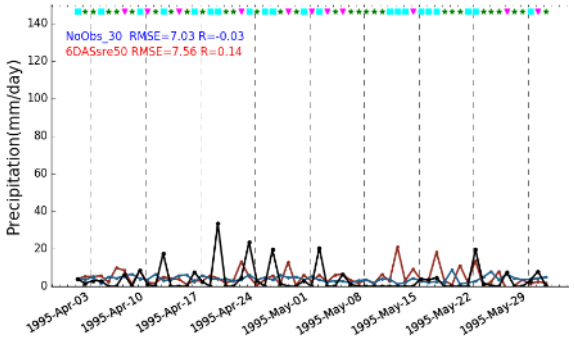
Sapporo



Obihiro

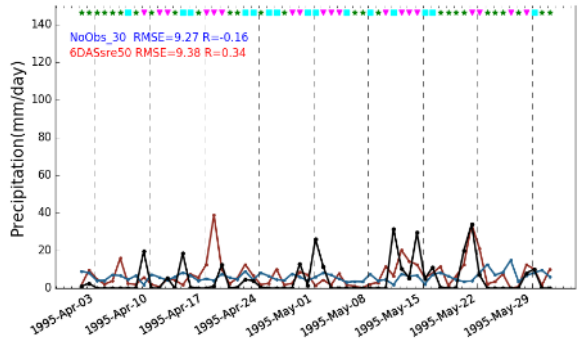


Aomori

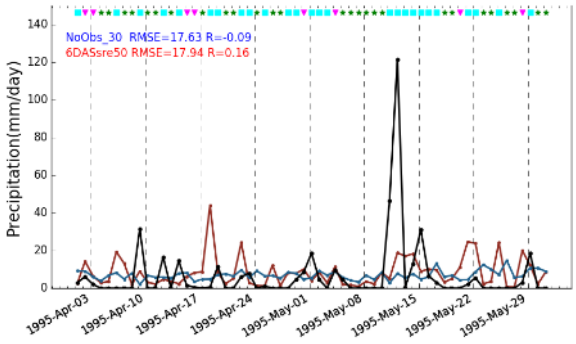


Morioka

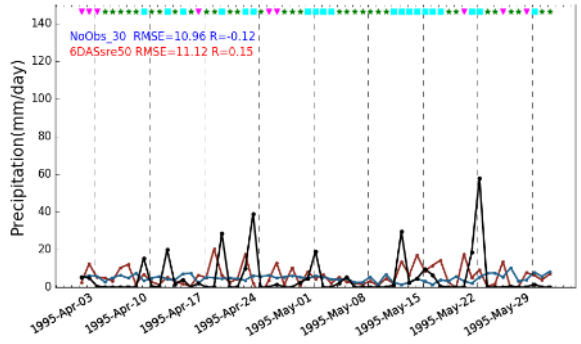
Fig. 5.4: Region A



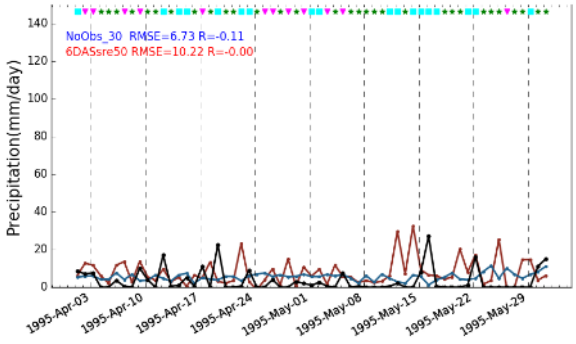
Yonago



Maizru

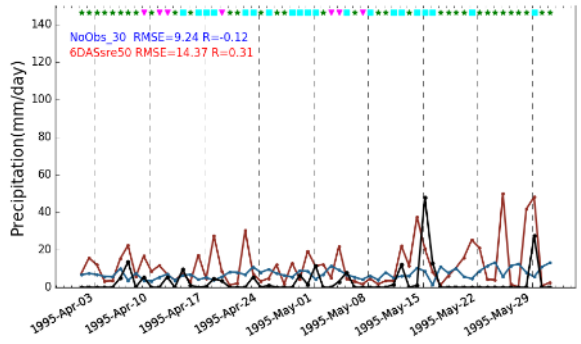


Wajima

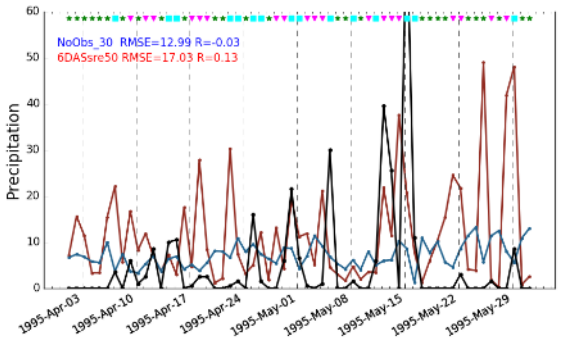


Takada

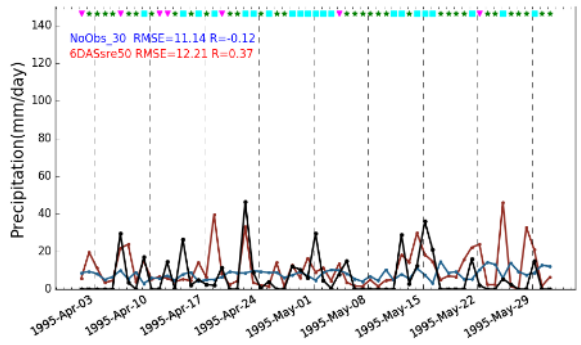
Fig. 5.5: Region B



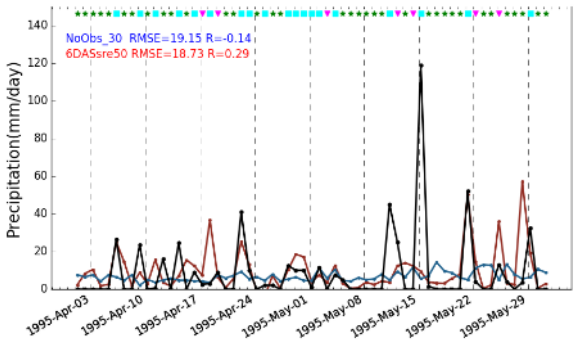
Kofu



Choshi

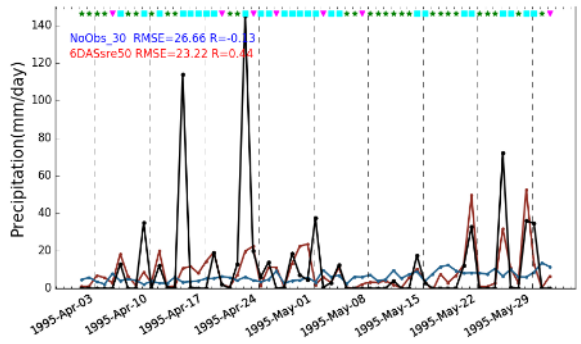


Nagoya

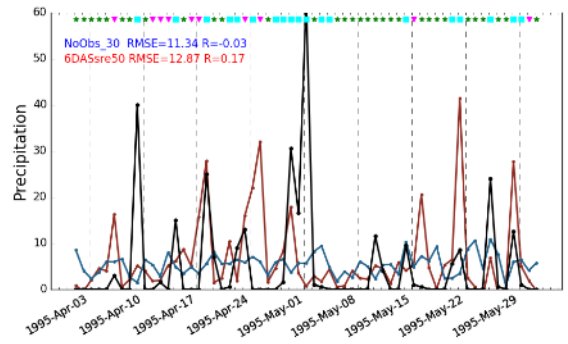


Shinomisaki

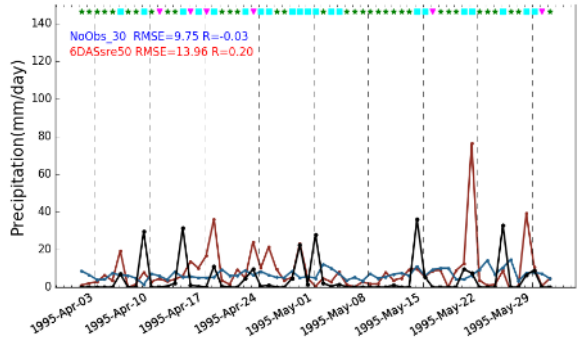
Fig. 5.6: Region C



Kagoshima



Nagasaki



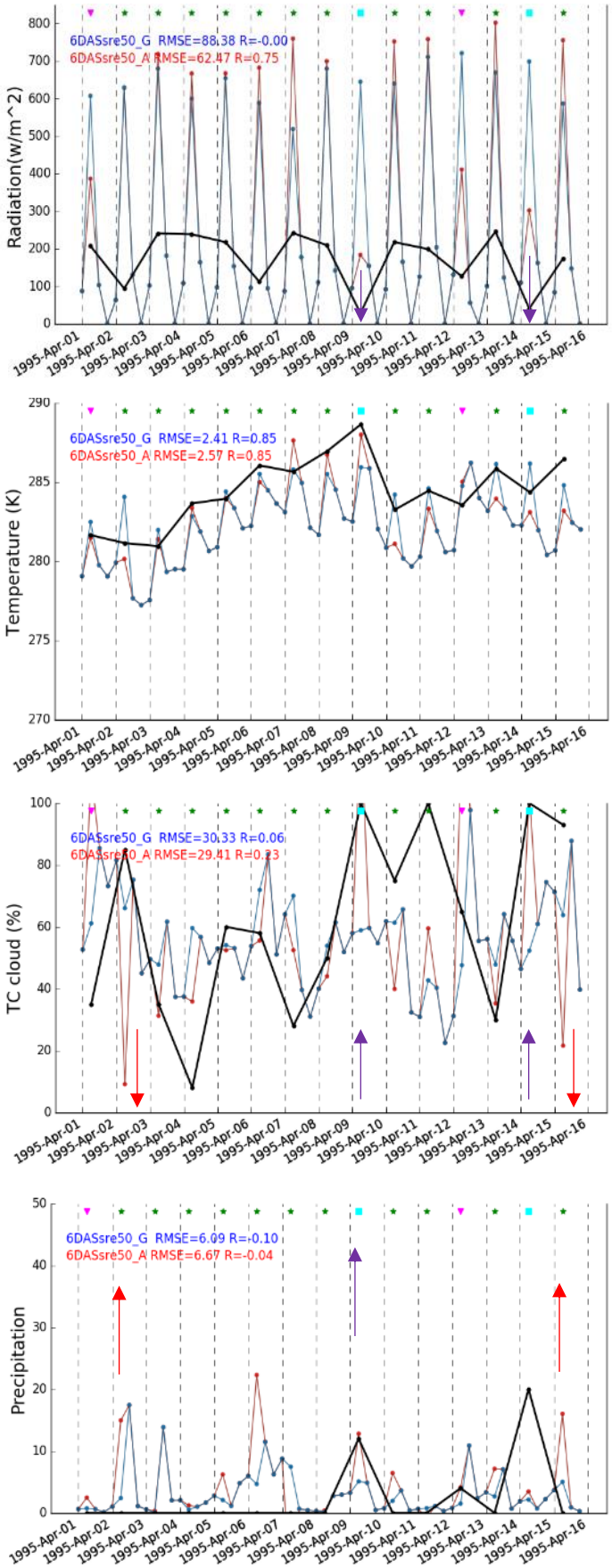
Oita

Fig. 5.7: Region D

5.1.3. Precipitation estimation skill of the assimilation system.

To investigate the skill of the GSM forecast, GSM's guess (forecast before data assimilation) and the analysis (after data assimilation) were compared. The **Fig. 5.8** shows the 6 hourly results of the guess simulation and the analysis simulation in the Hamada station from 1st April to 16th April. During this period there were two days with class 3 weather on April 09th and April 14th. Both days SR was decreased after the assimilation and 9th reduction is higher than 14th. In both days reduction in observed daily SR and presence of precipitation events were observed. Both days model TCC increased. However, both days model guess does not forecast a precipitation event, but the analysis could capture observed precipitation event on April 9th. Less reduction of Solar radiation on 14th might be a reason. During the sunny days, analysis tends to increase the guess solar radiation slightly while reducing TCC. Even though there is no significant change to the solar radiation during sunny days considerable variation in observed TCC can be seen. There is an unrealistic precipitation event on 6th April in the analysis. Diary indicated sunny during this day. However, SR was reduced and TCC was very high even though there is no precipitation event in observations. Artificial change to TCC and solar radiation may be the reason for the unrealistic precipitin event. Also, the guess has several unrealistic precipitation events, and it is complicated to find a reason. A probable reason may be due to poor SST quality.

There is an increasing trend in the temperature up to 9th April. The guess run could capture this up to 7th April and then it starts to decrease, but the assimilation run could capture the increasing temperature up to 9th April and the sudden drop on 10th April.



-- Guess (6hourly) – Analysis(6hourly) -- Observation (JMA) -daily

Fig. 5.8: **S**ix hourly model guess and analysis performance.

5.1.3.1. Unrealistic heavy precipitation

Two events with higher estimation of precipitation during weather class one and three are analyzed further.

Table 5-2: Events with heavy precipitation

Date	Weather Class	TCC (JMA)	Precipitation	Model Precipitation
May 25th	1 -Sunny	>70%	0	Very high
May 29th	3- Rainy	100%	10 mm/day	

These two events are identified from the times series variation of precipitation in the following figure as indicated by the red arrows. The Solar radiation and TCC variation on these days is typical. By checking the pressure map plots on May 25th and May 29th high TCC can be observed over Japan in reanalysis data and in the model simulations as well. The reanalysis data has the ExT cyclone towards north-west Japan and a small low-pressure area over east Japan. The assimilation introduces an excessively low pressure around the Kanto area, an ExT cyclone over Kanto region causing heavy precipitation while the observed precipitation is smaller.

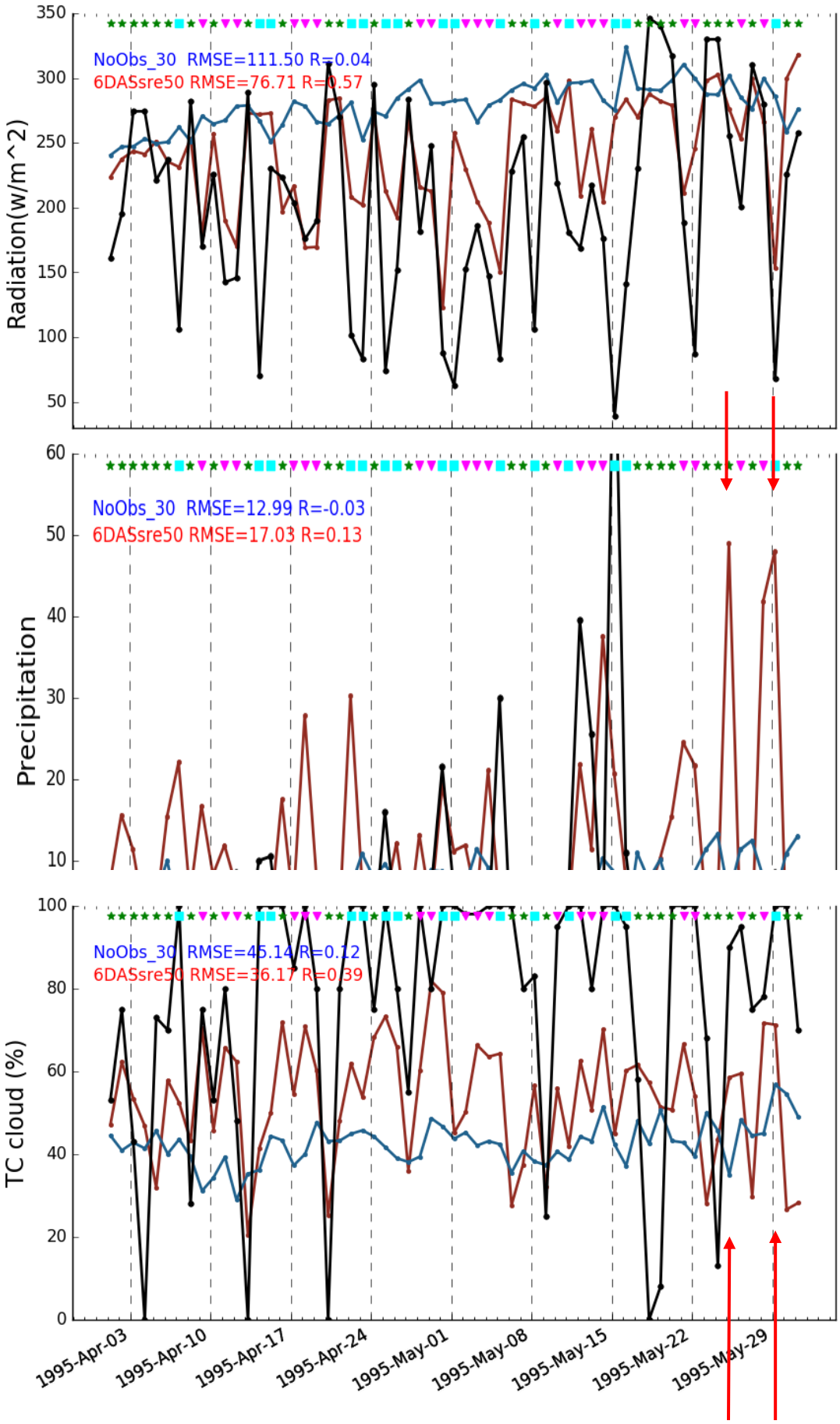
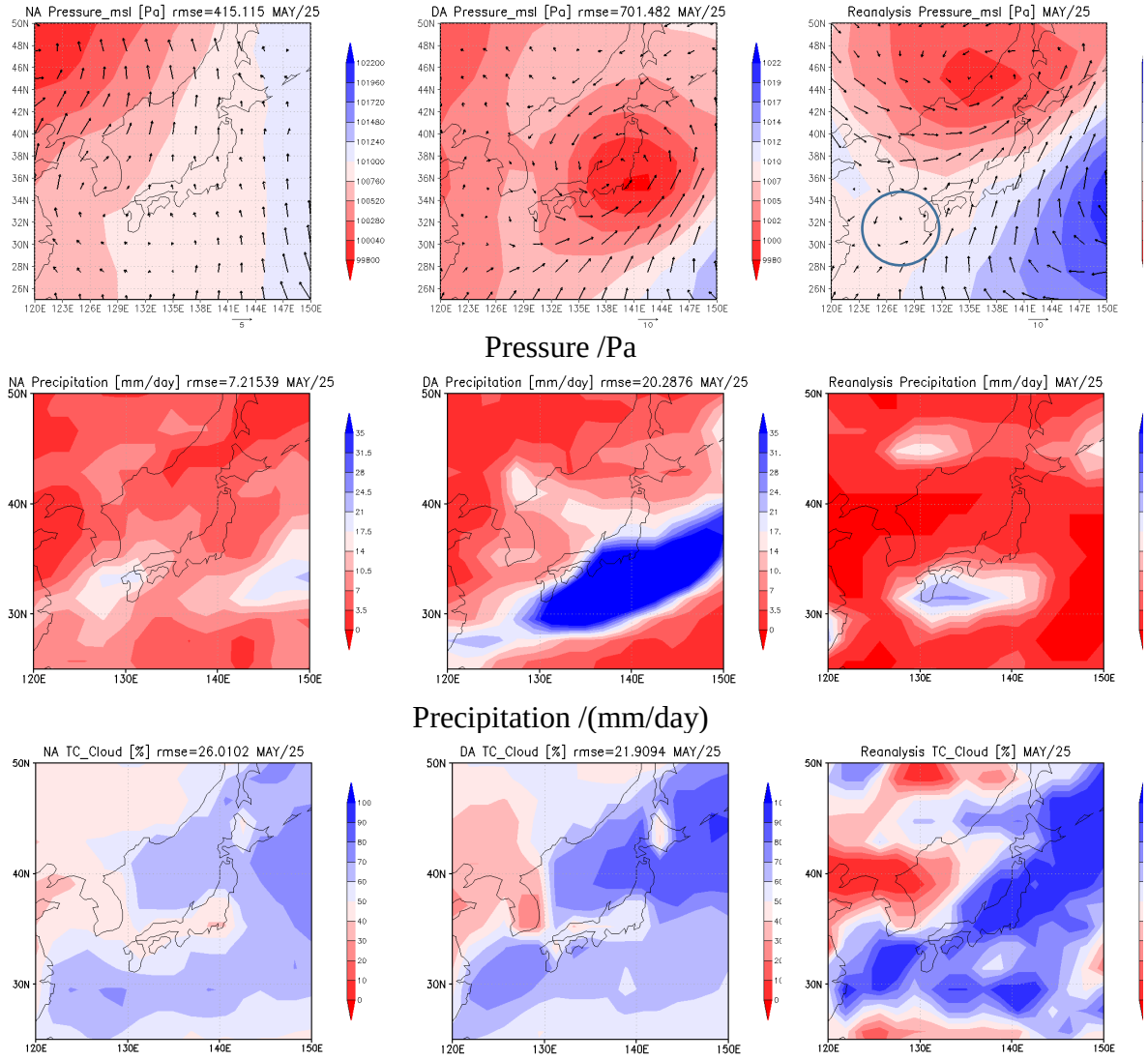
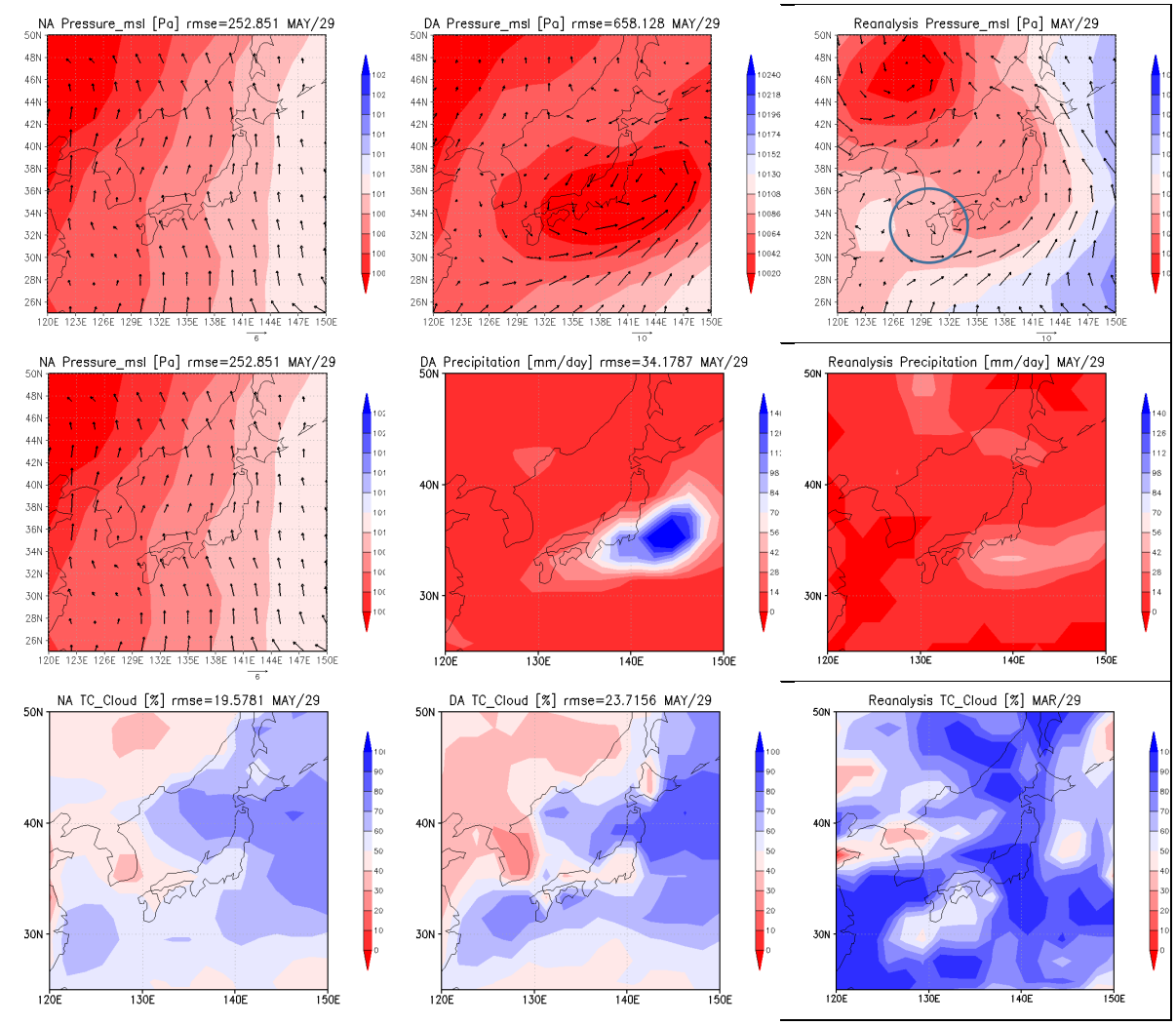


Fig. 5.9: Times series variation of model performance at Choshi station



(a) No Assimilation (b) Data assimilation (c) Reanalysis data

Fig. 5.10: Pressure, Precipitation and TCC variation on 25th May



(a) No Assimilation (b) Data assimilation (c) Reanalysis data

Fig. 5.11: Pressure, Precipitation and TCC variation on 29th May

In Conclusion, due to the data assimilation, the guess run has an improved skill, the analysis further improves the performances. However, some unrealistic precipitation events are occurring in the guess and in the assimilation, which may be due to the influence due to inaccurate weather class, influence due to resolution, a shift in the produced synoptic scale event and lack of observations to reproduce the exact event.

5.1.4. TCC assimilation

In this experiment performance of the model was checked in two aspects. Firstly, the influence of assimilation time (as explained in section 4.5.3) is investigated and then the performance of TCC assimilation was compared with SR assimilation as explained in Chapter 4. The results are shown in Fig. 4.2. The blue line is the cloud cover assimilation at 9 am. and the red line is the cloud cover assimilation at 3 pm.

By comparing RMSE and correlation coefficient values between the SR assimilation at 3 pm and TCC assimilation at 3 pm in Table 5-3. SR assimilation has a better performance in correlation coefficient in all the variables except temperature, in temperature correlation coefficients differ only by 0.01. However, TCC assimilation does not produce overestimation of precipitation as in SR assimilation. Control simulation has the lowest RMSE as it does not have a matching temporal variation. In the data assimilation when the model is constrained to produce lower SR and higher TCC similar to observed values, the precipitation tends to overestimate. As the SR assimilation has a higher constraint on these fields, it causes a higher RMSE than TCC assimilation. In summary, TCC assimilation has a lower skill than the SR assimilation to capture the observed trend.

Table 5-3: Summary of model performance for the three experiments

Variable	RMSE/R	SR	TCC 3 pm	TCC 9 am	Noobs
TCC	RMSE (%)	16.02	**15.62**	15.71	27.49
	R	0.72	0.69	**0.67**	0.30
SR	RMSE(W/m^2)	34.34	**39.82**	43.69	71.56
	R	0.89	**0.79**	0.73	0.65
Temp	RMSE(K)	2.48	**2.26**	2.27	1.96
	R	0.95	0.96	0.96	0.97
Precipitation	RMSE (mm/day)	12.30	**8.60**	8.97	6.03
	R	0.33	0.23	**0.32**	0.12

5.1.5. Precipitation information assimilation

The weather data has several other information in addition to weather type (sunny<>cloudy) such as;

- Rain and snow information (rain, fog, snow etc.)
- About Temperature (warm<>cold)
- Special info (Thunder, Typhoon, saium, storm)
- Wind (weak<>strong) and wind direction

The uncertainty of the model can be further reduced if this information can be utilized. However, it is a challenge to develop a sufficient relationship to be used in the model. For example, the temperature can be related to precipitation events as in (Mikami, 2008). However,

these relationships depend on the region and season which require individual examination. In this study, we utilized the information about dry days by assimilating minimal random precipitation amount when precipitation is zero. 1 mm precipitation was assimilated with specified observation error of 2 mm/day.

The information from the dry days intends to minimize the unrealistic precipitation events occur in the solar radiation assimilation. Fig. 5.12 shows the improvement in precipitation by incorporating dry day's information. From Fig. 5.12 (a) it is evident that information of dry days minimized the unrealistic precipitation incidents. The correlation coefficient increased to 0.67 from 0.45. RMSE decreased to 12 mm/day from 17 mm/day. The precipitation events on 12th April and 14th April could be captured after introducing the zero-precipitation information.

Fig. 5.12 (b) shows the simulation results at Choshi station and Fig. 5.12 (c) shows the simulation results at Wajima station. These two stations are located opposite sides of Japan. In both stations, correlation is improved considerably. However, it leads to further overestimation of the Precipitation in the two local stations even though average RMSE of all stations decreased.

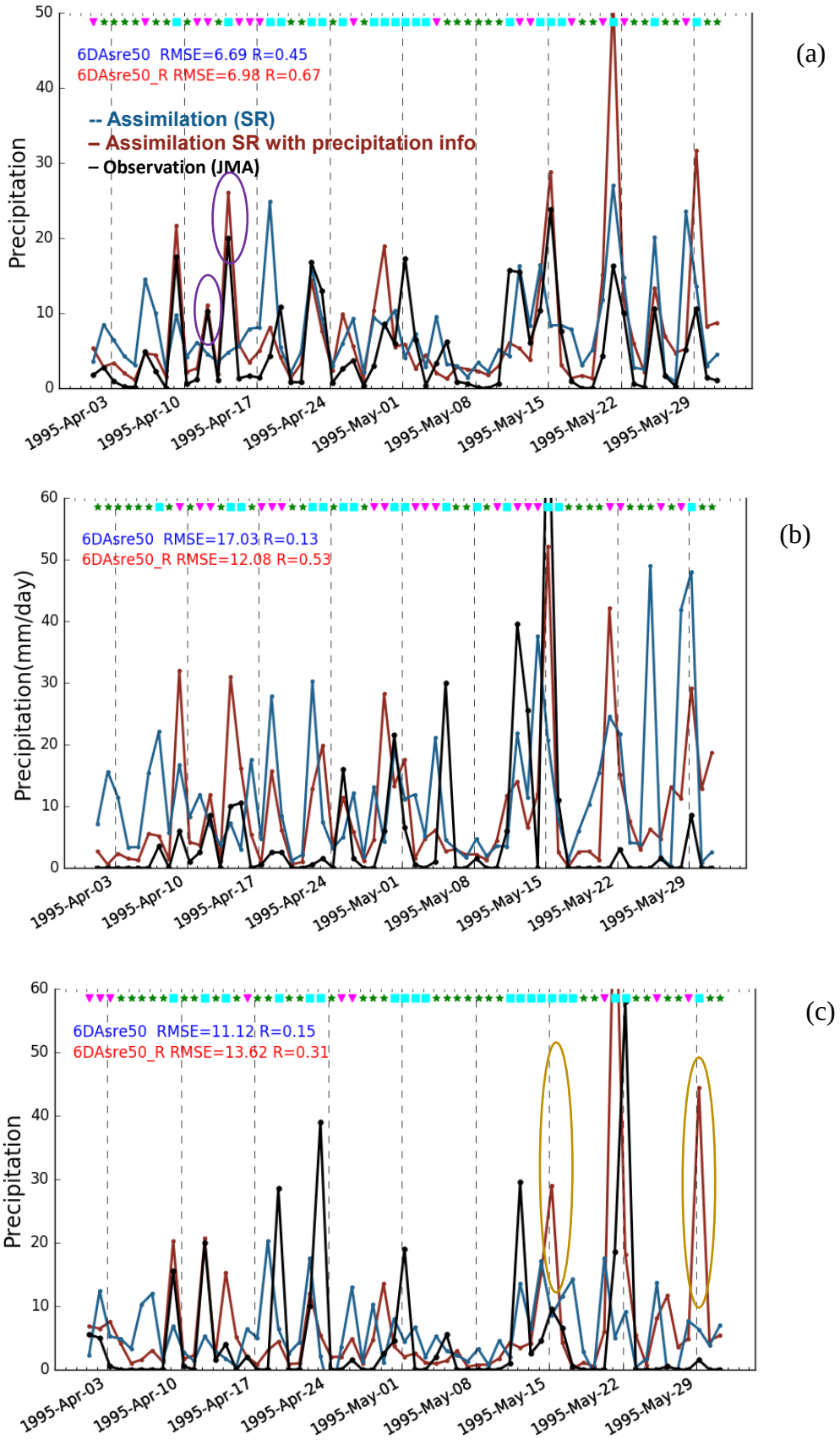


Fig. 5.12: Performance after introducing zero precipitation days to data assimilation observations; (a)- all the station average, (b) – at Choshi station, (c) – at Wajima station

5.1.6. Summary of model performance

Summary of the RMSE and R for different seasons are given in Fig. 5.13 and Fig. 5.14. According to seasonal results in summer, RMSE precipitation is higher due to frequent tropical cyclones and lack of the skill of the model to reproduce them. The model has a negative correlation in both with and without data assimilation simulation in temperature during the winter period.

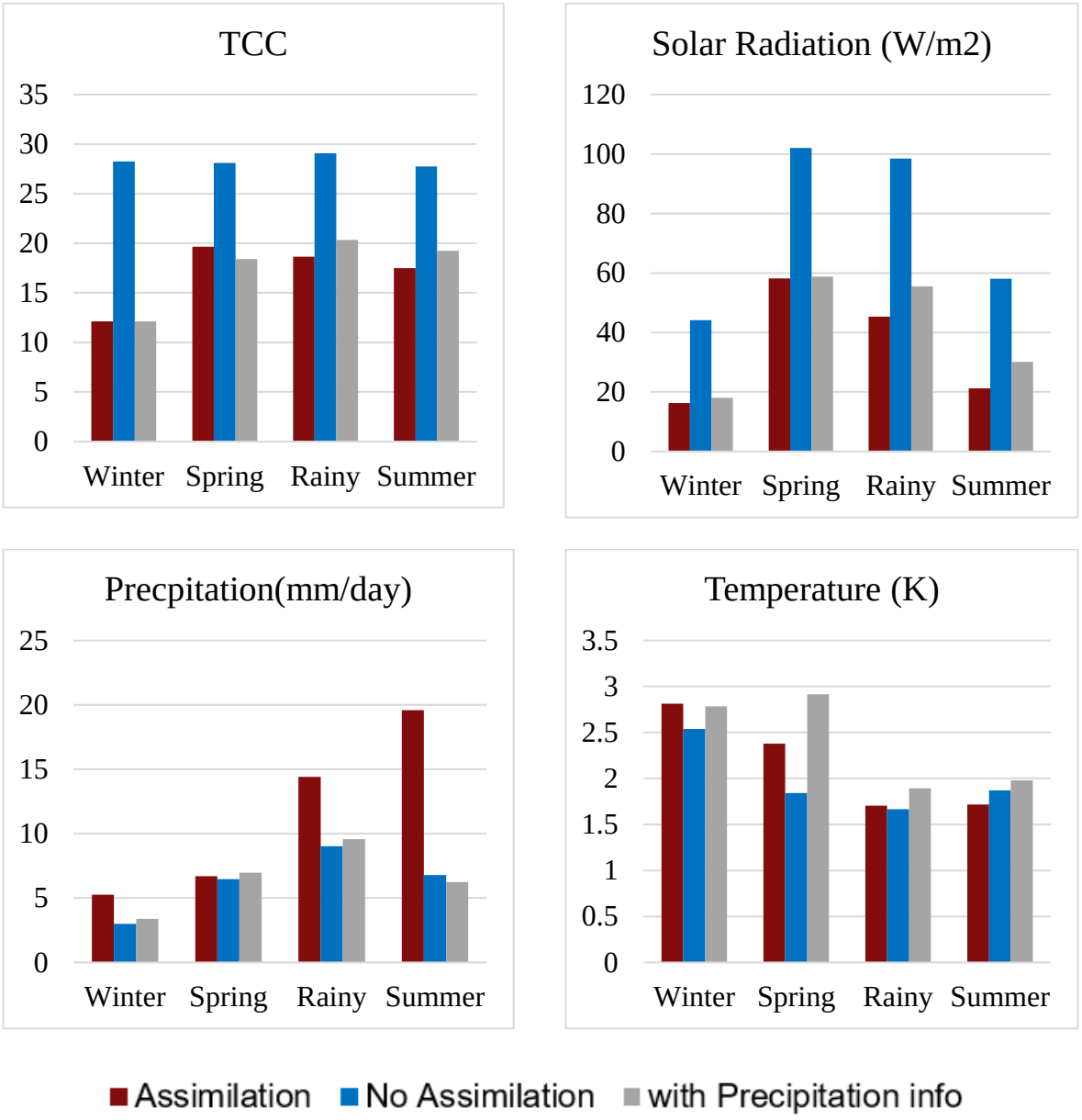


Fig. 5.13: RMSE values of Solar radiation assimilation, No assimilation and Solar radiation assimilation with precipitation information assimilation experiments in different seasons

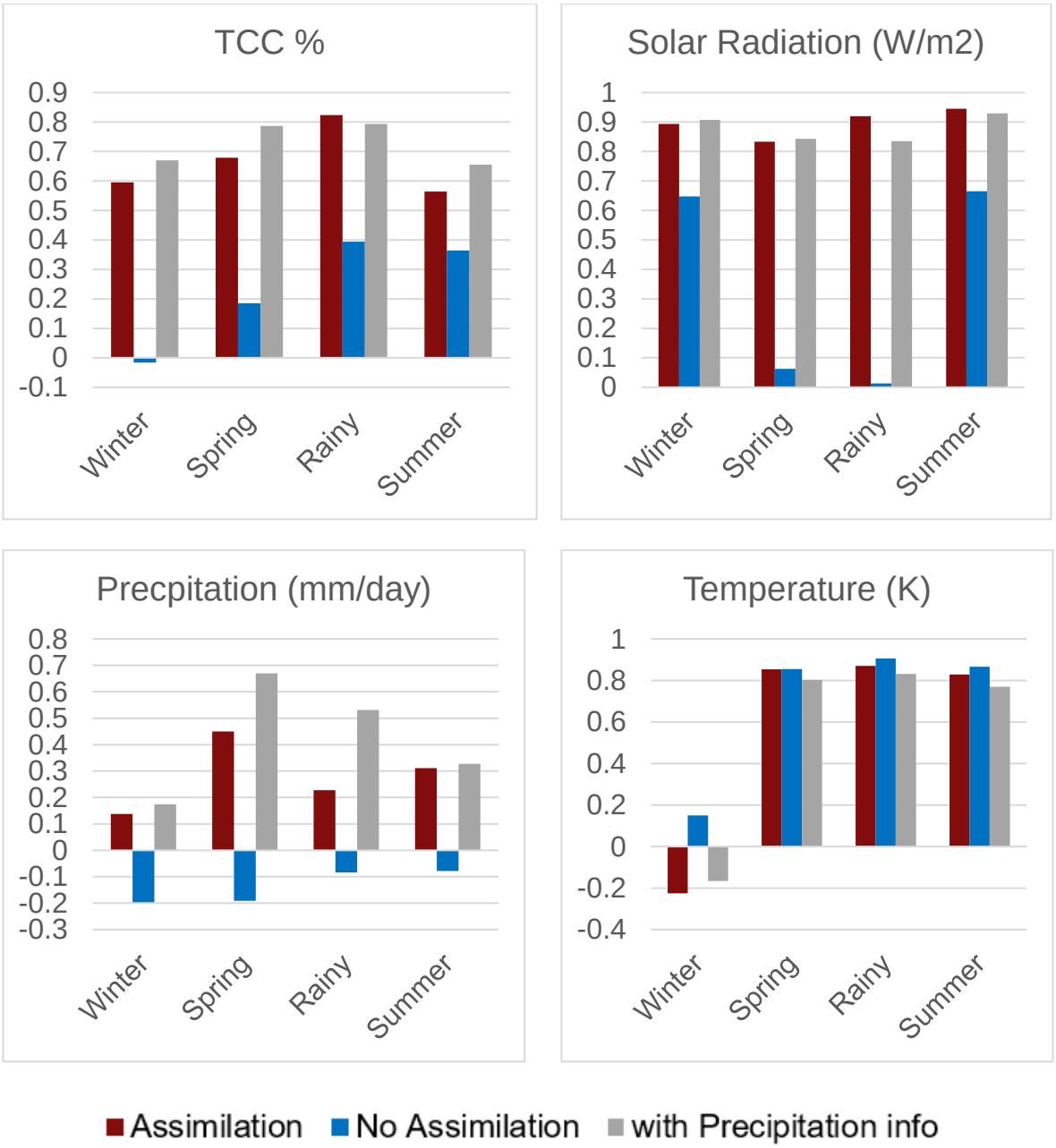


Fig. 5.14: RMSE values of Solar radiation assimilation, No assimilation and Solar radiation assimilation with precipitation information assimilation experiments in different seasons

Statistically significant was checked by caring out t-test to observation mean, and the results have higher significance levels as shown in Table 5-4. For instance, precipitation after data assimilation is significant at $p < .05$ in most of the season and least value is 5.8E-02 which also indicates it is near to $p < .05$.

Table 5-4: Statistical significance

Variable	P Value		T value	
	Data Assam.	No Assam.	Data Assam.	No Assam.
Winter (January -March)				
TCC (%)	1.3E-04	1.4E-24	4.0E+00	1.3E+01
Radiation(w/m^2)	4.3E-01	2.9E-11	-7.9E-01	-7.4E+00
Temperature (K)	3.5E-76	3.7E-94	4.6E+01	6.6E+01
Precipitation(mm/day)	9.9E-06	3.9E-01	-4.6E+00	8.6E-01
Spring (April-June)				
TCC (%)	2.7E-04	1.8E-09	3.8E+00	6.5E+00
Radiation(w/m^2)	6.2E-04	2.3E-12	-3.5E+00	-7.8E+00
Temperature (K)	8.2E-55	2.9E-58	2.8E+01	3.0E+01
Precipitation(mm/day)	5.8E-02	7.7E-01	-1.9E+00	-3.0E-01
Rainy season (June-Aug)				
TCC (%)	5.6E-09	6.4E-18	6.3E+00	1.0E+01
Radiation(w/m^2)	1.5E-03	6.2E-18	-3.3E+00	-1.0E+01
Temperature (K)	2.9E-64	8.4E-64	3.5E+01	3.4E+01
Precipitation(mm/day)	2.1E-04	4.2E-01	-3.8E+00	-8.1E-01
Summer (August- October)				
TCC (%)	6.3E-11	1.9E-25	7.2E+00	1.3E+01
Radiation(w/m^2)	4.6E-01	1.3E-07	-7.3E-01	-5.6E+00
Temperature (K)	2.6E-68	1.5E-70	3.8E+01	3.9E+01
Precipitation(mm/day)	1.2E-05	4.2E-02	-4.6E+00	-2.1E+00

5.1.7. Spatial climatology

Fig. 5.15 compare the model climatology with the observations. GSM has the skill to produce the seasonal climatology without data assimilation. The figure shows the spatial variation in 4 different seasons Winter (Jan-March), Spring (April-June), Rainy season (June-August), Summer (August - October). Even though the seasonal trend is captured in the no assimilation simulation, it has a bias which could be reduced with data assimilation. This is clear in winter where no assimilation run has mostly TCC less than 50% (red) while the assimilation run has TCC more than 50% in the China Sea side becoming closer to the observations.

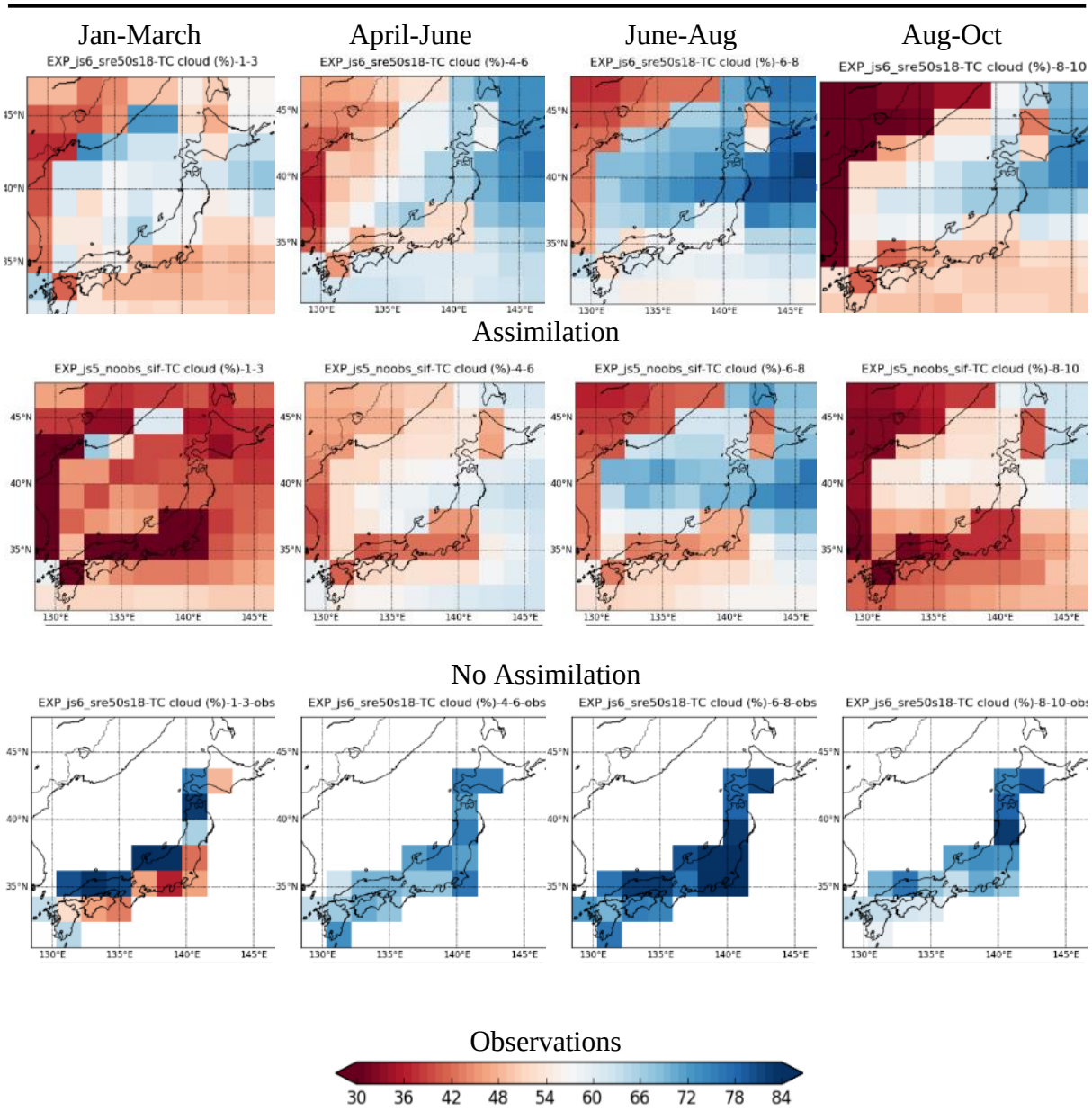


Fig. 5.15: Spatial variation of TCC in different seasons (%)

5.1.8. Long-term simulation results (1995-1999)

Experiments were carried out for several years to evaluate the model performance on different years and to compare the relative or the anomaly performance of the model. Fig. 5.16 shows the April anomaly variation during this period. Even the model has a bias model could follow the annual trend well with data assimilation except in temperature in 1996.

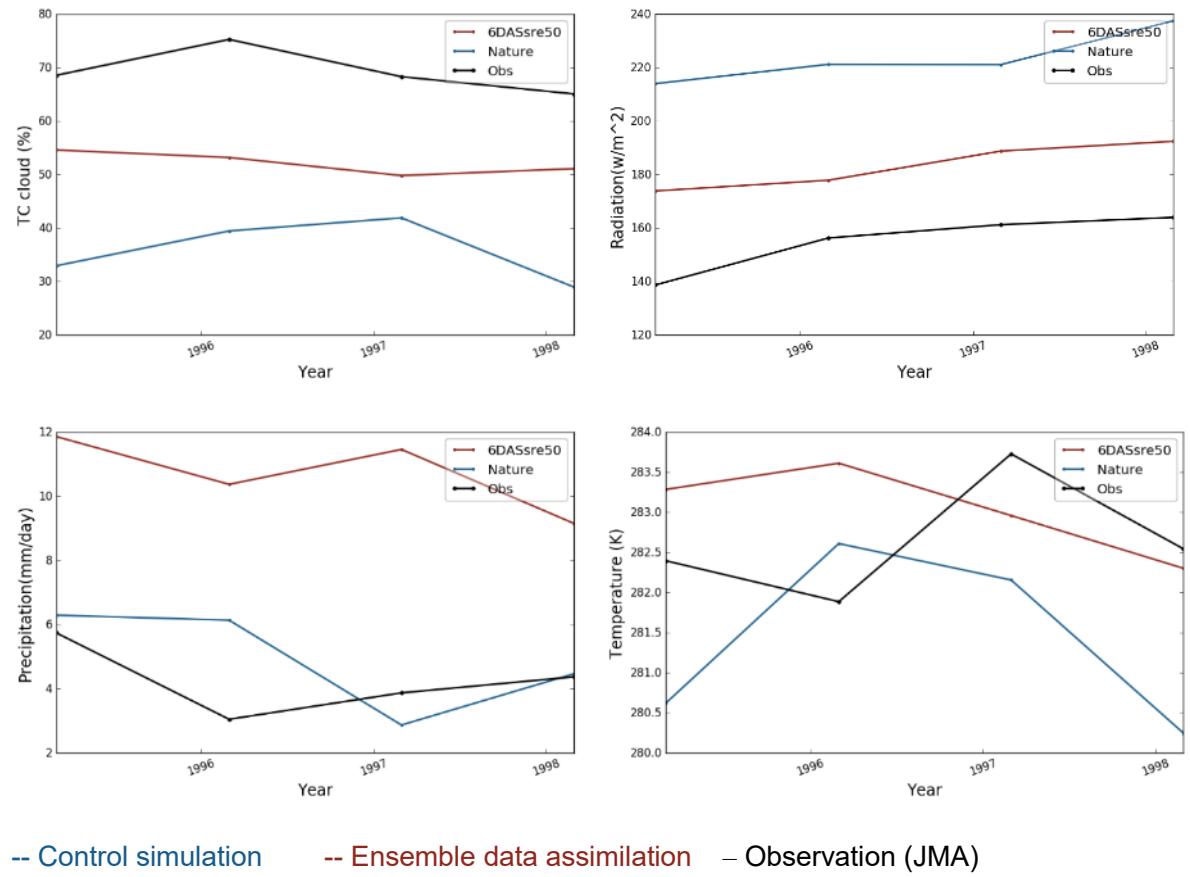


Fig. 5.16: March Anomaly variation from 1995 -1999

5.2. Patterns of consecutive daily values of weather classes

Occurrence of Each weather class in January is calculated for a 5-year period (1995-2000). From a similar analysis to all the months, it is possible to identify long-term impacts such as droughts. A lesser number of rainy days can indicate drought conditions and a higher number of rainy days can indicate the wet condition.

From the line graph in Fig. 5.17, 1998 has higher number rainy days (lower number of sunny days), and this was captured by model which produced higher rainfall in 1998 May.

Table 5-5: Different types of weather in Japan

Year	Rainy	Cloudy	Sunny	TCC (%)		Radiation (w/m2)		Temperature (K)		Precipitation/ (mm/day)	
				Model	Obs.	Model	Obs.	Model	Obs.	Model	Obs.
1995	**8**	6	45	46.7	46.0	134.6	129.6	281.2	279.5	**7.7**	**1.2**
1996	**9**	4	46	46.3	48.8	132.0	124.9	281.2	278.8	**7.3**	**1.3**
1997	**6**	7	46	43.1	41.3	139.9	137.3	281.2	280.2	**6.1**	**2.1**
1998	**18**	6	34	44.9	57.7	127.5	105.3	281.9	280.0	**8.4**	**4.4**
1999	**9**	4	46	43.2	40.2	133.1	132.9	282.0	280.0	**6.2**	**2.0**

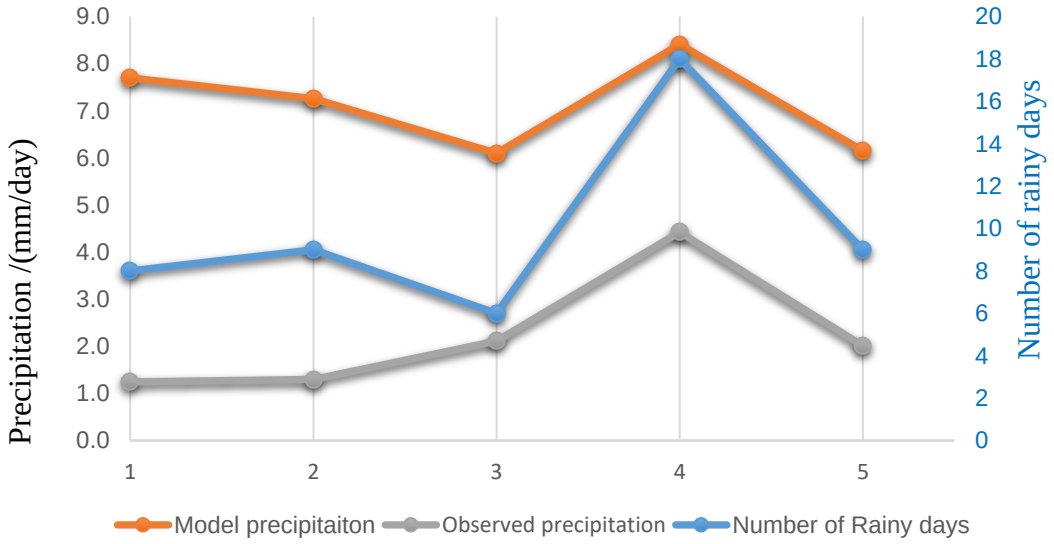


Fig. 5.17: Number of rainy days vs. observed precipitation and model precipitation.

5.3. Impact on different weather types

Japanese climate depends on different sources during various periods (Hiroyuki Kusu, 2013), (Kurashima, 2002), (Kodama, 1992). A summary of weather is given in below table.

Table 5-6: Different types of weather in Japan

Weather type	Period
Winter (continental high pressure and low oceanic pressure)	Winter- December, Jan, Feb
ExT Cyclone	Mostly in Spring, sometimes in Winter
Moving high pressure (Antic cyclones) from west to east	Spring/Autumn - March, April may, October, November December
Summer type	August September
Typhoon	Late summer – mostly in September and sometimes August
Baiu system	Early summer- June, July

The scale of these weather systems is different, the extratropical cyclones (ExT cyclones) occur in synoptic scale with slow moment while typhoon occurs in a smaller scale with rapid movement. The Baiu system produce local weather. Due to the model resolution and low assimilation interval, skill of the assimilation system capturing small-scale rapidly changing events such as typhoons and Baiu would be lower. On the other hand, assimilation system would be able to capture synoptic scale weather events more accurately. Following section would discuss the model performance during ExT cyclones and Typhoons.

5.3.1. The impact from ExT Cyclones

Fig. 5.18 shows the precipitation during May-June 1995. There are several precipitation events and found to be mostly due to the Ext Cyclones. ExT cyclones occurred during following data

- May 3rd - 4th - away from the mainland in the northward of Hokkaido
- May 7^{t-8th} – Anticyclone – strong over Onahama, Morioka area
- May 14th-16th - from west Japan to east Japan through mainland
- May 17-19- Anticyclone- from East Japan up to Kanto area
- May 21st-23rd from west Japan to east Japan through mainland
- May 23-25 Anticyclone- from East Japan
- May 25th Over Hokkaido
- May 29th Over Aomori

The pressure reduced to mean sea level and wind vectors of the key events are shown in Fig. 5.18. During May 7th there is a high-pressure zone over Japan, and several diaries catch this by indicating sunny weather. In reanalysis data pressure center is more towards eastern side however in the data assimilation simulation this was simulated shifted to little bit west Japan. There is a movement of an ExT cyclone from 14 to 16th May. The data assimilation simulation could capture this well while no assimilation simulation could not capture that event. A shift in the time and path of ExT cyclone can be observed which may be due to long assimilation interval, lack of observations outside and Japan and low model resolution. Again, during May 15-18, a large high-pressure zone moved over Japan and could be captured in the data assimilation simulation. During May 22 another ExT cyclone was occurred and could be captured. During May 25th there is an ExT in the northern side of Japan, data assimilation simulation could capture an ExT, but it was shifted towards central Japan creating unrealistic precipitation over that area.

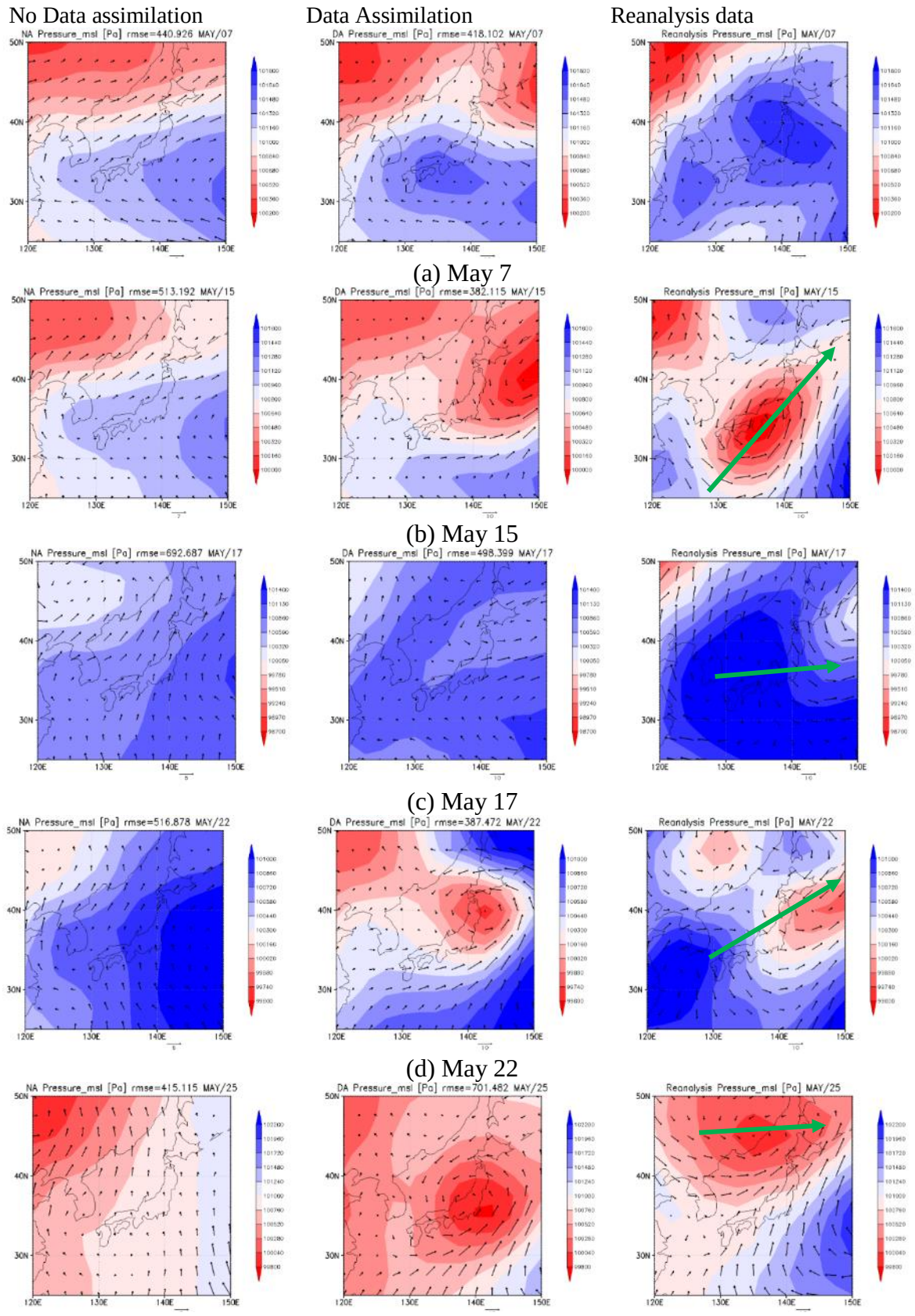


Fig. 5.18: Pressure distribution of ExT Cyclones and Anticyclones in May

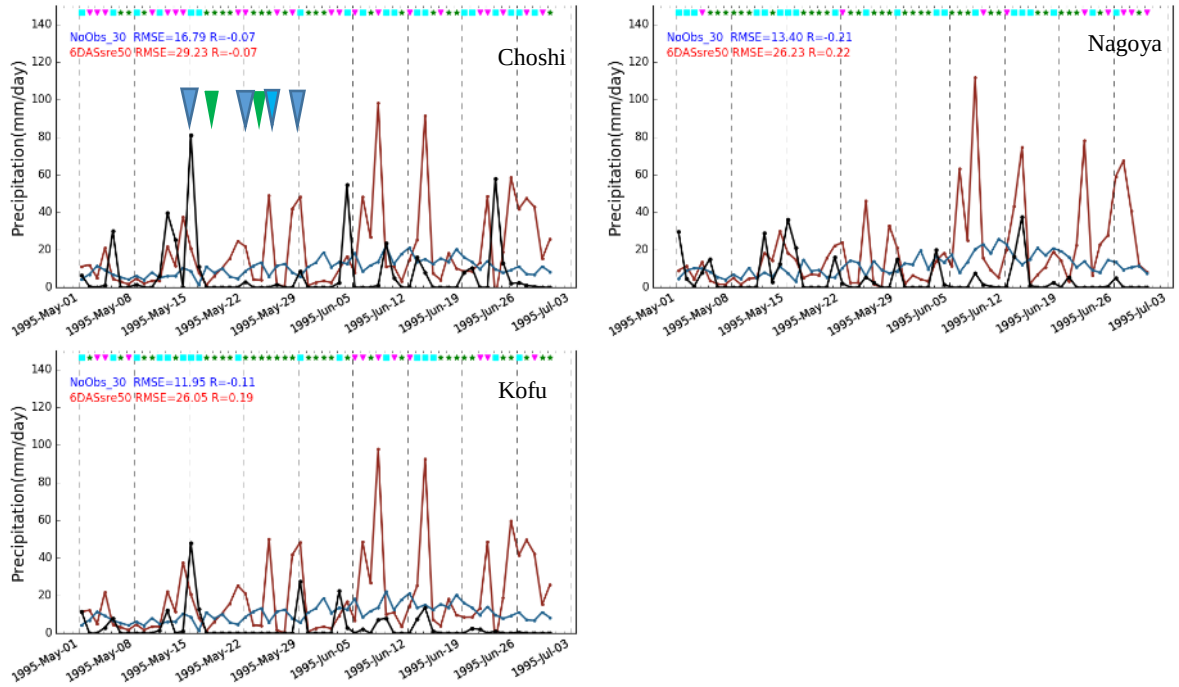


Fig. 5.19: Precipitation during ExT cyclone season.

Fig. 5.19 shows the precipitation during May and June period. Precipitation during ExT cyclones is visible in at least few stations that affected by the ExT cyclone. The May 15th ExT cyclone had strong pressure deference, and it caused the heaviest precipitation during May. The model also could capture this precipitation event. During other ExT cyclone model tend to overestimate than observations. Even though on May 25th ExT cyclone did not travel over central Japan, model simulate the ExT cyclones as traveling over central Japan causing a significant precipitation event over that region even though in reality there was no precipitation. The shift in the location of ExT cyclone is the reason for unrealistic precipitation and sometimes heavy precipitation forecast. In comparison, no assimilation simulation could not capture most of the weather events. The capability to produce ExT cyclones using uncertain description data would be an exciting result to future studies.

Fig. 5.20 shows the progression of ExT cyclone from 14th May to 17th May of the simulations with data assimilation and without data assimilation in comparison to the reanalysis data in detail. The low-pressure zone enters from south-west direction and move to central Japan on 15th May and then move to North-East direction. Thus, ExT cyclone impact on whole of Japan during this period. In the background, there is a high-pressure zone in East-South side and low pressure in West-North Direction. Without data assimilation, it could not capture the movement of the ExT cyclone at all. However, it could capture the generous low pressure and high-pressure zone in East-South side and in West-North Direction. On the other hand, with the data assimilation, we could see the passage of high pressure over Japan on 14th May and

movement of ExT cyclone from 15th to 17th, but there is a shift in the time in the movement. This may be because the weather data is assimilated only at a particular time step of the day (i.e., at 3.00 pm).

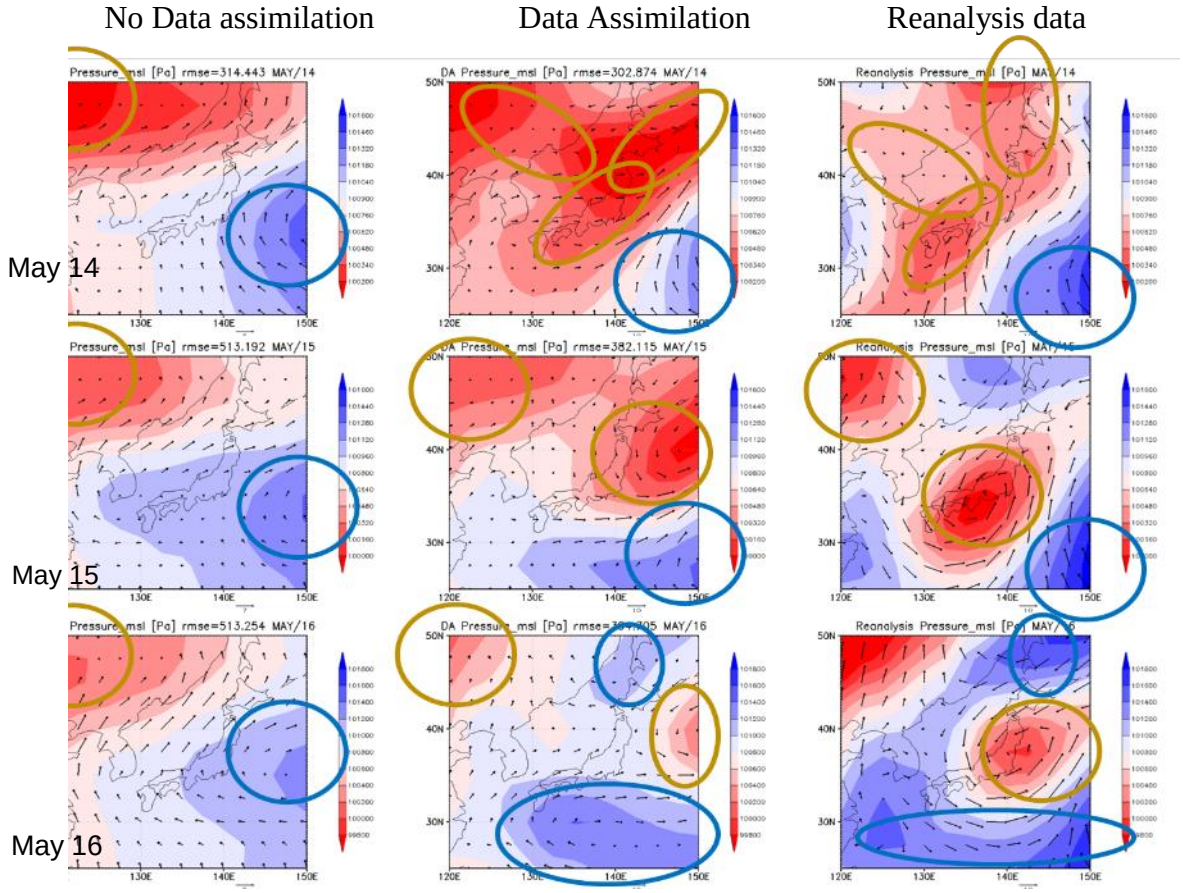


Fig. 5.20: Development of low-pressure area during the ExT cyclone 05/14-5/16

Fig. 5.21 shows the performance of other variables on 14th. Data assimilation system could capture the precipitation event, but it is overestimated and earlier than the actual event. This is because there is a shift in the location of the ExT cyclone in the simulation. Similarly, the pressure reduced to MSL, 10m Geopotential height at 500 level shows the zonal pressure difference which decides the path of the ExT cyclone. In TCC model could capture spatial distribution well. The China side has a lower TCC, and Japan seaside has higher TCC in North-East direction. Simulation without data assimilation could also capture this variation to some extent. Temperature and SR also has a similar variation in the two simulations in comparison to reanalysis data.

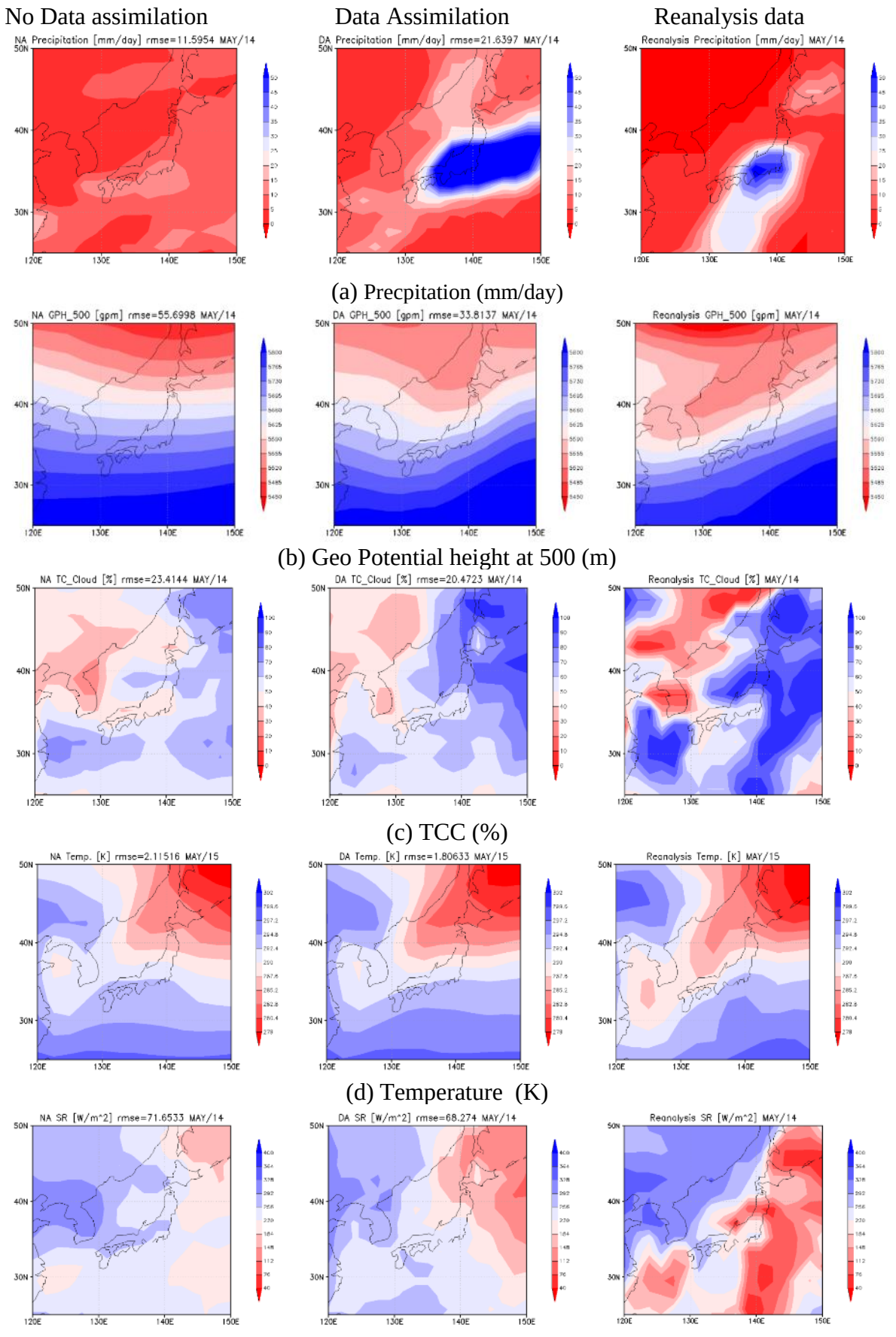


Fig. 5.21: Performance of other variables during the ExT cyclone 14-17 May

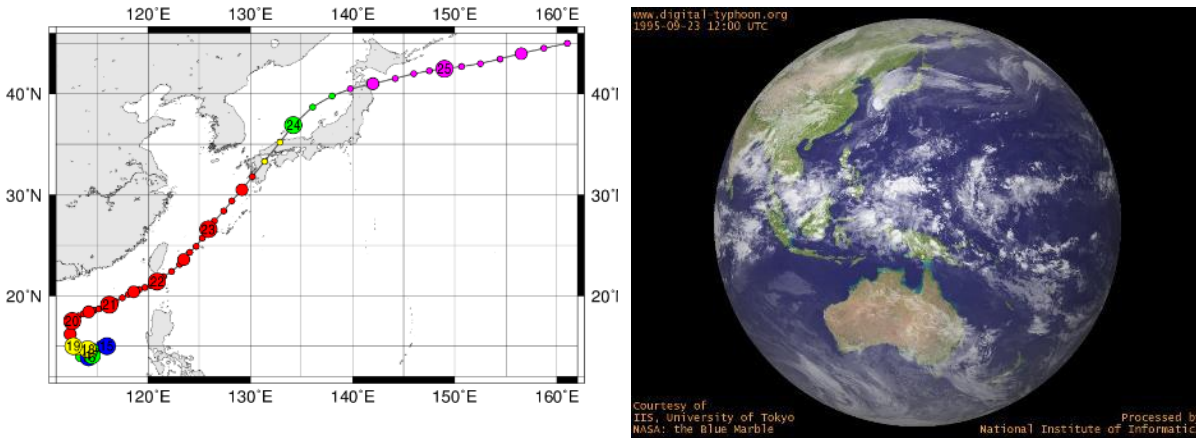
5.3.2. Impact from typhoons

Japan is located in the Pacific Ocean and faced by many typhoons which impact the weather from end of spring to the beginning of winter. Some diary data has information about the typhoons. However, as specific attention is not given to individual diaries, and only three weather classes are considered, all the precipitation events including typhoon are categories as class 3(rainy). However, as typhoon affects a large area, multiple diaries in different areas can indicate heavy rain and high cloudiness. Specific attention was given to average scale typhoon Ryan (i.e., Typhoon 199514). Table 5-7 shows the properties of Typhoon Ryan

Table 5-7: Properties of Typhoon Ryan

Property	Value
Minimum Pressure	940 (hPa)
Maximum Wind	85 (knots)
Largest Diameter of Storm Wind	370 (km)
Largest Diameter of Gale Wind	1000 (km)
Average Speed	23.1 (km/h)

The Typhoon entered Japan on 23rd September 1995 from the southern side and crossed to China seas side before crossing northern Japan again as shown in Fig. 5.22. As shown in Fig. 5.22 (b) the typhoon impacted on a large area and causing high cloudiness. In weather classes, this condition would be indicated by class 3.



(a) (b)

Fig. 5.22: (a) Track of Typhoon 199514 (RYAN), (b) Visual image of Ryan on 23rd September 1995 (Kitamot, 2015)

Fig. 5.23 shows precipitation during the typhoon (blue square s21st to 24th September) Ryan at different locations. As the typhoon enter Japan from southern side around Nagasaki, observations show very high precipitation of more than 100 mm during one day at Oita and Nagasaki stations. According to Table 5-8, all the stations have weather class 3 at least once during two days. Kochi, Nagasaki, Shnomaaki, Kagoshima areas show weather class 3 in the first two days and reduce in the third day when the Typhoon move up on the other hand in the northern stations, the first day is class 1 or 2, and then it became weather class 3 in the third day. Many stations have precipitation (more than 20 mm/day) at least once except the stations shown in yellow. However, the model shows higher precipitation in most of the stations probably because all the stations have weather class 3.

Table 5-8: Weather classes at observation stations

Station	**Weather Class**				
	22nd	23rd	24th	Observed Precipitation >20mm/day	Model Precipitation >40mm/day
Obihiro	-	-	-		
Sapporo	1	1	3	Y	Y
Aomori	1	1	3	N	Y
Morioka	1	2	3	N	Y
Onahama	2	2	3	N	Y
Choshi	1	1	3	N	N
Takada	1	3	3	N	Y
Kofu	2	2	3	N	N
Wajima	2	3	3	N	Y
Maizuru	3	3	3	Y	Y
Nagoya	3	3	3	Y	Y
Yonago	1	1	3	Y	Y
Oita	1	3	3	Y (>80)	Y
Hamada	3	3	3	Y (>40)	Y
Kochi	3	3	2	Y (>40)	Y
Nagazaki	3	3	2	Y (>80)	Y
Shinomisaki	3	3	2	Y	Y
Kagoshima	3	3	1	Y (>40)	Y

Fig. 5.24 shows the TCC and SR at Nagasaki station. SR goes to the minimum value in September during TCC peak. The model could capture this change very well.

Considering all the results. The model could capture the heavy precipitation, however as the model has a lower resolution model could not capture the typhoon itself. By comparing with other events, it is possible to judge manual that these results occurred due to a typhoon and the movement if from south to north. A limitation in this study is only three weather classes

are considered thus it is difficult to distinguish between heavy precipitation events and regular precipitation events. If the information about typhoons can also take into consideration, better results would be able to achieve.

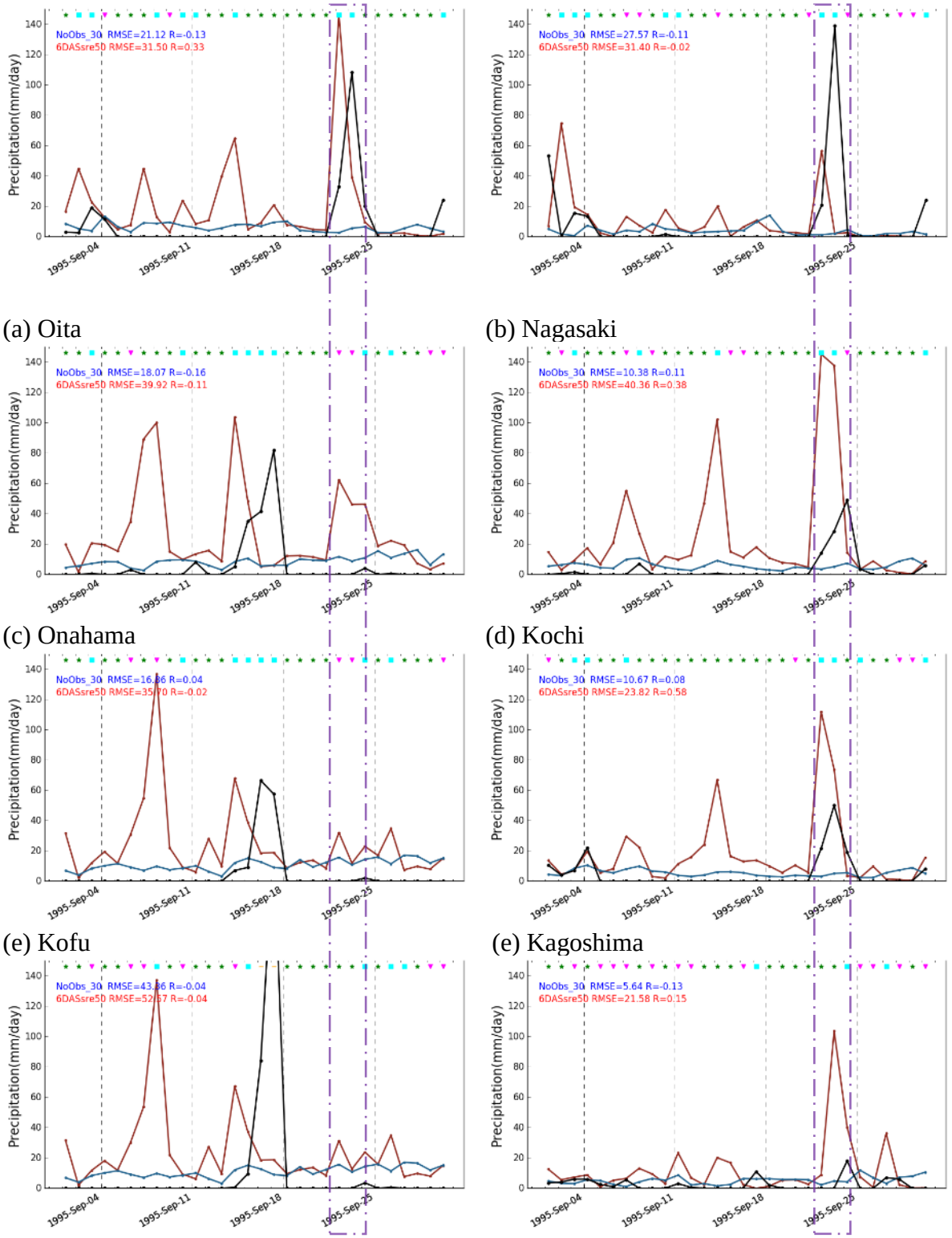


Fig. 5.23: Precipitation at different stations in September 1995

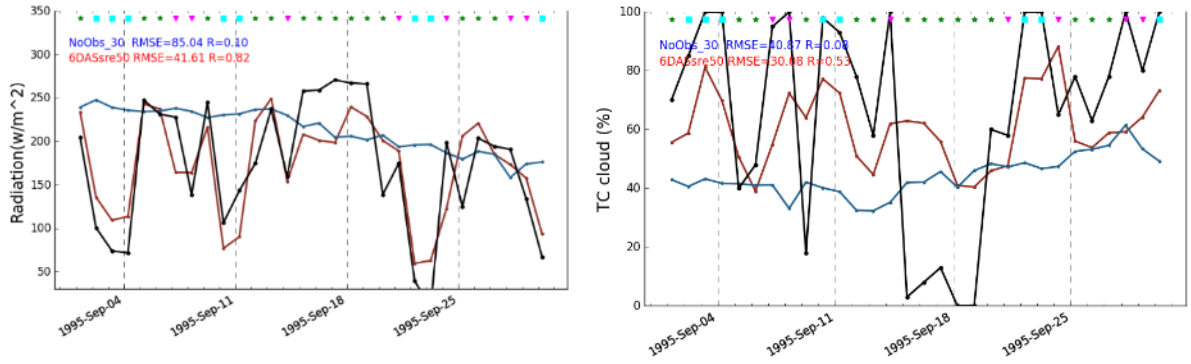


Fig. 5.24: TCC and SR at Nagasaki station

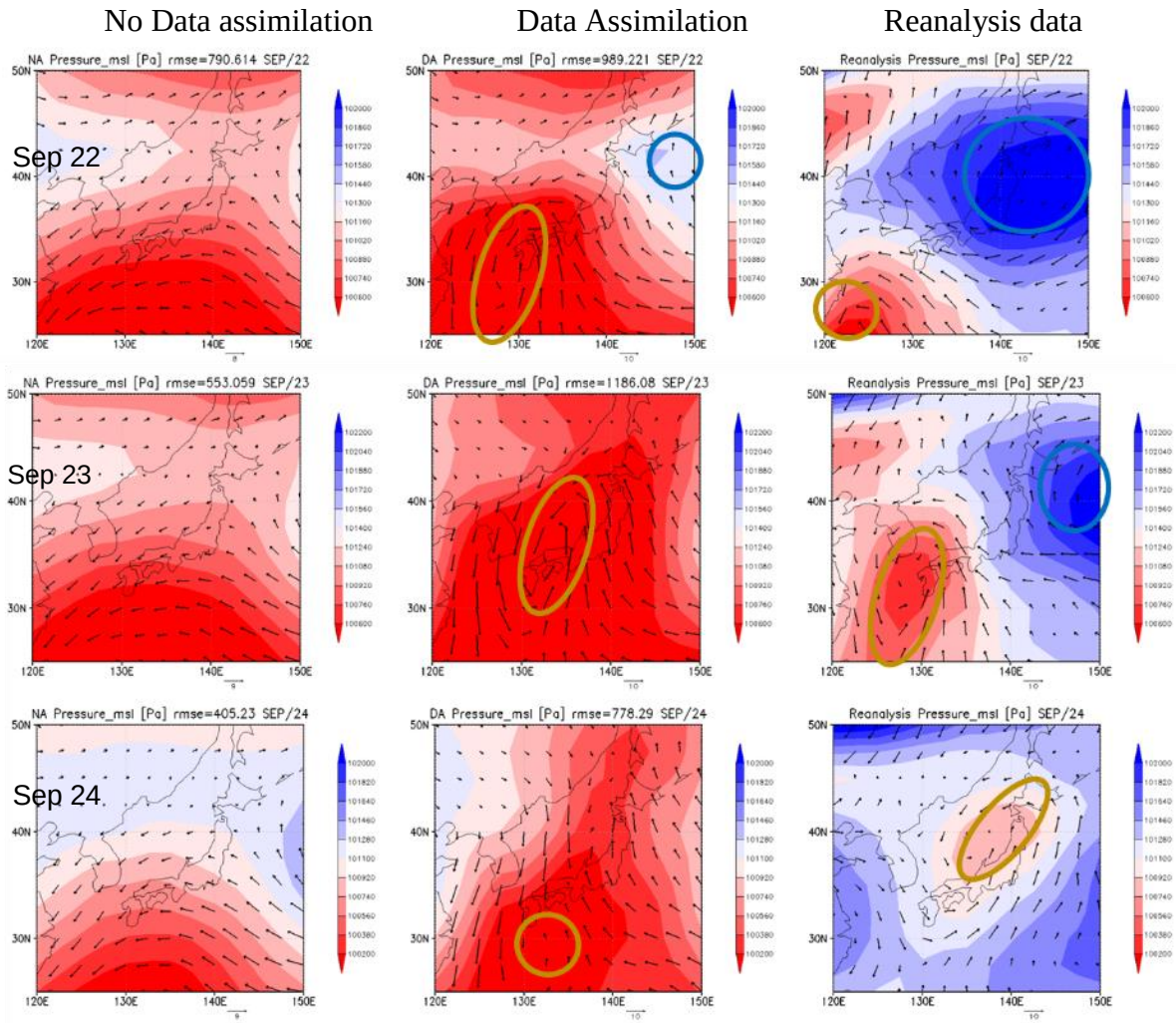


Fig. 5.25: Development of low-pressure area during the Tropical cyclone 09/22-9/23

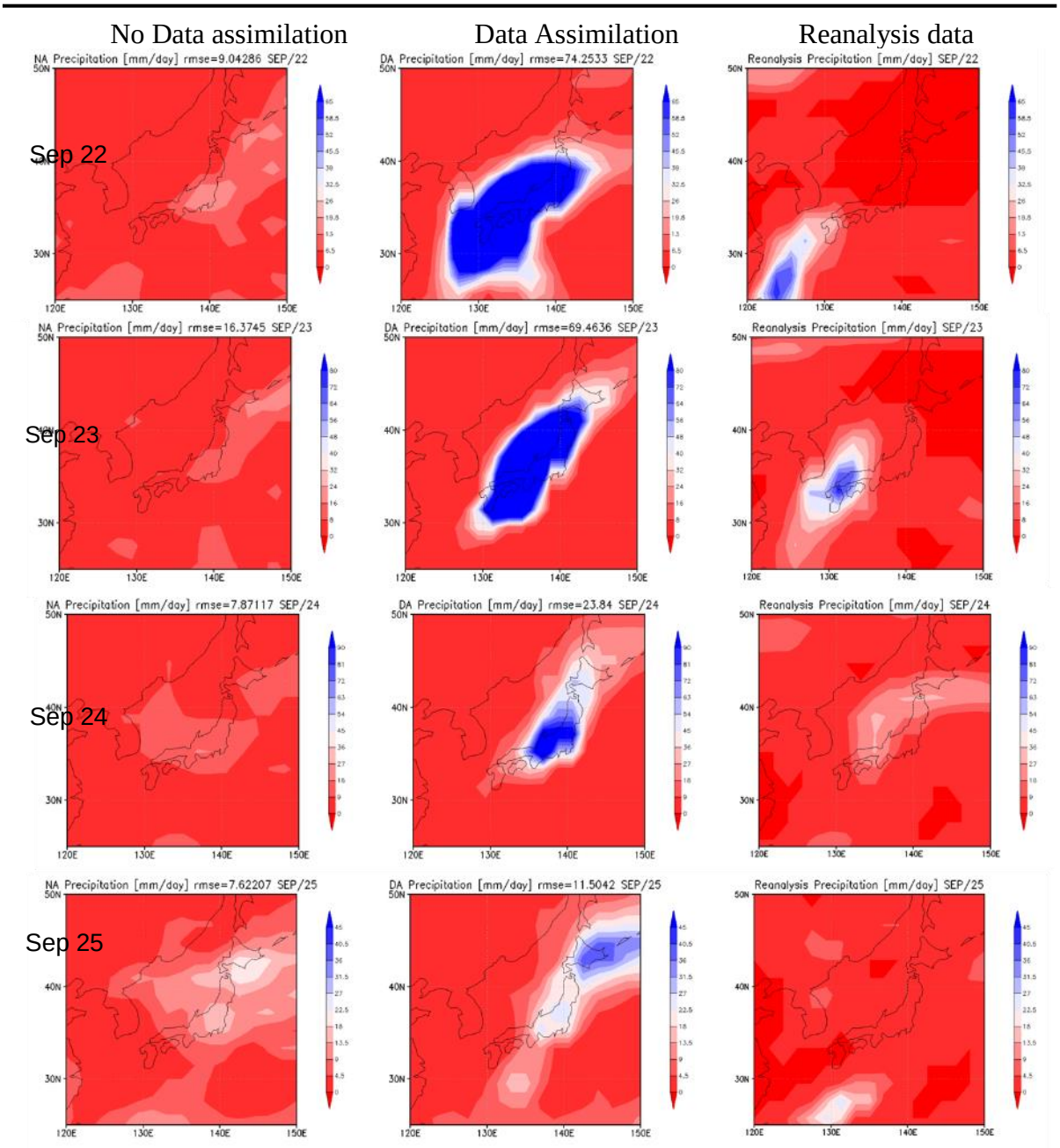
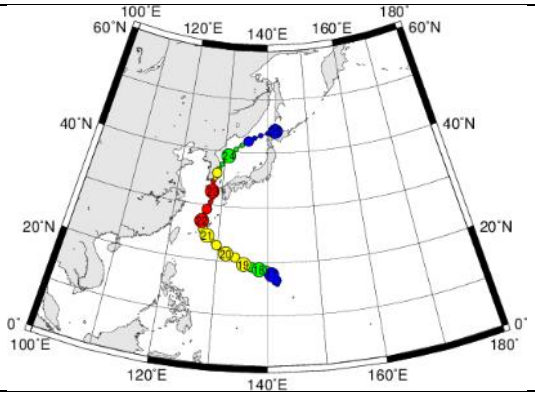
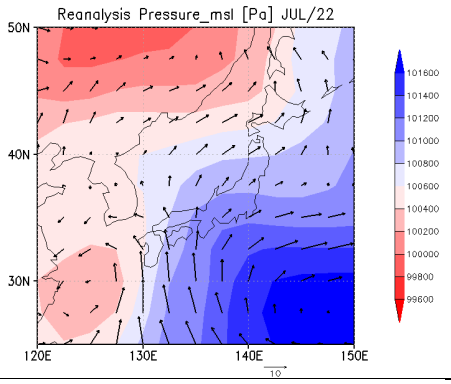
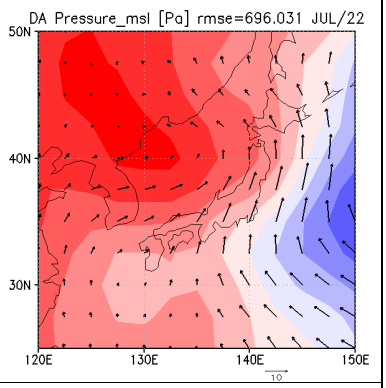
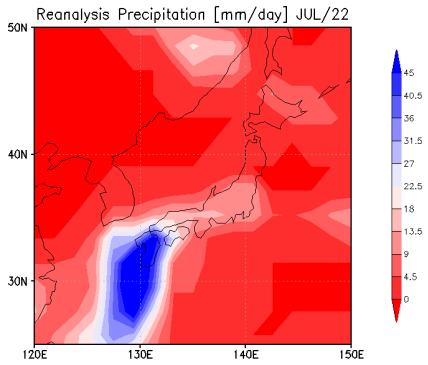
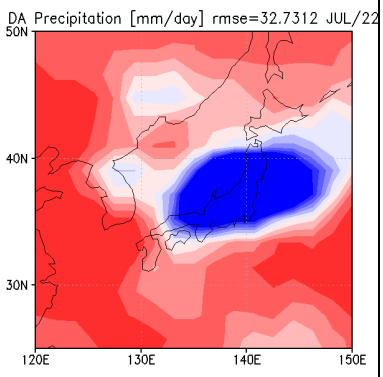
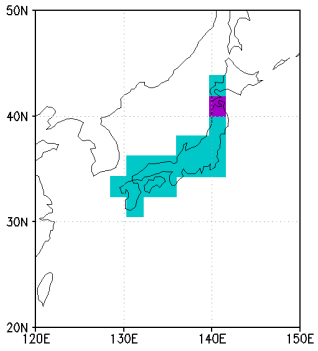
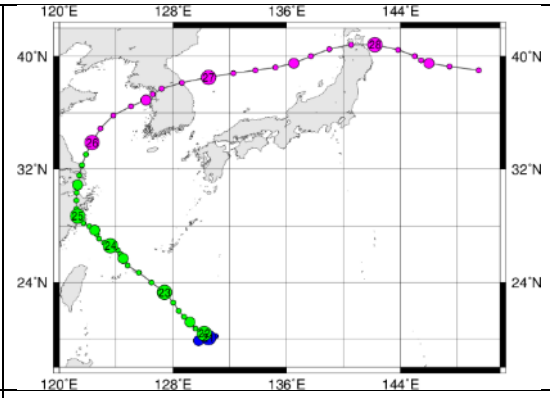
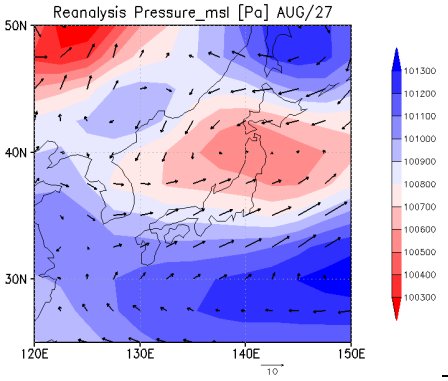
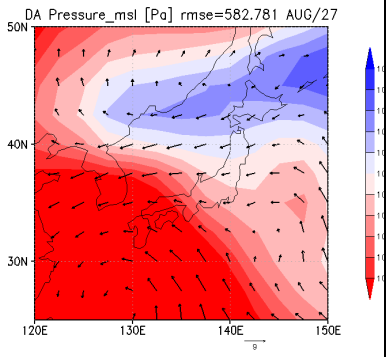
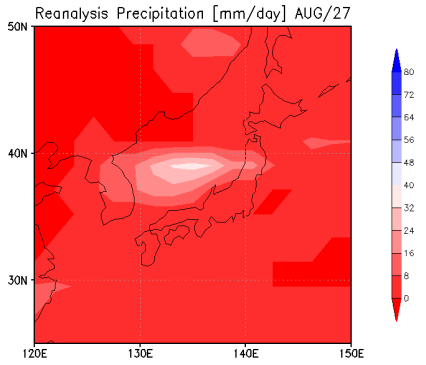
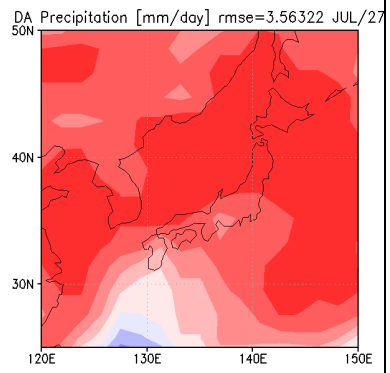
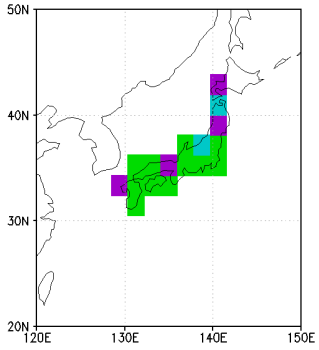
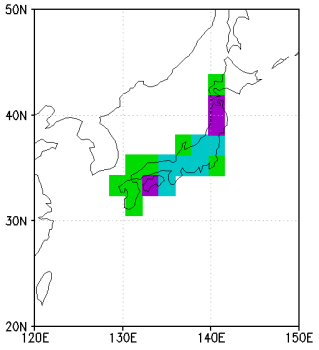


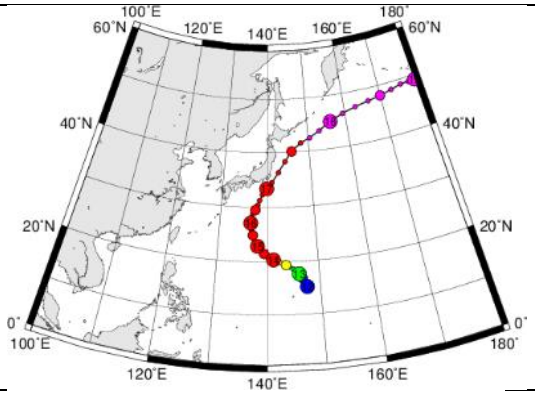
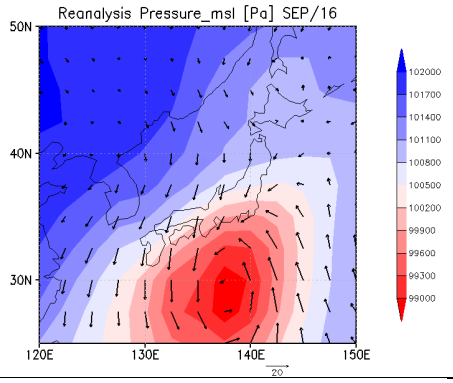
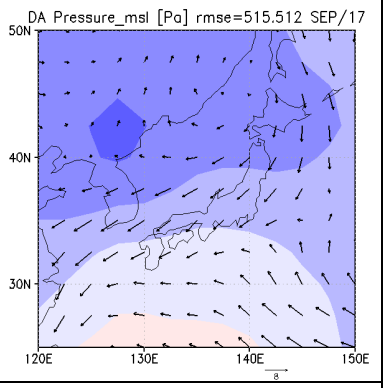
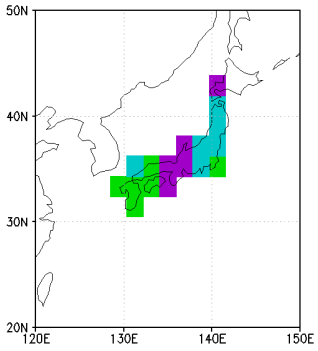
Fig. 5.26: Precipitation during the Tropical cyclone 09/22-9/23

More Typhoons are analyzed to investigate model skill further during the typhoons. The results are shown Table 5-9. According to the results, the model could not capture the Typhoon movement well in most of the cases. Mostly it could capture the precipitation events but not a low-pressure center as in the Typhoon.

Table 5-9: Typhoons in 1995 Early summer

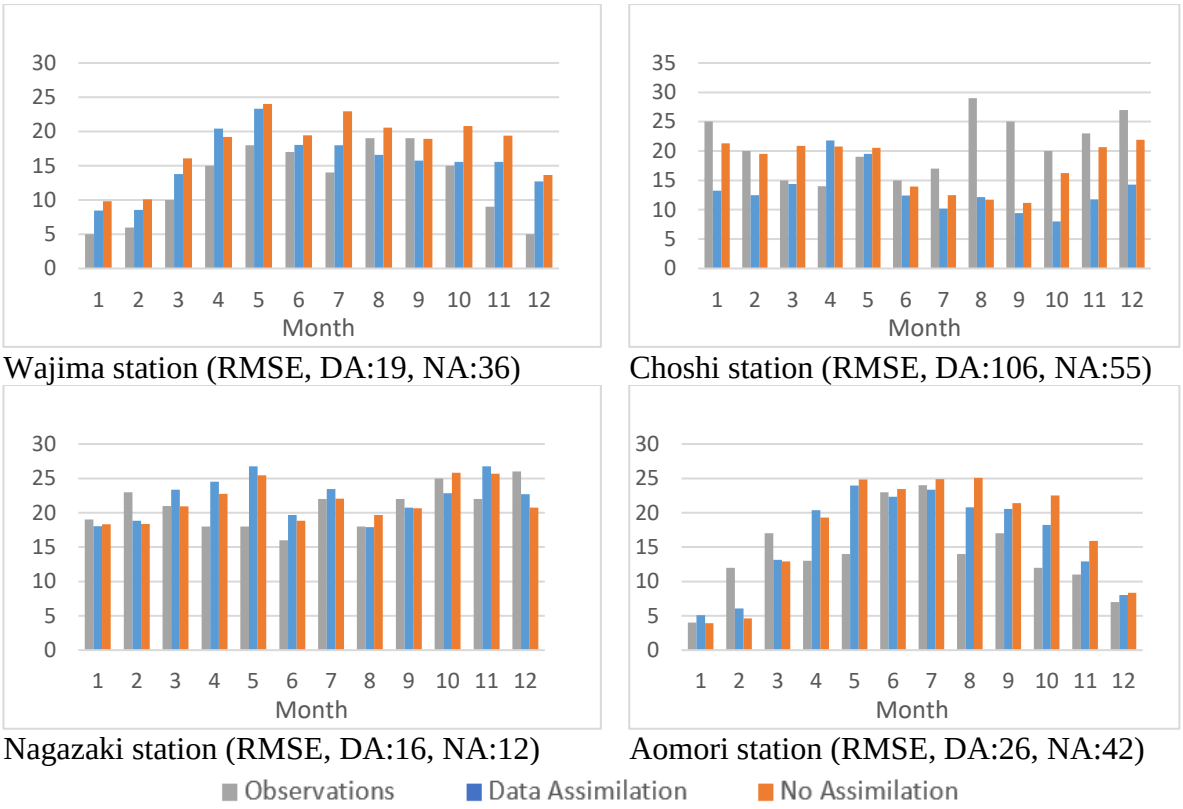
1995		
199503 (FAYE) 07/22		During this typhoon, the low-pressure center is not well captured. However, there is a small low-pressure center around lon 130 and lat 30 corresponding to the typhoon.
	Reanalysis	Model with DA
Pressure		
Precipitation		
Observed weather classes		Many diaries indicate class three during the Typhoon 

199507 (JANIS) 08/28		This typhoon arrives with less strength; reanalysis data shows the movement of the pressure centre. However, Data assimilation does not have a similar pressure pattern.
	Reanalysis	Model with DA
Pressure		
Precipitation		
Observed weather classes	09/17 	09/16 

199512 (OSCAR) 09/17		During these days, Typhoon travel in Japan sea side from west to east away from the mainland. Only Tokyo Chiba area got influenced. The model does not capture this event even though there is a slight low-pressure center in the south.
Pressure		
Observed weather classes		

5.4. No of zero precipitation days

Zero precipitation days in each month was compared between the observations and two model simulations (i.e. with and without data assimilation). As the ensemble mean is an average, it cannot produce precisely zero values. Hence zero precipitation days were calculated in each ensemble and averaged. The following figure shows the results of Wajima, Choshi, Nagasaki and Aomori stations representing four sides of Japan. Spring has the least difference in all stations as the model could capture synoptic scale weather well during the period. In overall data assimilation simulation produced better results at Wajima and Nagoya. At Choshi data assimilation reduce the number of dry days especially in summer which due to less accurate typhoon effects as discussed earlier. Annual comparison of zero precipitation days can help to identify long-term weather events such as droughts.



Wajima station (RMSE, DA:19, NA:36)

Choshi station (RMSE, DA:106, NA:55)

Nagazaki station (RMSE, DA:16, NA:12)

Aomori station (RMSE, DA:26, NA:42)

Fig1: No of zero precipitation dates in each month

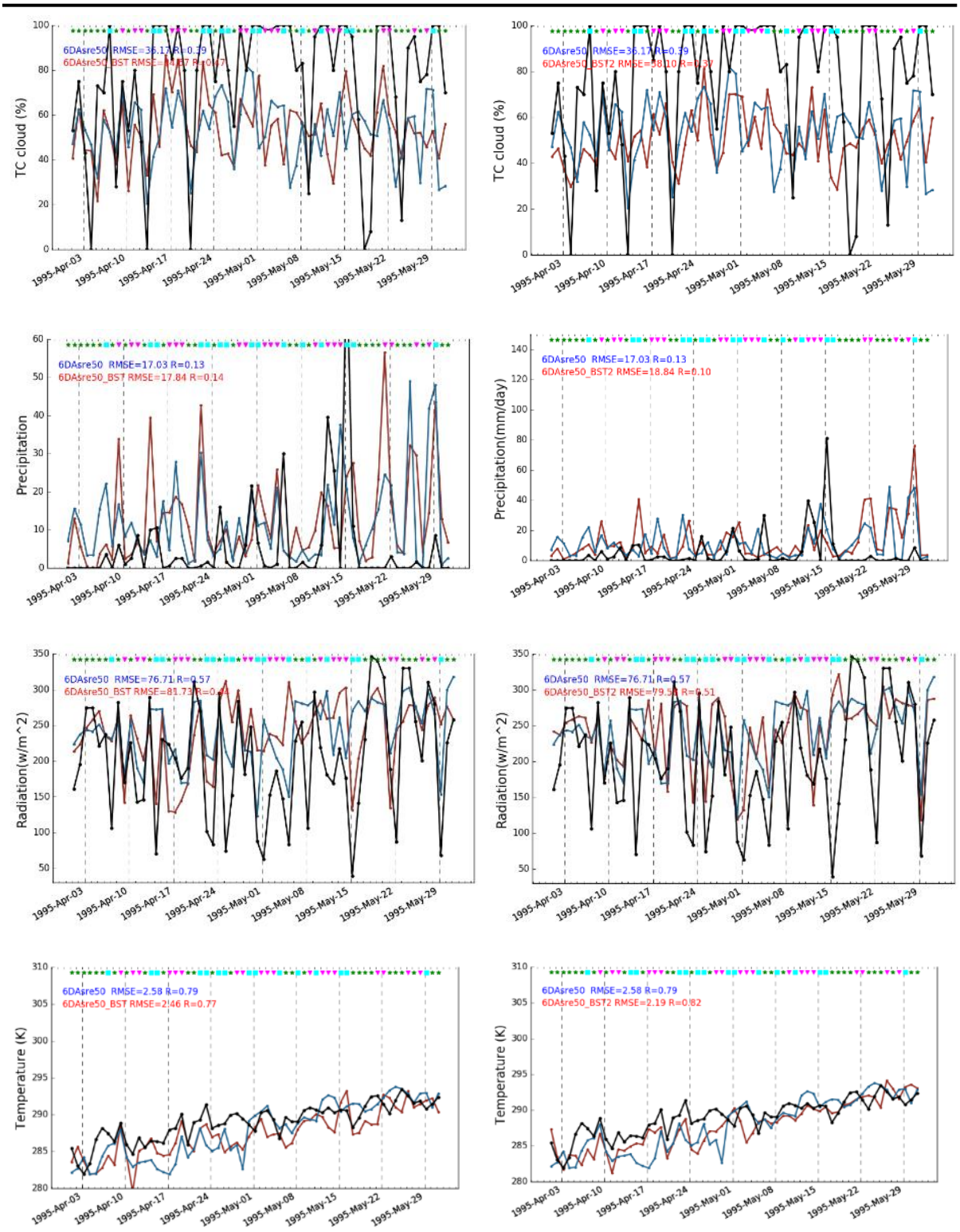
5.5. Bootstrap Experiment

Diary data do not cover all the regions in Japan. Sometimes a diary can be discontinuing for a short time. Hence it is essential to investigate whether it is possible to improve an area without diary data from available diary observations in the surrounding area. For that two kinds of experiments were carried out in recent period with the same conditions as in Chapter 5.

- Removing one observation (BST Choshi simulation)
- Removing several observations (BST2 simulation)
    - Only nine stations (Obihiro, Aomori, Onahama, Takada, Kofu, Maizuru, Shionomisaki, Hamada, Kagoshima)

For this experiment in 1995 , three weather classes data was used. In the results, the control simulation is the same as discussed in section 5.1.1. The results are given in Fig 5.27 (a) and (b) respectively. In the BST Choshi experiment, it could capture the SR to some extent however, R is lower (0.44) than the control run (0.57) and also the RMSE value increased from 81.7 W/m^2 from 76.71 W/m^2. TCC also has a similar performance in the BST Choshi run. In precipitation BST Choshi experiment overestimate the precipitation events causing to RMSE

to increase to 17.84 mm/day, and BST Choshi experiment also has similar unrealistic precipitation events as in control experiment. In BST2 experiments results of TCC and precipitation reduce further. However, in solar radiation BST2 experiments have slightly higher performance than BST 1 Choshi experiment. Change in temperature is small.

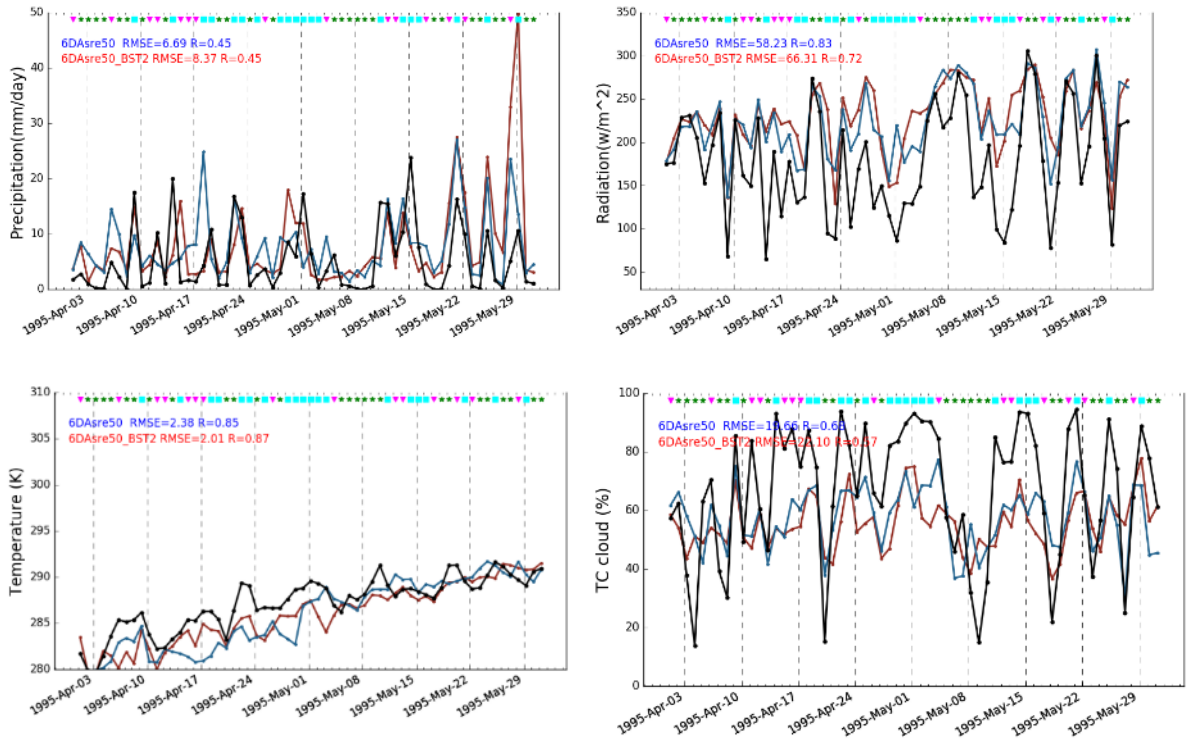


(a) BST Choshi simulation (b) BST2 simulation

-- BST experiments -- Regular Assimilation (control run) -- Observations

Fig 5.27: BST Choshi performance at Choshi station

Fig 5.28 and Table 5-10 shows the performance over all the stations. Removing Choshi station has a slight impact on the average performance. Interestingly it improves the performance in SR and TCC in both RMSE and correlation and in precipitation correlation improves while RMSE worsen. In temperature, correlation decreases 0.03 and RMSE improve by 0.02. This shows that impact of missing one station is minor and improvement in some variables can be due to the mismatch between the local weather and the climate in the grid as Choshi station stay very near to sea while most of the grid consists of land including Tokyo area. In the BST2 experiment where half of the observation removed, noticeable decrees in all the variables could be observed in average. For instance, SR RMSE was increased to 55 % from 58.2 %, and correlation decreased to 0.7 from 0.83. However, still results are better than no data assimilation. Without data assimilation, SR RMSE was high as 102 and correlation was 0.06, and in precipitation, correlation was negative while even half of the observations it is possible to have a closer correlation in precipitation and default data assimilation



-- Default data assimilation experiment -- Bootstrap experiment (BST2) -- Observations

Fig 5.28: Bootstrap simulations by neglecting half of the observations performance over all the stations

Table 5-10: Performance of the bootstrap simulations in all the observation stations average

Variable	RMSE/R	Default DA	BST Choshi	BST 2	Noobs
TCC	RMSE (%)	19.7	19.0	22.1	28.1
	R	0.68	0.74	0.57	0.19
SR	RMSE(W/m^2)	58.2	54	66	102
	R	0.83	0.86	0.7	0.06
Temp	RMSE(K)	2.38	2.36	2.0	1.84
	R	0.85	0.82	0.87	0.86
Precipitation	RMSE (mm/day)	6.7	7.6	8.4	6.47
	R	0.45	0.58	0.45	-0.19

5.6. Conclusion

In this chapter, the proposed model was evaluated with recent instrumental data in the 1990s. By assimilating SR derived from the weather classes, improvement in Solar radiation, TCC, Precipitation and Temperature could be observed. Assimilation of TCC also gave similar improvement even though it is less in overall accuracy than SR assimilation. Poor SST, Sea-Ice fraction and uncertainty of weather classes limit the accuracy of reconstruction. Precipitation was overestimated due to shift in the synoptic scale events and limited number of weather classes which doesn't distinguish between heavy and light precipitation events. Those events could be eliminated to some extent by assimilating information about sunny days (i.e. No precipitation days). From results of 1995-1999 it could be shown that model could capture the annual trend precipitation. In spring the model could capture synoptic scale events such as ExT cyclones even though location and time were little bit shifted. In the bootstrap experiment even without half of the observations the results were improved. Thus, assimilation of weather diary information can improve the surrounding areas as well. In summary assimilation of weather classes data enable the model to capture observed precipitation events and other variations indicating the advantage of using weather diary data in climate reconstructions.

Chapter 6

Weather Reconstruction Using 19th Century Diary Data

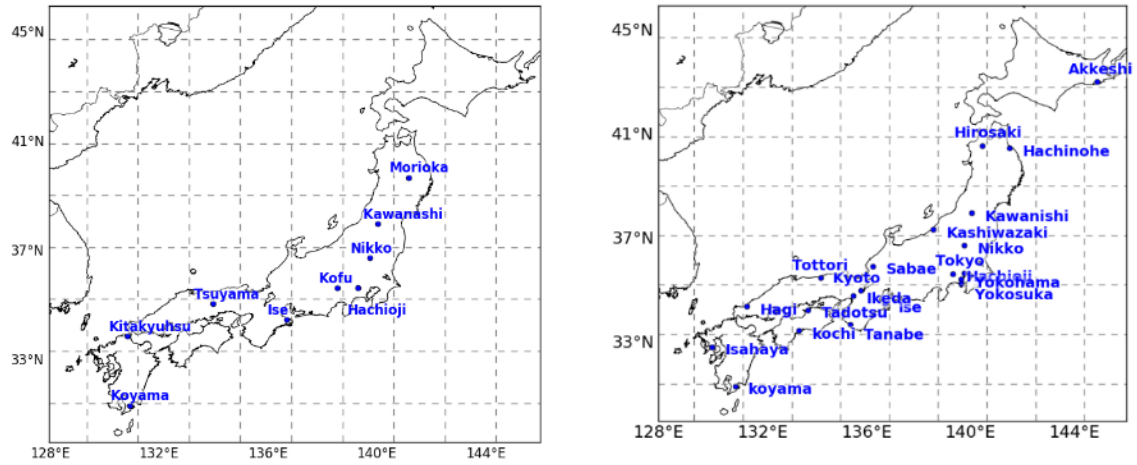
Abstract:

This chapter assimilate weather information from weather diaries in 19th century into the climate model for the first time in the historical data assimilation field using the settings and parameters identified from Chapter 2 and Chapter 3 with weather classes from other studies for 1830s and weather classes directly from derived from Historical Weather Data Base for 1860s as explained in Chapter 2. The model could capture weather types such as 'cloudy' and 'sunny' after data assimilation in 1830s similar to conditions of weather classes. Similar skill was observed in 1860s experiments. Due to the lack of instrumental data for daily comparison, monthly temperature from early instruments in Yokohama was used to check the model performance. The correlation coefficient in temperature without data assimilation and with assimilation were 0.96 and 0.94 respectively which are evenly high because the model can capture the seasonality. The model could produce ExT cyclones similar to 1995 when several diaries indicate rainy (weather class 3) during the spring period. By investigating precipitation anomaly from 1861 to 1864, 1861 May shown to be wet (19.0mm/month higher) and 1864 relatively dry year (13.5 mm/month less).

6.1. Experiments in 1830

In 1830s SR assimilation using weather classes derived from diary data Ichino, (2007) was done. Only eight diaries which have records of 1830s are available as shown in Fig 6.1(a). Fig 6.2 shows the model performances in the winter period, here the green line shows the weather class. As highlighted by the black box in end of January (22nd January to 01st February)

and in End of February (18th February 28th February), model variables could be well constrained using weather classes assimilation. In the first-period weather class is continuously 2 (i.e. cloudy) and this was well represented by model's low SR during the period. The TCC has a higher value except on 27th January. In the second box the weather class change to 2 from 1 and again to 2 after 4 days. This was well captured by the model with data assimilation while the control run without assimilation could only capture peak for few days. In overall, the model with weather classes assimilation has an evident temporal variation whereas the control run without assimilation could not capture the trend in a weather classes and has a more substantial standard deviation in the ensembles. As there are no validation data available in this period further investigation was done with late 19th century data (i.e. 1863) where some monthly instrumental data is available.



(a) 1830 (b) 1860

Fig 6.1: Availability of weather diary data in the 1830s and 1860s

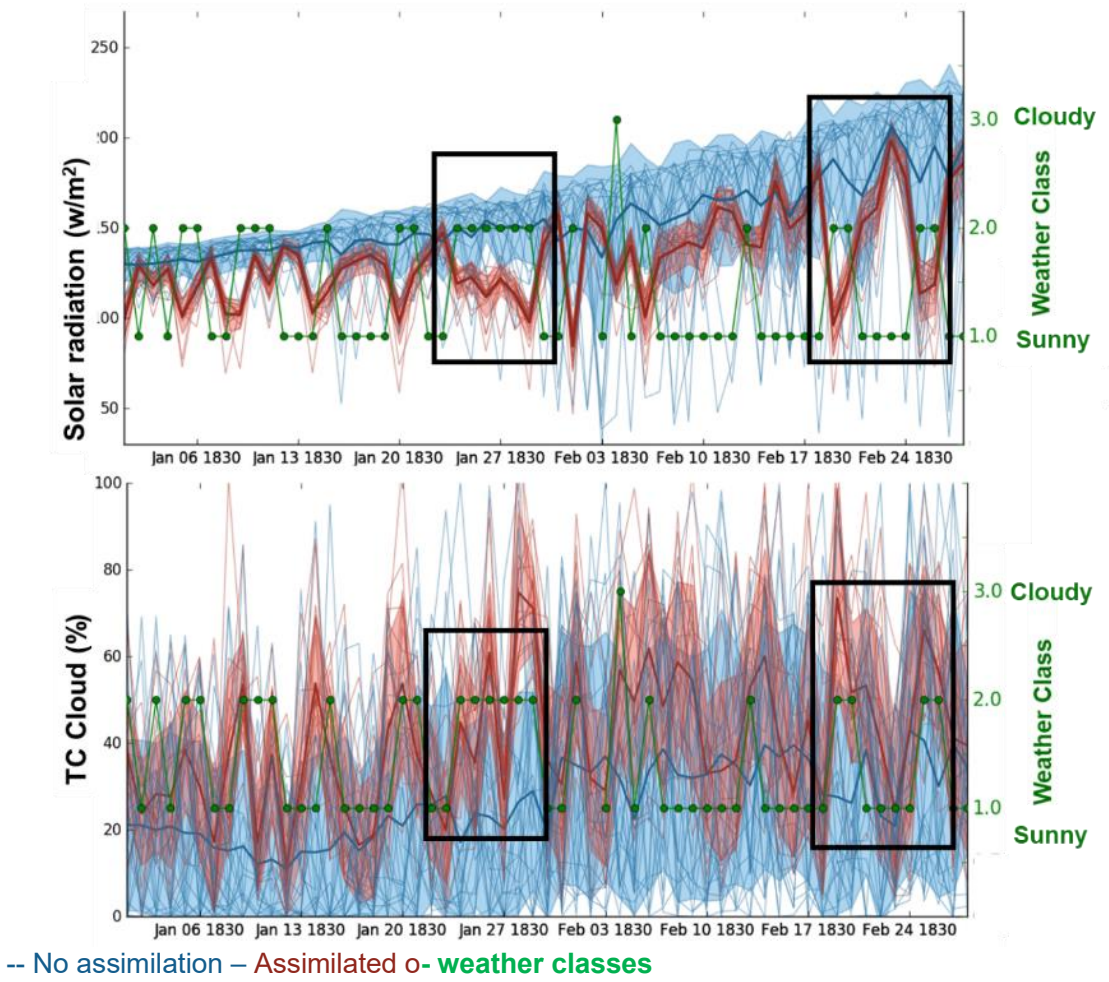
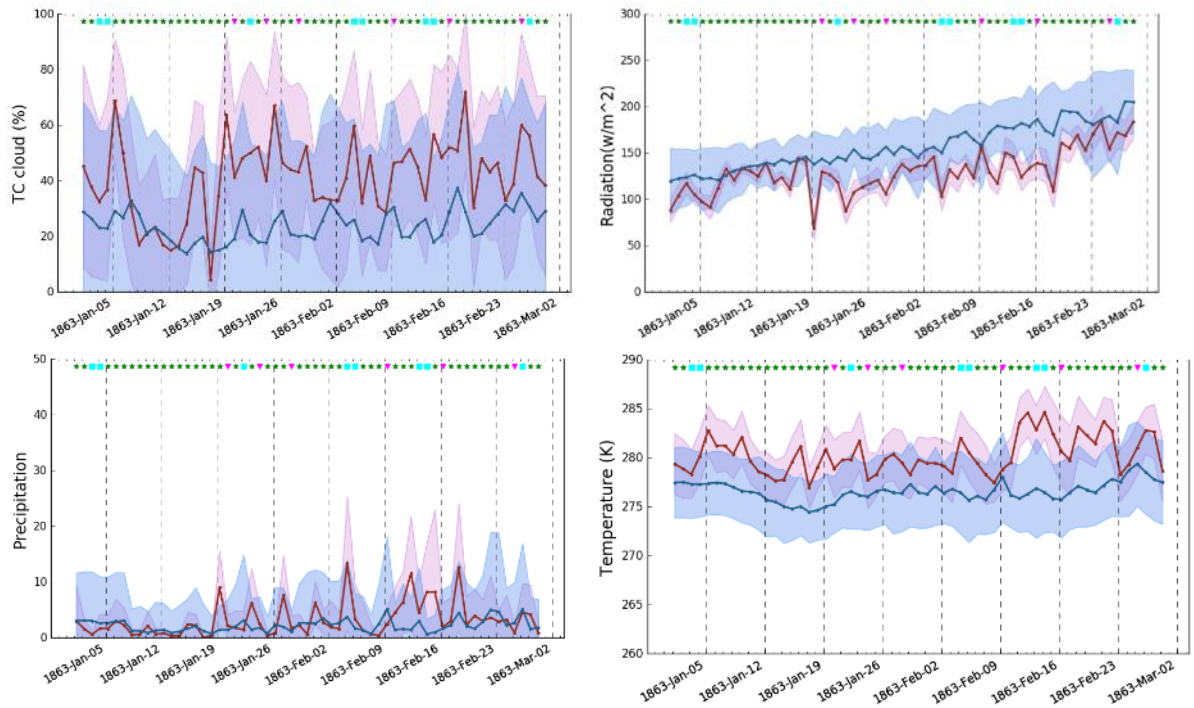


Fig 6.2: TCC (%) and SR(W/m2) variation at Choshi station in Jan-March in 1830

6.2. Experiments in the 1860s

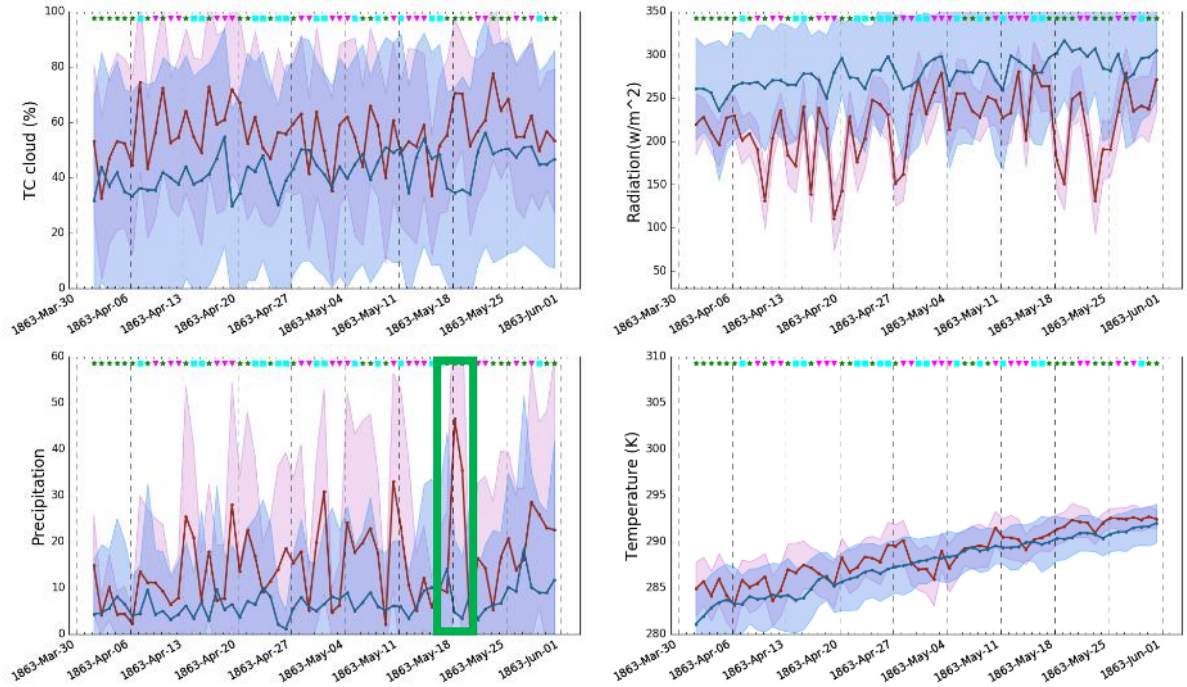
For this experiment, weather classes converted directly from HWDB as explained in section 2.2.2 was used. There are around 21 diary data in the 1860s as shown in Fig 6.1(b). However, several diaries overlap in the same model grid. In such circumstances, nearest observation to the model grid is used. The other document data can be used to evaluate observation uncertainty. The model performance is shown in Fig 6.3- Fig 6.6. By comparing the weather classes shown on the top of the figures and the model performance, it is clear that assimilation ensembles have a smaller standard deviation while the no assimilation run has a substantial uncertainty. Ensemble spread of TCC in the control run is quite high and the ensemble mean could not capture the weather classes on the other hand simulation with data assimilation has a narrow ensemble spread. Even though there are no direct instrumental observations in the 1860s, there is monthly temperature, and precipitation datasets and Comparison with those data will be discussed in the next section.



-- No assimilation – Assimilated

Fig 6.3: Model performance in 1863 winter at Choshi station. The standard deviation of the ensemble spread is indicated by the shaded color.

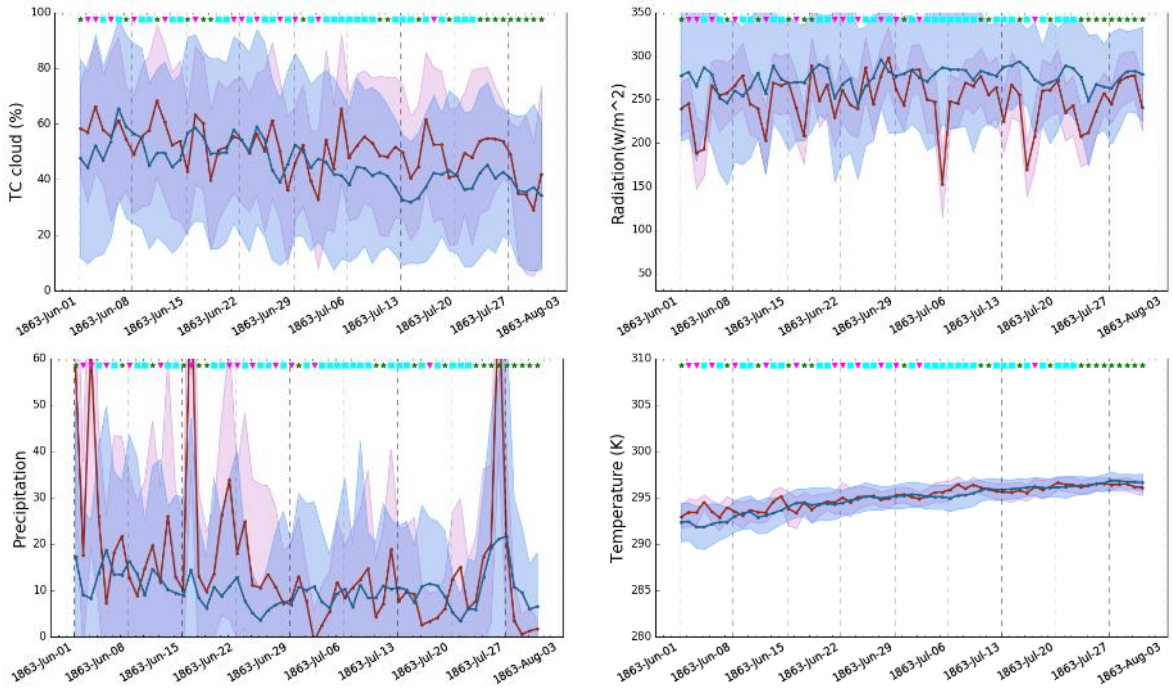
In winter Choshi station do not has large precipitation events. In spring there are several precipitation events, and this is captured in the ensemble mean of assimilation run whereas the control run without assimilation does not produce any precipitation events as experienced in1995.



-- No assimilation – Assimilated

Fig 6.4: Model performance in 1863 spring at Choshi station. The standard deviation of the ensemble spread is indicated by the shaded color.

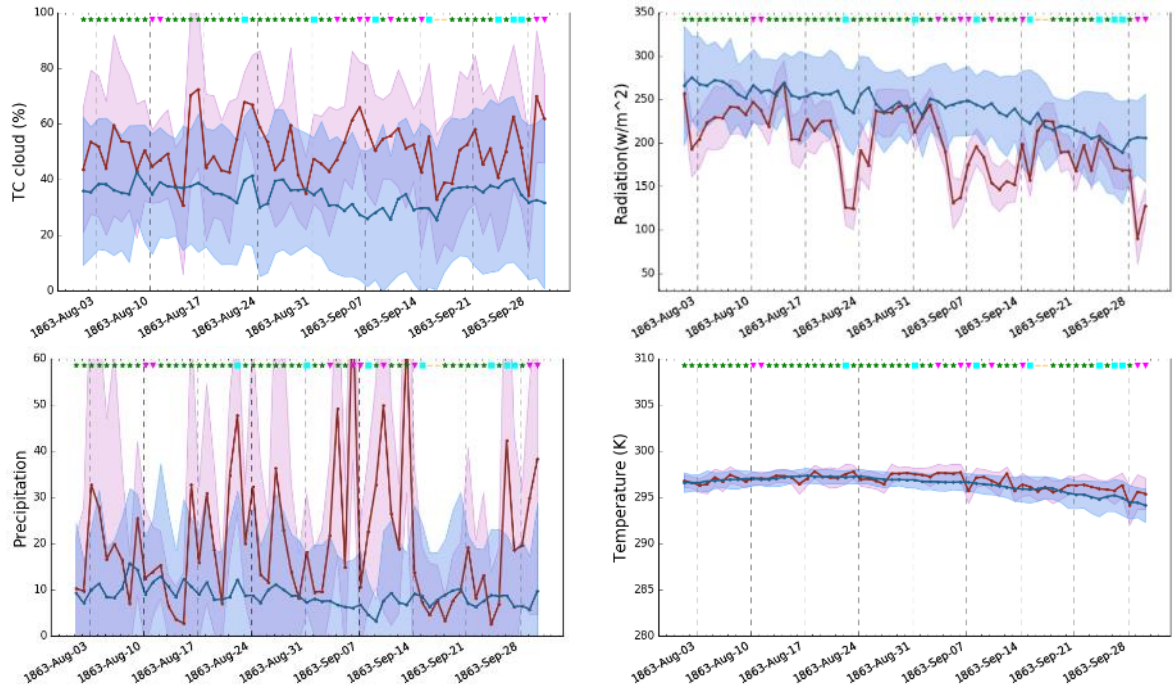
In the rainy season, many records indicate rain (Blue squares) model results show several intense precipitation events.



-- No assimilation – Assimilated

Fig 6.5: Model performance in 1863 rainy season at Choshi station. The standard deviation of the ensemble spread is indicated by the shaded color.

In the summer period, the sunny category is frequent and intense perception events could be observed during rainy and cloudy days.



-- No assimilation – Assimilated

Fig 6.6: Model performance in 1863 summer at Choshi station. The standard deviation of the ensemble spread is indicated by the shaded color.

6.2.1. The impact from ExT Cyclones

As highlighted by the green box in Fig 6.5, May 18th has higher precipitation. Fig. 6.7 shows the weather classes during 18th May 1863. Most of the diaries indicate rainy weather which may be mostly due to an Ext cyclone.

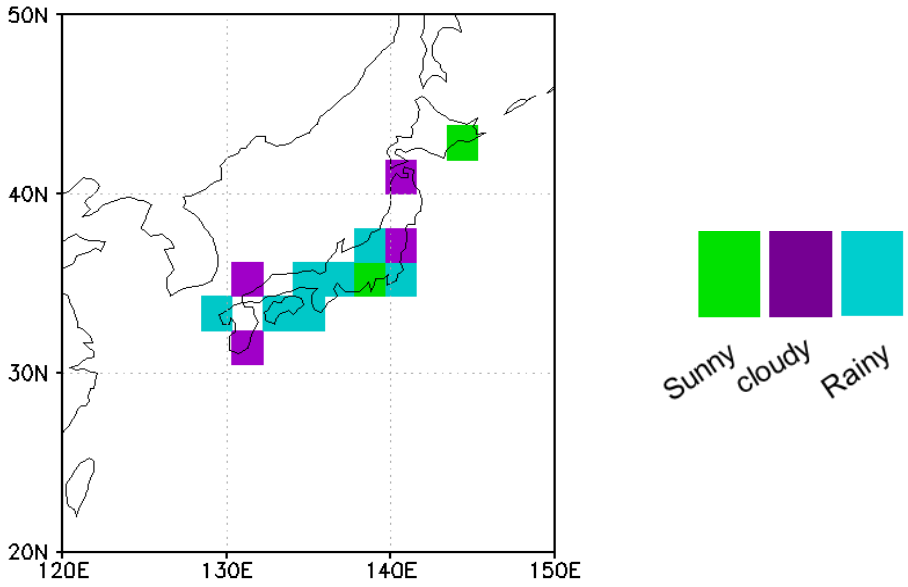


Fig. 6.7: Model weather classes in May 18th 1863

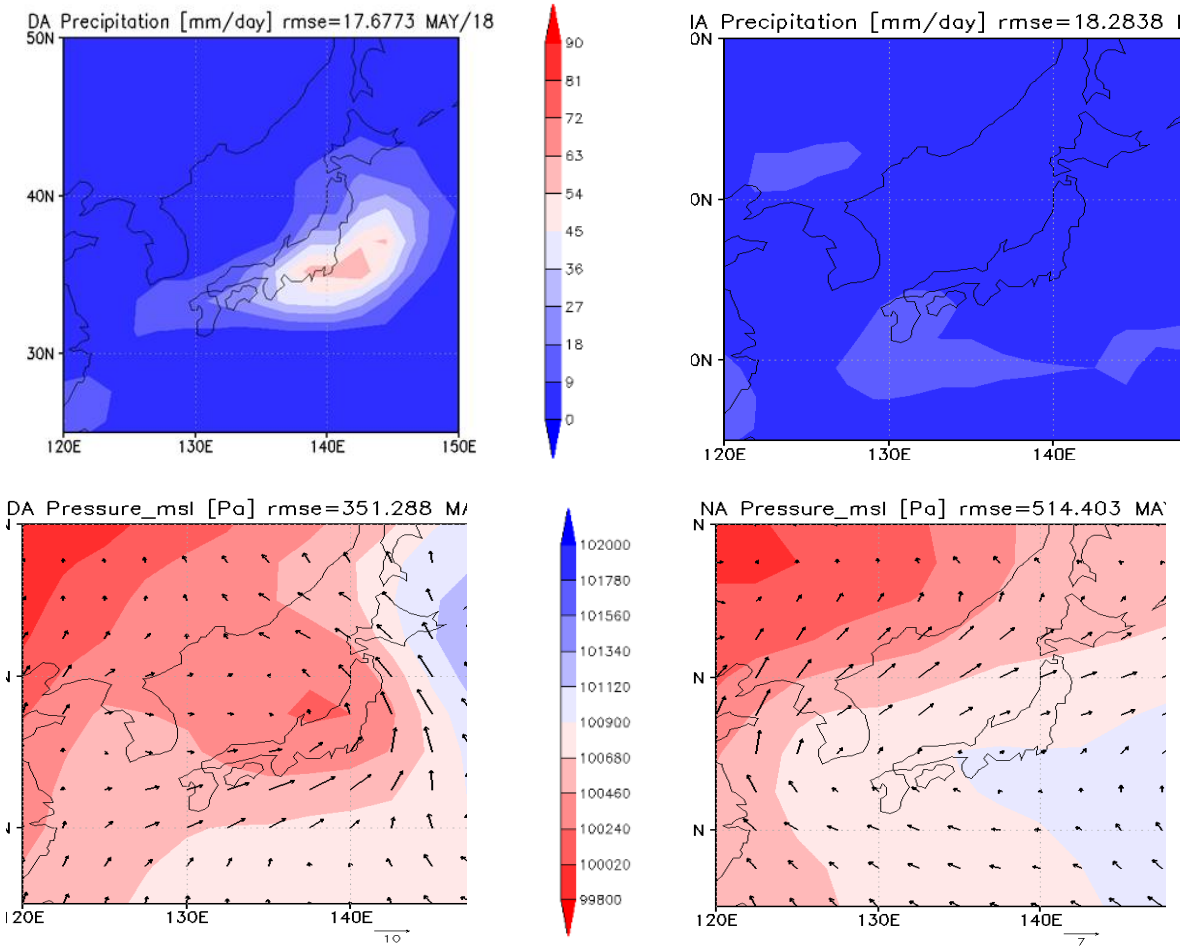


Fig. 6.8: Pressure and precipitation distribution of ExT Cyclones in May 18th 1863

6.3. Observed weather classes vs weather classes from the model.

As there are were no observations to compare in 1860s, model's weather classes can be compared with input weather classes to check wehather model could capture the weather classes. When assimilating weather descriptions, weather classes were converted to Solar radiation and assimilated to the model as explained in Chapter 2. A similar method was made to calculate weather classes backward from model Solar radiation values using data in recent period. Fig. 1 shows the relative solar radiation vs weather classes considering May data for five years from 1995-2000. This relationship was made to calculate weather classes from model simulations. As expected, solar radiation in the model has a clear correlation with observed weather classes.

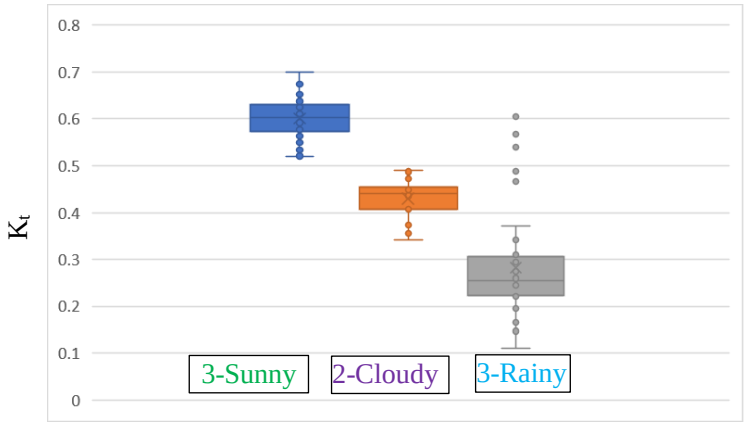


Fig 6.9: Relative solar radiation of model vs weather classes at Wajima station

The calculated Kt values at each month at each observation site can be used to calculate weather classes from the model. Kt values from Fig 6.9 was adjusted to maximize the equal number of stations for five years. Fig2. Shows the weather classes derived using the above method at Wajima station for five years. Results show assimilated solar radiation can represent the original weather classes well.

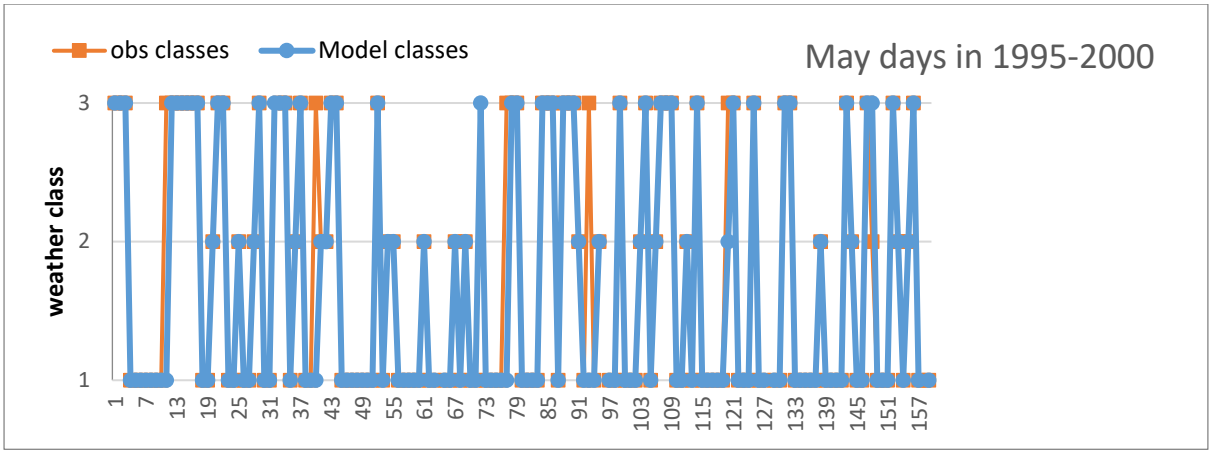


Fig 6.10: Weather classes in May from 1995-2000 at Wajima station

Table: Total classes from model and observations

	Weather class		
	3	2	1
Kt values	0.37	0.37-0.5	0.5
Model	45	21	88
Observations	42	21	91

This relationship was applied to Takada station in 1863, Kt values were increased by 0.12 to match the higher monthly mean of Takada station. Fig3 shows the model weather classes vs observed weather classes at Takada station in 1863 May. Even though the observation has several missing values a clear match can be seen with the model weather classes. Moreover, model indicate weather class 2 on 18th and 3 on 19th during the ExT cyclone.

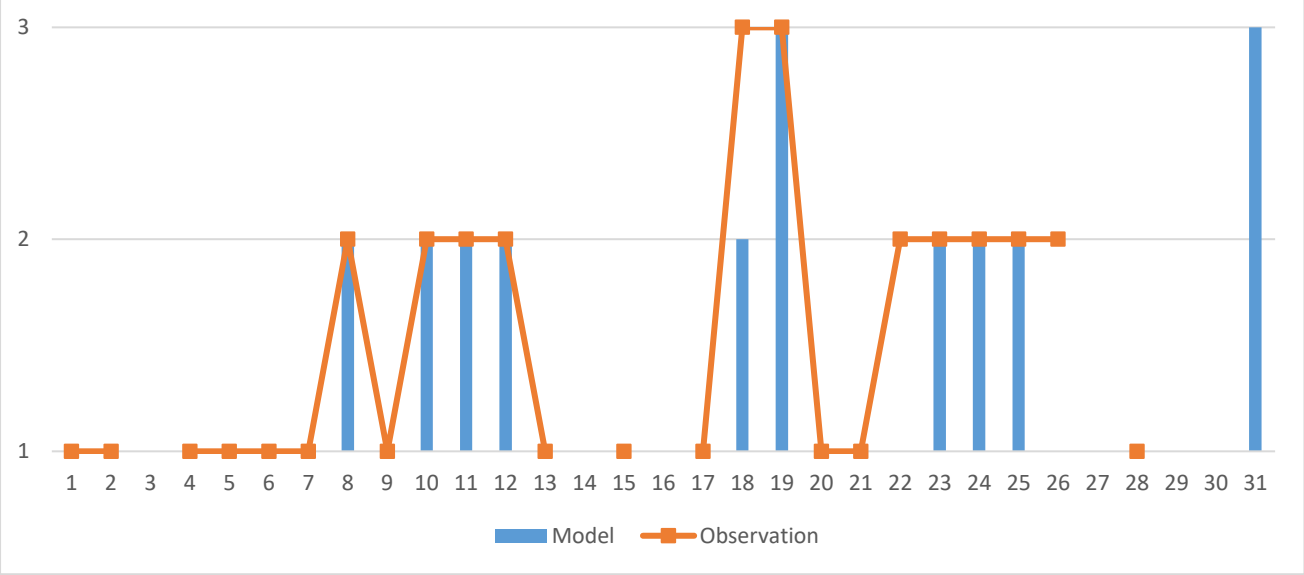


Fig 6.11: weather classes at Takada station in May 1863.

In Fig. 6.8 development of low-pressure center can be seen similar to 1995 by manually comparing pressure values at centre. It would be useful to make some method to identify ExT cyclones automatically in further studies.

6.4. Comparison with other proxies

There are several instrumental observations in the 1860s by different observers (i.e. P. Mourier (1865), P.A.L. Savatier (1866/12– 1868/01), J.C. Hepburn (1860, 1863–1869)) (Zaiki et al., 2006). However, these instrumental observations are not sometimes continuous and have a considerable uncertainty and do not follow standards. Hence only monthly Temperature and precipitation has been reconstructed from these early instrumental data (Zaiki et al., (2006), Hirano et al., (2018)). Fig 6.12 shows the assimilation and no assimilation run with the monthly observations. From 1995 observations it was found that the model has the skill to capture the seasonal climatology even without data assimilation. Hence here both experiments show similar performance and underestimation in summer was common in recent experiment as well.

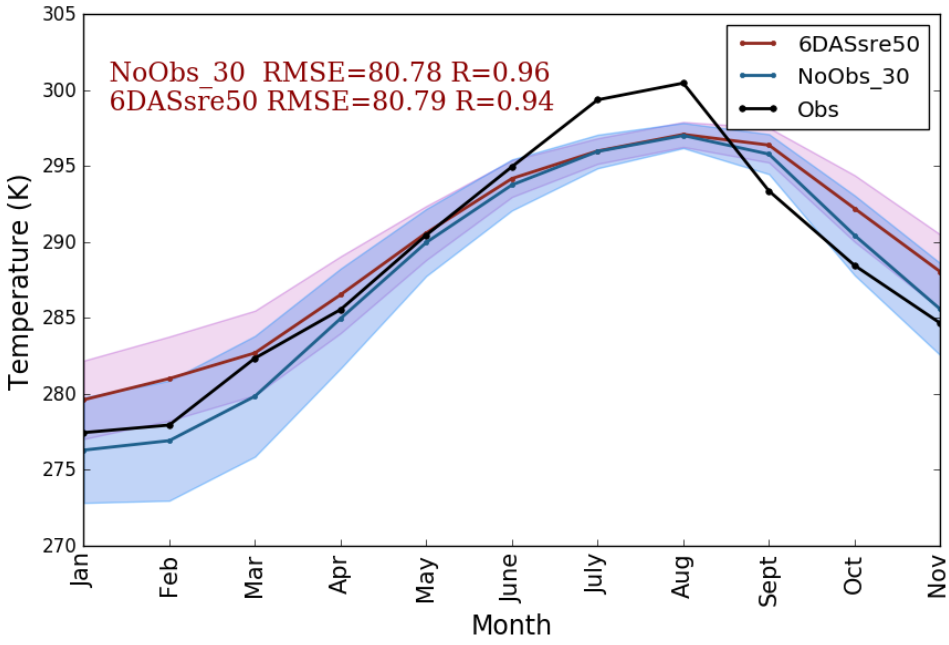


Fig 6.12: Comparison with observed monthly temperature reconstructed using instrumental observations in Yokohama (Zaiki et al., 2006)

6.5. Long-term trend

Due to the lack of instrumental data for daily comparison, monthly temperature from early instruments in Yokohama was used to check the model performance. The correlation

coefficient in temperature without data assimilation and with assimilation were 0.96 and 0.94 respectively which are evenly high because the model can capture the seasonality. The model could produce ExT cyclones similar to 1995 when several diaries indicate rainy (weather class 3) during the spring period. By investigating precipitation anomaly from 1861 to 1864, 1861 May shown to be wet (19.0mm/month higher) and 1864 relatively dry year (13.5 mm/month less).

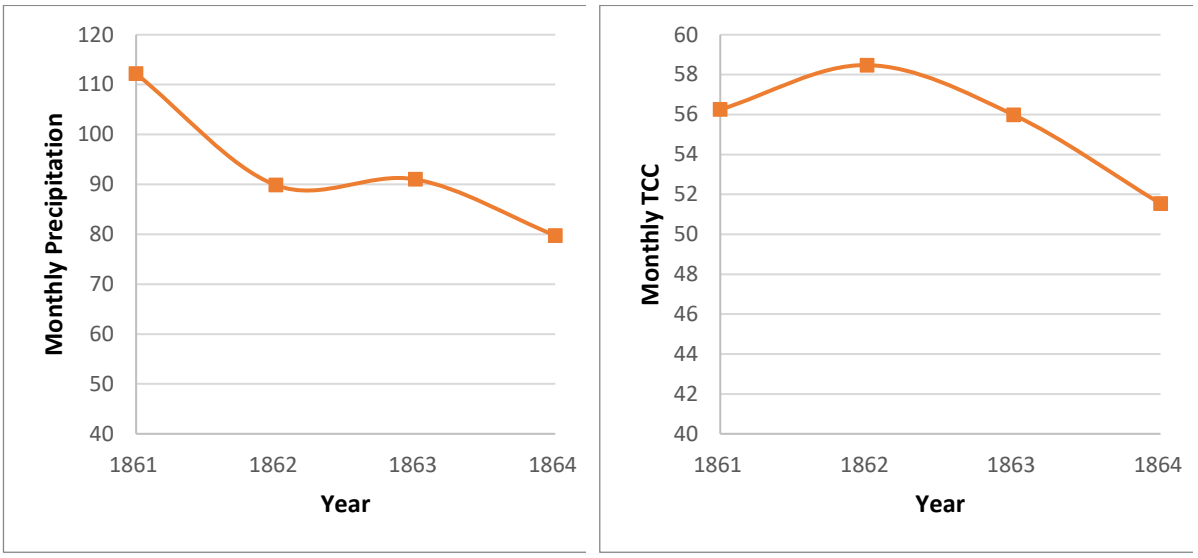


Fig 6.13: Annual variation of precipitation and TCC in Choshi station

6.6. Conclusion

Proposed weather reconstruction methodology could be successfully applied to 19th century using information from real weather diaries. This is the first study to use weather information from personal recording in a climate model. The performance of the model is was analyzed in different seasons and annually. Monthly performance was compared with other proxies. Weather classes calculated from model results show similar classes to observed weather classes. Model could produce synoptic scale weather events as seen in 1995. Further, annual trend was investigated to identify the long-term trend in monthly precipitation and TCC.

Chapter 7

Final Conclusion and Recommendations

Abstract:

The last chapter makes the conclusions and discusses recommendations for future studies. In this study, weather categories from qualitative description data could be converted to weather classes and assimilate into a general circulation model using a data assimilation scheme. The results showed that assimilation of weather classes using solar radiation improved the correlation of non-assimilated variables as well, and it was revealed that the resulted atmospheric distributions could capture the actual synoptic weather events. To further expand this research, more information from diaries such as wind direction and snow information can be utilized. This study used only the diary data in Japan. In future, this methodology can be applied globally when more digitized diaries are available in different regions.

7.1. Conclusion

The objective of this study is to develop a method to reconstruct weather during 18th and 19th centuries using personal diary data with a data assimilation scheme. For that GSM is used as the climate forecasting and LETKF was used as the data assimilation. Weather classes from HWDB were used for the past weather reconstruction and weather classes JMA description were used for validation the procedure.

Simplification of complex weather categories in HWDB could be successfully simplified into three simple weather classes and shown to be consistence with the weather classes from the other studies. Furthermore, the weather classes were shown to be correlated with SR and TCC.

Assimilation of weather classes using SR improves the correlation of non-assimilation variable as well. However, overestimation of precipitation events and a low correlation of temperature in winter period was noticed.

Information such as absence of precipitation could further improve the model performances removing the unrealistic rainfalls and improving the skill to capture more precipitation events.

7.2. Recommendations

In this study, the importance of using weather diary data was evaluated and practically shown with HWD in the 19th century. However the usage of weather diary data was limited to Japan, and this can be expanded by applying the same methodology to available weather documents found in other regions.

Information about the time of the weather events is not utilized due to the unavailability of usable format. If the assimilation can be done at the specific time slots using the records about the time, the results can be improved further.

Furthermore, empirical relationships between different variables can be investigated in area wise in monthly scale. With the aid of such relationship, by offline assimilation to the monthly averages of online assimilation may provide better results and would be an interesting area to investigate.

Even though the 19th-century model results are obtained, proper validation could not be done due to the unavailability of instrumental data. Other approaches such as validating with other proxy data and anomaly comparison can be made.

REFERENCES

Alpert, J.C., Kanamitsu, M., Calpan, P.M., Sela, J., White, G., Kalnay, E., 1988. Mountain induced gravity wave drag parameterization in the NMC medium-range forecast model, Proceedings of the 8th Conference on Numerical Weather Prediction. American Meteorological Society, Baltimore, Md, USA, pp. 726–733.

Anderson, J.L., 2009. Spatially and temporally varying adaptive covariance inflation for ensemble filters. Tellus A 61, 72–83. https://doi.org/10.1111/j.1600-0870.2008.00361.x

Aono, Y., Omoto, Y., 1994. Estimation of temperature at Kyoto since the 11th century using flowering data of cherry trees in old documents. J. Agric. Meteorol. 49, 263–272.

Bell, W.T., Ogilvie, A.E.J., 1978. Weather compilations as a source of data for the reconstruction of european climate during the medieval period. Clim. Change 1, 331–348. https://doi.org/10.1007/BF00135154

Bernhardt, J., 2015. Determining Regional Weather Patterns from a Historical Diary. Weather. Clim. Soc. 295–308. https://doi.org/10.1175/WCAS-D-15-0016.1

Bishop, C.H., Etherton, B.J., Majumdar, S.J., 2001. Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects. Mon. Weather Rev. 129, 420–436. https://doi.org/10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2

Buntgen, U., Tegel, W., Nicolussi, K., McCormick, M., Frank, D., Trouet, V., Kaplan, J.O., Herzig, F., Heussner, K.-U., Wanner, H., Luterbacher, J., Esper, J., 2011. 2500 Years of European Climate. Science (80-.). 331, 578–582.

Chou, M.-D., 1992. A Solar Radiation Model for Use in Climate Studies. J. Atmos. Sci. 49, 762–772. https://doi.org/10.1175/1520-0469(1992)049<0762:ASRMFU>2.0.CO;2

Chou, M.-D., Suarez, M.J., 1994. An efficient thermal infrared radiation parameterization for use in general circulation models. Nasa Tech. Memo 104606, 85.

Compo, G.P., Whitaker, J.S., Sardeshmukh, P.D., 2006. Feasibility of a 100-year reanalysis using only surface pressure data. Bull. Am. Meteorol. Soc. 87, 175–190. https://doi.org/10.1175/BAMS-87-2-175

D'Arrigo, R., Cook, E.R., Wilson, R.J., Allan, R., Mann, M.E., 2005. On the variability of ENSO over the past six centuries. Geophys. Res. Lett. 32, 1–4. https://doi.org/10.1029/2004GL022055

Dobrovolný, P., Moberg, A., Brázdil, R., Pfister, C., Glaser, R., Wilson, R., van Engelen, A., Limanówka, D., Kiss, A., Halíčková, M., Macková, J., Riemann, D., Luterbacher, J., Böhm, R., 2010. Monthly, seasonal and annual temperature reconstructions for Central Europe derived from documentary evidence and instrumental records since AD 1500. Clim. Change 101, 69–107. https://doi.org/10.1007/s10584-009-9724-x

Ek, M.B., Mitchell, K.E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., Tarpley, J.D., 2003. Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. J. Geophys. Res. 108, 8851. https://doi.org/10.1029/2002JD003296

Ensemble Prediction Systems [WWW Document], 2016. . Weather Predict. center(WPC). URL http://www.wpc.ncep.noaa.gov/ensembletraining/ (accessed 6.10.18).

Evensen, G., 1994. Inverse Methods and Data Assimilation in Nonlinear Ocean Models. Phys. D Nonlinear Phenom. 77, 108–129. https://doi.org/10.1016/0167-2789(94)90130-9

Franke, J., Brönnimann, S., Bhend, J., Brugnara, Y., Wanner, H., 2017. A monthly global paleo-reanalysis of the atmosphere from 1600 to 2005 for studying past climatic variations. Sci. Data 4, 170076. https://doi.org/10.1038/sdata.2017.76

Ge, Q.S., Zheng, J.Y., Hao, Z.X., Zhang, P.Y., Wang, W.C., 2005. Reconstruction of historical climate in China. Bull. Am. Meteorol. Soc. 86, 671–679. https://doi.org/10.1175/BAMS-86-5-671

Goosse, H., Brovkin, V., Fichefet, T., Haarsma, R., Huybrechts, P., Jongma, J., Mouchet, A., Selten, F., Barriat, P.Y., Campin, J.M., Deleersnijder, E., Driesschaert, E., Goelzer, H., Janssens, I., Loutre, M.F., Morales Maqueda, M.A., Opsteegh, T., Mathieu, P.P., Munhoven, G., Pettersson, E.J., Renssen, H., Roche, D.M., Schaeffer, M., Tartinville, B., Timmermann, A., Weber, S.L., 2010. Description of the Earth system model of intermediate complexity LOVECLIM version 1.2. Geosci. Model Dev. 3, 603–633. https://doi.org/10.5194/gmd-3-603-2010

Goosse, H., Guiot, J., Mann, M.E., Dubinkina, S., Sallaz-Damaz, Y., 2012. The medieval climate anomaly in Europe: Comparison of the summer and annual mean signals in two reconstructions and in simulations with data assimilation. Glob. Planet. Change 84–85, 35–47. https://doi.org/10.1016/j.gloplacha.2011.07.002

Hamill, T.M., Whitaker, J.S., Snyder, C., 2001. Distance-Dependent Filtering of Background Error Covariance Estimates in an Ensemble Kalman Filter. Mon. Weather Rev. 129, 2776–2790. https://doi.org/10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2

Hannak, A., Anderson, E., Barrett, L.F., Lehmann, S., Mislove, A., Riedewald, M., 2010. Tweetin ' in the Rain : Exploring Societal-scale Effects of Weather on Mood. Artif. Intell. 479–482. https://doi.org/10.1.1.222.9736

Harlim, J., Hunt, B., 2005. Local Ensemble Transform Kalman Filter: An Efficient Scheme for Assimilating Atmospheric Data. Preprint 1–18.

Hiroyuki Kusu, 2013. 学んでみると気候学はおもしろい. ベレ出版, Shikoku.

Hong, S.-Y., Pan, H.-L., 1998. Convective Trigger Function for a Mass-Flux Cumulus Parameterization Scheme. Mon. Weather Rev. 126, 2599–2620. https://doi.org/10.1175/1520-0493(1998)126<2599:CTFFAM>2.0.CO;2

Hunt, B.R., Kostelich, E.J., Szunyogh, I., 2007. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. Phys. D Nonlinear Phenom. 230, 112–126. https://doi.org/10.1016/j.physd.2006.11.008

Ichino, M., 2007. Climatorogical study on distribution of the frequencies of global solar radiation for each weather categories at Tokyo*. お茶の水地理 Vol.47, p.15-26.

Japan Meteorological Agency, 2016. Climate Change Monitoring Report 2015.

Junpei Hirano，Takehiko Mikami，Masumi Zaiki，unji Nishina, 2018. Analysis of precipitation data at Yokohama, Japan, from 1863 to 1869 observed by J.C. Hepburn. Chigaku Zasi.

Kanamitsu, M., Alpert, J.C., Campana, K.A., Caplan, P.M., Deaven, D.G., Iredell, M., Katz, B., Pan, H.-L., Sela, J., White, G.H., 1991. Recent Changes Implemented into the Global

Forecast System at NMC. Weather Forecast. https://doi.org/10.1175/1520-0434(1991)006<0425:RCIITG>2.0.CO;2

Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S.-K., Hnilo, J.J., Fiorino, M., Potter, G.L., 2002a. Ncep--doe amip-ii reanalysis (r-2). Bull. Am. Meteorol. Soc. 83, 1631–1643. https://doi.org/10.1175/BAMS-83-11

Kanamitsu, M., Kumar, A., Juang, H.M.H., Schemm, J.K., Wang, W., Yang, F., Hong, S.Y., Peng, P., Chen, W., Moorthi, S., Ji, M., 2002b. NCEP dynamical seasonal forecast system 2000. Bull. Am. Meteorol. Soc. 83, 1019–1037. https://doi.org/10.1175/1520-0477(2002)083<1019:NDSFS>2.3.CO;2

Kitamot, A., 2015. Debut of Himawari-8 - Next Generation Weather Satellite [WWW Document]. Natl. Inst. Informatics. URL http://www.digital-typhoon.org/ (accessed 6.11.18).

Kodama, Y., 1992. Large-scale common features of subtropical precipitation zones (the Baiu Frontal Zone , the SPCZ , and the SACZ) Part I: Characteristics of subtropical frontal zones. J. Meteorol. Soc. Japan 70, 813–836. https://doi.org/10.1248/cpb.37.3229

Kurashima, A., 2002. Daigaku tekisuto no kikou. 古今書院.

Lamb, H.H., 2005. Climate, History and the Modern World, Economic Geography. https://doi.org/10.2307/143446

Lawrence, J.R., 1998. Isotopic spikes from tropical cyclones in surface waters: Opportunities in hydrology and paleoclimatology. Chem. Geol. 144, 153–160. https://doi.org/10.1016/S0009-2541(97)00090-9

Lorrey, A.M., Chappell, P.R., 2015. The “Dirty Weather” diaries of Reverend Richard Davis: insights about early Colonial-era meteorology and climate variability for Northern New Zealand, 1839–1851. Clim. Past Discuss. 11, 3799–3851. https://doi.org/10.5194/cpd-11-3799-2015

Lunt, D.J., Elderfield, H., Pancost, R., Ridgwell, A., Foster, G.L., Haywood, A., Kiehl, J., Sagoo, N., Shields, C., Stone, E.J., Valdes, P., 2013. Warm climates of the past--a lesson for the future? Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. 371, 20130146–20130146. https://doi.org/10.1098/rsta.2013.0146

M Zaiki, G P Können, T Tsukahara, P D Jones, T Mikami, K.M., 2006. Recovery of 19th century Tokyo / Osaka meteorological data in Japan., Int. J. Climatol.

Mann, M.E., Woodruff, J.D., Donnelly, J.P., Zhang, Z., 2009. Atlantic hurricanes and climate over the past 1,500 years. Nature 460, 880–883. https://doi.org/10.1038/nature08219

McGregor, S., Timmermann, a., England, M.H., Elison Timm, O., Wittenberg, a. T., 2013. Inferred changes in El Niño-Southern Oscillation variance over the past six centuries. Clim. Past 9, 2269–2284. https://doi.org/10.5194/cp-9-2269-2013

Mika Ichino, Naomi Sakamoto, Kooti Masuda, T.M., 2001. The Method for Estimating Global Solar Radiation Based on Weather Records : Toward the Climatic Reconstruction in the Historical Period*. Tokyo Tenki 48, 823–830.

Mikami, T., 2008. Climatic variations in Japan reconstructed from historical documents. Weather 63, 190–193. https://doi.org/10.1002/wea.281

Miyoshi, T., 2011. The Gaussian Approach to Adaptive Covariance Inflation and Its

Implementation with the Local Ensemble Transform Kalman Filter. Mon. Weather Rev. 139, 1519–1535. https://doi.org/10.1175/2010MWR3570.1

Miyoshi, T., Yamane, S., 2007. Local Ensemble Transform Kalman Filtering with an AGCM at a T159/L48 Resolution. Mon. Weather Rev. 135, 3841–3861. https://doi.org/10.1175/2007MWR1873.1

MOEJ, 2012. Climate Change and Its Impacts in Japan.

Moorthi, S., Suarez, M.J., 1992. Relaxed Arakawa-Schubert. A Parameterization of Moist Convection for General Circulation Models. Mon. Weather Rev. 120, 978–1002. https://doi.org/10.1175/1520-0493(1992)120<0978:RASAPO>2.0.CO;2

NCEP_Reanalysis 2 data [WWW Document], n.d. . NOAA/OAR/ESRL PSD, Boulder, Color. USA. URL https://www.esrl.noaa.gov/psd/

O'Brien, M.J., Philipson, C.D., Tay, J., Hector, A., 2013. The Influence of Variable Rainfall Frequency on Germination and Early Growth of Shade-Tolerant Dipterocarp Seedlings in Borneo. PLoS One 8, 1–9. https://doi.org/10.1371/journal.pone.0070287

Okazaki, A., Yoshimura, K., 2017. Development and evaluation of a system of proxy data assimilation for paleoclimate reconstruction. Clim. Past 13, 379–393. https://doi.org/10.5194/cp-13-379-2017

Orszag, S. a., 1970. Transform Method for the Calculation of Vector-Coupled Sums: Application to the Spectral Form of the Vorticity Equation. J. Atmos. Sci. https://doi.org/10.1175/1520-0469(1970)027<0890:TMFTCO>2.0.CO;2

Ott, E., Hunt, B.R., Szunyogh, I., Zimin, A.V.A., Kostelich, E.J., Corazza, M., Kalnay, E., Patil, D.J., Yorke, J. a., 2002. A Local Ensemble Kalman Filter for Atmospheric Data Assimilation. Tellus A A 56, 66. https://doi.org/10.1111/j.1600-0870.2004.00076.x

Ott, E., Hunt, B.R., Szunyogh, I., Zimin, A. V., Kostelich, E.J., Corazza, M., Kalnay, E., Patil, D.J., Yorke, J.A., 2004. A local ensemble Kalman filter for atmospheric data assimilation. Tellus, Ser. A Dyn. Meteorol. Oceanogr. 56, 415–428. https://doi.org/10.1111/j.1600-0870.2004.00076.x

Rayner, N.A., Parker, D.E., Horton, E.B., Folland, C.K., Alexander, L.V., Rowell, D.P., Kent, E.C., Kaplan, A., 2003. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. J. Geophys. Res. 108, 4407. https://doi.org/10.1029/2002JD002670

Reichler, T.J., Roads, J.O., 2003. The role of boundary and initial conditions for dynamical seasonal predictability. Nonlinear Process. Geophys. 10, 211–232. https://doi.org/10.5194/npg-10-211-2003

Reynolds, R.W., Smith, T.M., Liu, C., Chelton, D.B., Casey, K.S., Schlax, M.G., 2007. Daily high-resolution-blended analyses for sea surface temperature. J. Clim. 20, 5473–5496. https://doi.org/10.1175/2007JCLI1824.1

Saha, S., Nadiga, S., Thiaw, C., Wang, J., Wang, W., Zhang, Q., Van den Dool, H.M., Pan, H.L., Moorthi, S., Behringer, D., Stokes, D., Pe??a, M., Lord, S., White, G., Ebisuzaki, W., Peng, P., Xie, P., 2006. The NCEP Climate Forecast System. J. Clim. 19, 3483–3517. https://doi.org/10.1175/JCLI3812.1

Seppä, H., Birks, H.J.B., 2001. July mean temperature and annual precipitation trends during the Holocene in the Fennoscandian tree-line area: Pollen-based climate reconstructions.

Holocene 11, 527–539. https://doi.org/10.1191/095968301680223486

Steiger, N.J., Steig, E.J., Dee, S.G., Roe, G.H., Hakim, G.J., 2017. Climate reconstruction using data assimilation of water isotope ratios from ice cores. J. Geophys. Res. Atmos. 122, 1545–1568. https://doi.org/10.1002/2016JD026011

Szunyogh, I., Satterfield, E. a, Elana, J., Gyarmati, G., Kalnay, E., Hunt, B.R., Eric, J., Kuhl, D.D., Ott, E., Yorke, J. a, 2007. The Local Ensemble Transform Kalman Filter and its implementation on the NCEP global model at the University of Maryland The LETKF and its implementation on the NCEP GFS model. Proc. ECMWF Work. flow-dependent Asp. data Assim. 1–18.

Takehiko MIKAMI, 1988. Climatic Reconstruction Based on Weather in Historical Records Times. Geogr. Rev. Jpn. 61, 14–22.

Tiedtke, M., 1983. The sensitivity of the time-mean large-scale flow to cumulus convection in the ECMWF model, in: ECMWF Workshop on Convection in Large-Scale Models. Shinfield Park, Reading, U. K., pp. 297–316.

Tippett, M.K., Anderson, J.L., Bishop, C.H., Hamill, T.M., Whitaker, J.S., 2003. Ensemble Square Root Filters. Mon. Weather Rev. 131, 1485–1490. https://doi.org/10.1175/1520-0493(2003)131<1485:ESRF>2.0.CO;2

Toride, K., Neluwala, P., Kim, H., Yoshimura, K., 2017. Feasibility Study of the Reconstruction of Historical Weather with Data Assimilation. Mon. Weather Rev. 145, 3563–3580. https://doi.org/10.1175/MWR-D-16-0288.1

Toth, Z., Kalnay, E., 1993. Ensemble Forecasting at NMC: The Generation of Perturbations. Bull. Am. Meteorol. Soc. https://doi.org/10.1175/1520-0477(1993)074<2317:EFANTG>2.0.CO;2

van Leeuwen, P.J., 2009. Particle Filtering in Geophysical Systems. Mon. Weather Rev. 137, 4089–4114. https://doi.org/10.1175/2009MWR2835.1

Wang, P.K., 1992. Recent Studies of the Reconstruction Climate in the Past Using Historical of East Asian Monsoon Literature of China. J. Meteorol. Soc. Japan 70–1, 423–446.

Weather diary records since the 18th century [WWW Document], 2017. . Japan-Asia Clim. Data Progr. URL http://jcdp.jp/historical-documents/ (accessed 5.28.18).

Whitaker, J.S., Hamill, T.M., 2002. Ensemble Data Assimilation without Perturbed Observations. Mon. Weather Rev. 130, 1913–1924. https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2

Whitaker, J.S., Hamill, T.M., Wei, X., Song, Y., Toth, Z., 2008. Ensemble Data Assimilation with the NCEP Global Forecast System. Mon. Weather Rev. 136, 463–482. https://doi.org/10.1175/2007MWR2018.1

Yoshimura, K., Kanamitsu, M., 2013. Incremental Correction for the Dynamical Downscaling of Ensemble Mean Atmospheric Fields. Mon. Weather Rev. 141, 3087–3101. https://doi.org/10.1175/MWR-D-12-00271.1

Yoshimura, K., Miyoshi, T., Kanamitsu, M., 2014. Observation system simulation experiments using water vapor isotope information. J. Geophys. Res. Atmos. 1–21. https://doi.org/10.1002/2014JD021662.Received

Yoshimura, M., 2006. Historical Weather Database on The web [WWW Document]. URL

http://tk2-202-10627.vs.sakura.ne.jp/index_hw.html (accessed 3.25.18).

Yoshimura, M., 1993. Historical Weather Data Base and Reconstruction the Climate in Historical Time. J. Geog. 102, 131–143.

Zaiki, M., Können, G.P., Tsukahara, T., Jones, P.D., Mikami, T., Matsumoto, K., 2006. Recovery of nineteenth-century Tokyo/Osaka meteorological data in Japan. Int. J. Climatol. 26, 399–423. https://doi.org/10.1002/joc.1253

Zhang, B., Cao, J., Bai, Y., Zhou, X., Ning, Z., Yang, S., Hu, L., 2013. Effects of rainfall amount and frequency on vegetation growth in a Tibetan alpine meadow. Clim. Change 118, 197–212. https://doi.org/10.1007/s10584-012-0622-2

Zhang, X.Z., Ge, Q.S., Fang, X.Q., Zheng, J.Y., Fei, J., 2013. Precipitation variations in Beijing during 1860-1897 AD revealed by daily weather records from the Weng Tong-He Diary. Int. J. Climatol. 33, 568–576. https://doi.org/10.1002/joc.3448