



UNIVERSITY OF THE PHILIPPINES

Master of Science in Meteorology

Crizzia Mielle M. de Castro

*Hyperspectral Unmixing as an Analog Forecasting Method during
Strong Monsoon Events in the Philippines*

Thesis Adviser:

Gerry Bagtasa, Ph.D.

Institute of Environmental Science and Meteorology
University of the Philippines Diliman

Date of Submission:
July 2022

Thesis Classification:
F

This thesis is available to the public

Institute of Environmental Science and Meteorology
College of Science
University of the Philippines
Diliman, Quezon City

ENDORSEMENT

This is to certify that this thesis entitled **Hyperspectral Unmixing as an Analog Forecasting Method during Strong Monsoon Events in the Philippines**, prepared and submitted by Crizzia Mielle M. de Castro in partial fulfillment of the requirements for the degree of Master of Science in Meteorology, is hereby accepted.

GERRY BAGTASA, Ph.D.
Thesis Adviser

OLIVIA C. CABRERA, Ph.D.
Thesis Reader

MAY T. LIM, Ph.D.
Thesis Co-Adviser

The Institute of Environmental Science and Meteorology endorses acceptance of this thesis in partial fulfillment of the requirements for the degree of Master of Science in Meteorology.

CHEERY L. RINGOR, Ph.D.
Director
Institute of Environmental Science
and Meteorology

The thesis is hereby officially accepted in partial fulfillment of the requirements for the degree of Master of Science in Meteorology.

GIOVANNI A. TAPANG, Ph.D.
Dean, College of Science

Ad astra per aspera

Acknowledgments

I would like to acknowledge the help of the following..

Sir Gerry and Ma'am May for providing me with guidance as a Master's student. Thank you so much for all your patience and understanding as I slowly discovered new things.

Ma'am Olive, Ma'am Jing, and Ma'am Flavy for all your helpful comments and suggestions as my thesis panellists.

IESM and our professors for all the support and new skills I learned during my Master's.

My batchmates **Joanne, Cy, Agatha, Marielle, and Phoebe** for making my Master's experience a little bit more fun despite the online classes set-up.

Everyone in **UPians and ELYU** for all the super fun game nights, movie nights, study dates, and chikahan sessions we had throughout our whole grad school experience. You kept me sane and happy during my most stressful moments.

Da, Somido, and Viel, my favorite low maintenance friends, for being there for me during my highest and lowest points. The long distance apart never managed to diminish our friendship. I love you all, always.

Liz, my favorite person and partner in life, for all the care and love you showered me despite being miles and miles apart. You are my rock, and being with you makes me feel so unbelievably happy and fulfilled. I wouldn't have gotten this far without you. I love you.

Isay, Kosmo, and all the stray cats that visit our house for offering me a unique kind companionship no person can ever match. You helped me feel at ease and de-stress whenever the world becomes a little too loud.

My family for the unwavering love and support in everything I do and accomplish. You helped me feel more comfortable and more sure of myself. I hope I have made you proud.

I would also like to acknowledge the financial support of the following...

DOST-SEI for my scholarship.

ABSTRACT

HYPERSPECTRAL UNMIXING AS AN ANALOG FORECASTING METHOD DURING STRONG MONSOON EVENTS IN THE PHILIPPINES

Crizzia Mielle M. de Castro
University of the Philippines (2022)

Adviser:
Gerry Bagtasa, Ph.D.
Co-Adviser:
May T. Lim, Ph.D.

The interaction between the monsoon winds and terrain causes rainfall along the coastal regions of western and eastern Philippine during the southwest monsoon (Habagat) and the northeast monsoon (Amihan) seasons, respectively. The destructive capabilities of heavy rainfall events arising from this interaction during active monsoon days necessitates an efficient and computationally cheap complement to our current rainfall forecasting methods.

This study applies the analog forecasting technique to predict strong monsoon rains. In analog forecasting, the rainfall distribution of different days with similar weather patterns should look similar. A day in the past with similar environmental conditions as today can be used to forecast today's rainfall. Hence, the goal of analog forecasting is to look at days in the past with similar environmental conditions as today in the hopes that rainfall then will be similar to today's rainfall. The environmental conditions used to compare historical rainfall days are the mean sea-level pressure (SLP) and relative humidity (RH) obtained from the (2001-2018) JRA-55 reanalysis dataset. They are chosen as SLP and RH are the main meteorological parameters that describe the air-land interaction that drive heavy monsoon rainfall in the country.

In this study, we test a novel approach using the hyperspectral unmixing method in the selection of historical days with similar weather patterns. The method selects past days with varying degrees of similarity to a target forecast day. A composite weighted-ensemble mean rainfall map is then produced as the rainfall forecast. Utilization of this method requires three steps: (1) reducing data dimension, (2) unmixing the patterns' signatures, and (3) estimating the weights of each pattern. Then, rainfall from the GPM IMERGE are compiled to construct rainfall forecast maps. As a point of comparison, we also selected weather patterns of historical rainfall days using the direct correlation method where past SLP maps most correlated to the target

forecast day are chosen as analogs. To assess the skill of the forecasts, accuracy scores were calculated using the observed rainfall obtained from PAGASA weather stations. Forecasts with a threat score of at least 0.25 are deemed acceptable. The SLP, RH, and satellite-derived rainfall dataset from 2001-2020 were randomly separated into a training phase (90%) and a testing phase (10%). The training phase data were used to find the optimal SLP domains, rainfall scaling, and the weather archetypes (combination of SLP, RH, and rainfall) characterizing strong monsoon events. The testing phase data were then used as target rainfall forecast days.

Based on the forecast skill scores of the testing phase, hyperspectral unmixing and direct correlation performed equally in predicting strong Habagat and Amihan rainfall. Both methods tend to underestimate rainfall directly caused by a tropical cyclone during the Habagat season. Both methods also tend to overestimate rainfall when the continental high over China extends southeastward during strong Amihan, but does not necessarily result to heavy rainfall along east Philippines. The median skill for strong Amihan rainfall forecasts have threat scores 0.3, probabilities of detection 0.35, and false alarm ratios 0.15. On the other hand, the median skill for strong Habagat rainfall forecasts have threat scores 0.5, probabilities of detection 0.8, and false alarm ratios 0.4. The aforementioned threshold assessment scores are the average scores when checking for heavy rainfall in the testing phase. Individually, 12 out of 13, and 15 out of 17 of strong Habagat and Amihan test forecasts, respectively, reach the standard threat score 0.25. While the result of this study does not show a really robust forecasting skills, some acceptable rainfall forecasts were observed. Unlocking its potential and overcoming its limiting inabilities to forecast TC-induced heavy rainfall and lack of heavy rainfall during a strong northeasterly flow requires further investigation.

Table of Contents

Acknowledgments	iii
Abstract	v
List of Figures	viii
List of Tables	viii
1 Introduction	1
1.1 Statement of the Problem	3
1.2 Significance	3
1.3 Objectives	4
1.4 Research Questions	4
1.5 Scope and Limitations	5
2 Review of Literature	6
2.1 Philippine Monsoons	6
2.1.1 Strong monsoon events in the Philippines	8
2.2 Analog Forecasting	9
2.3 Hyperspectral Unmixing	11
2.3.1 First step of hyperspectral unmixing: dimension reduction . .	12
2.3.2 Second step of hyperspectral unmixing: unmixing	13
2.3.3 Final step of hyperspectral unmixing: inversion	14
3 Theory behind Hyperspectral Unmixing	15
3.1 Theory behind hyperspectral unmixing's dimension reduction	15
3.1.1 Noise estimation	15
3.1.2 Signal subspace estimation	16
3.2 Theory behind hyperspectral unmixing's unmixing	17
3.2.1 Vertex Component Analysis	17
3.2.2 Minimum Volume Simplex Analysis	20
3.3 Theory behind hyperspectral unmixing's inversion	25
4 Methodology	28
4.1 Filtering days with strong monsoon events	29
4.2 Sea-level pressure, relative humidity, and rainfall datasets	30
4.3 Applying hyperspectral unmixing on the SLP and RH of training datasets	32

4.4	Predicting rainfall distribution of test set based on SLP and RH	33
4.5	Analog forecasting using direct correlation	34
4.6	Verifying accuracy of rainfall forecasts using hyperspectral unmixing and direct correlation	35
5	Results and Discussion	38
5.1	Selection of strong monsoon days	38
5.2	Determining the optimal SLP domain for analog forecasting from the training set	41
5.2.1	Relative humidity as second rainfall predictor	42
5.2.2	Applying the hyperspectral unmixing method on the three domains	44
5.2.3	Correlation between the composite and actual SLP maps of the training set for different SLP domains	49
5.2.4	Rainfall forecast accuracy scores of the training set for different SLP domains	52
5.3	Determining scaling between actual and forecast rainfall	56
5.4	Weather archetypes and rainfall forecasts obtained for strong Habagat events	57
5.4.1	SLP archetypes obtained through hyperspectral unmixing	58
5.4.2	Rainfall archetypes obtained through successive SLP-RH hyperspectral unmixing	58
5.4.3	Forecast accuracy assessment scores	58
5.4.4	“Best” rainfall forecast using hyperspectral unmixing	59
5.4.5	“Worst” rainfall forecast using hyperspectral unmixing	61
5.5	Weather archetypes and rainfall forecasts obtained for strong Amihan events	63
5.5.1	SLP and RH archetypes obtained through hyperspectral unmixing	64
5.5.2	Rainfall archetypes obtained through hyperspectral unmixing	66
5.5.3	Forecast accuracy assessment scores	68
5.5.4	“Best” rainfall forecast using hyperspectral unmixing	68
5.5.5	“Worst” rainfall forecast using hyperspectral unmixing	70
5.6	Assessment of rainfall forecasts using hyperspectral unmixing on strong monsoon events	71
6	Conclusions and Recommendations	74

List of Figures

1.1	Mean sea-level pressure (colored isobars) and wind vectors over north-west Mediterranean Sea illustrating the relationship between SLP and wind(Figure 2 in Grifoll et al., 2016)	3
2.1	Daily mean rainfall and wind distribution with the monsoon trough (in red dashed lines) over the Philippines from February to November. (Figures 2 & 3 in Matsumoto et al., 2020)	7
2.2	Total column water vapor flux (gray shading), 850 mb wind (arrows) and mean sea level pressure (dark gray lines) during strong Habagat (Figure 4 in Bagtasa, 2019)	9
2.3	Mean sea level pressure (isobars) and surface temperature (shading) during a winter monsoon surge in December 30, 2004. (Figure 3.3b in Chang et al., 2006)	10
2.4	Diagram illustrating how linear hyperspectral unmixing decomposes the satellite image's pixel spectra into their constituent archetypes and weights. (Figure 2 in Bioucas-Dias et al., 2012)	11
2.5	Diagram of a 2-dimensional simplex with three archetypes (green vertices) encompassing the hyperspectral data (red dots).	13
2.6	Comparison of hyperspectral data (a) with, and (b) without the pure pixel assumption.	14
3.1	Illustration of the VCA algorithm. VCA projects S_p along \mathbf{f}_1 and \mathbf{f}_2 , then sets the extremes as the archetypes. (figure from Nascimento and Dias, 2005)	19
4.1	Methodological framework in applying hyperspectral unmixing to perform analog rainfall forecasting of strong monsoon events	28
4.2	Domain for considering winds of strong Habagat (red box) and strong Amihan (blue box)	29
4.3	Sea level pressure domain for analysis of strong Amihan (red box) and strong Habagat (blue box) obtained from the JRA-55 dataset.	30
4.4	Relative humidity and PAGASA weather station domains for strong Habagat (red box) and strong Amihan (blue box), and the extent of GPM IMERGES rainfall bounded by the whole map	31
5.1	Monthly distribution of the training set and testing set for strong monsoon days	41

5.2	Mean relative humidity (in %) of all strong monsoon days with the domains of interest for RH and PAGASA weather stations (black box)	43
5.3	Mean rainfall (in mm) of all strong monsoon days with the domains of interest for RH and PAGASA weather stations (maroon box) and the PAGASA weather stations (black dots) within the domains	43
5.4	Normalized SLP archetypes obtained using hyperspectral unmixing for strong Habagat season in the small SLP domain	45
5.5	Composite SLP map for strong Habagat testing day August 24, 2004 as a weighted ensemble of the test set's weights and the training set's SLP archetypes obtained through hyperspectral unmixing on the small SLP domain	46
5.6	Relative humidity (in %) archetypes obtained using successive hyperspectral unmixing for strong Habagat season	47
5.7	Rainfall (in mm) archetypes obtained using successive hyperspectral unmixing for strong Habagat season	48
5.8	Composite rainfall map for strong Habagat testing day August 26, 2004 as a weighted ensemble of the training set's weights and the training set's rainfall archetypes obtained through hyperspectral unmixing on the small SLP domain	49
5.9	Comparing the correlation (corr) of the actual and composite SLP maps (in mb) of strong Habagat test day August 24, 2004 for three different SLP domains described in Table 5.3	50
5.10	Comparing the correlation (corr) of the actual and composite SLP maps (in mb) of strong Amihan test day January 2, 2002 for three different SLP domains described in Table 5.3	51
5.11	Actual and composite rainfall maps (in mm) of strong Habagat training day August 26, 2004 for three different SLP domains described in Table 5.3	54
5.12	Actual and composite rainfall maps (in mm) of strong Amihan training day January 8, 2004 for three different SLP domains described in Table 5.3	55
5.13	Actual and forecasts rainfall maps of strong Habagat testing day July 7, 2017 (first row), and strong Amihan testing day January 30, 2011 (second row) with and without the multiplicative scaling	57
5.14	Composite SLP (in mb) of “best” rainfall forecast for strong Habagat season on 07/14/2006 using analog forecasting with hyperspectral unmixing and direct correlation	60
5.15	“Best” rainfall forecast (in mm) for strong Habagat season on 07/14/2006 using analog forecasting with hyperspectral unmixing and direct correlation	60
5.16	Composite SLP (in mb) of “worst” underestimated rainfall forecast for strong Habagat season on 06/29/2004 using analog forecasting with hyperspectral unmixing and direct correlation	62
5.17	“Worst” underestimated rainfall forecast (in mm) for strong Habagat season on 06/29/2004 using analog forecasting with hyperspectral unmixing and direct correlation	62

5.18 Composite SLP (in mb) of “worst” overestimated rainfall forecast for strong Habagat season on 08/02/2014 using analog forecasting with hyperspectral unmixing and direct correlation	63
5.19 “Worst” overestimated rainfall forecast (in mm) for strong Habagat season on 08/02/2014 using analog forecasting with hyperspectral unmixing and direct correlation	63
5.20 Normalized SLP archetypes obtained using hyperspectral unmixing for strong Amihan	64
5.21 Relative humidity (in %) archetypes obtained using successive hyperspectral unmixing for strong Amihan season	65
5.22 Rainfall (in mm) archetypes obtained using successive hyperspectral unmixing for strong Amihan	67
5.23 Composite SLP (in mb) of “best” rainfall forecast for strong Amihan on 01/23/2019 using analog forecasting with hyperspectral unmixing and direct correlation	69
5.24 “Best” rainfall forecast for strong Amihan on 01/23/2019 using analog forecasting with hyperspectral unmixing and direct correlation	70
5.25 Composite SLP (in mb) of “worst” rainfall forecast for strong Amihan on 01/31/2020 using analog forecasting with hyperspectral unmixing and direct correlation	70
5.26 “Worst” rainfall forecast for strong Amihan on 01/31/2020 using analog forecasting with hyperspectral unmixing and direct correlation . .	71

Chapter 1

Introduction

The Philippines relies on seasonal monsoon rainfall for consistent crop yield and water supply for dams. However, too much rainfall also becomes disastrous to a large portion of the country's population. Intense monsoon rainfall can cause flash floods due to extreme surface runoff, flooding and inundation in low lying areas, lower crop yields due to drowned crops, landslides in mountainous regions, and property damage due to heavy flooding and strong winds. A fast and efficient forecasting method of rainfall amount and distribution during strong monsoon events can dampen its devastating effects.

Currently, dynamic numerical weather prediction (NWP) that serves as model guidance is ubiquitous in rainfall forecasting worldwide. Numerical weather prediction utilizes mathematical models to simulate interactions between the atmosphere and oceans to predict future weather based on current weather conditions. NWP provides a robust approach to weather data interpolation and handling missing data, while still maintaining accuracy and cost effectiveness (Basu and Halder, 2017). NWPs can also be adjusted to provide more location specific results. However, the complexity of NWP prediction models lengthens processing time, which necessitates using a substantial processing system. If the weather suddenly behaves in a way not expected by the NWP model, then the new forecast may not catch up in time to react accordingly. Uncertainties using NWP are associated with insufficient assumptions of initial and boundary conditions, and model bias dependent on which algorithms were implemented for forecasting. Other than dynamic NWP forecasting, statistical forecasting or analog forecasting provides an alternative method for rainfall prediction.

Analog forecasting presents a faster and computationally cheaper alternative to dynamic forecasting methods, the more common methods used in modern forecasting.

Analog forecasting relies on the notion that, "History repeats itself," (Bagtasa, 2021). If the weather patterns for today's daily forecast looks similar (or analogous) to the weather patterns from previous days, then by intuition, their rainfall distributions will probably look alike as well. To perform analog forecasting, we first need a robust pattern extraction technique that can handle high-dimension weather data. This is where hyperspectral unmixing, a satellite image processor, comes in. Hyperspectral unmixing assumes that each pixel spectra in a hyperspectral satellite image can be linearly decomposed into their constituent archetypes or patterns (Bioucas-Dias et al., 2012). These archetypes are universal for all pixels in a single image, and function similar to basis vectors.

Daily measurements of meteorological parameters comprise weather data, similar to how pixels with corresponding emission spectra comprise satellite images. The pixels' emission spectra are expressed in terms of intensity vs wavelength, while the daily meteorological parameters are expressed in terms of measurements vs location (latitude and longitude). Weather data can be modified to become structurally similar to hyperspectral satellite data, which leads to the main motivation of this study—to apply the hyperspectral unmixing method in weather data pattern extraction to perform analog rainfall forecasting of strong monsoon rainfall.

This study applies analog forecasting to two weather parameters, sea-level pressure (SLP) and relative humidity (RH), to predict rainfall due to strong monsoon winds. Winds that carry moisture govern the likely location distribution of rainfall, while relative humidity governs how much moisture is present in the air, and consequently, the likelihood of precipitation (Atkins, 1974).

Directly working with monsoon winds adds complexity to analysis due to its two dimensions, zonal and meridional wind. For simplicity, this study uses sea-level pressure instead, which already accounts for both wind speed and direction (illustrated in Fig. 1.1). Isobars—lines along a weather map with equal pressure—act as quasi-streamlines i.e., wind flow almost follows the isobars (Murphree and Van Den Dool, 1988). The pressure gradient or the changes in pressure distribution dictate how strong the winds are. A steep pressure gradient (tightly packed isobars) corresponds to strong winds, while a gentle pressure gradient (widely spaced isobars) corresponds to weaker winds. Previous studies in Switzerland (Horton et al., 2017), France (Ben Daoud et al., 2016), and Sweden (Wetterhall et al., 2005) have already shown the

effectiveness of using sea-level pressure and relative humidity in the analog rainfall forecasting.

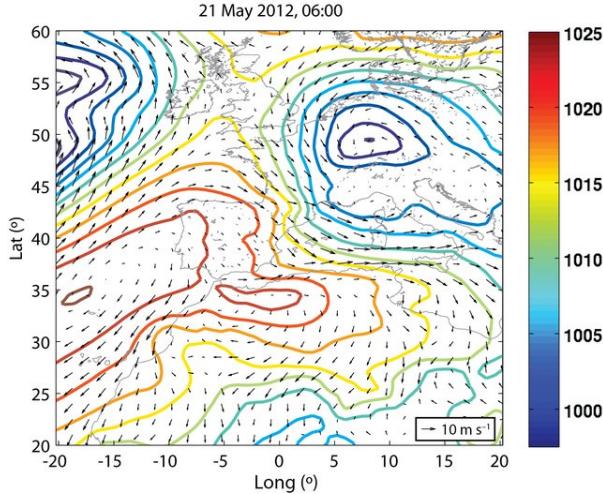


Figure 1.1: Mean sea-level pressure (colored isobars) and wind vectors over north-west Mediterranean Sea illustrating the relationship between SLP and wind (Figure 2 in Grifoll et al., 2016)

1.1 Statement of the Problem

This study implements analog forecasting, a method that relies on extracting patterns from historical measured data and finding similar patterns in current measured data (Zhao and Giannakis, 2016). This study attempts to predict rainfall distribution during strong monsoon events in the Philippines by applying analog rainfall forecasting using measurements of sea-level pressure and relative humidity. To perform analog forecasting on high spatial resolution weather data, we use the hyperspectral unmixing technique, which was shown to be faster and more efficient in handling high dimension data compared to other clustering algorithms (C. M. de Castro, 2020). The hyperspectral unmixing technique will produce a weighted-ensemble rain forecast that is used to generate analog rain forecasts.

1.2 Significance

A quarter of the Philippine workforce is in the agricultural sector with 18% land area of the whole country considered as permanent cropland, making agriculture a significant aspect of the country's economy. Agricultural practices in the Philippines

rely heavily on seasonal rainfall due to monsoon winds (Galacgac and Balisacan, 2009; Lansigan et al., 2000). Monsoon strength varies intra-seasonally as well as inter-annually. Weak monsoons tend to signify prolonged periods of drought, which result to lower crop and water resource yield. Strong monsoons, on the other hand, tend to signify prolonged periods of rainfall, which result to higher crop yield (Webster et al., 1998). However, strong monsoons also mean a higher chance of disastrous floods that could potentially cause severe economic and property damage. Predicting rainfall during strong monsoon events can aid the government and the public sector to prepare accordingly. The methods presented in this study can aid the impact based forecasts implemented by local government units (LGUs), which take into account histories of severe flooding and landslides specific to each region. Efficient and accurate forecasts can provide LGUs with informed decisions on how much to prepare for incoming heavy rainfall.

Analog forecasting is nothing new in predicting monsoon behavior (Xavier and Goswami, 2007), and extreme weather events (Bagtasa, 2021; Chattopadhyay et al., 2020). But, to the author's knowledge, hyperspectral unmixing is yet to be applied as a forecasting technique for predicting rainfall distributions. Weather and climate is dynamic, and discovering new forecasting methods could always bring something new to the field. Analog forecasting also provides a faster and computationally cheaper, yet still effective option in our forecasting arsenal.

1.3 Objectives

This study aims to achieve the following objectives

- Analyze Philippine weather archetypes obtained from hyperspectral unmixing applied to sea-level pressure and relative humidity
- Predict rainfall distributions of strong monsoon events in the Philippines using hyperspectral unmixing-based analog forecasting
- Assess proposed method's forecast capabilities using scores

1.4 Research Questions

This study aims to answer the following questions

- What are the dominant Philippine weather archetypes based on sea-level pressure and relative humidity?
- How do sea-level pressure and relative humidity affect rainfall distribution in the Philippines?
- Can hyperspectral unmixing be used as an effective method for analog rainfall forecasting during strong monsoon events?

1.5 Scope and Limitations

The forecast and analysis of rainfall distributions in the Philippines is limited to strong monsoon events from 2001-2018, and are primarily based on sea-level pressure and relative humidity measurements from reanalysis data. We include the following parameters in our analyses: sea-level pressure, relative humidity, and satellite-derived rainfall. This study does not include predicting if a strong monsoon event will occur, or what the onset and duration of monsoon seasons are. The lead time of rainfall forecasts using hyperspectral unmixing and rainfall forecasting are dependent on the availability of SLP and RH data. For instance, if we wish to forecast rainfall for the next three days, then analog forecasting would also need the SLP and RH predictions of the next three days.

Chapter 2

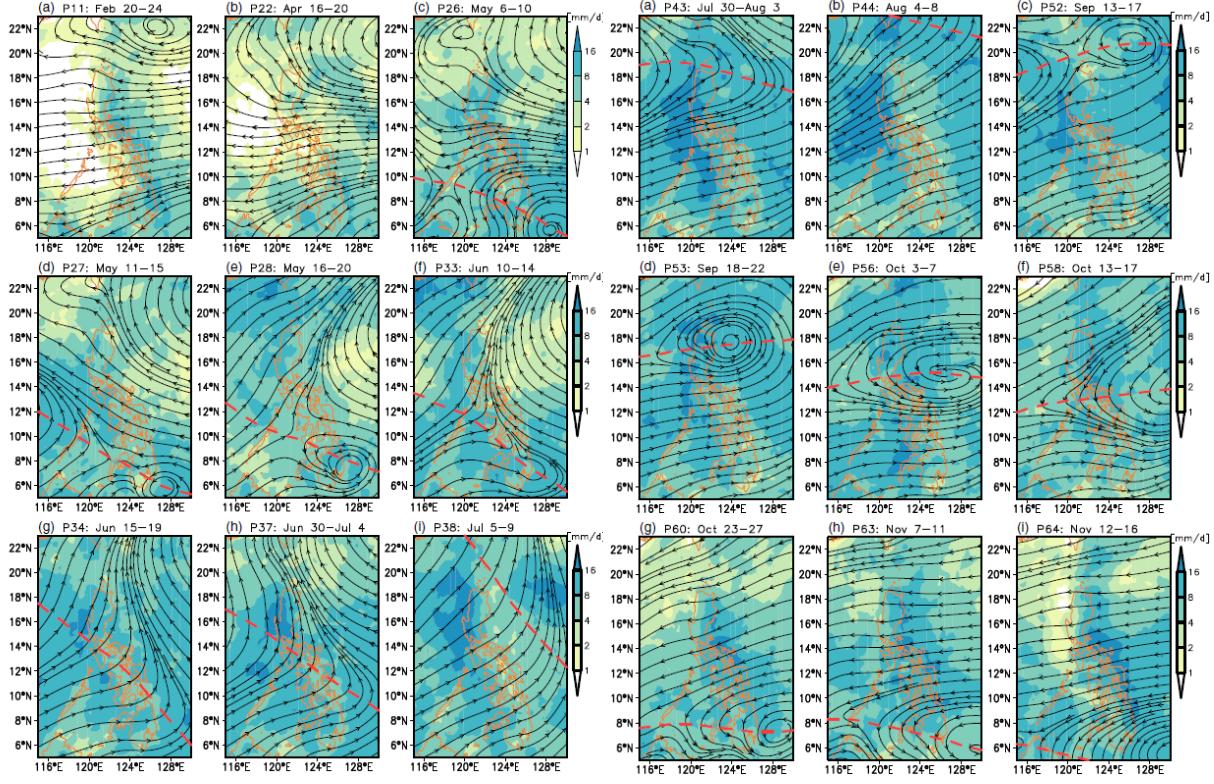
Review of Literature

2.1 Philippine Monsoons

The seasonal migration of the Intertropical Convergence Zone (ITCZ)—the area of convergence of the northern and southern hemisphere trade winds—dictates the seasonal wind shifts in the Philippines known as *monsoons*. More specifically, the portion of the ITCZ that directly affects the western North Pacific monsoons is known as the *monsoon trough*—the convergence zone of southwesterlies from the south and northeasterlies from the north. As a consequence of wind shift, the seasonal monsoon winds alter the rainfall distribution in the Philippines (Akasaka et al., 2007). Depending on the origin of the prevailing monsoon and its interaction with varying terrain, seasonal monsoons result in a sharp east-west rainfall distribution over the country.

The Philippines experiences two distinct seasonal monsoons: the western North Pacific boreal summer (Habagat) and winter (Amihan) monsoons (Bagtasa, 2020). Habagat or southwest monsoon brings warm moist air from the southwest of the Philippines during May/June to September. On the other hand, Amihan or northeast monsoon brings northeasterlies originating from the anticyclonic circulations of the Siberian High over mainland Asia and trade winds during October to March (Chang et al., 2006).

The Habagat season's onset in the Philippines begins around mid-May, felt earlier in the north than in south Philippines (Matsumoto et al., 2020). As Habagat progresses through June, the monsoon trough deepens and moves northeastward from Mindanao to Visayas (illustrated in Fig. 2.1). The monsoon trough traverses past the Philippines by early July, so the southwesterlies south of the monsoon trough fully encompass the Philippines by then. Habagat reaches its peak season around



(a) February to July

(b) July to November

Figure 2.1: Daily mean rainfall and wind distribution with the monsoon trough (in red dashed lines) over the Philippines from February to November. (Figures 2 & 3 in Matsumoto et al., 2020)

early August. As Habagat winds carry moisture from the west Philippine Sea and south China sea, the wind-terrain interaction between the monsoon winds and the topography of west Philippines causes precipitation to form over land. The rainfall distribution during Habagat peaks along Luzon's western coast.

By mid-September, the monsoon trough begins to move southward across the Philippines, signalling the Habagat's retreat and the Amihan's approach (Matsumoto et al., 2020). The northeasterlies begin to prevail in the northern Philippines, and the southwesterlies begin to recede. As Amihan winds carry moisture from the Philippine Sea, the same wind-terrain interactions as Habagat occurs over the eastern Philippines. Thus, rainfall distribution begins to increase along the eastern coast of the Philippines. By early October, the northeasterlies fully encompass the Philippines, indicating the end of Habagat and the start of Amihan (illustrated in Fig. 2.1). Amihan winds then prevail over the Philippines for the rest of the year until the monsoon trough returns northward around mid-May, repeating the monsoon seasonal cycle.

Despite a change in wind directions from Habagat to Amihan, the rainfall along west Luzon remains high for two months. During Amihan, the rain distribution peaks at northeast to central Philippines until mid-December. From late December to late March, the rain distribution during Amihan peaks in southeast Philippines, along Mindanao's northeast coasts.

2.1.1 Strong monsoon events in the Philippines

Around 43% of the annual rainfall in the Philippines occurs during Habagat season (Asuncion and Jose, 1980). Accumulated rainfall induced by Habagat winds peaks around mountainous regions (such as the Zambales and Cordillera mountain ranges), with prolonged heavy rainfall along western Luzon.

When strengthened, Habagat's westerlies accelerate, bringing more moisture from as far as the Indian Ocean towards Luzon. The moisture laden westerlies interact with mountain ranges in west Luzon, inducing wind-terrain interactions (orographic lifting) that result in excessive and prolonged rainfall. Heavy rainfall generated by strong Habagat winds can cause landslides around mountainous regions, and flash floods in low lying areas.

Strong Habagat can be characterized by a zonally-oriented eastward monsoon trough extending over northern Philippines deepened by tropical cyclones (TCs) in northern South China Sea (Bagtasa, 2019). When the trough is not extended, Habagat generates southwesterlies over the Philippines. However, upon its extension, Habagat winds begin to follow the trough's isobars, inducing much stronger zonal wind jets extending from the South China Sea to the Philippine Sea (illustrated in Fig. 2.2).

Heavier rainfall along western Philippines may also occur when the Habagat season coincides with TCs present in north to northeast Philippines (Bagtasa, 2019; Cayanan et al., 2011). TCs within northeast Philippines generate westerlies over Luzon. Habagat southwesterlies along Luzon's west coast overlap with the generated winds from said TCs, thereby strengthening Habagat's zonal wind. This phenomenon is commonly referred to locally as *enhanced Habagat*.

Outside of the Habagat season, heavy rainfall along the eastern Philippines may also occur during the Amihan season when strong monsoon winds coincide with a low system in east to southeast Philippines (Bagtasa, 2020). Strong Amihan winds occur

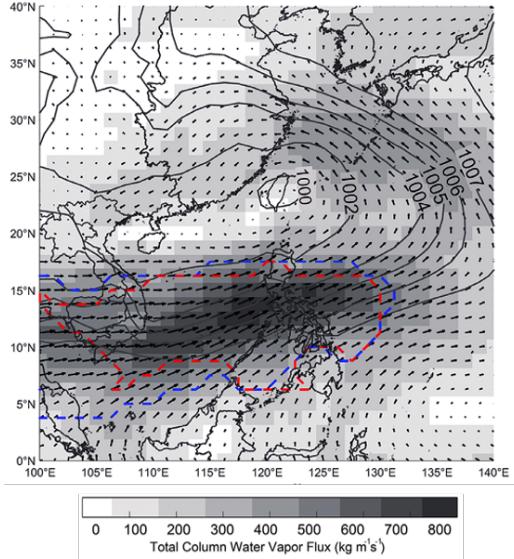


Figure 2.2: Total column water vapor flux (gray shading), 850 mb wind (arrows) and mean sea level pressure (dark gray lines) during strong Habagat (Figure 4 in Bagtasa, 2019)

during *cold surges* or *winter monsoon surges* (illustrated in Fig. 2.3). During winter monsoon surges, the Siberian High over mainland China strengthens and extends southeastward, thereby accelerating the northeasterlies that extend southward over the Philippines (Chang et al., 2006). A low system within east to southeast Philippines coinciding with the accelerated monsoon winds generates strong convergent northeasterlies along the Philippines' east coast. These convergent winds are locally known as the *tail-end of the cold front* or *shearline*—the interaction of cold air (from Amihan) and warm air (from the low system winds). The tail-end of the cold front accumulates moisture over the ocean, and upon reaching land, causes anomalously heavy rainfall along the Philippines' east coast.

The devastating effects of rainfall induced by strong monsoon winds necessitates supplementing and augmenting our current arsenal of forecasting techniques. In this study, we explore the potential of analog forecasting for predicting rainfall distribution during strong monsoon events.

2.2 Analog Forecasting

Analog forecasting methods formulate future predictions by searching for historical weather patterns similar or analogous to present day weather patterns (Chan et al., 2014). Then, the method bases the present day rainfall distribution on the analogous

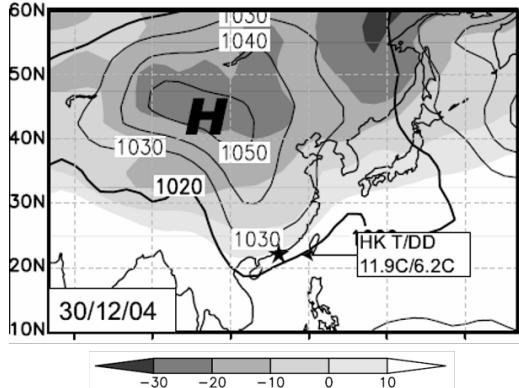


Figure 2.3: Mean sea level pressure (isobars) and surface temperature (shading) during a winter monsoon surge in December 30, 2004. (Figure 3.3b in Chang et al., 2006)

past rainfall (Ben Daoud et al., 2016). Generally, analog forecasting can be separated into three phases: (1) extracting patterns from historical weather data, (2) finding analogous patterns between historical and present data, and (3) predicting present day precipitation based on analogous historical data.

Previous studies used analog forecasting to predict rainfall distributions during TCs in the Philippines (Bagtasa, 2021) and Korea (Kim et al., 2020), to predict the effects of heavy precipitation events in Hong Kong (Chan et al., 2014), to forecast the subseasonal variability of the Indian summer monsoon (Xavier and Goswami, 2007), and to predict precipitation distribution around a large river basin in France (Ben Daoud et al., 2016). Analog forecasting is yet to be applied in forecasting rainfall distributions during strong monsoon events in the Philippines.

Previous studies typically implement clustering techniques (Kim et al., 2020), correlation analyses (Chan et al., 2014), and deep learning (Chattopadhyay et al., 2020) to perform the first and second phases in analog forecasting. This study uses hyperspectral unmixing, a pattern extraction technique typically applied in satellite image processing. We also implement a two-level successive analog forecasting method similar to the works of Horton et al. and Daoud et al. In this method, the first step is successively repeated using a different meteorological parameter each time, thereby narrowing down and refining our search for similar weather patterns.

2.3 Hyperspectral Unmixing

Hyperspectral unmixing eases analyses of satellite images with high spectral resolutions by decomposing the pixel spectra into their component archetypes (patterns) and weights (Bioucas-Dias et al., 2012, illustrated in Fig. 2.4). Archetypes represent and characterize the dominant patterns present in the satellite image, while the weights correspond to how much of each archetype comprises each pixel. Thus, instead of describing each individual pixel, hyperspectral unmixing summarizes the information contained in a satellite image into its constituent archetypes.

For example, a single pixel's spectral signature in an image may be a combination of 30% mountains spectra, 20% river spectra, and 50% grasslands spectra. Here, the percentages correspond to the weights for that single pixel, and the spectra correspond to the overall archetypes of the satellite image. Note that hyperspectral unmixing merely estimates the archetype spectra, so the characterization of the archetypes relies on the subjective judgement of the user. In our example, hyperspectral unmixing does not automatically know that the archetypes are mountains, rivers, and grasslands; the user subjectively determines this based on how their spectra look.

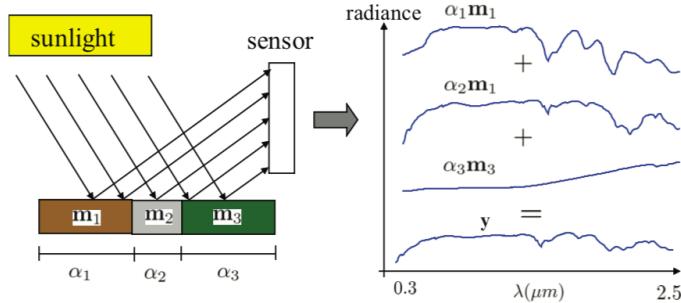


Figure 2.4: Diagram illustrating how linear hyperspectral unmixing decomposes the satellite image's pixel spectra into their constituent archetypes and weights. (Figure 2 in Bioucas-Dias et al., 2012)

Most studies on hyperspectral unmixing used the method for satellite image processing, such as mapping soil mineral composition (Clark et al., 2000), estimating crop production (Kanning et al., 2018), and monitoring water quality (Cheng et al., 2008). Some applications of hyperspectral unmixing in industrial image processing include assessing food product quality and safety (Mladenov et al., 2015), and monitoring ischemic wound development(Xu et al., 2012).

Only a few studies extended hyperspectral unmixing outside of image processing.

Some of these include inferring key biological tasks contained in tissue genes (Hart et al., 2015), and extracting Philippine senate election patterns (C. M. M. de Castro and Lim, 2020). In a similar manner, hyperspectral unmixing will be used to decompose the slp patterns during strong monsoon events and the decomposed slp patterns will be used to generate a weighted-ensemble rain maps. Despite its use in satellite images, hyperspectral unmixing is yet to be applied to meteorological data, much less in rainfall forecasting.

This study applies the linear hyperspectral unmixing model, which estimates \mathbf{M} and \mathbf{A} in the following equation

$$\mathbf{Y} = \mathbf{MA} + \mathbf{N} \quad (2.1)$$

In this equation, \mathbf{Y} contains the observed hyperspectral data, the mixing matrix \mathbf{M} contains the spectral signature of archetypes, the abundance matrix \mathbf{A} contains the archetype weights, and the noise matrix \mathbf{N} contains intrinsic model noise (Bioucas-Dias and Figueiredo, 2010; Bioucas-Dias and Nascimento, 2008; Li et al., 2015; Nascimento and Dias, 2005). The linear unmixing model applied in our study imposes two constraints: (1) the archetype weights must all be positive, and (2) the weights should sum to one per pixel or data point.

The model follows a general three-step process. First, dimension reduction determines the number of archetypes, and simplifies computation for the next steps. Next, unmixing estimates the archetypes contained in \mathbf{M} . Lastly, inversion estimates the corresponding weights per archetype contained in \mathbf{A} .

2.3.1 First step of hyperspectral unmixing: dimension reduction

The spatial resolution of weather data used in forecasting can reach up to hundreds of dimensions, making dimension reduction an essential first step in pattern extraction to ease computational complexity. To perform dimension reduction, the hyperspectral unmixing algorithm used in this study implements the Hyperspectral signal Subspace identification by minimum error (HySime) (Bioucas-Dias and Nascimento, 2008)

In HySime's linear unmixing model, the number of archetypes k correspond to the reduced number of dimension. This number must be much smaller than the initial number of dimensions of the observed hyperspectral data \mathbf{Y} . If k is too high, unnec-

essary information from higher dimensions is included in analysis, thereby increasing noise. If k is too low, important information contained in higher dimensions is not included, thereby increasing projection error. The optimal k estimated by HySime minimizes both noise and projection error. After determining k , HySime projects \mathbf{Y} to the k -dimensional subspace. The projected \mathbf{Y} with reduced dimensions is denoted as \mathbf{Y}' , and will be used for unmixing.

2.3.2 Second step of hyperspectral unmixing: unmixing

Previous studies developed varying methods to perform unmixing, but this study employs geometric unmixing; specifically, a minimum volume based algorithm. In geometric unmixing, (1) the vertices of a simplex enclosing the data represent the archetypes (illustrated in Fig. 2.5), and (2) performing transformations that preserve the lines and parallelisms on a simplex results to another simplex. A simplex generalizes a triangle or a tetrahedron in any dimension. For instance, a line is a 1-dimensional simplex, a triangle is a 2-dimensional simplex, and a tetrahedron is a 3-dimensional simplex. In geometric unmixing, the simplex spanning the data should be a k -dimension simplex.

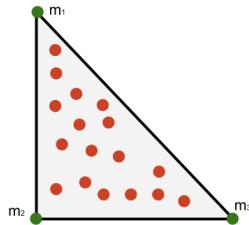


Figure 2.5: Diagram of a 2-dimensional simplex with three archetypes (green vertices) encompassing the hyperspectral data (red dots).

A minimum volume based geometric unmixing approach estimates the k -dimension simplex with the lowest possible volume such that it still surrounds most if not all of the data \mathbf{Y}' . Geometric unmixing can either be a pure pixel or a non-pure pixel algorithm.

Pure pixel geometric unmixing algorithms such as Vertex Component Analysis (VCA) (Nascimento and Dias, 2005) assume that the archetypes can already be found in a pure pixel in the data (illustrated in Fig. 2.6). For example, pure pixel algorithms expect that a satellite image contains at least one pixel containing almost purely trees

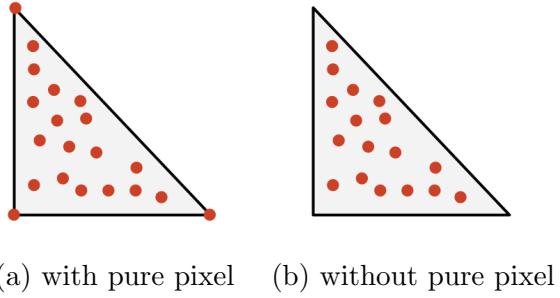


Figure 2.6: Comparison of hyperspectral data (a) with, and (b) without the pure pixel assumption.

or purely water. On the other hand, non-pure pixel algorithms such as Minimum Volume Simplex Analysis (MVSA) assume otherwise, the more realistic scenario (Li et al., 2015). This study implements both, with VCA providing the initial conditions for estimating \mathbf{M} in MVSA.

2.3.3 Final step of hyperspectral unmixing: inversion

This study uses Spectral unmixing by splitting and augmented Lagrangian (SUnSAL) to perform inversion and solve for \mathbf{A} (Bioucas-Dias and Figueiredo, 2010). To estimate \mathbf{A} , SUnSAL fits the original observed hyperspectral data \mathbf{Y} with the known spectral signatures of the archetypes \mathbf{M} obtained after unmixing. SUnSAL simplifies complicated convex optimization problems using the alternating direction method of multipliers (ADMM).

Chapter 3

Theory behind Hyperspectral Unmixing

Here, we discuss the mathematical theories and computations behind each step of hyperspectral unmixing, which is divided into three sections: 3.1 dimension reduction, 3.2 unmixing, and 3.3 inversion.

3.1 Theory behind hyperspectral unmixing's dimension reduction

This study uses HySime to perform dimension reduction, which involves two primary steps: noise estimation and signal subspace estimation (Bioucas-Dias and Nascimento, 2008).

3.1.1 Noise estimation

HySime performs noise estimation through multiple regression, an effective method for hyperspectral data, because of the high correlation between neighboring spectral bands. Multiple regression assumes that the i th band can be explained by the linear combination of all spectral bands excluding itself such that

$$\mathbf{z}_i = \mathbf{Z}_{\partial_i} \boldsymbol{\beta}_i + \boldsymbol{\xi}_i \quad (3.1)$$

where \mathbf{z}_i represents intensities along the i th spectral band (columns of \mathbf{Y}), \mathbf{Z}_{∂_i} contains the spectral bands' intensities excluding i , $\boldsymbol{\beta}_i$ is the regression vector, and $\boldsymbol{\xi}_i$ is the modeling error vector. Noise estimation calculates $\hat{\boldsymbol{\xi}}_i$, given by

$$\hat{\boldsymbol{\xi}}_i = \mathbf{z}_i - \mathbf{Z}_{\partial_i} \hat{\boldsymbol{\beta}}_i \quad (3.2)$$

Under linear multiple regression, the least squares estimator of $\boldsymbol{\beta}_i$ is given by

$$\hat{\boldsymbol{\beta}}_i = (\mathbf{Z}_{\partial_i}^T \mathbf{Z}_{\partial_i})^{-1} \mathbf{Z}_{\partial_i}^T \mathbf{z}_i \quad (3.3)$$

Calculating $(\mathbf{Z}_{\partial_i}^T \mathbf{Z}_{\partial_i})^{-1}$ for all $i = 1, \dots, L$ is too complicated so HySime overcomes this by calculating the inverse $(\mathbf{Z}^T \mathbf{Z})$ only once.

$$(\mathbf{Z}_{\partial_i}^T \mathbf{Z}_{\partial_i})^{-1} = (\mathbf{Z}^T \mathbf{Z})_{\partial_i, \partial_i}^{-1} - \frac{(\mathbf{Z}^T \mathbf{Z})_{\partial_i, i}^{-1} (\mathbf{Z}^T \mathbf{Z})_{i, \partial_i}^{-1}}{(\mathbf{Z}^T \mathbf{Z})_{i, i}^{-1}} \quad (3.4)$$

where $\mathbf{Z} = \mathbf{Y}^T$. The following table summarizes the meaning of the subscripts in Eq. 3.4.

Table 3.1: Meaning of subscripts in Eq. 3.4

subscript	meaning
δ_i, δ_i	remove i th row and column
δ_i, i	i th column after removing i th row
i, δ_i	i th row after removing i th column
i, i	i th row and column

Finally, HySime can estimate the noise $\hat{\boldsymbol{\xi}}_i$ by combining the previous three equations. HySime also computes for the correlation matrix as

$$\hat{\mathbf{R}}_n := 1/N \sum_i \left(\hat{\boldsymbol{\xi}}_i \hat{\boldsymbol{\xi}}_i^T \right) \quad (3.5)$$

3.1.2 Signal subspace estimation

Signal subspace estimation determines the number of dimensions spanned by the archetypes and weights $\mathbf{x} = \mathbf{MA}$. To do this, HySime finds the minimum mean square error (MSE) between the original signal \mathbf{x} , and a noisy projection of \mathbf{x} given by $\mathbf{y} = \mathbf{x} + \mathbf{n}$. Similar to Eq. 3.5 the signal correlation matrix is

$$\hat{\mathbf{R}}_x := 1/N \sum_i \left((\mathbf{y}_i - \hat{\boldsymbol{\xi}}_i) (\mathbf{y}_i - \hat{\boldsymbol{\xi}}_i)^T \right) \quad (3.6)$$

where \mathbf{y}_i denotes the rows of \mathbf{Y} . The sample correlation matrix is

$$\widehat{\mathbf{R}}_y \equiv (\mathbf{Z}^T \mathbf{Z}) / N \quad (3.7)$$

The final MSE minimization problem between the original signal and its noisy projection is given by

$$(\widehat{k}) = \arg \min_k (c + \sum_{j=1}^k \underbrace{\mathbf{e}_{i_j}^T \widehat{\mathbf{R}}_y \mathbf{e}_{i_j} + 2\mathbf{e}_{i_j}^T \widehat{\mathbf{R}}_n \mathbf{e}_{i_j}}_{\delta_{i_j}}) \quad (3.8)$$

where c is an irrelevant constant, \mathbf{e}_i are the eigenvectors of $\widehat{\mathbf{R}}_y$, and $k < L$ is the reduced signal subspace dimension spanned by the archetypes and weights. The first term in the summation represents the projection error, while the second term represents the noise error. The index j stands for the j th dimension in the k -dimensional subspace.

The right hand side can be minimized by considering only the negative values of δ_i for $i = 1, \dots, L$. The number of terms $\delta_i < 0$ dictates the signal subspace dimension k . The eigenvectors \mathbf{e}_i corresponding to the $\delta_i < 0$ arranged in ascending order denote the signal subspace \mathbf{E}_k . To project the observed data into the signal subspace, HySime uses $\mathbf{Y}_k = \mathbf{E}_k \mathbf{E}_k^T \mathbf{Y}$.

In other words, HySime reduces the L -dimension data by projecting \mathbf{Y} to a subspace with dimension $k < L$ as \mathbf{Y}_k . From here on, hyperspectral unmixing uses the reduced observed data \mathbf{Y}_k instead of the original \mathbf{Y} for simplicity of calculations.

3.2 Theory behind hyperspectral unmixing's unmixing

This study uses two algorithms for unmixing: Vertex Component Analysis (VCA), a pure pixel unmixing, and Minimum Volume Simplex Analysis (MVSA), a non-pure pixel unmixing. VCA provides the initial conditions for MVSA.

3.2.1 Vertex Component Analysis

VCA projects data from the subspace spanned by the current archetypes to an orthogonal subspace, and sets the extremes of the projection as new archetypes (Nascimento

and Dias, 2005). VCA repeatedly orthogonally projects data until all archetypes are exhausted.

To account for illumination variability due to surface topography, VCA adds a scaling factor γ to Eq. 2.1 such that

$$\mathbf{Y}_k = \mathbf{X} + \mathbf{N} = \mathbf{M}\gamma\mathbf{A} + \mathbf{N} \quad (3.9)$$

The signal x contained in the set

$$S_x = \{\mathbf{x} \in \Re^L : \mathbf{x} = \mathbf{M}\mathbf{A}, \mathbf{1}^T \mathbf{A} = 1, \boldsymbol{\alpha} \succeq 0\} \quad (3.10)$$

belongs to a k -dimensional simplex. To account for the scaling factor γ , the observed data now belongs to the convex cone (also a simplex) defined by

$$C_p = \{\mathbf{Y}_k \in \Re^L : \mathbf{Y}_k = \mathbf{M}\gamma\mathbf{A}, \mathbf{1}^T \mathbf{A} = 1, \mathbf{A} \succeq 0, \gamma \geq 0\} \quad (3.11)$$

When C_p is projected into a properly chosen hyperplane (subspace with one less dimension than the surrounding space), we get another simplex with vertices corresponding to the vertices of S_x .

The projection of C_p onto the chosen hyperplane $\mathbf{r}^T \mathbf{u} = 1$ is given by the simplex

$$S_p = \{\mathbf{Y}_p \in \Re^L : \mathbf{Y}_p = \mathbf{Y}_k / (\mathbf{r}_k^T \mathbf{u}), \mathbf{r} \in C_p\} \quad (3.12)$$

In Fig. 3.1, VCA first projects the simplex S_p along the random initial vector \mathbf{f}_1 . VCA sets \mathbf{m}_a , the extreme of the projection (farthest from the origin), as the first endmember. In the next iteration, VCA projects the simplex S_p along the orthogonal vector to \mathbf{m}_a , vector \mathbf{f}_2 . Then, VCA sets \mathbf{m}_b , the extreme of the new projection, as the next endmember. The next projection vector should then be orthogonal to both \mathbf{m}_a and \mathbf{m}_b , and so on until all k endmembers are obtained.

Finding the projection matrix

To obtain the projection matrix \mathbf{U}_d , VCA implements two different dimension reduction algorithms depending on the signal-to-noise ratio (SNR). If $\text{SNR} > \text{SNR}_{th}$, VCA applies singular value decomposition (SVD), while if $\text{SNR} < \text{SNR}_{th}$, VCA applies principal component analysis (PCA). The threshold is given by

$$\text{SNR}_{th} = 15 + 10 \log_{10}(k) \text{dB} \quad (3.13)$$

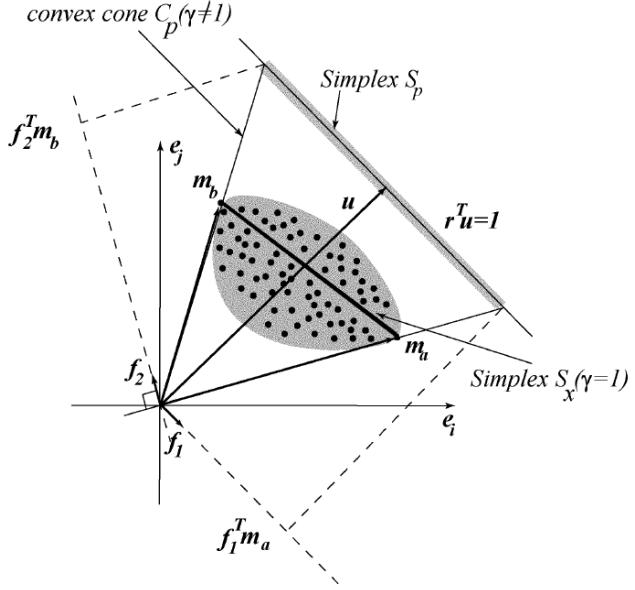


Figure 3.1: Illustration of the VCA algorithm. VCA projects S_p along \mathbf{f}_1 and \mathbf{f}_2 , then sets the extremes as the archetypes. (figure from Nascimento and Dias, 2005)

For simplicity, SNR is set as $\text{SNR} = 1$ (Ringnér, 2008). Eq. 3.13 always exceeds 1, we always use PCA in this study. Since VCA projects C_p to a hyperplane, the number of dimensions is set at $d = k - 1$. PCA obtains \mathbf{U}_d from $(\mathbf{Y}_k - \bar{\mathbf{y}}_k)(\mathbf{Y} - \bar{\mathbf{y}}_k)^T / k$, where $\bar{\mathbf{y}}_k$ denotes the sample mean per column of \mathbf{Y}_k , and $\mathbf{Y}_k \equiv [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_k]$.

We use $\mathbf{X}_p = \mathbf{U}_d^T (\mathbf{Y}_k - \bar{\mathbf{y}}_k)$ to project C_p into the d -dimensional hyperplane. In the original dimension k , the simplex S_p is given by

$$S_p := \begin{bmatrix} \mathbf{X}_p \\ \mathbf{c} \end{bmatrix} \quad (3.14)$$

where \mathbf{c} is a $1 \times N$ vector with all elements given by $c := \arg \max_{j=1 \dots N} \|[\mathbf{X}_p] :, j\|$. The index $:; j$ indicates \mathbf{X}_p 's j th column.

Estimating archetype signatures

VCA initializes the matrix containing the archetype signatures as

$$\mathbf{M} := [\mathbf{e}_u | 0 | \dots | 0] \quad (3.15)$$

where \mathbf{M} is a $k \times k$ matrix, and $\mathbf{e}_u = [0, \dots, 0, 1]^T$. Then, VCA sets a random vector \mathbf{f} orthonormal to the columns of \mathbf{M} given by

$$\mathbf{f} := ((\mathbf{I} - \mathbf{M}\mathbf{M}^\#) \mathbf{w}) / (\|(\mathbf{I} - \mathbf{M}\mathbf{M}^\#) \mathbf{w}\|) \quad (3.16)$$

where \mathbf{w} is a zero-mean random Gaussian $k \times 1$ vector, and the operation $M^\#$ stands for the pseudoinverse of M . We use $v = \mathbf{f}^T S_p$ to project S_p along \mathbf{f} . Matrix \mathbf{M} stores the pixel spectral signature corresponding to the maximum of the projection, the data point with the highest $|v|$, as the next endmember. VCA exhausts the number of archetypes k by iterating all steps starting from Eq. 3.16 k times. The final matrix \mathbf{M} corresponds to the estimated mixing matrix containing the pure pixels in the scene. This repeated iteration is illustrated in Fig.

3.2.2 Minimum Volume Simplex Analysis

MVSA, unlike VCA, employs soft constraints that allows negative weights, thereby violating the positivity constraint (Li et al., 2015). However, to maintain simplicity, this study employs hard positive constraints. To efficiently solve the volume maximization problem, MVSA implements a sequence of quadratically constrained subproblems using the interior point method. Thus, MVSA can unmix large and complex data.

MVSA Preprocessing

Similar to the γ factor in VCA, MVSA also takes into account pixel-dependent scale factors such as spectral variability (variations in surface topography). To remove the effects of pixel-dependent scale factors, MVSA projects \mathbf{Y}_k onto the hyperplane that best represents \mathbf{Y}_k in the least squares sense. The orthogonal projection of the vectors of \mathbf{Y}_k is given by

$$\mathbf{y}_{p,i} \leftarrow \bar{\mathbf{y}} + \mathbf{E}_{p-1}^T (\mathbf{y}_{k,i} - \bar{\mathbf{y}}), \quad i \in \{1, 2, \dots, N\} \quad (3.17)$$

where the columns of \mathbf{E}_{p-1} contain the eigenvalues of the sample covariance matrix $(\mathbf{Y}_k - \bar{\mathbf{Y}}_k) (\mathbf{Y}_k - \bar{\mathbf{Y}}_k)^T / N$, $\mathbf{y}_{k,i}$ denote the rows of \mathbf{Y}_k , $\mathbf{y}_{p,i}$ denotes the rows of \mathbf{Y}_p , and

$$\bar{\mathbf{y}} \equiv \frac{1}{N} \sum_{i=1}^N \mathbf{y}_{k,i}; \quad \bar{\mathbf{Y}}_k = \bar{\mathbf{y}} \mathbf{1}_N^T \quad (3.18)$$

From here on, we denote $\mathbf{y}_{p,i}$ as the orthogonal projection of $\mathbf{y}_{k,i}$.

MVSA inference criterion

MVSA neglects the noise after data projection. Thus, the minimum volume simplex problem can be written as

$$\widehat{\mathbf{M}} = \arg \min_{\mathbf{M}} |\det(\mathbf{M})| \text{ s.t. : } \mathbf{QY}_p \geq 0, \quad \mathbf{1}_k^T \mathbf{QY}_p = \mathbf{1}_N^T \quad (3.19)$$

where $\mathbf{Q} \equiv \mathbf{M}^{-1}$. Since $|\det(M)|$ represents the volume of the simplex spanned by the archetypes, Eq. 3.19 estimates \mathbf{M} corresponding to the minimum simplex volume that satisfies the constraints.

Rewriting Eq. 3.19 in terms of \mathbf{Q} , we get

$$\widehat{\mathbf{Q}} = \arg \max_{\mathbf{Q}} \log |\det(\mathbf{Q})| \quad (3.20)$$

since $\det(\mathbf{Q}) = 1 / \det(\mathbf{M})$. The constraints don't change.

Constraint Reduction

MVSA assumes that \mathbf{Y}_p has linearly independent rows and/or columns (depending on its size), so $\mathbf{Y}_p \mathbf{Y}_p^T$ is invertible. Thus, we can rewrite the sum-to-one constraint in Eq. 3.19 as

$$\mathbf{1}_k^T \mathbf{QY}_p = \mathbf{1}_N^T \Leftrightarrow \mathbf{1}_k^T \mathbf{Q} = \mathbf{q}_k \quad (3.21)$$

where $\mathbf{q}_k \equiv \mathbf{1}_N^T \mathbf{Y}_p^T (\mathbf{Y}_p \mathbf{Y}_p^T)^{-1}$. By rewriting the sum-to-one constraint, the number of equality constraints in the optimization problem decreased from kN to k .

Minorize-maximization optimization

MVSA adopts the minorize-maximization framework. The framework builds a minorizer of the objective function, and iteratively maximizes it (Lange et al., 2000). In our case, the objective function is given by $f(\mathbf{x}) \equiv \log |\det(\mathbf{Q})|$, and the iterative procedure is given by

$$\begin{aligned} \mathbf{x}^{(t+1)} &= \arg \max_{\mathbf{x}} \phi(\mathbf{x}; \mathbf{x}^{(t)}) \\ \text{s.t.: } \mathbf{A}_I \mathbf{x} &\geq \mathbf{b}_I, \quad \mathbf{A}_E \mathbf{x} = \mathbf{b}_E \end{aligned} \quad (3.22)$$

where

$$\begin{aligned} \mathbf{A}_I &\equiv (\mathbf{Y}_p^T \otimes \mathbf{I}) \in \mathbb{R}^{kN \times k^2} \\ \mathbf{A}_E &\equiv (\mathbf{I} \otimes \mathbf{1}_k^T) \in \mathbb{R}^{k \times k^2} \\ \mathbf{b}_I &\equiv \mathbf{0} \in \mathbb{R}^{kN} \\ \mathbf{b}_E &\equiv \mathbf{q}_p \in \mathbb{R}^p \end{aligned} \quad (3.23)$$

In Eq. 3.22, we stack the columns of \mathbf{Q} into one column \mathbf{x} . The symbol \otimes denotes the Kronecker operator, $\phi(\mathbf{x}; \mathbf{x}^{(t)})$ denotes the minorizer of $f(x)$, and t denotes the iteration index. MVSA uses the following quadratic function as the minorizer for f .

$$\phi(\mathbf{x}; \mathbf{x}^{(t)}) \equiv f(\mathbf{x}^{(t)}) + \mathbf{c}^{(t)T} \mathbf{x} + \frac{1}{2} \mathbf{x}^T \mathbf{G}^{(t)} \mathbf{x} \quad (3.24)$$

where

$$\begin{aligned} \mathbf{c}^{(t)} &\equiv \mathbf{g}^{(t)} - \mathbf{G}^{(t)} \mathbf{X}^{(t)} \\ \mathbf{G} &\equiv \min \{\lambda_{\min}(\mathbf{H}), -v\} \mathbf{I}. \end{aligned} \quad (3.25)$$

Here, $\mathbf{g}(x)$ and $H(x)$ denote the gradient and Hessian (matrix of second order partial derivatives) of f , respectively. The notation $\lambda_{\min}(\mathbf{H})$ stands for the smallest eigenvalue of \mathbf{H} , and v is a small positive number. To maximize the minorizer function $\phi(\mathbf{x}; \mathbf{x}^{(t)})$, MVSA solves the following convex optimization problem, the core step of the algorithm.

$$\begin{aligned} &\max \mathbf{c}^T \mathbf{x} + \frac{1}{2} \mathbf{x}^T \mathbf{G} \mathbf{x} \\ \text{s.t.: } &\mathbf{A}_I \mathbf{x} \geq \mathbf{b}_I, \quad \mathbf{A}_E \mathbf{x} = \mathbf{b}_E \end{aligned} \quad (3.26)$$

The convex optimization problem evolved from the nonconvex optimization problem in Eq. 3.19.

Interior Point Method

MVSA uses the interior-point method to solve Eq. 3.26. The solution must satisfy the Karush-Kuhn-Tucker (KKT) conditions indicated in the interior point method, which are enumerated as follows (Nocedal and Wright, 2000)

$$\begin{aligned} \mathbf{G} \mathbf{x} - \mathbf{A}_I^T \boldsymbol{\lambda} + \mathbf{A}_E^T \boldsymbol{\mu} + \mathbf{c} &= 0 \\ \mathbf{A}_I \mathbf{x} - \mathbf{b}_I &\geq 0 \\ \mathbf{A}_E \mathbf{x} - \mathbf{b}_E &= 0 \\ (\mathbf{A}_I \mathbf{x} - \mathbf{b}_I)_i \lambda_i &= 0, i = 1 \dots n_I \equiv Nk \\ \boldsymbol{\lambda} &\geq 0 \end{aligned} \quad (3.27)$$

where $\boldsymbol{\lambda} \equiv [\lambda_1, \dots, \lambda_{n_I}]^T$, and $\boldsymbol{\mu} \in \mathbb{R}^p$ denote the Lagrange multipliers of the inequality and equality constraints, respectively, and n_I denotes the number of inequality constraints. The first equation represents the maximizing stationarity condition; the second and third represent the primal feasibility conditions; the fourth represents

the complementary slackness condition; and the last represents the dual feasibility condition. Eq. 3.27 can be rewritten by introducing a slack vector $\mathbf{s} \equiv [s_1, \dots, s_{n_I}]^T$ such that the interior point method can solve it. Slack variables/vectors convert inequality constraints to equality constraints through addition and subtraction.

$$\begin{aligned} \mathbf{G}\mathbf{x} - \mathbf{A}_I^T \boldsymbol{\lambda} + \mathbf{A}_E^T \boldsymbol{\mu} + \mathbf{c} &= 0 \\ \mathbf{A}_I \mathbf{x} - \mathbf{s} - \mathbf{b}_I &= 0 \\ \mathbf{A}_E \mathbf{x} - \mathbf{b}_E &= 0 \\ s_i \lambda_i &= 0, i = 1 \dots n_I \\ \boldsymbol{\lambda}, \mathbf{s} &\geq 0 \end{aligned} \tag{3.28}$$

To calculate the roots in the system of equations from Eq 3.28, the interior point method uses Newton's method. Newton's method for nonlinear systems of equations uses the following equation

$$J_F(\mathbf{x}_n)(\mathbf{x}_{n+1} - \mathbf{x}_n) = -F(\mathbf{x}_n) \tag{3.29}$$

where $J_F(\mathbf{x}_n)$ is the Jacobian matrix (first-order partial derivatives) for $F(\mathbf{x}_n)$. In our case, $F(\mathbf{x}_n)$ denotes the left-hand side of Eq. 3.28, and the unknown variables \mathbf{x}_n denote \mathbf{x} , $\boldsymbol{\mu}$, \mathbf{s} , and $\boldsymbol{\lambda}$. MVSA's predictor corrector interior point algorithm applies Newton's method twice. First, Newton's method determines the affine step. Second, Newton's method corrects the affine step. Newton's method applied to the affine step is given by

$$\begin{bmatrix} \mathbf{G} & \mathbf{A}_E^T & \mathbf{0} & -\mathbf{A}_I^T \\ \mathbf{A}_I & \mathbf{0} & -\mathbf{I} & \mathbf{0} \\ \mathbf{A}_E & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Lambda & \mathbf{S} \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x}^{\text{aff}} \\ \Delta \boldsymbol{\mu}^{\text{aff}} \\ \Delta \mathbf{s}^{\text{aff}} \\ \Delta \boldsymbol{\lambda}^{\text{aff}} \end{bmatrix} = \begin{bmatrix} -\mathbf{r}_d \\ -\mathbf{r}_I \\ -\mathbf{r}_E \\ -\Lambda \mathbf{S} \mathbf{e} \end{bmatrix} \tag{3.30}$$

where

$$\begin{aligned} \mathbf{r}_d &\equiv \mathbf{G}\mathbf{x} - \mathbf{A}_I^T \boldsymbol{\lambda} + \mathbf{A}_E^T \boldsymbol{\mu} + \mathbf{c} \\ \mathbf{r}_I &\equiv \mathbf{A}_I \mathbf{x} - \mathbf{s} - \mathbf{b}_I \\ \mathbf{r}_E &\equiv \mathbf{A}_E \mathbf{x} - \mathbf{b}_E \\ \Lambda &\equiv \text{diag}(\lambda_1, \dots, \lambda_{n_I}) \\ \mathbf{S} &\equiv \text{diag}(s_1, \dots, s_{n_I}) \\ \mathbf{e} &\equiv [1, \dots, 1]^T \end{aligned} \tag{3.31}$$

Newton's method applied again to correct the affine step is given by

$$\begin{bmatrix} \mathbf{G} & \mathbf{A}_E^T & 0 & -\mathbf{A}_I^T \\ \mathbf{A}_I & 0 & -\mathbf{I} & 0 \\ \mathbf{A}_E & 0 & 0 & 0 \\ 0 & 0 & \Lambda & \mathbf{S} \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x} \\ \Delta \mu \\ \Delta \mathbf{s} \\ \Delta \lambda \end{bmatrix} = \begin{bmatrix} -\mathbf{r}_d \\ -\mathbf{r}_I \\ -\mathbf{r}_E \\ -\Lambda \mathbf{S} \mathbf{e} - \Delta \Lambda^{\text{aff}} \Delta \mathbf{S}^{\text{aff}} \mathbf{e} + \sigma \rho \mathbf{e} \end{bmatrix} \quad (3.32)$$

where

$$\begin{aligned} \Delta \Lambda^{\text{aff}} &\equiv \text{diag}(\Delta \lambda_1^{\text{aff}}, \dots, \Delta \lambda_{n_I}^{\text{aff}}) \\ \Delta \mathbf{S}^{\text{aff}} &\equiv \text{diag}(\Delta s_1^{\text{aff}}, \dots, \Delta s_{n_I}^{\text{aff}}) \\ \rho &\equiv \frac{\mathbf{s}^T \boldsymbol{\lambda}}{n_I} \\ \sigma &= \left(\frac{\rho_{\text{aff}}}{\rho} \right)^3 \end{aligned} \quad (3.33)$$

The quantity ρ , the duality measure, measures the average value of the pairwise products $s_i \lambda_i$. For MVSA, $s_i \lambda_i$ converges to a small average value $s_i \lambda_i = \sigma \rho$ instead of at 0, a less aggressive approach to Newton's method. The quantity $\sigma \in [0, 1]$, the centering parameter, dictates how strict the algorithm applies the condition $s_i \lambda_i = 0$. A lower σ results to systems of equations closer to the affine Newton step in Eq. 3.30.

Normal Equations

We reformulate the system of equations 3.30 and 3.32 to more compact symmetric coefficient matrices to ease computational difficulty. Since \mathbf{A} and \mathbf{S} are nonsingular matrices (i.e. matrices whose inverse exist), we can multiply the fourth equation in Eqs. 3.30 and 3.32 by \mathbf{S}^{-1} and isolate $\Delta \lambda$. We get

$$\Delta \lambda = -\mathbf{S}^{-1} \Lambda (\Lambda^{-1} \mathbf{r}_{\Lambda S} + \Delta \mathbf{s}) \quad (3.34)$$

where $\mathbf{r}_{\Lambda S}$ represents the fourth element in the right hand side of Eqs. 3.30 and 3.32. We can get $\Delta \mathbf{s}$ from the second equation.

$$\Delta \mathbf{s} = \mathbf{A}_I \Delta \mathbf{x} + \mathbf{r}_I \quad (3.35)$$

The third equation remains unchanged.

$$\mathbf{A}_E \Delta \mathbf{x} = -\mathbf{r}_E \quad (3.36)$$

By plugging in $\Delta\lambda$ from Eq. 3.34 to the first equation, we get

$$(\mathbf{G} + \mathbf{A}_I^T \mathbf{S}^{-1} \Lambda \mathbf{A}_I) \Delta \mathbf{x} + \mathbf{A}_E^T \Delta \boldsymbol{\mu} = -\mathbf{r}_d + \mathbf{A}_I^T \mathbf{S}^{-1} \Lambda (-\mathbf{r}_I - \Lambda^{-1} \mathbf{r}_{\Lambda S}) \quad (3.37)$$

Eqs. 3.34-3.37 are known as the "normal equations," of the system of Eqns. 3.30 and 3.32.

Solving the normal equations

By combining Eqs. 3.36 and 3.37, we can formulate a $(k^2 + k) \times (k^2 + k)$ linear system,

$$\Delta \mathbf{m} = \begin{bmatrix} \mathbf{K} & \mathbf{A}_E^T \\ \mathbf{A}_E & 0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{r}_h \\ -\mathbf{r}_E \end{bmatrix} \quad (3.38)$$

where \mathbf{r}_h denotes the right hand side of Eq. 3.37, and $\mathbf{K} = \mathbf{G} + \mathbf{A}_I^T \mathbf{S}^{-1} \Lambda \mathbf{A}_I$. We can solve for $\Delta \mathbf{x}$ and $\Delta \boldsymbol{\mu}$ as $\Delta \mathbf{x} = \Delta \mathbf{m} [: k^2]$, and $\Delta \boldsymbol{\mu} = \Delta \mathbf{m} [k^2 + 1 :]$ from Eq. 3.38. Here, we used *Python*'s array index notation.

MVSA first applies the normal equations on the affine Newton step in Eq. 3.30 to solve for $\Delta \mathbf{x}^{\text{aff}}$, $\Delta \boldsymbol{\mu}^{\text{aff}}$, $\Delta \mathbf{s}^{\text{aff}}$, and $\Delta \lambda^{\text{aff}}$. Based on Eqs. 3.30 and 3.32, $\Delta \mathbf{a}^{\text{aff}} = \Delta \mathbf{a}$. We can then use $\Delta \mathbf{x}$ to solve for $\Delta \mathbf{s}$ in Eq. 3.35, and use $\Delta \mathbf{s}$ to solve for $\Delta \lambda$ in Eq. 3.34. MVSA initializes the values for the unknown variables as $\boldsymbol{\mu} = \mathbf{s} = \boldsymbol{\lambda} = [1, \dots, 1]^T$, while the estimated mixing matrix \mathbf{M} VCA obtained initializes the values for x .

MVSA iterates through the interior point method until a stopping criterion is achieved. Arbitrarily, the iterations stop When $\sigma < 10^{-8}$ and $\rho < 10^{-8}$, and the final solution \mathbf{x} for Eq. 3.26 is obtained. Recall that $\mathbf{x} \equiv \text{vec}(\mathbf{Q})$, and $\mathbf{Q} \equiv \mathbf{M}^{-1}$. Thus, we have an estimate for the final mixing matrix

$$\mathbf{M} = [\text{vec}^{-1}(\mathbf{x})]^{-1} \quad (3.39)$$

3.3 Theory behind hyperspectral unmixing's inversion

This study uses Spectral unmixing by splitting and augmented Lagrangian (SUnSAL) to perform inversion (Bioucas-Dias and Figueiredo, 2010). SUnSAL solves the following general constrained sparse regression problem

$$\begin{aligned} & \min_{\mathbf{A}} (1/2) \|\mathbf{MA} - \mathbf{Y}_k\|_2^2 + \lambda \|\mathbf{A}\|_1 \\ & \text{subject to: } \mathbf{A} \geq 0, \quad \mathbf{1}^T \mathbf{A} = 1 \end{aligned} \tag{3.40}$$

The operators $\|\dots\|_1$ and $\|\dots\|_2$ stand for the Manhattan and Euclidean norms, respectively, and $\lambda \geq 0$ controls the relative weight between the two norms.

Alternating direction method of multipliers

First, we consider an unconstrained problem with the form

$$\min_{\mathbf{A} \in \mathbb{R}^n} f_1(\mathbf{A}) + f_2(\mathbf{G}\mathbf{A}) \tag{3.41}$$

where \mathbf{G} has linearly independent rows and columns, and f_1 and f_2 are closed proper convex functions. A proper convex function $f(x)$ satisfies the following conditions: (1) $f(x) < +\infty$ for at least one x , and (2) $f(x) > -\infty$ for every x (Rockafellar, 1970). By applying ADMM, the convergence of the sequence \mathbf{A}_k solves Eq. 3.41.

$$\mathbf{A}_{k+1} = \arg \min_{\mathbf{A}} f_1(\mathbf{A}) + \frac{\mu}{2} \|\mathbf{G}\mathbf{A} - \mathbf{u}_k - \mathbf{d}_k\|_2^2 \tag{3.42}$$

$$\mathbf{u}_{k+1} = \arg \min_{\mathbf{u}} f_2(\mathbf{u}) + \frac{\mu}{2} \|\mathbf{G}\mathbf{A}_{k+1} - \mathbf{u} - \mathbf{d}_k\|_2^2 \tag{3.43}$$

$$\mathbf{d}_{k+1} = \mathbf{d}_k - (\mathbf{G}\mathbf{A}_{k+1} - \mathbf{u}_{k+1}) \tag{3.44}$$

where $\mu > 0$ is some arbitrary constant.

ADMM for SUnSAL

By rewriting Eq. 3.40 so that we can translate it to Eq. 3.41, we have

$$\min_{\mathbf{A}} (1/2) \|\mathbf{MA} - \mathbf{Y}_k\|_2^2 + \lambda \|\mathbf{A}\|_1 + \iota_{\{\mathbf{1}\}}(\mathbf{1}^T \mathbf{A}) + \iota_{\mathbb{R}_+^n}(\mathbf{A}) \tag{3.45}$$

where $\iota_{\{s\}}(X)$ is the indicator function given by

$$\iota_{\{s\}}(\mathbf{A}) = \begin{cases} 0, & \mathbf{A} \in S \\ \infty, & \mathbf{A} \notin S \end{cases} \tag{3.46}$$

In terms of the ADMM problem, Eq. 3.45 can be interpreted as

$$f_1(\mathbf{M}) \equiv \frac{1}{2} \|\mathbf{AA} - \mathbf{Y}_k\|_2^2 + \iota_{\{\mathbf{1}\}}(\mathbf{1}^T \mathbf{A}) \tag{3.47}$$

$$f_2(\mathbf{A}) \equiv \lambda \|\mathbf{A}\|_1 + \iota_{\mathbb{R}_+^n}(\mathbf{A}) \tag{3.48}$$

$$G \equiv I \quad (3.49)$$

The solution of Eq. 3.42 is then given by

$$A_{k+1} \leftarrow B^{-1}w - C \left(1^T B^{-1}w - 1 \right) \quad (3.50)$$

where

$$B \equiv M^T M + \mu I \quad (3.51)$$

$$C \equiv B^{-1}1 \left(1^T B^{-1}1 \right)^{-1} \quad (3.52)$$

$$w \equiv M^T Y_k + \mu (u_k + d_k) \quad (3.53)$$

We can rewrite Eq. 3.43 as

$$u_{k+1} \leftarrow \arg \min_u (1/2) \|u - \nu_k\|_2^2 + (\lambda/\mu) \|u\|_1 + \iota_{\mathbb{R}_+^n}(u) \quad (3.54)$$

where $\nu_k \equiv A_{k+1} - d_k$. To solve this, we use

$$u_{k+1} \leftarrow \max \{0, \text{soft}(\nu_k, \lambda/\mu)\} \quad (3.55)$$

where $\text{soft}(\nu_k, \lambda/\mu) = \text{sign}(\nu_k) (|\nu_k| - \lambda/\mu)_+$, and

$$(x)_+ = \begin{cases} x, & x \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (3.56)$$

SUnSAL iterates through Eqs. 3.50, 3.54, and 3.44 until A converges. The convergence criterion can be either a maximum number of iterations, or a minimized difference between the previous and current values of A.

Chapter 4

Methodology

The flowchart in Fig. 4.1 summarizes how this study applied hyperspectral unmixing to perform analog rainfall forecasting on strong monsoon events. We start by determining strong monsoon event days and apply hyperspectral unmixing on the SLP to obtain the archetypes of those days. From those archetypes, we find days with analogous archetypes and produce rainfall maps depending on the weights of the corresponding archetypes. We then calculate the forecast accuracy scores. The source codes of this study can be accessed through this link:

<https://github.com/cmdecastro/MSThesis>.

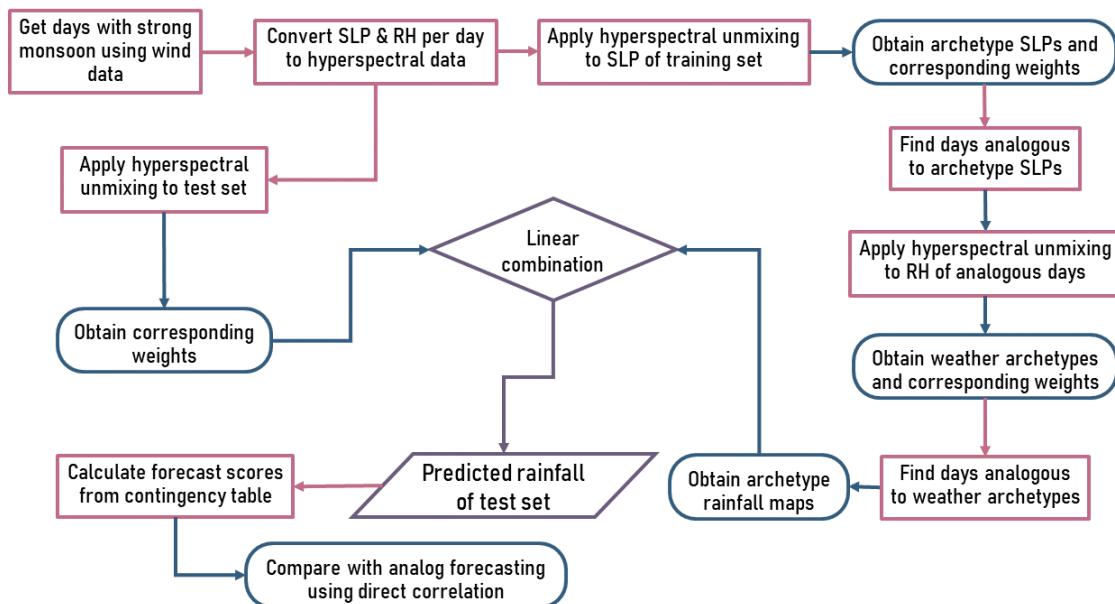


Figure 4.1: Methodological framework in applying hyperspectral unmixing to perform analog rainfall forecasting of strong monsoon events

4.1 Filtering days with strong monsoon events

The mean daily winds were obtained from the NCEP Reanalysis data provided by the NOAA/OAR/ESRL PSL, Boulder, Colorado, USA with spatial resolution $2.5^{\circ} \times 2.5^{\circ}$. Analysis was performed from 2001 to 2018 (Kobayashi et al., 2015).

To quantitatively determine strong Habagat days, the daily mean zonal wind along northwest Philippines bounded by latitude $12^{\circ}N$ to $18^{\circ}N$ and longitude $117.5^{\circ}E$ to $120^{\circ}E$ (illustrated in Fig 4.2) was calculated. Days with mean positive zonal wind in the upper 95th percentile are considered strong Habagat days.

On the other hand, the daily mean meridional wind along northeast of the Philippines bounded by latitude $12.5^{\circ}N$ to $20^{\circ}N$ and longitude $122.5^{\circ}E$ to $127.5^{\circ}E$ were calculated to determine strong Amihan days (illustrated in Fig 4.2). Days with negative mean meridional wind in the upper 95th percentile were considered strong Amihan days. Strong Amihan days with TCs over southeast Philippines were manually removed from the analysis, since forecasting rainfall due to TCs that coincide with strong Amihan winds is beyond the scope of this study. The same was not applied to TCs over northeast Philippines, since doing so would leave us with very few data points. From the study of Bagtasa, 2020, about 80% of strong Habagat events include the presence of a TC within northeast Philippines.

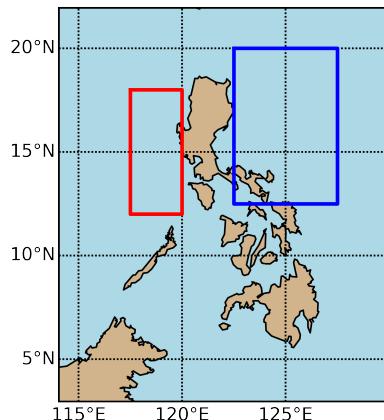


Figure 4.2: Domain for considering winds of strong Habagat (red box) and strong Amihan (blue box)

4.2 Sea-level pressure, relative humidity, and rainfall datasets

The 2001-2020 daily mean sea-level pressure were obtained from the JRA-55 reanalysis dataset in the domain ($5 - 22.5^{\circ}N$, $111.25 - 128.75^{\circ}E$) for Habagat and ($5 - 30^{\circ}N$, $115 - 135^{\circ}E$) for Amihan (downloaded from <https://rda.ucar.edu/> and illustrated in Fig. 4.3). The SLP domain for Habagat covers 225 tiles or pixels, while the SLP domain for Amihan covers 357 pixels.

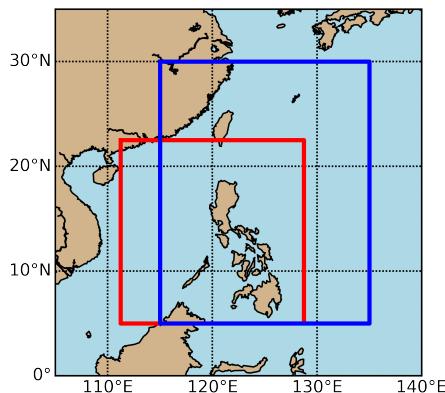


Figure 4.3: Sea level pressure domain for analysis of strong Amihan (red box) and strong Habagat (blue box) obtained from the JRA-55 dataset.

The 2001-2020 daily mean relative humidity (RH) was obtained from the same source in the domain ($11.25 - 18.75^{\circ}N$, $118.75 - 121.25^{\circ}E$) for Habagat and ($5 - 12.5^{\circ}N$, $123.75 - 126.25^{\circ}E$) for Amihan (illustrated in Fig. 4.4). We chose these specific domains for RH, since these are the regions of interest for heavy rainfall over the Philippines due to strong Amihan and Habagat winds. From here on, we denote each $1.25^{\circ} \times 1.25^{\circ}$ region as a tile. The RH domains for both Habagat and Amihan cover 21 tiles.

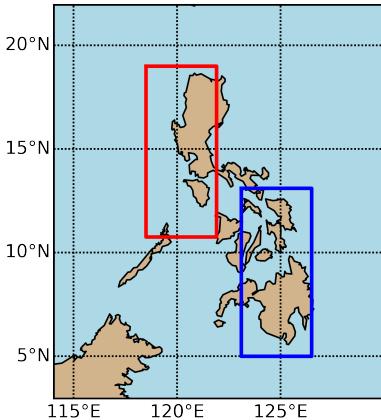


Figure 4.4: Relative humidity and PAGASA weather station domains for strong Haba-gat (red box) and strong Amihan (blue box), and the extent of GPM IMERGES rainfall bounded by the whole map

Lastly, the 2001-2020 daily mean rainfall distribution was obtained from the Global Precipitation Measurement (GPM) Integrated Multi-Satellite Retrievals for GPM (IMERGE) (downloaded from <https://disc.gsfc.nasa.gov/>) with spatial resolution $0.10^\circ \times 0.10^\circ$ in the domain ($3 - 22^\circ N$, $114 - 130^\circ E$). Note that though the rainfall domain encompasses parts of Malaysia and Brunei (illustrated in Fig. 4.4), they are excluded from discussion since this study only focuses on Philippine rainfall. Excluding rainfall over the ocean and rainfall over Malaysia and Brunei, the rainfall domain covers 2466 pixels.

Converting SLP and RH to Hyperspectral Data

Before we can apply hyperspectral unmixing on the daily SLP and RH, they must first be converted to hyperspectral data. To convert into hyperspectral data, the two-dimensional SLP and RH maps were first flattened into one-dimensional arrays starting from the upper-left corner of the domain, and traversing latitude-wise (or row-wise). This was done for all selected strong monsoon days. To maintain uniformity, the SLP and RH were normalized by dividing all values by the maximum value per day. In the end, each strong monsoon day corresponds to two normalized one-dimensional arrays containing the SLP and RH measurements for each tile.

Similar to the work of Horton et al., 2017, the strong monsoon days were then randomly separated into two groups: 90% of the strong monsoon days form the training set, while the remaining 10% of the strong monsoon days form the test

set. Horton et al., 2017 grouped one out of every five years as the validation period and the remaining as the calibration period to reduce the potential bias introduced by climate change. Randomizing the separation between the two sets also accounts for short term variabilities and climate oscillations such as the El Niño-Southern Oscillation (ENSO). All analyses were also separated between strong Amihan events and strong Habagat events, to reduce the high variability of the behaviours of SLP and RH across the whole year while transitioning from one monsoon season to the next.

4.3 Applying hyperspectral unmixing on the SLP and RH of training datasets

The MATLAB code from Li et al., 2015 was used to perform HySime, MVSA, and SuNSAL on the training set's SLP. Initially, the number of dimensions correspond to the number of tiles in an SLP map. In our case, we are dealing with at least 100 dimensions, which is too computationally expensive and complicated. Thus, HySime reduces the number of dimensions the succeeding steps need to analyze to a more manageable few. The number of SLP archetypes corresponded to the final number of reduced dimensions. MVSA estimates the normalized SLP corresponding to each archetype. From here on, we indicate said SLP as the archetype SLPs. Lastly, SuNSAL fits the archetype SLPs with the actual SLPs to calculate the corresponding weights of the archetype SLPs per day. After applying hyperspectral unmixing to the training set's daily mean SLP, we get a set of SLP archetypes and their corresponding weights per training day. Based on hyperspectral unmixing's working equation 2.1, \mathbf{M} contains the SLP archetypes, \mathbf{A} contains the corresponding weights, and \mathbf{Y} contains the actual SLP per day. Here, hyperspectral unmixing assumes that the actual SLP maps per strong monsoon day is just an ensemble of universal archetypes depending on the monsoon season (Habagat or Amihan).

Next, we apply the same process to the RH. However, we will not consider all training days, instead only the days with SLPs that match best with the archetype SLPs. The corresponding weights of the archetype SLPs per day dictates how similar the daily actual SLPs and the archetype SLPs are. For each SLP archetype, only training days that reached a set weight threshold were considered. The threshold was arbitrarily set as 0.4.

After applying hyperspectral unmixing twice—once to SLP, and then again on RH—we have narrowed down the final group of weather archetypes, characterized by both SLP and RH, that describe the training days. The second run of hyperspectral unmixing also obtained the weather archetypes’ corresponding weights per training day whose RH measurements were considered for the second run.

Getting the archetype rainfall maps per archetype SLPs

After using hyperspectral unmixing to extract the weather archetypes from the training set, we now move on to performing analog forecasting. Each weather archetype can be characterized by a group of analogous days from the training days of the second run. Again, the corresponding weights from hyperspectral unmixing dictate how analogous the second run’s training days are to the weather archetypes. The threshold was arbitrarily set to 0.4 again. Each group of analogous days have corresponding rainfall maps per day. The corresponding mean composite archetype rainfall maps were created from these groups of rainfall maps. In the end, each weather archetype combining SLP and RH patterns describe a corresponding archetype composite rainfall map.

4.4 Predicting rainfall distribution of test set based on SLP and RH

Hyperspectral unmixing was first applied to the test set to predict their rainfall distribution. From here, SuNSAL estimates the weights of the weather archetypes for each day in the test set. These weights can then be multiplied to the corresponding archetype rain maps to form a composite rain map per day.

Applying HSU on the SLP and RH of test set

Similar to the training set, the MATLAB code for hyperspectral unmixing from Li et al., 2015 was also applied to the test set. However, unmixing, the second step, was skipped since the weather archetypes were already previously obtained from the training phase. First, HySime projected the test set into the same subspace as the training set, thereby reducing the test set to the same number of dimensions as the training set. Then, SuNSAL fitted the weather archetypes from the training phase to

the actual SLP and RH measurements from the test set. In the end, hyperspectral unmixing obtained the corresponding weights of the training set's weather archetypes for each day in the test set. Hyperspectral unmixing also obtained the corresponding weights of the training set's SLP archetypes for each actual SLP of the testing days.

Creating composite rainfall and SLP maps of test set

The composite rainfall maps for the test set were obtained by forming a linear combination of the test set's weights and the training set's archetype rain maps. In other words, the test set's weights are multiplied by their corresponding archetype rain map, and the obtained rain maps are added up. The composite rain maps obtained were then compared with the actual rain maps in the test set. The same methods discussed in this subsection were also applied to form composite SLP maps of the test set. The composite SLP maps of the test set were obtained by forming a linear combination of the test set's weights and the training set's SLP archetypes.

4.5 Analog forecasting using direct correlation

To test the viability of using hyperspectral unmixing as a new method for analog forecasting, analog rainfall forecasting using the more traditional direct correlation was also applied. Here, the same training and test sets as hyperspectral unmixing were used for both Habagat and Amihan. Every daily mean SLP map from the test set was compared to all daily mean SLP maps in the training set to find the three most correlated (analogous) days. Each day in the test set would then correspond to three analogous days from the training set. The composite SLP and rainfall maps of these three days denote the forecasts for the test set. The correlation between SLP maps of the test and training sets are calculated as

$$R = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{\left(\sum_m \sum_n (A_{mn} - \bar{A})^2\right)\left(\sum_m \sum_n (B_{mn} - \bar{B})^2\right)}} \quad (4.1)$$

where A and B correspond to the two 2D arrays being compared, and \bar{A} and \bar{B} correspond to their overall mean. This equation was adapted from MATLAB's function for calculating the 2D correlation between two images.

4.6 Verifying accuracy of rainfall forecasts using hyperspectral unmixing and direct correlation

The obtained rainfall forecasts were compared to the observed rainfall data from PAGASA weather stations within the same regions of interest specified by the RH domains in Fig. 4.4 for Habagat and Amihan seasons. Other weather stations not affected by strong Habagat or strong Amihan-induced rains were disregarded, since they skew the contingency table towards more correct negatives.

The closest actual latitude-longitude location of each weather station was located in the GPM IMERGE rainfall domain, and then an $N \times N$ grid centered on the station was formed. This is done to account for spatial shifts in the resulting rainfall forecasts. The observed rainfall per station was compared to the closest forecast rainfall within each $N \times N$ grid. This study investigated two cases: $N = 3$ and $N = 5$. To verify the accuracy of predictions, forecast scores were calculated based on a contingency table (Gordon and Shaykewich, 2000). The contingency table implements the following format

Table 4.1: General schematic of a contingency table for verification of forecasts

		Observed	
		Yes	No
Forecast	Yes	hits	false alarm
	No	miss	correct negative

and abides by the following definitions

- hit - forecast and observed \geq HRT
- miss - forecast $<$ HRT, and observed \geq HRT
- false alarm - forecast \geq HRT, and observed $<$ HRT
- correct negative - forecast and observed $<$ HRT

Here, HRT corresponds to the heavy rain threshold, which designates the conditions for heavy rain and is set arbitrarily. Since this study serves as a proof of concept for a novel method, we only investigated two cases: rain vs no rain (HRT = 0.5mm), and heavy vs light rain (HRT = 30mm for Habagat and HRT = 20mm for Amihan). From the contingency table, the following forecast scores were calculated: bias score

(frequency bias), probability of detection (POD, hit rate), false alarm ratio (FAR), and the threat score (TS, critical success rate).

The bias score measures the ratio of the frequency of forecast events to the frequency of observed events, and is given by

$$BIAS = \frac{\text{hits} + \text{false alarms}}{\text{hits} + \text{misses}} \quad (4.2)$$

The probability of detection measures what fraction of the observed events were correctly forecast, and is given by

$$POD = \frac{\text{hits}}{\text{hits} + \text{misses}} \quad (4.3)$$

The false alarm ratio measures what fraction of the predicted events are false alarms, and is given by

$$FAR = \frac{\text{false alarms}}{\text{hits} + \text{false alarms}} \quad (4.4)$$

Lastly, the threat score measures the fraction of observed and/or forecast events that were correctly predicted, and is given by

$$TS = \frac{\text{hits}}{\text{hits} + \text{misses} + \text{false alarms}} \quad (4.5)$$

Analysis of forecast vs actual rainfall

We handpicked a select few forecasts from the test set for further analysis, and discuss the results for both methods, hyperspectral unmixing and direct correlation. The following criteria were used to determine which sample forecasts will be discussed farther

- "best" forecast for Habagat
- "worst" forecast for Habagat
 - rainfall due to strong Habagat
 - rainfall due to direct effects of a TC
- "best" forecast for Amihan
- "worst" forecast for Amihan

- overestimated rainfall

The TS of all rainfall forecasts for both the training and test sets were also compared to the standard value 0.25 (Wang et al., 2021). Rainfall forecasts with TS at least 0.25 when checking for heavy rainfall (HRT 30mm for Habagat, and 20mm for Amihan) are deemed acceptable.

Chapter 5

Results and Discussion

5.1 Selection of strong monsoon days

From 2001-2020, 132 days were considered as strong Habagat days and 267 days were considered as strong Amihan days based on the criteria discussed in Section 4.1. Tables 5.1 and 5.2 list down the strong Habagat and Amihan event days, respectively. Note that no strong Habagat events were found in 2010, 2019, and 2020 as no days were within the selection criteria. Strong Habagat events are concentrated only within the months of June to September, which is the active season of the Philippines' southwest monsoon, while strong Amihan events are concentrated only within the months of November to March, which is the active season of the Philippines' northeast monsoon.

Table 5.1: List of all strong Habagat days included in analysis separated into the test set (in boldface) and the training set (in non-boldface)

	June	July	August	September
2001		4, 5		25
2002		6, 7, 8, 9, 10, 11, 13		
2003			19	1
2004	29 , 30	1, 2	11, 12, 23, 24 , 25, 26, 27, 28	
2005		18	4, 5, 12	22
2006		6, 7, 8, 11, 12, 13, 14 , 24, 25		
2007			7, 8, 9, 10, 17, 18	
2008		27 , 28	22	
2009	3, 4, 5	17	6, 7, 8, 9	
2010				
2011	24, 25	17, 18	27, 28, 29, 30, 31	
2012	17, 18, 19	21, 30, 31	1, 2, 3, 6, 7 , 8, 26, 27	
2013			19 , 20, 21, 22	20 , 21, 22
2014		23	1 , 2 , 5, 6 , 7, 9	20
2015		6, 7, 8, 10, 11, 15, 16, 17	21, 22, 23	
2016			12, 13, 14, 15, 16, 17	
2017		28, 29, 30		
2018	10, 11, 12, 13, 14	21, 22	10, 11, 12, 13, 14, 15, 21, 22, 23, 24, 25	15

Table 5.2: List of all strong Amihan days included in analysis separated into the test set (in boldface) and the training set (in non-boldface)

	Jan	Feb	Mar	Nov	Dec
2001	15, 16	14, 15, 16, 17		18, 19	22, 24, 25
2002	2, 3, 8, 9	19	8, 9	3, 4, 5, 9, 26	9, 10, 13
2003	1, 4, 5, 28				15, 19, 20, 26, 27
2004	24, 25	10, 15, 16	7, 8	27	6, 29
2005	1, 14, 15, 16		5, 6, 13		5, 6, 15, 17, 18, 22
2006	7	3, 4, 8, 9			17, 18, 19, 22, 23, 28, 29, 30
2007	6, 7, 8, 28, 29, 30, 31	1, 2, 3		30	4, 5, 31
2008	1, 2, 3, 16, 17	10, 13 , 17	4		5, 6, 7, 18, 19
2009	1, 2, 3, 9 , 10, 11, 12, 13, 14 , 15			29, 30	1, 2, 3 ,4
2010	12, 13, 14 15, 16 , 17		10, 27, 28, 29	26	7, 26, 27, 31
2011	1, 7, 16, 17, 30 , 31	12	16, 25, 29	21, 24	16, 17, 24, 25, 26
2012	2, 4, 11, 25	8, 9, 11, 18, 19	25, 26 , 27 , 28		23, 24, 31
2013	14, 15, 16, 17, 18, 19, 28		3, 4	19, 29	27
2014	10, 11, 15, 17, 18, 19, 21, 22, 23	20			17, 18, 22
2015	1, 2, 8, 9, 10, 22 , 23, 31	1, 2			17, 18
2016	24, 25	6, 7, 8, 16, 24, 25, 29	1	1, 2, 3	16, 17, 30, 31
2017		10, 11, 13, 14			20, 31
2018	11, 12, 13	4, 11, 13	10		28, 29
2019	22, 23 , 24, 25, 26, 27			28, 29	7, 8, 9, 31
2020	1, 2, 14, 30, 31	1, 9, 17, 18, 19, 20, 22		24, 25, 29, 30	31

The monthly distributions of strong Habagat and Amihan days are illustrated in Figure 5.1. Based on the monthly distribution, strong Habagat events are more likely to occur during July and August and strong Amihan events are more likely to occur during December and January. The strong monsoon days for each season were separated into two sets: 90% of days as training set, and the remaining 10% as the test set. Training and testing days were randomly selected. Of the 132 strong Habagat days, 119 were considered as training days, while 13 were considered as testing days. Conversely, of the 267 strong Amihan days, 240 were considered as training days, while 27 were considered as testing days.

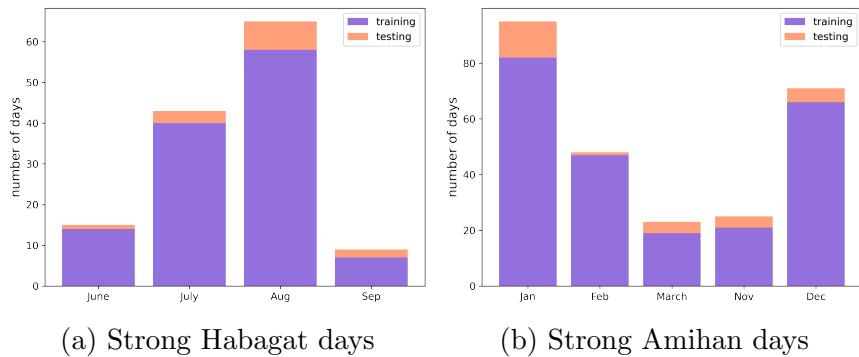


Figure 5.1: Monthly distribution of the training set and testing set for strong monsoon days

5.2 Determining the optimal SLP domain for analog forecasting from the training set

We first selected a domain where SLP is assumed to be the main driver of monsoon rainfall. To utilize SLP for analog rainfall forecasting, the SLP patterns that characterize strong monsoon rainfall events need to be included in the domain. In strong Habagat rainfall events, the depth of the Asian monsoon trough needs to be present and properly depicted in the domain. On the other hand, in strong Amihan rainfall events, the continental high over the Asian main continent that produces strong northerlies needs to be included. In this study, the SLP domains were varied through trial-and-error before arriving to the final domains used to perform the analog rainfall forecasts. For the Habagat season, the SLP domain was varied within the longitude range ($100^{\circ}E - 140^{\circ}E$) and latitude range ($0^{\circ} - 35^{\circ}N$). On the other hand, for the Amihan season, the SLP domain was varied within the longitude range

($110^{\circ}E - 140^{\circ}E$) and latitude range ($5^{\circ}E - 40^{\circ}E$). The SLP domain of interest for Habagat season extends more westward to account for TCs over the northern south China Sea that can partly cause the eastward extension of the monsoon trough, which can strengthen the monsoon winds. Conversely, the SLP domain of interest for the Amihan season is extended more northeastward to account for the continental high over China, which can strengthen the winter monsoon winds.

Three domains with varying sizes, summarized in Table 5.3, were tested as the SLP domain for the analog forecasting of strong Amihan and Habagat rainfall cases. Here, the domains vary in increments of 5, except for the Habagat season's small domain, which was chosen as the smallest area to encompass the monsoon trough to the northwest of the Philippines. We expect to lose information as the domain becomes smaller, however, that is compensated by more detailed archetypes as the smaller domain will contain less variability.

Table 5.3: Domain extent of the created SLP domains to determine the optimal SLP domain for hyperspectral unmixing for strong Habagat and Amihan events

	large domain		medium domain		small domain	
	<i>latitude</i>	<i>longitude</i>	<i>latitude</i>	<i>longitude</i>	<i>latitude</i>	<i>longitude</i>
Habagat	0-35	100-140	5-30	105-135	5-22.5	111.25-128.75
Amihan	5-40	110-140	5-35	115-135	5-30	115-135

5.2.1 Relative humidity as second rainfall predictor

Since rainfall is driven by wind-terrain interaction, SLP, which characterizes wind flow, is considered as the main environmental or dynamical predictor of rainfall. Other than SLP, however, the amount of moisture in air also influences the resulting rainfall an environment can produce. Hence, the use of an additional environmental predictor to characterize weather patterns in the analog forecasting method - relative humidity (RH). Since RH is carried by the wind towards the region of precipitation, RH is considered only in the region where precipitation is significant. In addition, in the skill assessment of the analog forecasting method, only the region where there is heavy rainfall will the comparison with observed data will be conducted. Therefore, the region where RH is calculated and rainfall assessment is done will be in the same domain. Here, we present how well the domains of interest for RH and PAGASA

weather stations capture regions with high RH and heavy rainfall in the Philippines for each strong monsoon season. To check the accuracy of rainfall forecasts using the hyperspectral unmixing method, the rainfall forecasts are compared with actual rainfall measurements from the PAGASA weather stations within the region of assessment. Figures 5.2 and 5.3 show the mean RH and rainfall, respectively, of all strong monsoon days for Habagat and Amihan season.

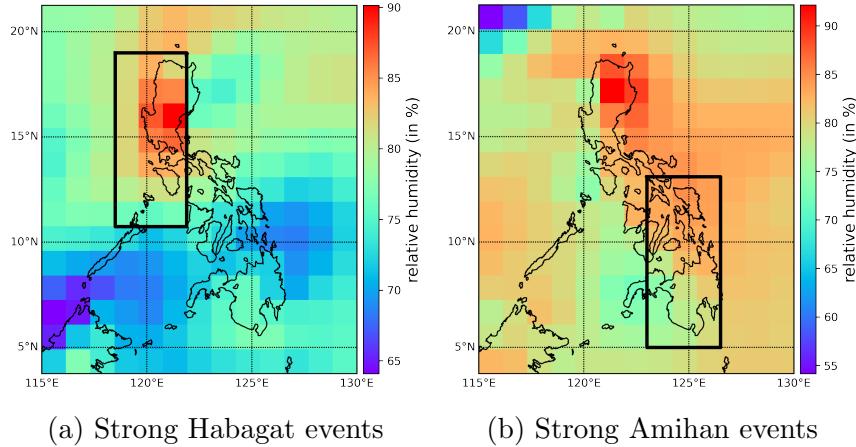


Figure 5.2: Mean relative humidity (in %) of all strong monsoon days with the domains of interest for RH and PAGASA weather stations (black box)

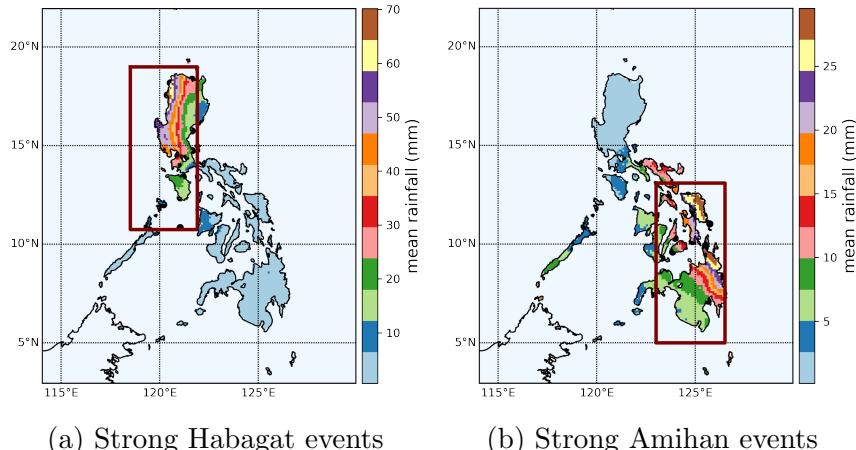


Figure 5.3: Mean rainfall (in mm) of all strong monsoon days with the domains of interest for RH and PAGASA weather stations (maroon box) and the PAGASA weather stations (black dots) within the domains

Based on Figs. 5.2a and 5.3a, the domains of interest for RH and PAGASA weather stations for strong Habagat events encompasses the regions with high RH at $\geq 85\%$, and the heaviest rainfall at ≥ 30 mm. The domain of PAGASA weather stations for

strong Habagat events also covers regions of heavy rainfall over west Luzon, especially Ilocos and central Luzon. On the other hand, based on Figs. 5.2b and 5.3b, the domains of interest for RH and PAGASA weather stations for strong Amihan events encompasses regions with high RH at 80% – 85%, and regions with the heaviest rainfall at ≥ 15 mm. The domain of PAGASA weather stations for strong Amihan events also covers regions of heavy rainfall over east Philippines, especially eastern Visayas and CARAGA.

5.2.2 Applying the hyperspectral unmixing method on the three domains

We then applied the hyperspectral unmixing method to the SLP and RH on the three domains for all heavy monsoon rainfall days. The goal is to find which of the created domains will produce the most similar weather pattern (SLP and RH) from the reconstructed weighted-mean archetypes. Note that the same training set, test set, and RH domain illustrated in Fig. 5.2 are used for all three SLP domains. The SLP and RH from different domains must first be converted to hyperspectral data before the MATLAB code from Li et al., 2015 can be used to perform HySime, MVSA, and SuNSAL.

After applying the method on the SLP data, the hyperspectral unmixing obtained different numbers of SLP archetypes with their corresponding weights per training day for each investigated domain. For instance, the six SLP archetypes obtained by hyperspectral unmixing for strong Habagat season's small domain are shown in Fig. 5.4.

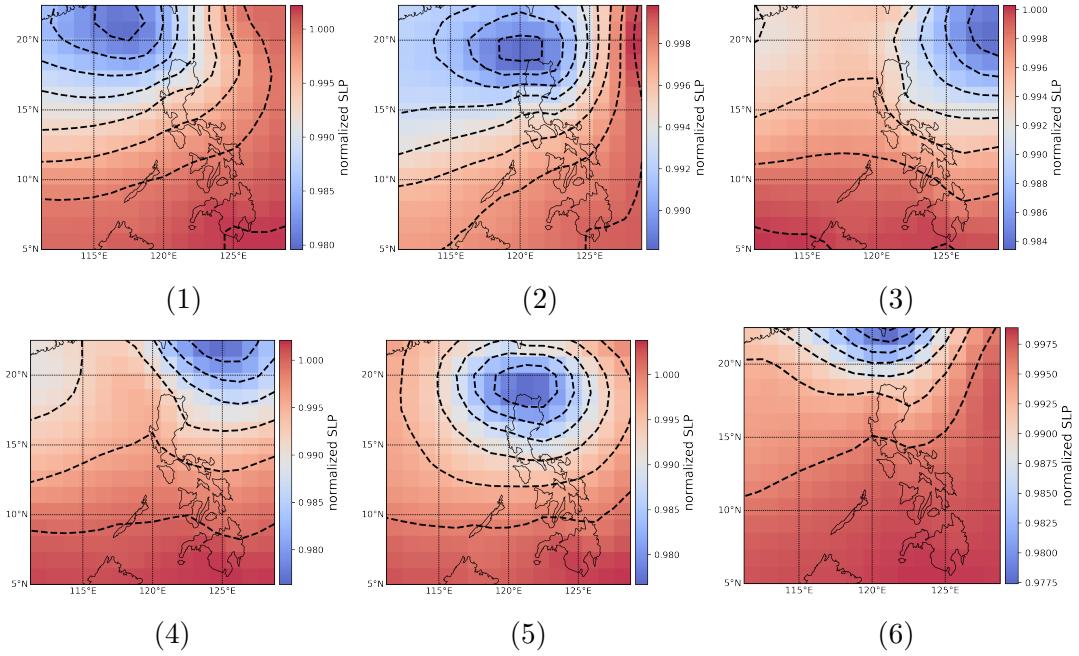


Figure 5.4: Normalized SLP archetypes obtained using hyperspectral unmixing for strong Habagat season in the small SLP domain

The reconstructed composite SLP maps of the test set were then obtained by applying hyperspectral unmixing to the test set's SLP patterns. Again, during the testing phase, unmixing, hyperspectral unmixing's second step, is skipped, since the training phase already obtained the SLP archetypes. The composite SLP maps of the test set is just the weighted-ensemble mean of the test set's corresponding weights and the training set's SLP archetypes. For instance, Figure 5.5 shows the composite SLP map for strong Habagat testing day August 24, 2004 as a weighted ensemble mean of the test set's weights and the training set's SLP archetypes.

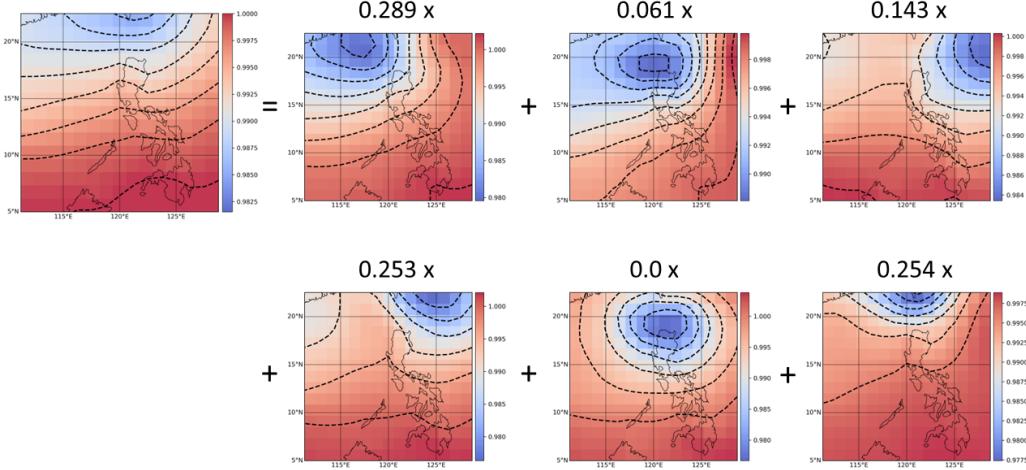


Figure 5.5: Composite SLP map for strong Habagat testing day August 24, 2004 as a weighted ensemble of the test set’s weights and the training set’s SLP archetypes obtained through hyperspectral unmixing on the small SLP domain

After obtaining the SLP archetypes, hyperspectral unmixing was applied to the training set’s RH. However, recall that not all training days are considered for the second run of hyperspectral unmixing on RH—only the days with actual SLPs that match best with the archetype SLPs. SLP serves as the primary characterization of strong monsoon weather patterns, and RH serves as the secondary characterization, thereby filtering the weather archetypes even further. We denote the chosen days as the filtered training set.

After applying to the filtered training set’s RH, hyperspectral unmixing obtained the RH archetypes describing strong monsoon events. For example, Fig. 5.6 shows the 12 RH archetypes obtained by hyperspectral unmixing for strong Habagat season’s smallest domain. Again, we narrow down the filtered training set even further by finding the days with actual RHs that match best with the archetype RHs. The rainfall of the filtered days dictate the overall rainfall archetypes of the training set. Thus, for each archetype RH, we have a corresponding rainfall archetype. For instance, Fig. 5.7 shows the 12 rainfall archetypes obtained after finding the 12 RH archetypes for strong Habagat season’s smallest domain previously shown in Fig. 5.6.

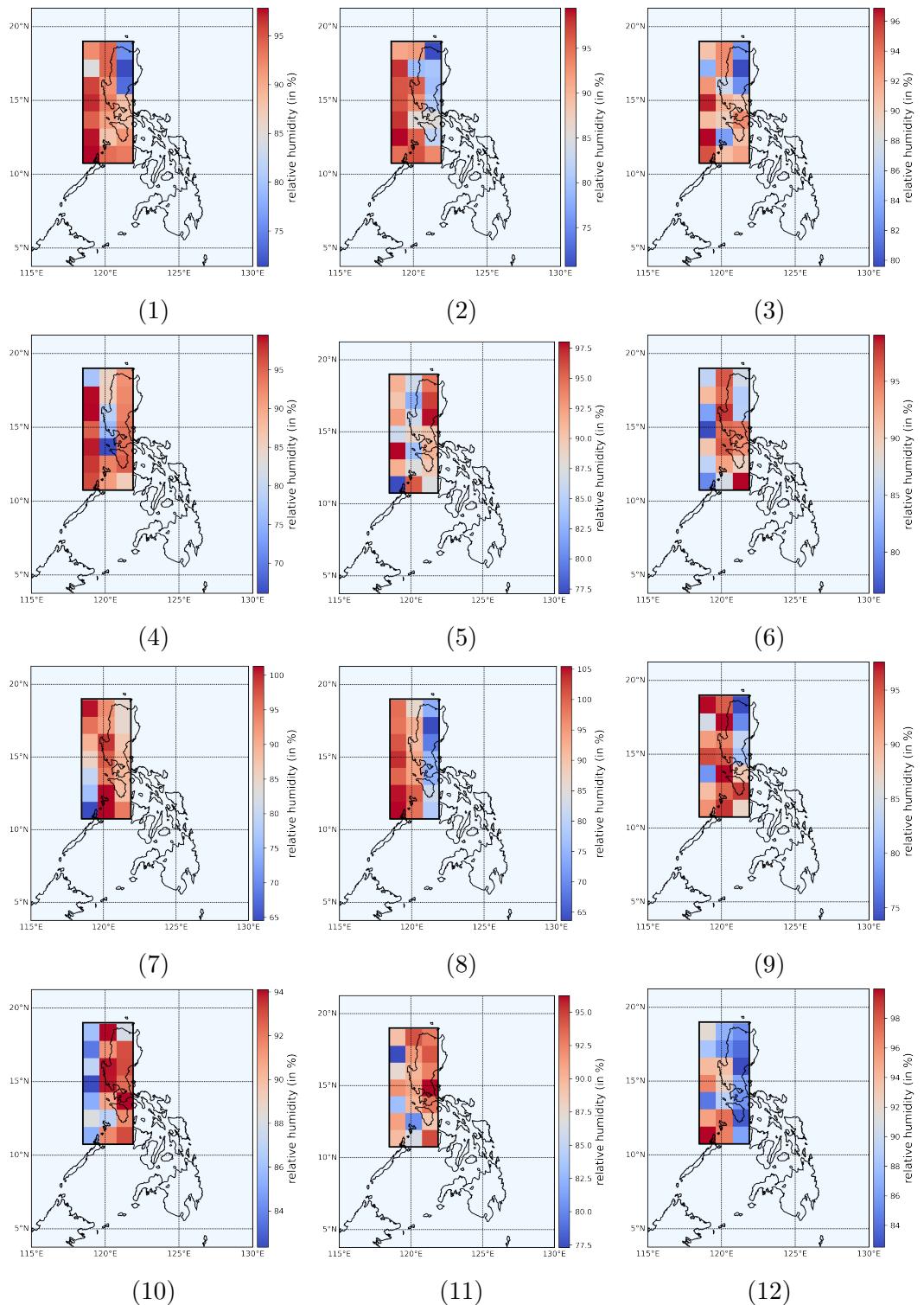


Figure 5.6: Relative humidity (in %) archetypes obtained using successive hyperspectral unmixing for strong Habagat season

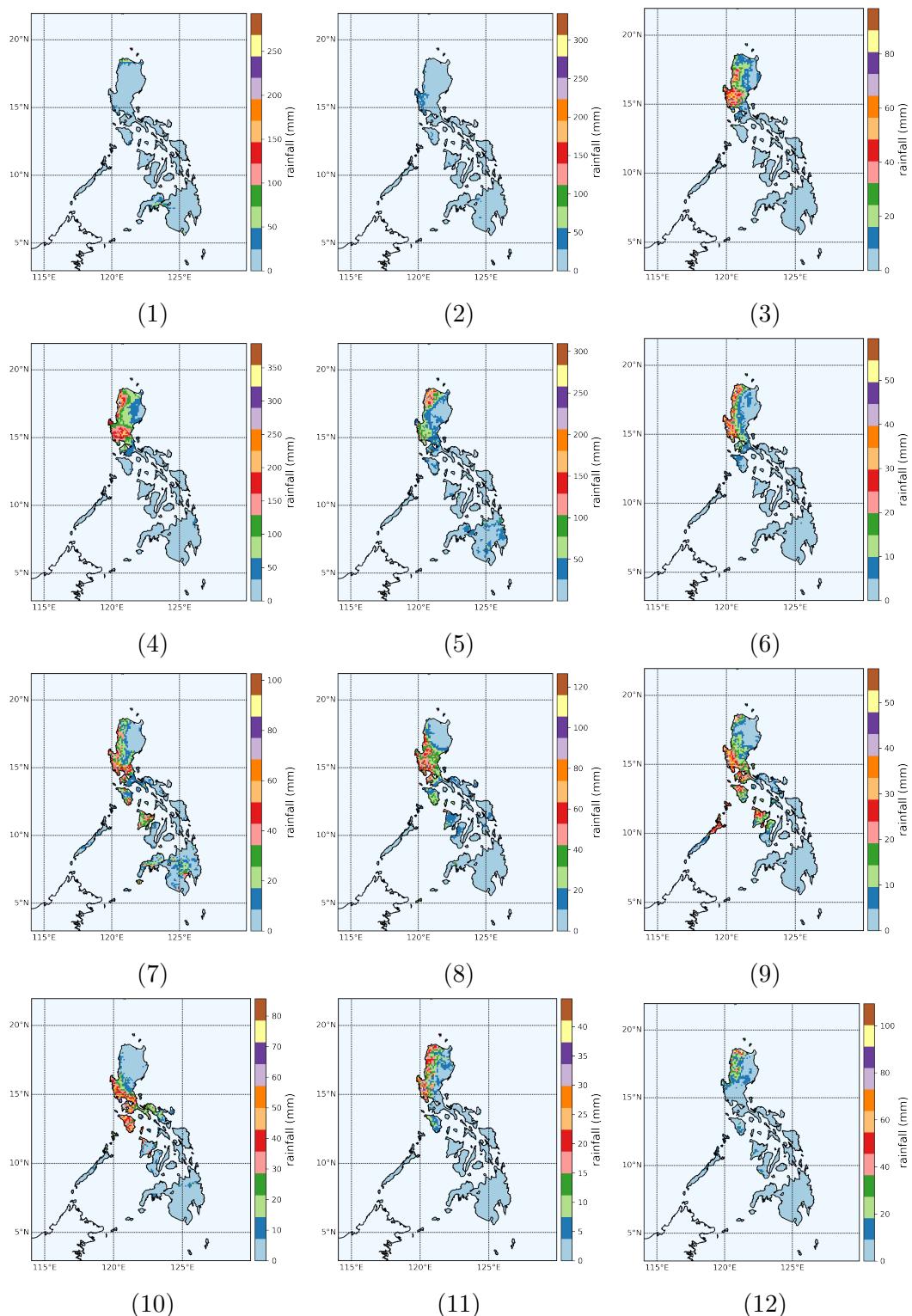


Figure 5.7: Rainfall (in mm) archetypes obtained using successive hyperspectral unmixing for strong Habagat season

The weighted ensemble mean of the rainfall archetypes and their corresponding weights per filtered training day comprise the training days' rainfall predictions. For

instance, Figure 5.8 shows the composite rainfall map for strong Habagat training day August 26, 2004 as a weighted ensemble mean of the filtered training set's weights and rainfall archetypes.

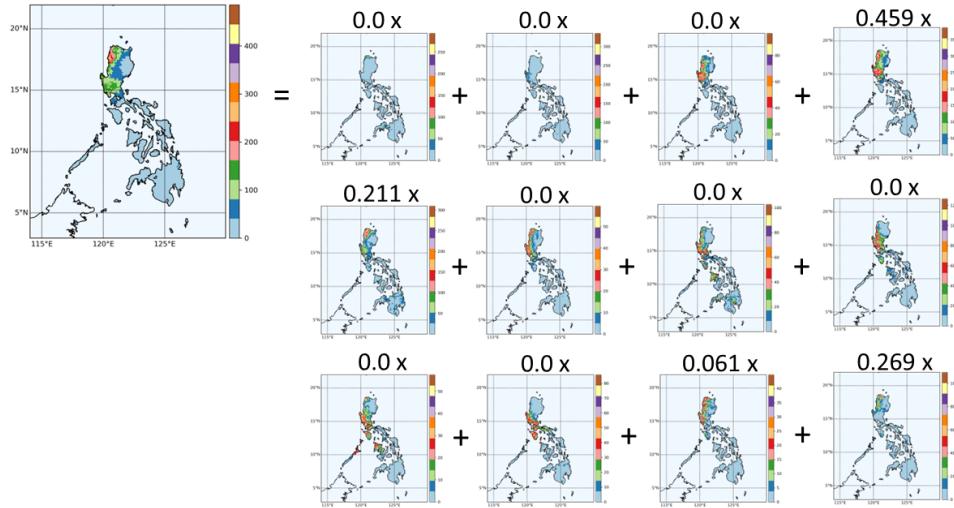


Figure 5.8: Composite rainfall map for strong Habagat testing day August 26, 2004 as a weighted ensemble of the training set's weights and the training set's rainfall archetypes obtained through hyperspectral unmixing on the small SLP domain

Lastly, the rainfall forecast accuracy scores of the training set were calculated by comparing the rainfall forecasts with the actual observed rainfall from PAGASA weather stations. The aforementioned methods for applying hyperspectral unmixing to SLP and RH are detailed in Sections 4.2, 4.3, 4.4, and 4.6.

5.2.3 Correlation between the composite and actual SLP maps of the training set for different SLP domains

We used the test set's composite SLP maps to determine which domain best captures the actual SLP maps. To compare how closely the composite SLP maps captured the actual SLP maps of the test set, their correlation was calculated using Eqn. 4.1. The mean correlation of SLP maps per domain determines which domain works best with hyperspectral unmixing. The mean correlation of SLP maps and the number of SLP archetypes per domain are summarized in Table 5.4. Sample composite SLP maps comparing their correlation with the actual SLP maps for different domains are also shown in Figures 5.9 and 5.10.

Table 5.4: Mean correlation (corr) between the composite and actual SLP maps of training sets for three domains detailed in Table 5.3, together with the corresponding number of SLP archetypes (archs) after hyperspectral unmixing

domain	large		medium		small	
	corr	archs	corr	archs	corr	archs
Habagat	0.715	22	0.723	14	0.939	6
Amihan	0.942	13	0.987	8	0.994	6

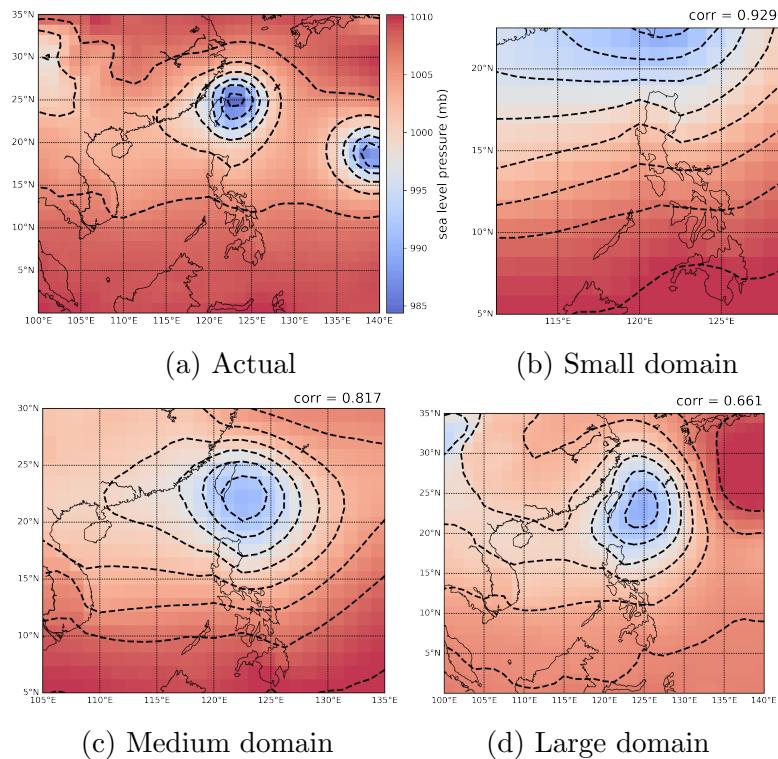


Figure 5.9: Comparing the correlation (corr) of the actual and composite SLP maps (in mb) of strong Habagat test day August 24, 2004 for three different SLP domains described in Table 5.3

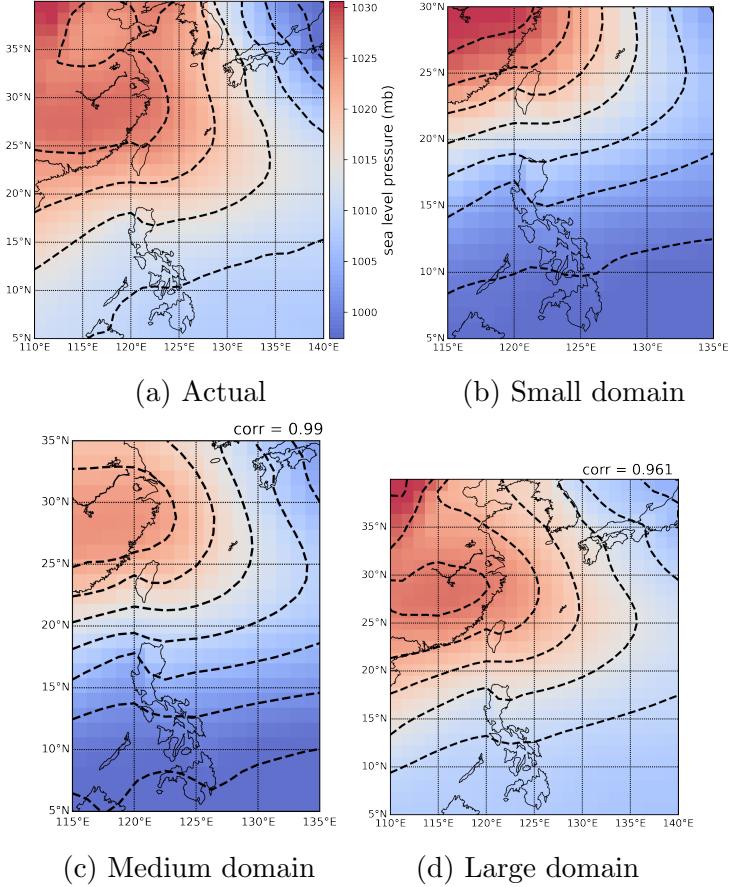


Figure 5.10: Comparing the correlation (corr) of the actual and composite SLP maps (in mb) of strong Amihan test day January 2, 2002 for three different SLP domains described in Table 5.3

Generally, the mean correlation increases as the SLP domain area size decreases, so initially, we would think the smaller the domain size, the better. However, decreasing the domain area further also means losing some important information (characterization of strong monsoon events) from the surrounding SLP patterns. Decreasing the SLP domain further for Habagat season means excluding the parts of the monsoon trough over northwest Philippines or southern China, which is a significant factor in strong Habagat events. Conversely, decreasing the SLP domain further for Amihan season excludes the continental high pressure to the north that determines strong northeasterly flow.

In addition, the number of archetypes increases as the SLP domain size increases. Increasing the domain size may mean more unnecessary information that can contribute to noise. For instance, the largest SLP domain for Habagat season covers Indonesia, Malaysia, and Thailand, which don't necessarily affect strong Habagat events

in the Philippines. Thus, hyperspectral unmixing extracts more SLP archetypes to account for the SLP over these other regions. A smaller domain results to fewer, more concise SLP archetypes that focus on regions of interest for strong monsoon events in the Philippines. The goal, then, is to choose the smallest possible domains that encompass the monsoon trough at northwest Philippines for Habagat season, and the continental high for Amihan season. The results in Table 5.4 show that the highest correlation with actual SLP is found in the small domain. In the next subsection, we look at how these three domains produce the best analog rainfall forecasts.

5.2.4 Rainfall forecast accuracy scores of the training set for different SLP domains

To determine which domain size yields the best rainfall forecast, we look at the rainfall forecast accuracy scores of the training set for the different domains. Table 5.5 displays the mean assessment scores obtained from the rainfall forecasts of the training set for the three domains of strong Habagat events, while Table 5.6 displays the same for strong Amihan events.

Table 5.5: Summary of mean assessment scores of all strong Habagat season training days for three SLP domains with different grid sizes and rain thresholds

domain	grid size	threshold (mm)	TS	BIAS	POD	FAR
large	$N = 3$	0.5	0.761	1.34	0.973	0.212
		30	0.343	1.40	0.520	0.402
	$N = 5$	0.5	0.772	1.34	0.980	0.208
		30	0.416	1.31	0.560	0.343
medium	$N = 3$	0.5	0.784	1.34	0.998	0.215
		30	0.450	1.60	0.694	0.405
	$N = 5$	0.5	0.785	1.34	0.998	0.214
		30	0.554	1.51	0.762	0.297
small	$N = 3$	0.5	0.794	1.36	1	0.206
		30	0.464	1.75	0.733	0.388
	$N = 5$	0.5	0.794	1.36	1	0.206
		30	0.548	1.60	0.782	0.316

Table 5.6: Summary of mean assessment scores of all strong Amihan season training days for three SLP domains with different grid sizes and rain thresholds

domain	grid size	threshold (mm)	TS	BIAS	POD	FAR
large	$N = 3$	0.5	0.615	1.82	0.929	0.347
		20	0.177	1.88	0.427	0.735
	$N = 5$	0.5	0.649	1.76	0.961	0.322
		20	0.248	1.65	0.508	0.631
medium	$N = 3$	0.5	0.615	2.07	0.999	0.385
		20	0.323	1.76	0.671	0.570
	$N = 5$	0.5	0.618	2.06	1	0.382
		20	0.443	1.53	0.745	0.441
small	$N = 3$	0.5	0.625	2.02	0.991	0.369
		20	0.294	1.18	0.530	0.570
	$N = 5$	0.5	0.628	2.02	0.991	0.367
		20	0.438	0.934	0.631	0.347

To determine the optimal SLP domain for rainfall forecasting, we focus primarily on the threat score (TS), the false alarm ratio (FAR), and the probability of detection (POD). A higher TS means more forecast events are correctly predicted. Next, a lower FAR means that fewer predictions are false alarms. Lastly, a higher POD means that more predictions are correct hits. In addition, N refers to the grid neighbor distance from the grid point that corresponds to the location of a PAGASA station. Thus, we want to pick the domain with the highest TSs, the lowest FARs, and the highest probabilities of detection. From the assessment results, we see that increasing the grid neighbor distance from $N = 3$ to $N = 5$ generally improves the assessment scores. Also, increasing the threshold from 0.5 mm to 30 mm (for strong Habagat events) or 20 mm (for strong Amihan events) reduces the assessment scores. The first threshold merely checks for the presence of rain, while the second sets a specific threshold for heavy rainfall. Since we obtained better assessment scores when using a higher grid size of $N = 5$, all discussions in the remaining sections of this chapter will only focus on assessment scores that use $N = 5$.

For strong Habagat rainfall, the TSs, PODs, and FARs remain consistently around 0.8, 1.0, and 0.3, respectively, when checking for the presence of rain (threshold 0.5 mm) in the three domains. Thus, we turn our attention to the assessment scores when checking for heavy rainfall (threshold 30 mm). The small and medium domains provide the highest TS at 0.5 versus the 0.4 of the large domain. Furthermore, all

three domains have FARs at around $0.3 - 0.4$. Lastly, the small domain has the highest POD at 0.8 versus the 0.7 of the medium domain, and the 0.6 of the large domain. Thus, since the small SLP domain consistently provides the highest TSs and the highest PODs, in addition to having the highest correlation with the actual domain-wide SLP in Table 5.4, the small SLP domain works best for strong Habagat season. Sample rainfall forecasts for the three domains of strong Habagat training day August 26, 2004 is shown in Figure 5.11. Both the small and large domains were able to capture the spread of the rainfall over central and north Luzon, while the medium domain was only able to capture the rainfall over Ilocos region.

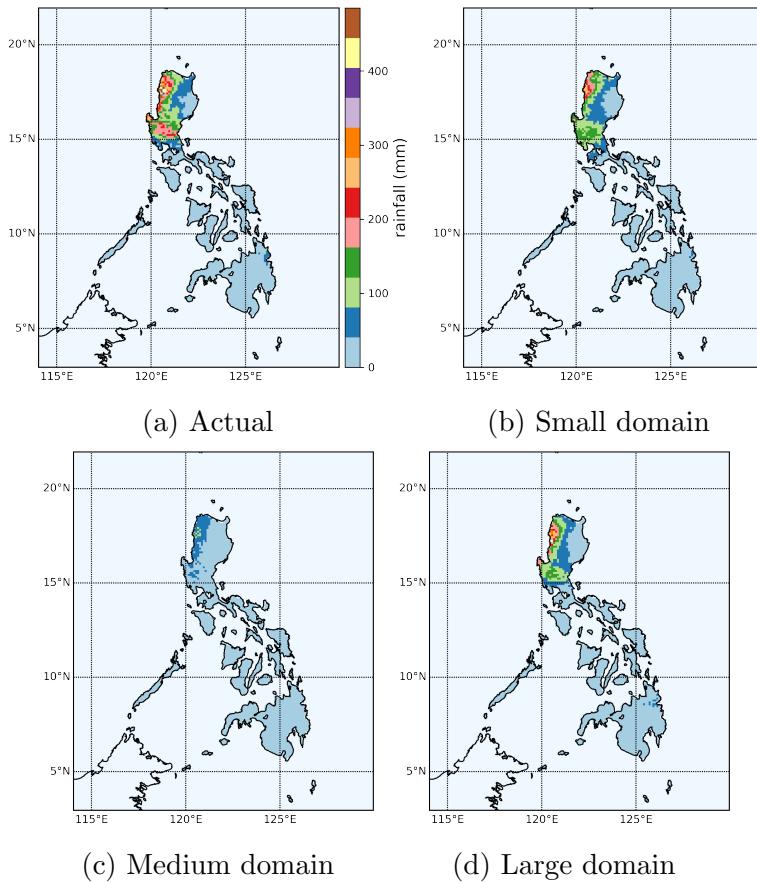


Figure 5.11: Actual and composite rainfall maps (in mm) of strong Habagat training day August 26, 2004 for three different SLP domains described in Table 5.3

For strong Amihan rainfall, the TSs, PODs, and FARs remain consistently around 0.6, 1.0, and 0.4 respectively, when checking for the presence of rain (threshold 0.5 mm) in the three domains. Thus, just like for strong Habagat rainfall, we focus on the assessment scores when checking for heavy rainfall (threshold 20 mm). The TS increases from 0.2 to 0.4, and the FAR decreases from 0.6 to 0.3 as the SLP domain

decreases. However, the medium domain considerably has the highest POD at 0.7 among the three. Thus, since the small SLP domain provides the highest TSs and lowest FARs, and the second highest POD, the small SLP domain also works best for strong Amihan events. Sample rainfall forecasts for the three domains of strong Amihan training day January 8, 2004 are shown in Figure 5.12. The small domain was able to capture the rainfall over southeast Mindanao, while the medium and large domains weren't. Furthermore, the medium and large domains overestimated the rainfall over Bicol region.

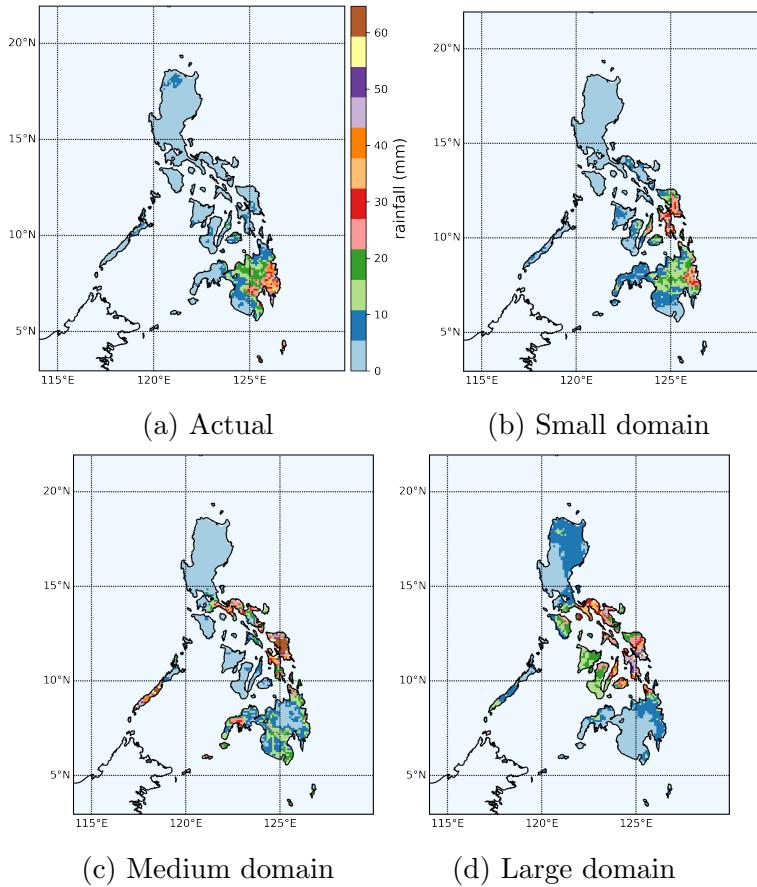


Figure 5.12: Actual and composite rainfall maps (in mm) of strong Amihan training day January 8, 2004 for three different SLP domains described in Table 5.3

In the end, the following small SLP domains: $(5 - 22.5^{\circ}N, 111.25 - 128.75^{\circ}E)$ for Habagat and $(5 - 30^{\circ}N, 115 - 135^{\circ}E)$ for Amihan were chosen since they encompass the necessary regions of interest while maintaining high SLP correlations and the best forecast accuracy scores.

From here on, we only present and discuss the results of the analog rainfall fore-

casting using the small SLP domains illustrated in Fig. 4.3. This domain was deemed to represent the best rainfall forecasting results for both strong Amihan and Habagat. The forecasting results from other investigated domains are excluded from analysis and discussion, since their forecasting accuracies were sub-par.

5.3 Determining scaling between actual and forecast rainfall

Upon initial investigation of the actual and forecast rainfall for the test set, we found that rainfall forecasts produced by hyperspectral unmixing tends to underestimate rainfall amount. Thus, a scaling factor was calculated from the training set to account for this bias. We obtained the scaling between the actual rainfall and forecast rainfall by getting the overall mean ratio between the two for all training days.

Since the heavy rainfall during strong Habagat events are concentrated over Luzon, only the rainfall within the latitudes $12.05^{\circ}N - 17.05^{\circ}N$ and longitudes $120.05^{\circ}E - 121.95^{\circ}E$ was considered. Conversely, since the heavy rainfall during strong Amihan season concentrates in east Philippines, only the rainfall within the latitude $5.55^{\circ}N - 11.95^{\circ}N$ and longitude $123.05^{\circ}E - 126.95^{\circ}E$ was considered. The rainfall scaling factor from derived only from these regions. Rainfall over water was also disregarded. For each training day, each tile in the actual rainfall map is divided by the corresponding tile in the forecast rainfall map. After this, the mean ratio between the actual and forecast rainfall maps per training day was calculated. The multiplicative scalings between the actual and forecast rainfall were determined to be 1.336 for strong Habagat events and 1.802 for strong Amihan events. Thus, all rainfall forecasts for the test set and training set were multiplied by their corresponding multiplicative scalings to match the actual rainfall better. Sample rainfall forecasts comparing the results with and without the multiplicative scalings are shown in Figure 5.13.

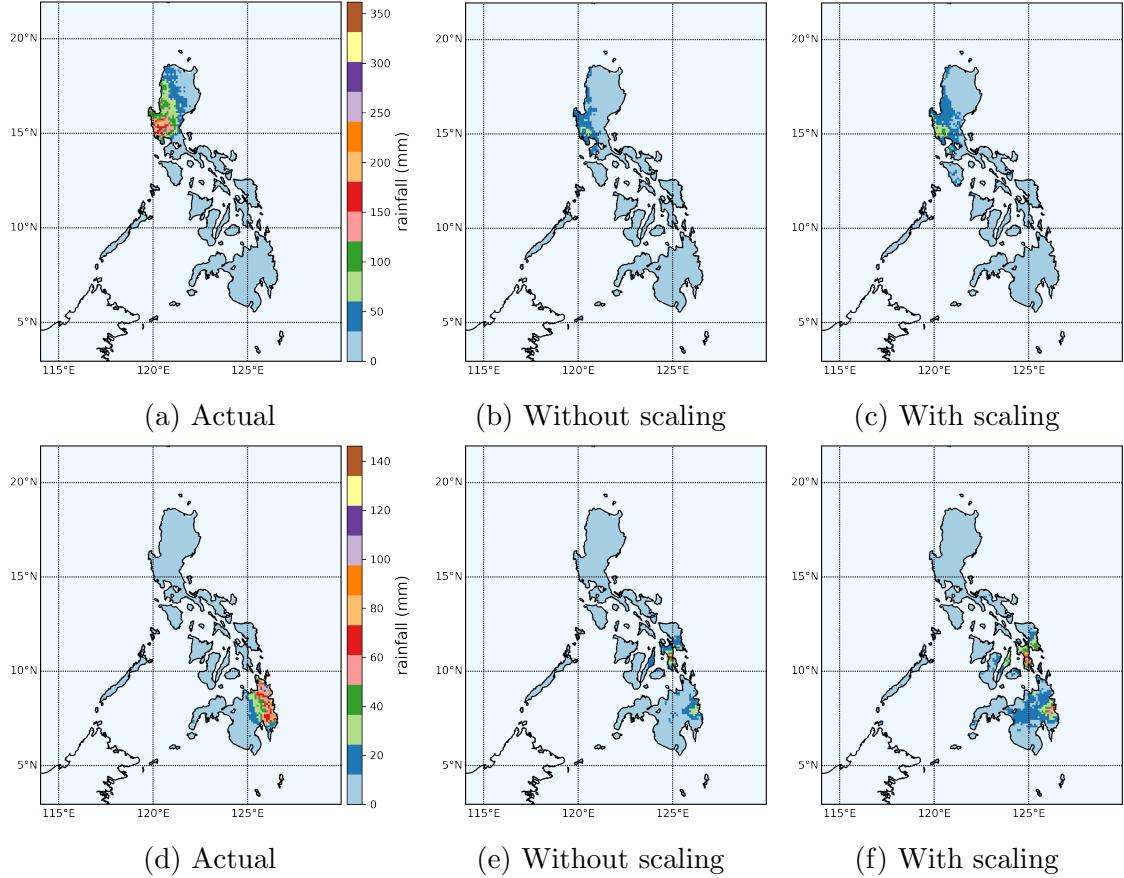


Figure 5.13: Actual and forecasts rainfall maps of strong Habagat testing day July 7, 2017 (first row), and strong Amihan testing day January 30, 2011 (second row) with and without the multiplicative scaling

5.4 Weather archetypes and rainfall forecasts obtained for strong Habagat events

This section presents the results after applying hyperspectral unmixing on the SLP and RH of the training sets for the smallest domain of strong Habagat events. This section also presents strong Habagat events' sample rainfall forecasts and the overall assessment scores for both hyperspectral unmixing and direct correlation. Here, hyperspectral unmixing's performance is compared with the more traditional method of direct correlation. In the direct correlation method, each daily mean SLP map from the test set was compared to all daily mean SLP maps in the training set to find the analogous days. Those with high correlation were deemed analogous. The rainfall maps of the analogous days denote the forecasts for the test set. The methods for direct correlation are detailed in Section 4.5.

5.4.1 SLP archetypes obtained through hyperspectral unmixing

Figure 5.4 shows the six archetype SLPs hyperspectral unmixing obtained using the smallest domain for strong Habagat events. Different combinations of the six archetype SLPs form ensembles of the actual SLP measurements of the training days. Archetypes 1 and 2 describe a low pressure area to the northwest of Luzon over the northern South China Sea, an indication of the eastward extension of the monsoon trough induced by a TC over the northern South China Sea similar to Fig. 2.2. The remaining four archetypes describe a low pressure system over north of the Philippines, indicating the movement of a TC coming from the east of Luzon Strait in archetypes 3 and 4 towards northern Luzon in archetype 5, or towards Taiwan in archetype 6. Archetypes 3 through 6 could also indicate the movement of a TC from northeast Philippines towards Luzon, and then curving upwards to Taiwan.

5.4.2 Rainfall archetypes obtained through successive SLP-RH hyperspectral unmixing

After successively applying hyperspectral unmixing twice (one for SLP then another for RH) to the training set, we obtained 12 rainfall archetypes show in Fig 5.7 corresponding to 12 weather archetypes (combining SLP and RH) for strong Habagat events. Similar to the SLP archetypes, different combinations of the 12 rainfall archetypes form ensembles of the actual rainfall maps of the test set. Archetypes 1 and 2 describe very little rain over Luzon. The remaining 10 archetypes describe heavy rainfall over Luzon in varying degrees of intensity and spatial distribution. Archetypes 3 and 4 describe heavy rainfall covering almost all of Luzon, while the others describe heavy rainfall concentrated over east Luzon. The heavy rainfall over east Luzon is distributed as follows: Ilocos region and Central Luzon in archetypes 5 and 6, mostly Central Luzon in archetypes 7, 8, 9, and 10, Ilocos region and Cagayan Valley in archetype 11, and mostly Ilocos region in archetype 12.

5.4.3 Forecast accuracy assessment scores

The mean assessment scores obtained from the rainfall forecasts of the test set using hyperspectral unmixing and direct correlation given different parameters are displayed in Table 5.7. The PODs for hyperspectral unmixing are closer to 1.000 than

those using direct correlation, indicating more correct hits from hyperspectral unmixing. Based on their TSs at 0.543 and 0.536, respectively, when checking for heavy rainfall (threshold 30 mm), both hyperspectral unmixing and direct correlation correctly forecast the rainfall for around half the cases with reasonable skill. However, hyperspectral unmixing's FAR at 0.362 is higher than direct correlation's at 0.257, indicating that hyperspectral unmixing is more likely to forecast false alarms when checking for heavy rainfall.

Table 5.7: Summary of mean assessment scores of all strong Habagat events testing days for hyperspectral unmixing and direct correlation with different heavy rain thresholds

method	threshold (mm)	TS	BIAS	POD	FAR
hyperspectral unmixing	0.5	0.713	1.45	1.0	0.287
	30	0.543	1.76	0.851	0.362
direct correlation	0.5	0.722	1.43	0.996	0.275
	30	0.536	1.24	0.757	0.257

5.4.4 “Best” rainfall forecast using hyperspectral unmixing

The testing day July 14, 2006 was chosen as the “best” rainfall forecast for strong Habagat events using hyperspectral unmixing, since hyperspectral unmixing clearly captured the regions of heavy rainfall, specifically over Ilocos region and Zambales-Pangasinan area. When checking for heavy rainfall, it also has higher TS at 0.636 and POD at 1.0 versus the mean TS at 0.543 and POD at 0.851 from Table 5.7. Figure 5.14 compares the composite SLP maps obtained by hyperspectral unmixing and direct correlation to the actual SLP maps for strong Habagat testing day July 14, 2006. While Figure 5.15 compares the rainfall forecast obtained by hyperspectral unmixing and direct correlation to the actual rainfall for the same date.

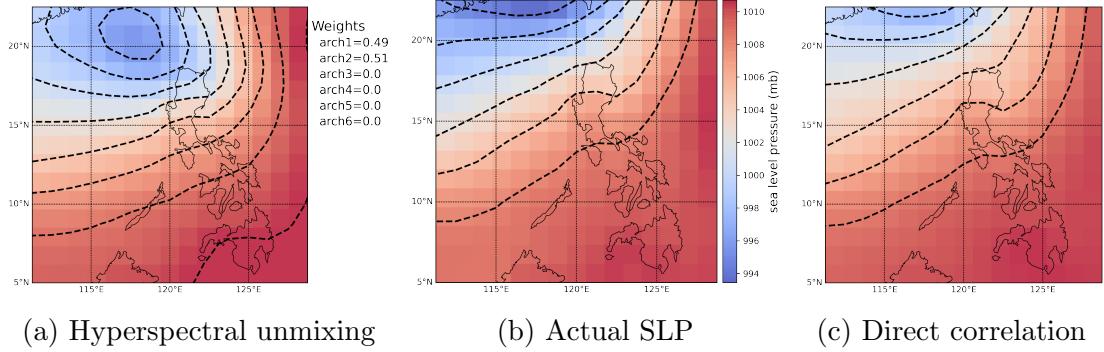


Figure 5.14: Composite SLP (in mb) of “best” rainfall forecast for strong Habagat season on 07/14/2006 using analog forecasting with hyperspectral unmixing and direct correlation

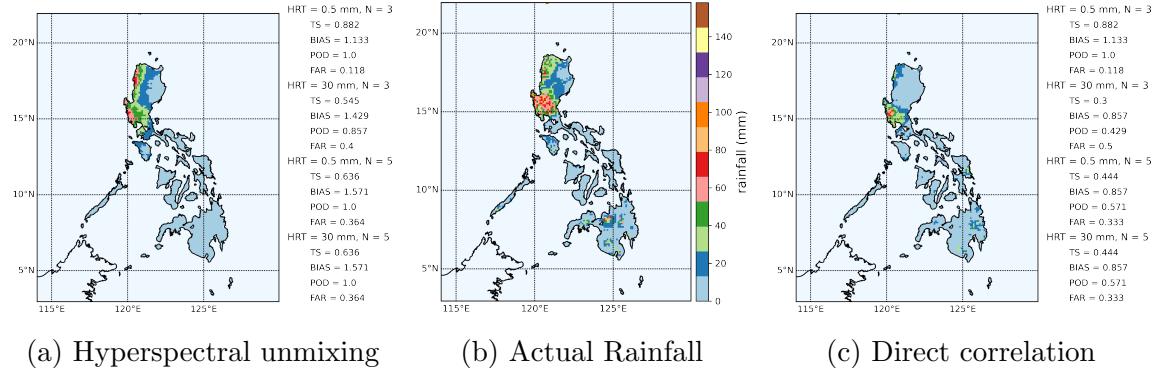


Figure 5.15: “Best” rainfall forecast (in mm) for strong Habagat season on 07/14/2006 using analog forecasting with hyperspectral unmixing and direct correlation

Figure 5.14a shows that the composite SLP did not fully capture the actual SLP for July 14, 2006. The low pressure area over southern mainland China extends farther southeastward in the composite compared to the actual. However, the distance of the isobars across western Philippines was better captured in the hyperspectral unmixing than the direct correlation. These isobars dictate the strength and direction of the strong monsoon flow. This resulted in the rainfall forecast in Fig. 5.15a which captures the high rainfall over Zambales and Pangasinan, and the decreasing rainfall towards central Luzon. In contrast, the forecast in Fig. 5.14c obtained through direct correlation resulted in rainfall more concentrated in southwest Luzon. The rainfall forecast in Fig. 5.15c only captured the rainfall over Zambales, but not the rainfall over Ilocos. Direct correlation also produced a lower TS at 0.444 and POD at 0.571 versus hyperspectral unmixing at 0.636 and 1.0, respectively, when checking for heavy rainfall or heavy rain threshold (HRT) = 30 mm.

5.4.5 “Worst” rainfall forecast using hyperspectral unmixing

The testing day June 29, 2004 was chosen as the “worst” rainfall forecast for strong Habagat events using hyperspectral unmixing. Figure 5.16 compares the composite SLP maps obtained by hyperspectral unmixing and direct correlation to the actual SLP maps for strong Habagat testing day June 29, 2004. On the other hand, Figure 5.17 compares the rainfall forecast obtained by hyperspectral unmixing and direct correlation to the actual rainfall for the same date.

Fig. 5.16a shows that a TC is present over northeast Philippines resulting to much heavier rainfall in the northern Luzon region compared to other test days. In this case, the composite SLP could not fully capture the low pressure system from the TC, so the forecast rainfall in Fig. 5.17a at ≤ 50 mm is much lower than the expected ≥ 100 mm. Looking closer at the archetype SLPs comprising the composite map, we see that more archetypes are associated with the low pressure systems over mainland China than archetypes associated with a TC over northeast Philippines. Most likely, the archetype SLPs are not enough to fully characterize a TC as strong as the one during June 29, 2004. In June 29, 2004, typhoon Mindulle (locally, typhoon Igme) struck Luzon peaking at 232 kph winds on June 28. Thus, hyperspectral unmixing has difficulty predicting rainfall directly caused by a TC.

Direct correlation captured the low pressure around the center of the TC closer to the actual than hyperspectral unmixing did. Also, hyperspectral unmixing predicts rainfall at 50 mm, while direct correlation predicts heavier rainfall at 100 – 250 mm, which is closer to the actual rainfall at 100 – 500 mm. Similar to the July 14, 2006 hyperspectral unmixing forecast, the forecast from hyperspectral unmixing on this event also has higher TSs at 1.000 and PODs at 1.000, and lower FARs at 0.000 when compared to the mean TS at 0.543, POD at 0.851, and FAR at 0.362 from Table 5.7. However, the assessment scores merely checks for the presence of rain and rainfall at 30 mm, which does not necessarily reflect the extremely heavy rainfall (at least 100 mm) directly caused by a TC. Thus, in this case, the assessment scores are not enough to judge whether the forecast is “good” or “bad.”

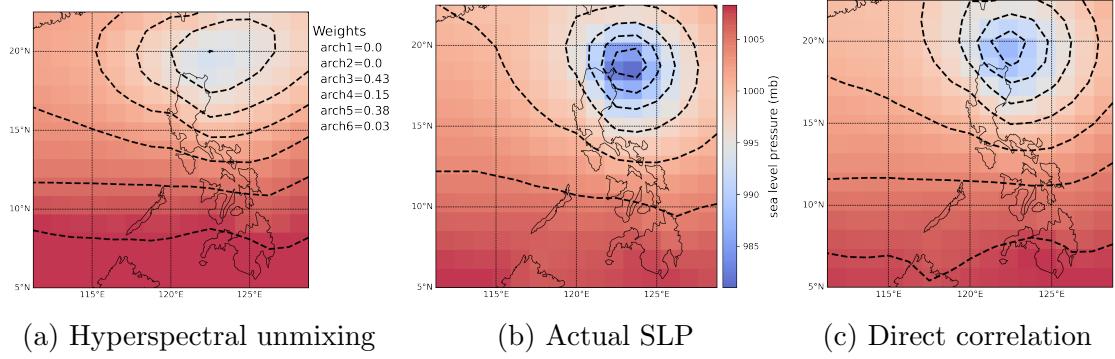


Figure 5.16: Composite SLP (in mb) of “worst” underestimated rainfall forecast for strong Habagat season on 06/29/2004 using analog forecasting with hyperspectral unmixing and direct correlation

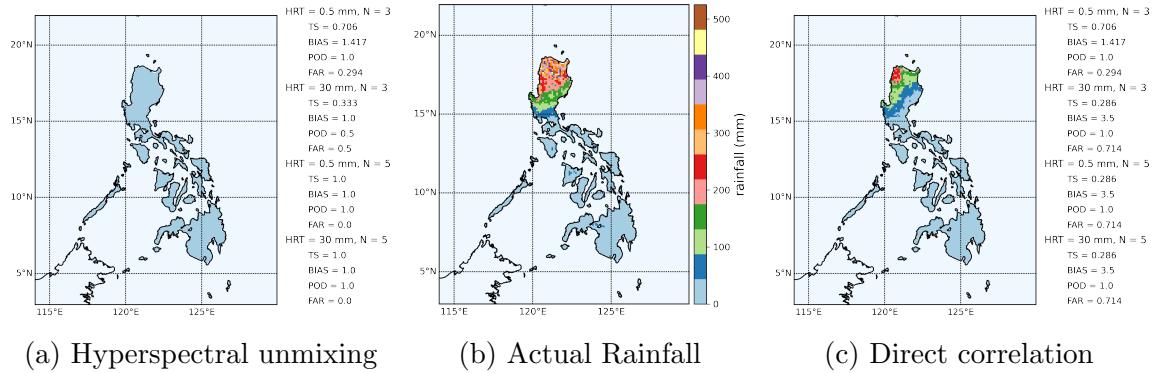


Figure 5.17: “Worst” underestimated rainfall forecast (in mm) for strong Habagat season on 06/29/2004 using analog forecasting with hyperspectral unmixing and direct correlation

Similarly, Figure 5.18 compares the composite SLP maps obtained by hyperspectral unmixing and direct correlation to the actual SLP maps for strong Habagat testing day August 2, 2014. August 2, 2014 has a lower TS at 0.364, and a higher FAR at 0.636 when compared to the mean TS at 0.543, and FAR at 0.362 from Table 5.7. It is apparent that the composite SLP from hyperspectral unmixing has a deeper low across northern Luzon with closely spaced isobars indicating strong winds. Indeed, the strong winds depicted in the hyperspectral unmixing forecasts overestimated rainfall mostly over northwest of Luzon at 30–80 mm versus the actual rainfall at 15 – 40 mm as shown in Fig. 5.18 and Fig. 5.19.

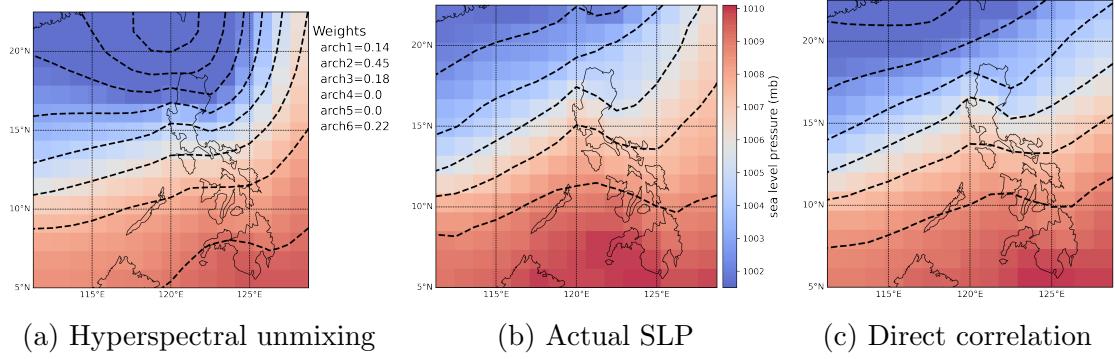


Figure 5.18: Composite SLP (in mb) of “worst” overestimated rainfall forecast for strong Habagat season on 08/02/2014 using analog forecasting with hyperspectral unmixing and direct correlation

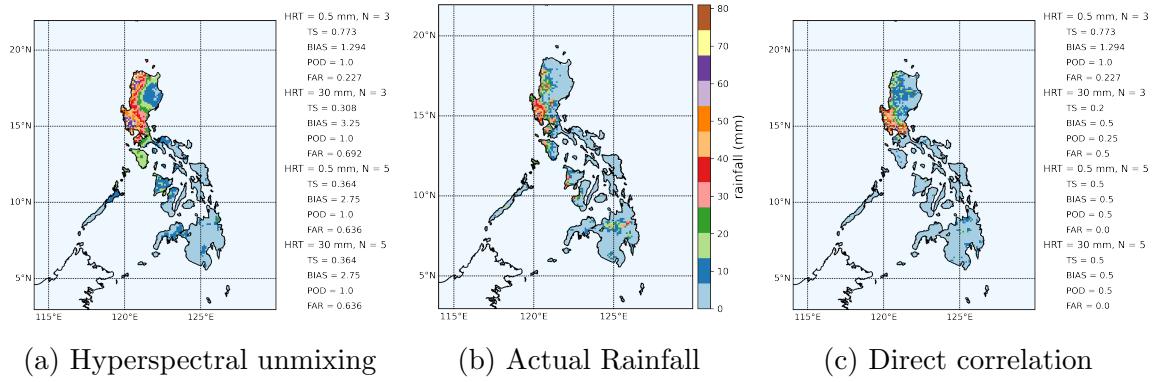


Figure 5.19: “Worst” overestimated rainfall forecast (in mm) for strong Habagat season on 08/02/2014 using analog forecasting with hyperspectral unmixing and direct correlation

In contrast, the direct correlation method has higher TS at 0.500 and lower FARs at 0.000 compared to that of hyperspectral unmixing at 0.364 and 0.636, respectively. Direct correlation captured the actual SLP better as shown in Fig. 5.18c and the corresponding rainfall distribution was also forecasted reasonably well as shown in Fig. 5.19c.

5.5 Weather archetypes and rainfall forecasts obtained for strong Amihan events

This section presents the results of applying hyperspectral unmixing on the SLP and RH of the training sets for the smallest domain of strong Amihan events. This section also presents strong Amihan events’ sample rainfall forecasts and the overall

assessment scores for both hyperspectral unmixing and direct correlation.

5.5.1 SLP and RH archetypes obtained through hyperspectral unmixing

Hyperspectral unmixing obtained six archetype SLPs shown in Figure 5.20 for strong Amihan events. All six archetypes describe a high pressure area over mainland China, with varying degrees of coverage and intensity. Archetypes 1, 2, and 3 indicate a high pressure area concentrated only over mainland China. Archetype 4 indicates an eastward extension of the high pressure area, while archetypes 5 and 6 indicate a southeastward extension of the high pressure area. Hyperspectral unmixing also obtained 11 RH archetypes shown in Figure 5.21 for strong Amihan events.

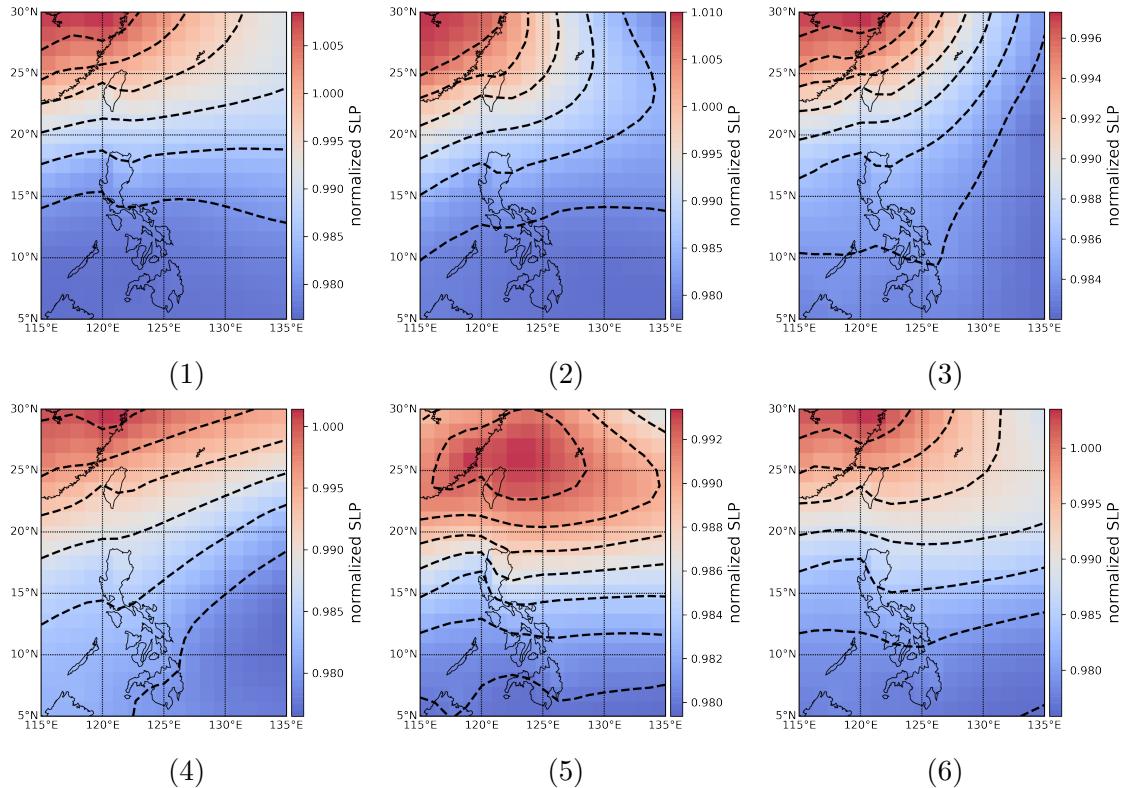


Figure 5.20: Normalized SLP archetypes obtained using hyperspectral unmixing for strong Amihan

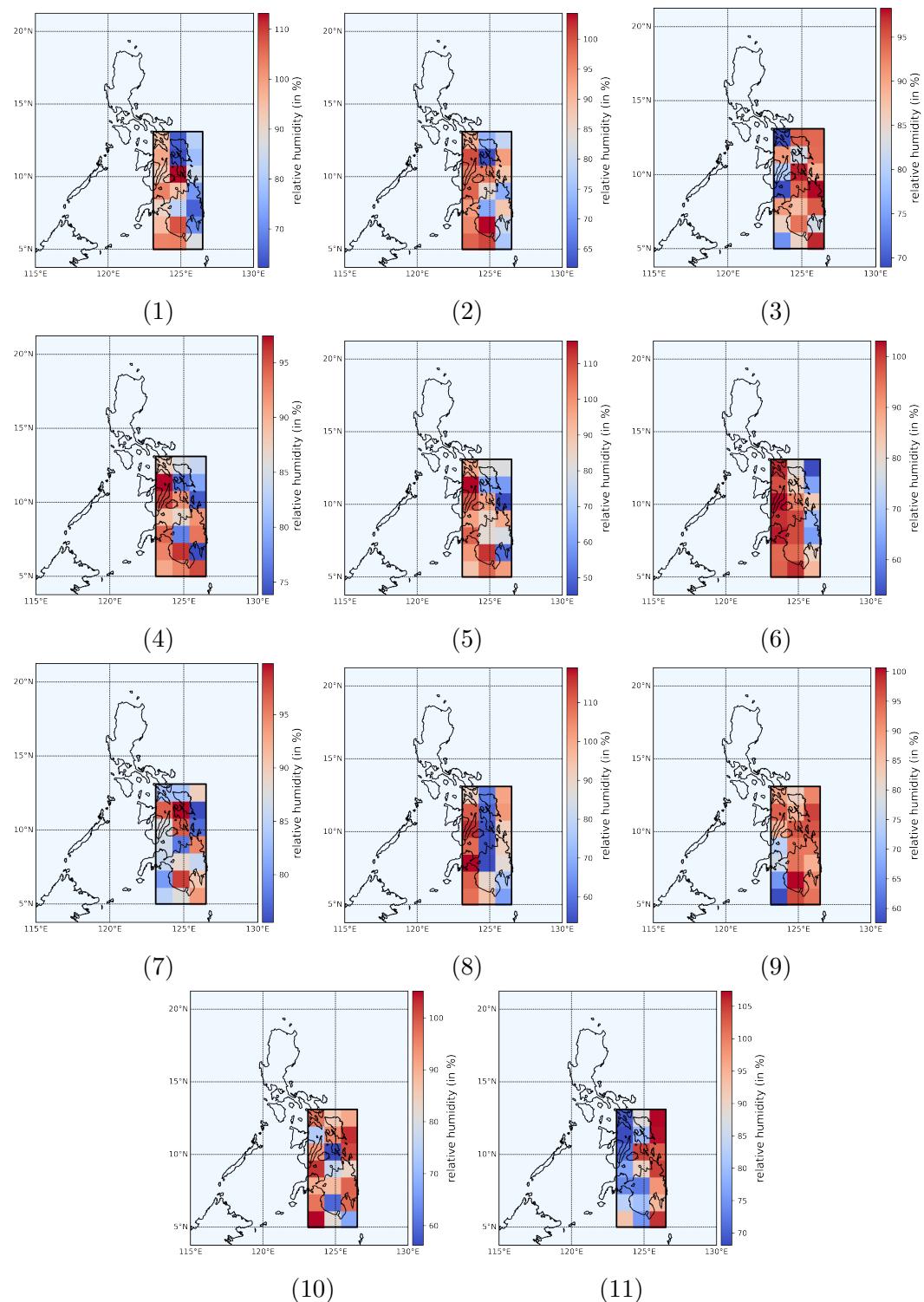


Figure 5.21: Relative humidity (in %) archetypes obtained using successive hyperspectral unmixing for strong Amihan season

5.5.2 Rainfall archetypes obtained through hyperspectral unmixing

Hyperspectral unmixing obtained 11 rainfall archetypes shown in Figure 5.22 corresponding to the 11 RH archetypes for strong Amihan events. Archetypes 1-4 describe very little rain over Visayas and Mindanao. The remaining seven describe heavy rainfall over Visayas and Mindanao in varying degrees of intensity and spatial distribution. Archetypes 5-8 describe heavy rainfall concentrated over Eastern Visayas, archetype 9 describes heavy rainfall over Bicol region, archetype 10 describes heavy rainfall covering almost all of Mindanao and parts of eastern Visayas, and archetype 11 describes heavy rainfall covering Zambales, BARMM, and SOCCKSARGEN.

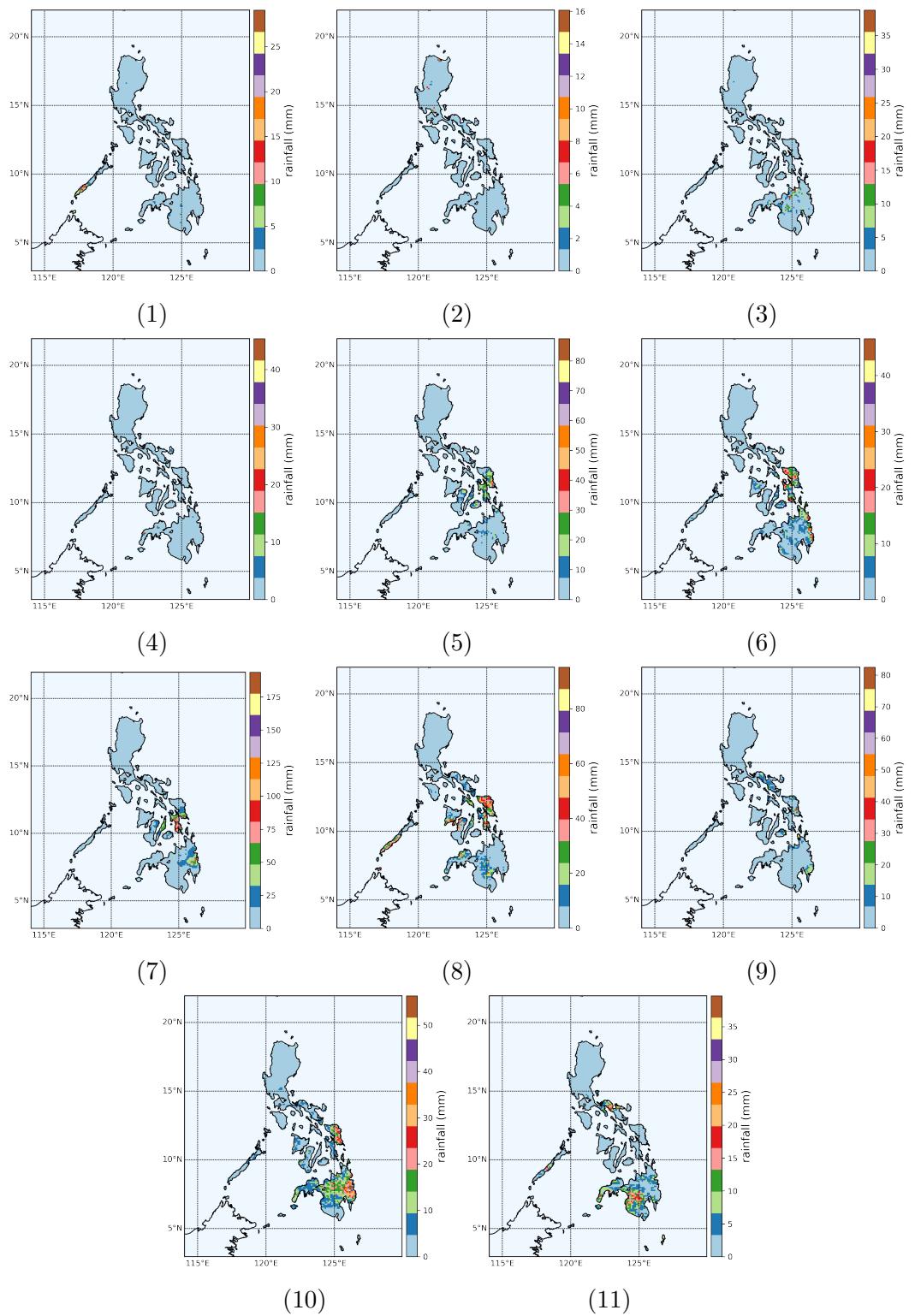


Figure 5.22: Rainfall (in mm) archetypes obtained using successive hyperspectral unmixing for strong Amihan

5.5.3 Forecast accuracy assessment scores

The mean assessment scores obtained from the rainfall forecasts of the test set using hyperspectral unmixing and direct correlation given different parameters are displayed in Table 5.8. Both hyperspectral unmixing and direct correlation can only correctly forecast heavy rainfall around a third of the time from strong Amihan events' 0.348 and 0.331 TS when checking for heavy rainfall (threshold 20 mm). The FAR at 0.250 and 0.029 for strong Amihan events are significantly lower than strong Habagat events at 0.287 and 0.362, so hyperspectral unmixing is less likely to overestimate rainfall during strong Amihan. On the other hand, direct correlation obtained an FAR at 0.459 for strong Amihan events, which is much higher than strong Habagat events at 0.257 when checking for heavy rainfall. Thus, direct correlation is more likely to overestimate strong Amihan rainfall than strong Habagat rainfall. The PODs for both hyperspectral unmixing and direct correlation remain close to 1.000 when checking for the presence of rain (threshold 0.5 mm). However, direct correlation obtains a higher POD at 0.558 than hyperspectral unmixing at 0.354 when checking for heavy rainfall (threshold 20 mm).

Table 5.8: Summary of mean assessment scores of all strong Amihan events testing days for hyperspectral unmixing and direct correlation with different heavy rain thresholds

method	threshold (mm)	TS	BIAS	POD	FAR
hyperspectral unmixing	0.5	0.749	1.42	0.997	0.250
	20	0.348	0.375	0.354	0.029
direct correlation	0.5	0.758	1.28	0.966	0.222
	20	0.331	1.26	0.558	0.459

5.5.4 “Best” rainfall forecast using hyperspectral unmixing

The testing day January 23, 2019 was considered the “best” rainfall forecast for strong Amihan events using hyperspectral unmixing, since hyperspectral unmixing captured the heavy rainfall over central Mindanao and eastern Visayas. Figure 5.23 compares the composite SLP maps obtained by hyperspectral unmixing and direct correlation to the actual SLP maps for strong Habagat testing day January 23, 2019. On the other hand, Figure 5.24 compares the rainfall forecast obtained by hyperspectral unmixing

and direct correlation to the actual rainfall for the same date.

Figure 5.23a shows the location of the high pressure ridge over mainland China extending over East China Sea. The ridge generally matches between the composite and actual, but the actual high is stronger than that of hyperspectral unmixing. Looking at the archetype SLPs in Fig. 5.20, the fifth archetype matches the actual SLP closely (Fig 5.23), but the attributed weight of 0.32 is not enough for the strong high pressure to manifest in the composite. However, the rainfall forecast for January 23, 2019 in Fig. 5.24a still captures the regions of heavy rainfall over Samar and Mindanao. As a result, hyperspectral unmixing has higher TSs at 0.500 and PODs at 0.500, and lower FARs at 0.000 when compared to the mean TS at 0.348, POD at 0.354, and FAR at 0.029 from Table 5.8. On the other hand, direct correlation seemed to better capture the actual continental high pressure over China than hyperspectral unmixing. However, direct correlation only captured the heavy rainfall over eastern Visayas and missed some of the heavy rainfall over Mindanao as shown in Fig. 5.24c.

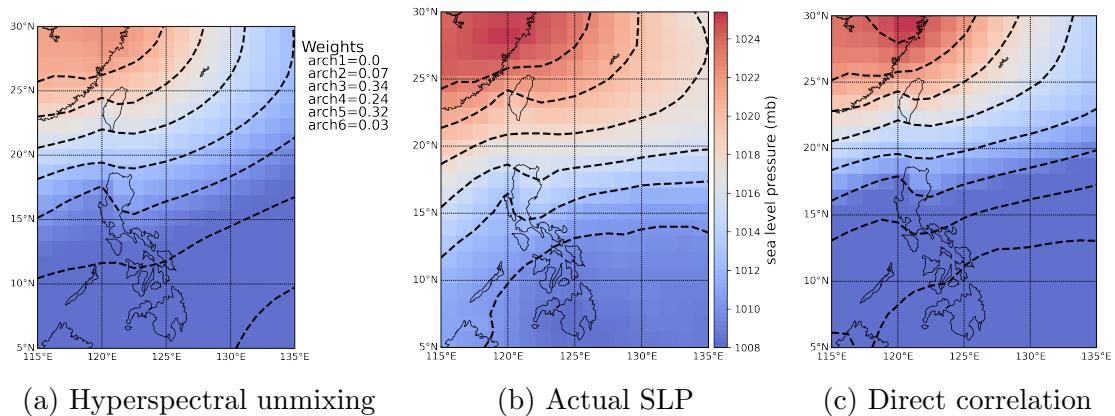


Figure 5.23: Composite SLP (in mb) of “best” rainfall forecast for strong Amihan on 01/23/2019 using analog forecasting with hyperspectral unmixing and direct correlation

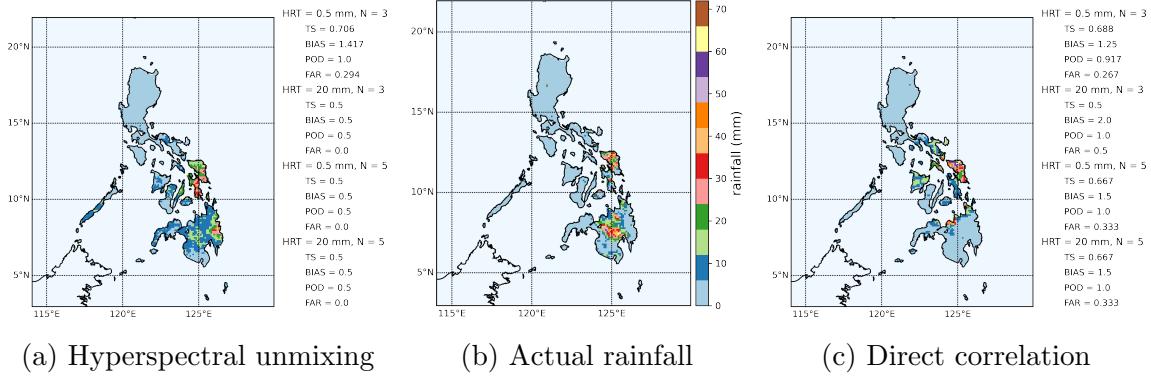


Figure 5.24: “Best” rainfall forecast for strong Amihan on 01/23/2019 using analog forecasting with hyperspectral unmixing and direct correlation

5.5.5 “Worst” rainfall forecast using hyperspectral unmixing

The testing day January 31, 2020 was chosen as the “worst” rainfall forecast for strong Amihan events using hyperspectral unmixing. The hyperspectral unmixing forecast has 0.000 TSs and 0.000 POD, lower than the mean TS at 0.348 and POD at 0.354 from Table 5.8. Figure 5.25 compares the composite SLP maps obtained by hyperspectral unmixing and direct correlation to the actual SLP maps for strong Habagat testing day January 31, 2020. It is very apparent that there is a weak pressure gradient across the Philippines and the Philippine Sea in the actual SLP, but both hyperspectral unmixing and direct correlation SLPs have higher gradients across the Luzon Strait region. This leads to higher forecast rainfall for both methods compared to the actual as shown in Fig. 5.26.

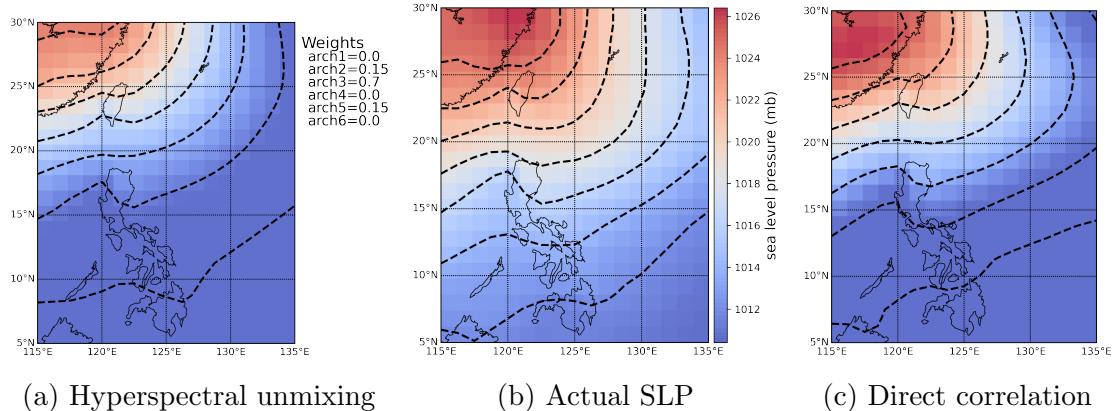


Figure 5.25: Composite SLP (in mb) of “worst” rainfall forecast for strong Amihan on 01/31/2020 using analog forecasting with hyperspectral unmixing and direct correlation

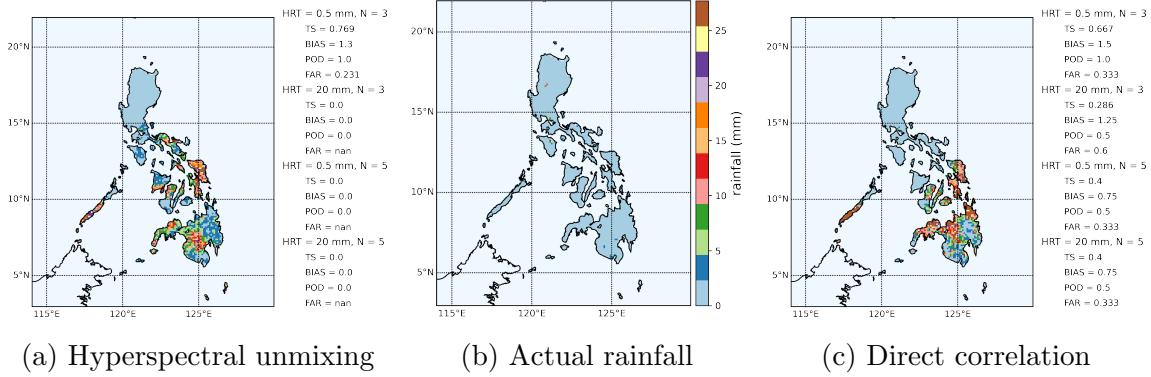


Figure 5.26: “Worst” rainfall forecast for strong Amihan on 01/31/2020 using analog forecasting with hyperspectral unmixing and direct correlation

Both hyperspectral unmixing and direct correlation were able to capture the actual continental high pressure over China for January 31, 2020, but both overestimated the rainfall at ≥ 5 mm, instead of at ≤ 5 mm. There is very little rainfall over the eastern regions of the Philippines during January 31, 2020 despite having an extended and strong continental high pressure, the criteria we’re looking for in a strong Amihan event. Both hyperspectral unmixing and direct correlation cannot account for cases such as this.

5.6 Assessment of rainfall forecasts using hyperspectral unmixing on strong monsoon events

Rainfall forecasts that have $TS \geq 0.25$ when checking for heavy rainfall ($HRT = 30\text{mm}$ for Habagat, $HRT = 20\text{mm}$ for Amihan) are deemed acceptable based on the assessment score of model rain forecast from the study of Wang et al., 2021. Tables 5.9 and 5.10 show the assessment scores of all strong Habagat and Amihan test rainfall forecasts using hyperspectral unmixing. The dates are arranged from the highest to lowest TS when checking for heavy rainfall. Here, we compare the TS when checking for heavy rain versus the standard value 0.25 to determine the number of acceptable forecasts. For strong Habagat, 12 out of 13 of the forecasts reach the standard, while for strong Amihan, 15 out of 17 of the forecasts reach the standard. Based on the methods and parameters presented in this study, hyperspectral unmixing can achieve acceptable predictions at about 90% of its forecasts for both strong Habagat and Amihan rainfall. The acceptable test forecasts of hyperspectral unmixing show potential in augmenting our current forecasts of monsoon rains with a computationally

cheap and fast method through analog forecasting. However, note that at the current point the methods stand, uncertainty and inaccuracy still persists in about 10% of the test forecasts. Even the acceptable test forecasts also entail uncertainty in the correct hits with PODs less than 1.0, and in the false alarms with FARs greater than 0.

Table 5.9: Assessment scores of all strong Habagat rainfall test forecasts for hyperspectral unmixing for different rain thresholds

date	threshold 0.5 mm				threshold 30 mm			
	TS	BIAS	POD	FAR	TS	BIAS	POD	FAR
20040629	0.71	1.42	1	0.29	1	1	1	0
20010925	0.6	1.67	1	0.4	1	1	1	0
20130819	0.89	1.12	1	0.11	0.69	1.25	0.92	0.27
20060714	0.88	1.13	1	0.12	0.64	1.57	1	0.36
20120807	0.86	1.17	1	0.14	0.59	0.93	0.71	0.23
20150717	0.57	1.77	1	0.43	0.55	1.83	1	0.45
20140806	0.75	1.33	1	0.25	0.5	1.5	0.83	0.44
20040828	0.59	1.7	1	0.41	0.5	0.5	0.5	0
20080727	0.8	1.25	1	0.2	0.38	0.57	0.43	0.25
20140802	0.77	1.29	1	0.23	0.36	2.75	1	0.64
20140801	0.79	1.26	1	0.21	0.36	2.8	1	0.64
20130920	0.56	1.8	1	0.44	0.33	1.67	0.67	0.6
20040824	0.5	2	1	0.5	0.18	5.5	1	0.82

Table 5.10: Assessment scores of all strong Amihan rainfall test forecasts for hyperspectral unmixing for different rain thresholds

date	threshold 0.5 mm				threshold 20 mm			
	TS	BIAS	POD	FAR	TS	BIAS	POD	FAR
20151218	0.65	1.55	1	0.35	1	1	1	0
20111224	0.5	2	1	0.5	1	1	1	0
20120327	0.81	1.23	1	0.19	0.67	0.67	0.67	0
20120326	1	1	1	0	0.6	0.6	0.6	0
20080213	0.87	1.15	1	0.13	0.57	0.57	0.57	0
20110130	0.88	1.13	1	0.12	0.5	0.5	0.5	0
20100116	0.85	1.18	1	0.15	0.5	0.5	0.5	0
20190123	0.65	1.33	0.92	0.31	0.5	0.5	0.5	0
20021109	0.82	1.22	1	0.18	0.5	0.5	0.5	0
20090113	0.73	1.36	1	0.27	0.5	0.5	0.5	0
20180113	0.76	1.31	1	0.24	0.4	0.4	0.4	0
20090114	0.64	1.56	1	0.36	0.33	1	0.5	0.5
20201130	1	1	1	0	0.33	0.33	0.33	0
20161101	0.88	1.14	1	0.13	0.29	0.29	0.29	0
20151217	0.81	1.23	1	0.19	0.25	0.25	0.25	0
20130303	0.75	1.33	1	0.25	0.2	0.2	0.2	0
20020102	1	1	1	0	0.2	0.2	0.2	0

Chapter 6

Conclusions and Recommendations

This work presents the use of the hyperspectral unmixing method in the selection of analog environmental patterns using SLP and RH to produce a composite weighted-ensemble analog rainfall forecasts for strong monsoon rainfall in the Philippines. To the author's knowledge, this is the first study to utilize this method in analog rainfall forecasting. The results show the initial frameworks of applying hyperspectral unmixing to weather forecasting.

The hyperspectral unmixing was initially applied to a training dataset which obtained the optimal domain size to produce rainfall archetypes characterized by strong monsoon events. The optimal domains are bounded by ($5 - 30^{\circ}\text{N}$, $115 - 135^{\circ}\text{E}$) and ($5 - 22.5^{\circ}\text{N}$, $111.25 - 128.75^{\circ}\text{E}$) for strong Amihan and Habagat events, respectively. These domains are the smallest possible domains that still encompass the Philippines and the parts of the monsoon trough to the northwest of the Philippines, and a continental high pressure to the north that determine strong Habagat and strong Amihan winds, respectively. For comparison, we also used the direct correlation method in the selection of past analog weather. The method provides comparable rainfall forecasting accuracy scores with direct correlation for both strong Amihan and Habagat events. Analog rainfall forecast of the testing phase during strong Habagat rainfall events yield a mean TS of 0.543, a mean POD of 0.851, and a mean FAR of 0.362. For strong Amihan rainfall events, analog forecasts have a mean TS of 0.348, mean POD of 0.354, and mean FAR of 0.029. The mean TSs of the testing phase for both strong Amihan and Habagat reach the standard value of 0.25. Individually, 12 out of 13, and 15 out of 17 of strong Habagat and Amihan test forecasts, respectively, reach the standard.

The "best" strong Habagat rainfall forecasts were able to capture the expected

heavy rainfall across western Luzon, especially over Ilocos, Pangasinan, and Zambales. Conversely, the “best” strong Amihan rainfall forecasts were able to capture the expected heavy rainfall over east Visayas and Mindanao, especially Samar, Leyte, and CARAGA. However, both hyperspectral unmixing and direct correlation were unable to accurately forecast heavy direct rainfall from TCs. Both also overestimated rain during days with strong Amihan, but not necessarily heavy rainfall along east Philippines.

The method shows potential in forecasting if it can be improved further. Analog forecasting is ultimately a statistical method, so intuitively, increasing the number of data points also increases its accuracy. Possibly, the temporal scope of the strong monsoon events can be extended to years before 2000 to increase the number of data points considered for the training and test set. Increasing or decreasing the spatial resolution of the SLP, RH, and rainfall could also affect the accuracy of forecasts. Determining the optimal spatial resolution of meteorological parameters can be investigated in future studies. The methods could also be extended to other meteorological parameters besides SLP and RH, such as geopotential height or sea surface temperature. Different combinations of meteorological parameters, besides SLP and RH, can be used to compare historical rainfall days in analog forecasting, which can determine the environmental conditions that best describe strong monsoon events. The study of Wang et al., 2021 investigated different rainfall assessment thresholds between 0.1 mm to 150 mm in increments of tens. Future studies utilizing the methods we presented can also investigate how varying the rainfall threshold affects the resulting assessment scores.

We demonstrated in this study that hyperspectral unmixing can perform analog rainfall forecasting of extreme weather events, specifically strong monsoon events. Thus, there is value in investigating the opposite—weak monsoon events that can cause drought. There is also value in investigating the days leading up to or after a strong monsoon day. Through this, we can explore the evolution in SLP and RH, and consequently the rainfall forecasts, a few days before and/or after a strong monsoon day. Recall also that in this study, only the associated weights for RH from hyperspectral unmixing are utilized in formulating rainfall forecasts. The associated weights for SLP are excluded. Future studies can investigate the value of including the weights associated with the first stage of analog forecasting in predicting the final

rainfall forecasts. Lastly, future studies can also modify and improve the methods such that hyperspectral unmixing and analog forecasting can overcome the high variability of SLP and RH, which prevents the methods from considering all monsoon days. We would expect future studies to still observe the same SLP and rainfall archetypes presented in this study, but also archetypes corresponding to transition months. The archetypes should look something like the wind and rainfall distributions shown in Fig. 2.1.

Bibliography

- Akasaka, I., Morishima, W., & Mikami, T. (2007). Seasonal march and its spatial difference of rainfall in the Philippines. *International Journal of Climatology*, 27(6), 715–725. [https://doi.org/https://doi.org/10.1002/joc.1428](https://doi.org/10.1002/joc.1428) <https://doi.org/10.1002/joc.1428>
- Asuncion, J., & Jose, A. (1980). A study of the characteristics of the northeast and southwest monsoons in the philippines. *National Research Council of the Philippines Assisted Project*.
- Atkins, M. J. (1974). The objective analysis of relative humidity. *Tellus*, 26(6), 663–671.
- Bagtasa, G. (2019). Enhancement of Summer Monsoon Rainfall by Tropical Cyclones in Northwestern PhilippinesEnhancement of Summer Monsoon Rainfall by Tropical Cyclones in Northwestern Philippines. *Journal of the Meteorological Society of Japan. Ser. II*, 97. <https://doi.org/10.2151/jmsj.2019-052>
- Bagtasa, G. (2020). Influence of Madden–Julian Oscillation on the Intraseasonal Variability of Summer and Winter Monsoon Rainfall in the Philippines. *Journal of Climate*, 33, 1–38. <https://doi.org/10.1175/JCLI-D-20-0305.1>
- Bagtasa, G. (2021). Analog forecasting of tropical cyclone rainfall in the Philippines. *Weather and Climate Extremes*, 32, 100323. <https://doi.org/https://doi.org/10.1016/j.wace.2021.100323>
- Basu, A., & Halder, A. (2017). Importance of Numerical Weather Prediction in Variable Renewable Energy Forecast.
- Ben Daoud, A., Sauquet, E., Bontron, G., Obled, C., & Lang, M. (2016). Daily quantitative precipitation forecasts based on the analogue method: Improvements and application to a French large river basin. *Atmospheric Research*, 169, 147–159. <https://doi.org/https://doi.org/10.1016/j.atmosres.2015.09.015>
- Bioucas-Dias, J. M., & Figueiredo, M. A. T. (2010). Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing. *2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, 1–4. <https://doi.org/10.1109/WHISPERS.2010.5594963>
- Bioucas-Dias, J. M., & Nascimento, J. M. P. (2008). Hyperspectral Subspace Identification. *IEEE Transactions on Geoscience and Remote Sensing*, 46(8), 2435–2445. <https://doi.org/10.1109/TGRS.2008.918089>
- Bioucas-Dias, J. M., Plaza, A., Dobigeon, N., Parente, M., Du, Q., Gader, P., & Chanussot, J. (2012). Hyperspectral Unmixing Overview: Geometrical, Statistical, and Sparse Regression-Based Approaches. *IEEE Journal of Selected*

- Topics in Applied Earth Observations and Remote Sensing*, 5(2), 354–379. <https://doi.org/10.1109/JSTARS.2012.2194696>
- Cayanan, E. O., Chen, T.-C., Argete, J., Yen, M., & Nilo, P. D. (2011). The effect of tropical cyclones on southwest monsoon rainfall in the Philippines. *Journal of the Meteorological Society of Japan*, 89, 123–139.
- Chan, S. T., Fong, C. H., & Chan, M. Y. (2014). An analog forecast system for precipitation forecasting. *Proc. 28th Guangdong–Hong Kong–Macao Seminar on Meteorological Science and Technology*.
- Chang, C.-P., Wang, Z., & Hendon, H. (2006). The asian winter monsoon. In *The asian monsoon* (pp. 89–127). Springer.
- Chattopadhyay, A., Nabizadeh, E., & Hassanzadeh, P. (2020). Analog Forecasting of Extreme-Causing Weather Patterns Using Deep Learning. *Journal of Advances in Modeling Earth Systems*, 12(2), e2019MS001958. <https://doi.org/https://doi.org/10.1029/2019MS001958>
- Cheng, Y.-B., Ustin, S. L., Riaño, D., & Vanderbilt, V. C. (2008). Water content estimation from hyperspectral images and MODIS indexes in Southeastern Arizona. *Remote Sensing of Environment*, 112(2), 363–374. <https://doi.org/https://doi.org/10.1016/j.rse.2007.01.023>
- Soil Moisture Experiments 2004 (SMEX04) Special Issue
- Clark, G. N., G. C., Rockwell, B. W., Clark, R. N., Cunningham, C. G., Sutley, S. J., Gent, C. A., Mcdougal, R. R., Livo, K. E., & Kokaly, R. F. (2000). Mineral mapping in the marysvale volcanic field, utah using aviris data.
- de Castro, C. M. (2020). *Extracting philippine voting patterns through hyperspectral unmixing*. National Institute of Physics.
- de Castro, C. M. M., & Lim, M. T. (2020). Extracting voting patterns across three Philippine senate elections using hyperspectral unmixing. *Proceedings of the Samahang Pisika ng Pilipinas*, 38, SPP–2020–2A–06. <https://proceedings.spp-online.org/article/view/SPP-2020-2A-06>
- Galacgac, E. S., & Balisacan, C. M. (2009). Traditional weather forecasting for sustainable agroforestry practices in Ilocos Norte Province, Philippines. *Forest Ecology and Management*, 257(10), 2044–2053. <https://doi.org/https://doi.org/10.1016/j.foreco.2009.01.002>
- Traditional forest-related knowledge in Asia
- Gordon, N., & Shaykewich, J. (2000). *Guidelines for Performance Assessment of Public Weather Services*.
- Grifoll, M., Navarro, J., Pallares, E., Ràfols, L., Infantes, M., & Palomares, A. (2016). Ocean–atmosphere–wave characterisation of a wind jet (Ebro shelf, NW Mediterranean Sea). *Nonlinear Processes in Geophysics*, 23, 143–158. <https://doi.org/10.5194/npg-23-143-2016>
- Hart, Y., Sheftel, H., Hausser, J., Szekely, P., Ben-Moshe, N. B., Korem, Y., Tendler, A., Mayo, A. E., & Alon, U. (2015). Inferring biological tasks using Pareto analysis of high-dimensional data. *Nature Methods*, 12(3), 233–235. <https://doi.org/10.1038/nmeth.3254>
- Horton, P., Obled, C., & Jaboyedoff, M. (2017). The analogue method for precipitation prediction: finding better analogue situations at a sub-daily time step.

- Hydrology and Earth System Sciences*, 21(7), 3307–3323. <https://doi.org/10.5194/hess-21-3307-2017>
- Kanning, M., Kühling, I., Trautz, D., & Jarmer, T. (2018). High-resolution uav-based hyperspectral imagery for lai and chlorophyll estimations from wheat for yield prediction. *Remote Sensing*, 10, 2000. <https://doi.org/10.3390/rs10122000>
- Kim, H.-J., Moon, I.-J., & Kim, M. (2020). Statistical prediction of typhoon-induced accumulated rainfall over the Korean Peninsula based on storm and rainfall data. *Meteorological Applications*, 27(1), e1853. <https://doi.org/https://doi.org/10.1002/met.1853>
- Kobayashi, S., OTA, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., & Takakashi, K. (2015). The JRA-55 Reanalysis: General Specifications and Basic Characteristics. *Journal of the Meteorological Society of Japan. Ser. II*, 93(1), 5–48. <https://doi.org/10.2151/jmsj.2015-001>
- Lange, K., Hunter, D. R., & Yang, I. (2000). Optimization Transfer Using Surrogate Objective Functions. *Journal of Computational and Graphical Statistics*, 9(1), 1–20. <https://doi.org/10.2307/1390605>
- Lansigan, F. P., de los Santos, W. L., & Coladilla, J. O. (2000). Agronomic impacts of climate variability on rice production in the Philippines. *Agriculture, Ecosystems Environment*, 82(1), 129–137. [https://doi.org/https://doi.org/10.1016/S0167-8809\(00\)00222-X](https://doi.org/https://doi.org/10.1016/S0167-8809(00)00222-X)
- Li, J., Agathos, A., Zaharie, D., Bioucas-Dias, J. M., Plaza, A., & Li, X. (2015). Minimum Volume Simplex Analysis: A Fast Algorithm for Linear Hyperspectral Unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 53(9), 5067–5082. <https://doi.org/10.1109/TGRS.2015.2417162>
- Matsumoto, J., Olaguera, L. M., Nguyen-Le, D., Kubota, H., & Villafuerte II, M. (2020). Climatological seasonal changes of wind and rainfall in the Philippines. *International Journal of Climatology*, 40. <https://doi.org/10.1002/joc.6492>
- Mladenov, M., Penchev, S., & Dejanov, M. (2015). Complex assessment of the quality of foodstuffs through the analysis of visual images, spectrophotometric and hyperspectral characteristics. *IFAC-PapersOnLine*, 48(24), 60–65. <https://doi.org/https://doi.org/10.1016/j.ifacol.2015.12.057>
16th IFAC Conference on Technology, Culture and International Stability
TECIS 2015
- Murphree, T., & Van Den Dool, H. (1988). Calculating tropical winds from time mean sea level pressure fields. *Journal of the atmospheric sciences*, 45(21), 3269–3282.
- Nascimento, J. M. P., & Dias, J. M. B. (2005). Vertex component analysis: a fast algorithm to unmix hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(4), 898–910. <https://doi.org/10.1109/TGRS.2005.844293>
- Nocedal, J., & Wright, S. (2000). *Numerical Optimization*, 2nd ed. Springer.
- Ringnér, M. (2008). What is principal component analysis? *Nature biotechnology*, 26(3), 303–304. <https://doi.org/10.1038/nbt0308-303>
- Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press. <https://books.google.com.ph/books?id=GV6YDwAAQBAJ>

- Wang, Y., He, J., Chen, Y., & Min, J. (2021). The Potential Impact of Assimilating Synthetic Microwave Radiances Onboard a Future Geostationary Satellite on the Prediction of Typhoon Lekima Using the WRF Model. <https://doi.org/10.3390/rs13050886>
- Webster, P. J., Magaña, V. O., Palmer, T. N., Shukla, J., Tomas, R. A., Yanai, M., & Yasunari, T. (1998). *Monsoons: Processes, predictability, and the prospects for prediction* (tech. rep. No. C7). <https://doi.org/10.1029/97JC02719>
- Wetterhall, F., Halldin, S., & Xu, C.-y. (2005). Statistical precipitation downscaling in central Sweden with the analogue method. *Journal of Hydrology*, 306(1), 174–190. <https://doi.org/https://doi.org/10.1016/j.jhydrol.2004.09.008>
- Xavier, P. K., & Goswami, B. N. (2007). An Analog Method for Real-Time Forecasting of Summer Monsoon Subseasonal Variability. *Monthly Weather Review*, 135(12), 4149–4160. <https://doi.org/10.1175/2007MWR1854.1>
- Xu, R. X., Allen, D. W., Huang, J., Gnyawali, S., Melvin, J., Elgarably, H., Gordillo, G., Huang, K., Bergdall, V., Litorja, M., Rice, J. P., Hwang, J., & Sen, C. K. (2012). Developing digital tissue phantoms for hyperspectral imaging of ischemic wounds. *Biomedical optics express*, 3(6), 1433–1445. <https://doi.org/10.1364/boe.3.001433>
- Zhao, Z., & Giannakis, D. (2016). Analog forecasting with dynamics-adapted kernels. *Nonlinearity*, 29(9), 2888–2939. <https://doi.org/10.1088/0951-7715/29/9/2888>