THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National Polytechnique de Toulouse (Toulouse INP)

Discipline ou spécialité :

Surfaces Interfaces Continentales Hydrologie

Présentée et soutenue par :

M. ANTHONY MUCIA

le mardi 19 octobre 2021

Titre :

Assimilation de données satellitaires pour le suivi et la prévision des sécheresses agricoles et des ressources en eau

---

Ecole doctorale :

Sciences de l'Univers de l'Environnement et de l'Espace (SDU2E)

Unité de recherche :

Groupe de Météorologie de Moyenne Echelle (CNRM-GMME)

Directeur(s) de Thèse :

M. CLÉMENT ALBERGEL

M. JEAN-CHRISTOPHE CALVET

Rapporteurs :

M. JEAN-PIERRE WIGNERON, INRA VILLENAVE D'ORNON

MME CATHERINE OTTLE, CNRS PARIS

Membre(s) du jury :

M. LIONEL JARLAN, CESBIO, Président

M. CLÉMENT ALBERGEL, CENTRE NATIONAL DE RECHERCHES METEO, Membre

M. GIANPAOLO BALSAMO, ECMWF READING, Membre

M. JEAN-CHRISTOPHE CALVET, CENTRE NATIONAL DE RECHERCHES METEO, Membre

MME FABIENNE MAIGNAN, CNRS PARIS, Membre

Assimilation de Données Satellitaires pour le Suivi et la Prévision des Sécheresses Agricoles et des Ressources en Eau

Résumé

Le suivi et la prévision des sècheresses concernent divers porteurs d'enjeux. Le suivi de l'étendue, de la gravité et de l'impact des sécheresses est nécessaire pour atténuer leurs effets. Les deux approches les plus utilisées pour le suivi des sécheresses sont la modélisation numérique et l'utilisation de données satellitaires. Les modèles représentent les processus et sont capables de simuler les échanges d'énergie et d'eau à la surface. Ils peuvent néanmoins souffrir d'une représentation trop simpliste de ces processus, de conditions initiales incorrectes et de défauts du forçage atmosphérique. Les données satellitaires permettent d'accéder à de nombreuses variables à l'échelle mondiale, de manière répétée dans le temps et à des échelles spatiales de plus en plus précises. Elles peuvent cependant être discontinues dans le temps et l'espace et toutes les variables des surfaces terrestres ne sont pas observables depuis l'espace. De plus elles sont représentatives d'un instant précis, et contrairement aux modèles numériques, n'offrent pas la possibilité de faire de la prévision. Afin d'améliorer le suivi des sécheresses, il est possible de combiner les modèles numériques et les observations satellitaires en utilisant des techniques d'assimilation de données. L'assimilation permet d'obtenir de meilleures conditions initiales et par conséquent de meilleures prévisions. Ce travail de thèse a pour objectif d'étudier l'impact de conditions de surface améliorées par l'assimilation d'observations satellitaires sur la prévisions des épisodes de sécheresses et leurs impacts sur l'agriculture et les ressources en eau. Le système d'assimilation de données pour les surfaces continentales (LDAS-Monde) développé au CNRM est utilisé. Des observations satellitaires sont assimilées dans le modèle de surface ISBA dans une série d'expériences sur les USA ainsi que sur plusieurs sous-domaines. La capacité du système à représenter et prévoir les variables de surface liées à la végétation et aux sécheresses est évaluée. L'impact de l'assimilation de trois variables différentes est analysé : l'indice de surface foliaire (« LAI »), l'humidité superficielle du sol (« SSM ») et l'épaisseur optique de la végétation dans le domaine spectral des micro-ondes (« VOD »). L'impact de l'assimilation est analysé grâce à l'utilisation de données indépendantes d'évapotranspiration, de production primaire brute de la végétation et d'humidité du sol. Sur l'état du Nebraska, le système LDAS-Monde permet de représenter la variabilité interannuelle du LAI mais aussi des rendements agricoles du maïs, y compris lors d'épisodes de sécheresse prolongés. LDAS-Monde a été amélioré et pourvu d'une capacité de prévision à courte et moyenne échéance (15 jours) en utilisant les prévisions atmosphériques du CEPMMT (ou « ECMWF »). La capacité du système à prévoir les variables de surfaces jusqu'à 15 jours d'échéances a été montrée, sur une période de deux ans. L'importance des conditions initiales sur la qualité des prévisions a été mise en évidence. Une série d'expériences d'assimilation a été réalisée dans laquelle le VOD a été utilisé comme proxy du LAI. Cela améliore beaucoup l'échantillonnage temporel car le VOD est disponible plus fréquemment que le LAI. Après une comparaison approfondie des produits de LAI, différentes expériences assimilant le LAI, le VOD et le SSM, de manière conjointe ou séparée ont été réalisées. Ces expériences confirment l'apport de l'assimilation conjointe d'observations liées à la végétation et de l'humidité superficielle du sol. L'amélioration des conditions initiales est ensuite utilisée dans une étude de cas prospective sur la mise en place d'une fonction de transfert du système actuel vers un système d'alerte précoce des sécheresses.

Assimilation of Satellite Data for the Monitoring and Prediction of Agricultural Droughts and Water Resources

Abstract

The monitoring and prediction of droughts and its impacts is of utmost importance to stakeholders around the world. Being able to closely track the extent, severity, and impacts of drought events leads to better response and mitigation, while reducing the effects. The two most widespread approaches to do this are using numerical simulations with land surface models (LSMs) and using satellite Earth observations (EOs). Both of these approaches allow for the tracking of drought extent and severity, but both also have drawbacks. While LSMs are able to simulate consistent spatial and temporal fluxes of the land surface, they may suffer from imperfect model physics, poor initial conditions, and from the quality of the atmospheric forcing used as input. EOs are able to monitor variables globally and provide observational data at an unrivaled scale, but they can suffer from temporal gaps in coverage and a limited range of observed variables, as well as not having the ability to forecast. To facilitate better tracking and to allow the prediction of droughts, LSMs can be combined with satellite observations through a data assimilation process, improving their accuracy and providing better initial conditions for forecasts. This thesis uses the land data assimilation system LDAS-Monde developed by CNRM, the research department of the French Meteorological service (Météo-France), to investigate the impact of improving land surface conditions through data assimilation on the potential to forecast drought events and impacts to agriculture and water resources. EOs are assimilated into the ISBA (Interactions between Soil, Biosphere, and Atmosphere) LSM for a number of experiments over the contiguous United States and various sub-domains to analyze the system's capability to represent and forecast land surface variables connected to vegetation and drought. This thesis explores the assimilation of three satellite observation datasets: leaf area index (LAI), surface soil moisture (SSM), and vegetation optical depth (VOD). Results are assessed against independent datasets of evapotranspiration, gross primary production, and soil moisture. Over the U.S. state of Nebraska, it is found that LDAS-Monde is able to represent the inter-annual variability of LAI and corn yield, including the impact of significant drought years, providing a good basis for the system's potential. Using ECMWF fifteen day atmospheric forecasts, LDAS-Monde has been strengthened with forecast capacity. The system proved skillful over the two year experimental period, providing potentially useful forecasts of land surface variables up to two weeks in advance. The use of LDAS-Monde in forecast mode also showed that the initial conditions are critical to accurate surface forecasts. The use of LAI in LDAS-Monde led to strongest impact. Therefore, to test and define improvements to initial conditions with data assimilation, a series of new experiments were made. Assessed among them was the use of VOD as an LAI proxy, which constrains the model far more frequently. Comparisons of the LAI and VOD satellite products over different sub-domains and vegetation types were first performed. Additional experiments with the separate and joint assimilation of vegetation and soil moisture are also of note. These different assimilation scenarios lend proof that when following certain land surface variables, the joint assimilation of vegetation and soil moisture provides more accurate results. Improvements to the initial conditions are then taken into consideration, and a prospective case study is introduced to transition into a drought early warning system.

Acknowledgements

The support and funding of the Make Our Planet Great Again (MOPGA) program, as well as of Météo-France, made this thesis possible, and I am sincerely grateful for their funding of this PhD.

Merci à tous les membres de mon jury de thèse d'avoir donné de temps pour lire, critiquer, et évaluer ce travail.

Il y a trop de personnes que je voudrais remercier pour leur soutien, leur aide, et leur compagnie. Ce fut un privilège et un plaisir de vous connaître tous.

Je tiens tout d'abord à exprimer mes plus vifs remerciements à mes directeurs de thèse, Clément Albergel et Jean-Christophe Calvet, pour leur aide et leurs conseils tout au long de ce processus. Merci pour votre patience et de m'avoir fait confiance pour cette thèse.

Clément, dès mon arrivée à l'aéroport de Toulouse, tu m'as accueilli et aidé à me lancer dans ce tout nouveau pays. Dans les premiers mois de la thèse, tu m'as poussé à devenir un expert de LDAS-Monde, et tu as toujours été disponible pour toute question ou discussion. Des trucs et astuces python, bash et vim aux "nitty-gritty" détails de l'assimilation de données, j'ai tellement appris de toi et j'ai apprécié énormément tes conseils ces dernières années.

Jean-Christophe, merci de m'avoir aidé à trouver un appartement dans la ville et à développer de nombreuses connaissances approfondies dans l'analyse des résultats scientifiques. Ton aide a été inestimable au cours de ce processus de thèse. J'ai appris quelque chose de nouveau tous les jours en parlant avec toi, et je t'en suis infiniment reconnaissant.

Bertrand, ton aide tout au long de ce processus de thèse à été très appréciée. Je pouvais toujours passer ton bureau pour des petites questions sur des problèmes d'HPCs, ou de te demander de traduire une partie de documents administratifs. Même si ton nom ne figurait malheureusement pas parmi mes encadrants, ton aide a rendu cette thèse possible, merci ! Et les nombreux pause cafés et "afterworks" avec toi étaient toujours des points forts ! Tes futurs thésards auront beaucoup de chance de t'avoir comme directeur !

Aux autres membres de l'équipe VEGEO, je ne peux pas exprimer pleinement ma gratitude pour tous les moments partagés. Catherine, merci pour ton assistance technique, les moments aux pauses café, et pour m'avoir accueilli dans ton bureau ces dernièrs mois de travail ! Xavier, ton esprit et ton humour ont constamment illuminé le couloir ! Chloe, Mickael, Yongjun, Remy, et Daniel J, ce fut une joie de passer de nombreuses heures de travail et de déjeuner avec vous ! Moustapha, tu m'as accueilli et m'as aidé à faire la transition vers cette nouvelle vie, et je ne peux pas te remercier suffisamment ! Daniel S., merci pour toute l'aide pour chercher un appartement, me mettre à jour avec les utilitaires d'LDAS-Monde, ou les nombreux échanges de code python ! C'était une joie de partager le bureau avec toi, et merci pour avoir toléré mes marmonnements et éteint le ventilateur quand j'oublierais !

Les nouveaux thésards, Timothée et Adele (et Sophie!), je vous souhaite le meilleur pour vos thèse ces trois années prochaines ! Profitez bien, et tirez le meilleur parti de cette opportunité !

Aux équipes SURFACE et VILLE, les pauses cafés de matin de nos trois équipes sont des moments inoubliables. J'ai toujours attendu avec impatience ces pauses bien méritées presque tous les jours. Elles m'ont donné certains des meilleurs moments pour écouter et parler en français, et je peux leur attribuer une grande partie des améliorations que j'ai apportées à mon français. Merci à tous d'avoir réussi à écouter mes tentatives ratées, et parfois absurdes, dans cette langue assez compliquée, et pour toutes les corrections et réponses à mes questions en français ! Malak et Gaëtan, je vous souhaite le meilleur pour votre dernière année de thèse, et ce fut un plaisir de partager près de deux ans avec vous. Merci aux nombreux collegues de ces deux équipes, les stagiaires, les thésards, les permanents, ou les post-doctorants. Vous êtes trop nombreux pour tous vous nommer, mais chaque expérience avec vous a été positive !

Patrick, Simone, Thibault, Stéphane, et Delphine, être parmi les lève-tôt avait ses privilèges, l'un d'eux étant de partager un café tôt le matin avec vous tous ! Toujours essayer de se rappeler à qui c'était le tour d'acheter le café à la machine était un si bon problème à résoudre. Delphine, tu as toujours été quelqu'un à qui je pouvais m'adresser pour de l'aide, merci pour les nombreuses fois où tu as répondu aux appels téléphoniques pour moi en français, tout en étant heureuse de m'aider. Je ne pouvais pas demander un meilleur groupe de personnes avec qui passer la matinée.

Aaron, what luck that the only other American in the building is just a few doors down from me! Whether it was talking about basketball, football (the 'real' football of course), or the ever-fun subject of American politics, it was a pleasure to be able to connect with you and it was a joy to get to know you. And the hikes with your family are among my most cherished memories here!

A tous les doctorants et stagiaires du PhD lunch, le déjeuner avec vous tous m'a permis de vivre certains de mes plus beaux moments ici au CNRM. Cela m'a donné un environnement social dans lequel apprendre la langue et me faire des amis, ce qui était exactement ce dont j'avais besoin après mon arrivée en France. Zied et Damien, vous étiez tous les deux incroyablement accueillants, et dès le premier déjeuner ensemble, vous vous êtes toujours assuré de m'inclure dans les événements.

William, Sarah, Martin et Nicolas, nos interactions quasi quotidiennes au déjeuner et après (au moins avant COVID !) m'ont aidé à traverser des jours difficiles. Jouer au billard par la suite était une façon incroyable de se détendre et de passer du bon temps. Je suis tellement reconnaissant pour ces moments partagés, et pour ces amitiés que j'ai nouées ici à Toulouse.

To the "English Lunch" group, meeting all of the regulars and enjoying a lunch once a week, speaking in English without shame, was always a highlight! In particular, my fellow anglophones Michael, Ross, and Alistair, it was great to always be at ease in conversation with you, and the numerous evenings we spent socializing were always appreciated. I am glad to have connected with you upon my arrival so that we were able to spend more time together!

Aunt Peggy, you have given me so much support and encouragement throughout this time. In every email and Skype call that we had, you showed so much excitement for me and for my work. Thank you!

And finally, to both my parents, thank you so much for the love and support through these years. Even before making the decision to come to France, you have always supported me and have always been there for me. Your support for this move and to spend three years away from home was needed for this big change. Having a video call every weekend made helped maintain some regularity, especially during the lockdowns. Those calls were always something that I looked forward to, whether it was catching up on the Husker's latest heartbreaking loss or talking about to goings-on in Omaha. Mom, Dad, I am so lucky to have you!

Contents

Résumé **iii**

Abstract **v**

Acknowledgements **vii**

Table of Contents **xiii**

List of Figures **xvii**

List of Tables **xx**

Abbreviations **xxi**

Introduction Générale **1**

General Introduction **5**

1 Scientific Context **9**
1.1 Importance of Monitoring and Forecasting Drought 10
1.1.1 Definitions of Drought 12
1.2 Land-Atmosphere Interactions 14
1.2.1 Energy Balance 14
1.2.2 The Water Cycle 15
1.2.3 Biogeochemical Cycles 16
1.3 Land Surface Models 19
1.4 Earth Observations 22
1.5 Use of Earth Observations in Land Surface Models 24
1.6 Objectives of this Thesis and Work Plan 26

2 Methodology **29**
2.1 SURFEX Modelling Platform 30
2.1.1 ISBA LSM 30
2.1.2 LDAS-Monde 34
2.1.3 Assimilated Observations 36
2.1.4 LDAS-Monde in Forecast Configuration 44
2.2 Assessment Observations 46
2.2.1 Assimilated Observations 46
2.2.2 United States Climate Reference Network 47
2.2.3 ALEXI Evapotranspiration 47
2.2.4 FLUXCOM Gross Primary Production 48
2.3 Experimental Setups 49
2.4 Experiment Analysis and Assessment 50

2.5 Summary of Chapter 2 51

3 Importance of Initial Conditions in Forecasting LSVs 53
3.1 Model Initial Conditions 54
3.2 Mucia et al., 2020 55
3.2.1 Article Introduction and Context 55
3.2.2 Towards Assimilating VOD 79
3.2.3 Mucia et al., 2020 Summary and Conclusions 80
3.3 Summary of Chapter 3 80

4 Improving Initial Conditions Using the LDAS 83
4.1 Vegetation Optical Depth as an LAI Proxy 84
4.1.1 Transforming VOD into an LAI Proxy 84
4.1.2 Comparison over CONUS 84
4.1.3 Anomaly Comparison of Corn Yield to VOD and LAI over Nebraska 87
4.1.4 Relationship Between VOD and LAI over Dominant Vegetation Types 87
4.2 Impact of Assimilating VOD as an LAI proxy 98
4.2.1 Analysis over CONUS using Satellite-Derived Observations . . 101
4.2.2 USCRN Soil Moisture 105
4.3 Impact of Jointly Assimilating Vegetation Variables and SSM 112
4.4 Summary of Chapter 4 114

5 Forecasting Impacts of Extreme Events on LSVS 117
5.1 Drought Forecast Alert System Case Study 118
5.1.1 Requirements of the Case Study 118
Time Frame 118
Selection of Drought Event 119
5.1.2 LDAS-Monde Experimental Parameters 121
Improving Initial Conditions with VOD assimilation 121
Joint Assimilation of Vegetation and Soil Moisture Observations 121
5.2 Linking LDAS-Monde Forecasts to a Drought Indicator 121
5.2.1 Indicators Used in Drought Alerts 123
5.2.2 Proof of Methodology 123
5.2.3 Comparison to the United States Drought Monitor and other Drought Indices 129
5.2.4 Impact of Forecast LSVs on the LDAS-Monde Drought Indicators 130
5.3 Summary of Chapter 5 130

6 Conclusions et Perspectives 133
6.1 Résumé des résultats 133
6.2 Perspectives 136

7 Conclusions and Prospects 139
7.1 Summary of Results 139
7.2 Prospects 142

A Appendix A **145**
A.1 LDAS Experimentation over Nebraska 145
A.1.1 Potential High Resolution LDAS 145
A.1.2 Assessment of LDAS-Monde, ERA5, and ECOCLIMAP over Nebraska 148
A.2 Summary of Appendix A 156

B Appendix B **157**
B.1 LAI versus VODC 157
B.2 LAI versus Matched VODC 161

C Appendix C **165**
C.1 VOD vs LAI Comparison over Subdomains 165

D Appendix D **173**
D.1 Impact of Assimilating VOD as an LAI proxy - Subdomains 173
D.2 Impact of Jointly Assimilating Vegetation Variables and SSM - Subdomains 187

Bibliography **191**

List of Figures

1 Schéma de la motivation et du plan de travail de la thèse 4
2 Diagram of the Motivation and Work Plan of the Thesis 7

1.1 USDM Categories of Drought 12
1.2 Relationships between types of Drought 13
1.3 The Water Cycle 16
1.4 The Carbon Cycle 17
1.5 Basic LSM Components and their Development Timeline 20
1.6 LSM Exchanges, Description and their Development Timeline 21
1.7 Evolution of Land Surface Models 22
1.8 The Electromagnetic Spectrum 23

2.1 The SURFEX Modelling Platform 30
2.2 Differences between ECOCLIMAP-II and ECOCLIMAP-SG 35
2.3 Maps of Number of Observations for LAI and VOD 39
2.4 Time Series of VOD Components 40
2.5 Correlation Analysis of L and X-band VOD to FLUXNET GPP . . . 41
2.6 LAI versus VODX & VODC over Lincoln, Nebraska 41
2.7 VODCA sensors for C, X, and Ku Bands 43
2.8 LDAS-Monde configuration 45
2.9 Map of Assimilated Observations 46
2.10 Map of Domains and Sub-domains 50

3.1 VOD Comparisons to LDAS-Monde OL and SEKF LAI 80

4.1 LAI vs VODC Density Scatter Plot - CONUS 86
4.2 LAI vs VODX Density Scatter Plot - CONUS 86
4.3 Time series of LAI, VOD, and Corn Yield Anomalies over Nebraska . 87
4.4 LAI vs VODX Density Scatter Plot over Dominant Vegetation: ECOCLIMAP-II 91
4.5 LAI vs VODX Density Scatter Plot over Dominant Vegetation: ECOCLIMAP-SG 92
4.6 LAI vs VODX Density Scatter Plot over Dominant Vegetation: ECOCLIMAP-II 96
4.7 LAI vs VODX Density Scatter Plot over Dominant Vegetation: ECOCLIMAP-SG 97
4.8 Monthly Correlations between Various LDAS-Monde Experiments and Satellite Derived Observations over CONUS 101
4.9 Probability Distribution Functions of Correlation Distributions between various LDAS-Monde Experiments and Satellite Derived Observations over CONUS 103
4.10 Average correlations between USCRN stations and limited LDAS-Monde Experiments at 5, 20, 50, and 100cm 107

4.11 Average correlations between USCRN stations and all LDAS-Monde experiments at 5, 20, 50, and 100cm 108
4.12 Probability Distribution Functions of the Distribution of Correlation Differences between OL and Various LDAS-Monde Assimilation Scenarios for USCRN 110
4.13 Maps of Normalized Information Contribution (NIC) Correlation for WG3 (5cm) and WG_20 (20cm) 111
4.14 Maps of Normalized Information Contribution (NIC) Correlation for WG6 (50cm) and WG8 (100cm) 112
4.15 Monthly Correlations between LDAS-Monde Experiments in Joint Assimilation configuration and Satellite Derived Observations over CONUS 113

5.1 Time Series of % Area in Drought over the US North Central Region . 120
5.2 Domain Differences between LDAS-Monde "Midwest" and USDM "North Central" Regions 124
5.3 RZSM Percent Area in Drought Categories 126
5.4 ET Percent Area in Drought Categories 126
5.5 LAI Percent Area in Drought Categories 127
5.6 LAI Percent Area in Drought Categories - SEKF Ranking Against SEKF 127
5.7 LAI, RZSM, and ET Blend Percent Area in Drought Categories 129
5.8 RZSM and ET Blend Percent Area in Drought Categories 129

A.1 Map of ECOCLIMAP-SG C4 Crop Fraction at 0.25°, 0.1°, and 0.01° spatial resolutions 147
A.2 Time Series of LAI and Corn Yield Anomalies over Nebraska 148
A.3 Precipitation for Lincoln, Nebraska 151
A.4 Precipitation for Grand Island, Nebraska 152
A.5 Spatial Comparison of ECOCLIMAP-II and ECOCLIMAP-SG C4 Crop Fraction over Nebraska 154
A.6 USDA Corn Cover Frequency over Nebraska 155
A.7 Time Series of LAI and Corn Yield Anomalies over Nebraska ECOCLIMAP-II vs ECOCLIMAP-SG 155

B.1 LAI vs VODC Heatmap over Dominant Vegetation: ECOCLIMAP-II 159
B.2 LAI vs VODC Heatmap over Dominant Vegetation: ECOCLIMAP-SG 160
B.3 LAI vs Matched VODC Heatmap over Dominant Vegetation: ECOCLIMAP-II 162
B.4 LAI vs Matched VODC Heatmap over Dominant Vegetation: ECOCLIMAP-SG 163

C.1 LAI vs VODC Density Scatter Plot - California 166
C.2 LAI vs VODX Density Scatter Plot- California 166
C.3 LAI vs VODC Density Scatter Plot - Midwest 167
C.4 LAI vs VODX Density Scatter Plot - Midwest 167
C.5 LAI vs VODC Density Scatter Plot- Northeast 169
C.6 LAI vs VODX Density Scatter Plot - Northeast 169
C.7 LAI vs VODC Density Scatter Plot - Southern Plains 170
C.8 LAI vs VODX Density Scatter Plot - Southern Plains 170
C.9 LAI vs VODC Density Scatter Plot - Nebraska 171
C.10 LAI vs VODX Density Scatter Plot - Nebraska 171

D.1 Monthly Correlations between Various LDAS-Monde Experiments and Satellite Derived Observations over California 173
D.2 Probability Distribution Functions of Correlation Distributions between various LDAS-Monde Experiments and Satellite Derived Observations over California 175
D.3 LAI Correlation Differences from OL 176
D.4 Monthly Correlations between Various LDAS-Monde Experiments and Satellite Derived Observations over the Midwest 177
D.5 Probability Distribution Functions of Correlation Distributions between various LDAS-Monde Experiments and Satellite Derived Observations over the Midwest 179
D.6 Monthly Correlations between Various LDAS-Monde Experiments and Satellite Derived Observations over the Northeast 181
D.7 Probability Distribution Functions of Correlation Distributions between various LDAS-Monde Experiments and Satellite Derived Observations over the Northeast 183
D.8 Monthly Correlations between Various LDAS-Monde Experiments and Satellite Derived Observations over the Southern Plains 185
D.9 Probability Distribution Functions of Correlation Distributions between various LDAS-Monde Experiments and Satellite Derived Observations over the Southern Plains 186
D.10 Monthly Correlations between LDAS-Monde Experiments in Joint Assimilation configuration and Satellite Derived Observations over California 188
D.11 Monthly Correlations between LDAS-Monde Experiments in Joint Assimilation configuration and Satellite Derived Observations over the Midwest 189
D.12 Monthly Correlations between LDAS-Monde Experiments in Joint Assimilation configuration and Satellite Derived Observations over the Northeast 189
D.13 Monthly Correlations between LDAS-Monde Experiments in Joint Assimilation configuration and Satellite Derived Observations over the Southern Plains 190

List of Tables

2.1 Details of atmospheric forcing datasets used in this thesis 33
2.2 Details of LDAS-Monde Assimilation Parameters 34
2.3 Details and parameters of experiments used in this thesis 49

4.1 Correlation coefficients between VODCA VODX and VODC versus CGLS LAI and ISBA LAI over the various subdomains 85
4.2 Seasonal and total correlations of VODX versus LAI for ECOCLIMAP-II patches 90
4.3 Seasonal and total correlations of VODX versus LAI for ECOCLIMAP-SG patches 90
4.4 Seasonal and total correlations of Matched VODX versus LAI for ECOCLIMAP-II patches 94
4.5 Seasonal and total correlations of Matched VODX versus LAI for ECOCLIMAP-SG patches 95
4.6 List of experiment names and their assimilated observations analyzed in Chapter 4 100
4.7 Average correlations scores between USCRN in situ soil moisture observations and LDAS-Monde soil moisture at 5, 20, 50, and 100cm depths. Bolded values indicate the highest score at each depth. *WG_20 is a weighted average of WG4 and WG5 in order to directly compare to 20cm observations from USCRN. 108
4.8 Number of degraded (red), neutral (black), and improved (green) USCRN stations after assimilation using NIC R between OL and various LDAS-Monde experiments at 5, 20, 50, and 100cm depths. Stations are considered improved if the NICR is greater than 3, degraded if the score is less than 3, and neutral if it is between -3 and 3. *WG_20 is a weighted average of WG4 and WG5 in order to directly compare to 20cm observations from USCRN. 109

5.1 Atmospheric Forecast Options for Case Study 119

A.1 Correlation Coefficients between Observed, Model, and Analysis LAI and Corn Yield over Nebraska 149
A.2 Correlation Coefficients between Model and Analysis LAI versus Corn Yield and LAI observations over Nebraska for two ECOCLIMAP versions. Values outside parentheses are correlations from runs with ECOCLIMAP-II and values within parentheses are from runs with ECOCLIMAP-SG. 156

B.1 Seasonal and total correlations of VODC versus LAI for ECOCLIMAP-II patches 158

B.2 Seasonal and total correlations of VODC versus LAI for ECOCLIMAP-SG patches 158
B.3 Seasonal and total correlations of Matched VODC versus LAI for ECOCLIMAP-II patches 161
B.4 Seasonal and total correlations of Matched VODC versus LAI for ECOCLIMAP-SG patches 161

List of Abbreviations

AGB	Above Ground Biomass
ALEXI	Atmosphere-Land Exchange Inverse
ASCAT	Advanced Scatterometer
CAM	Crassulacean Acid Metabolism
CCI	Climate Change Initiative
CDF	Cumulative Density Function
CGLS	Copernicus Global Land Services
CNRM	Centre National de Recherches Météorologiques (French National Centre for Meteorological Research)
CONUS	Contiguous United States
CRPS	Continuous Ranked Probability Score
CTRIP	CNRM Total Runoff Integrating Pathways
DA	Data Assimilation
DM or USDM	United States Drought Monitor
ECMWF	European Centre for Medium Range Forecasts
EKF	Extended Kalman Filter
EM	Electromagnetic
EnKF	Ensemble Kalman Filter
ENS	15-day Ensemble Forecast from ECMWF
EO	Earth Observation
ERTS-1	Earth Resources Technology Satellite
ESA	European Space Agency
ET	Evapotranspiration
FAR	False Alarm Rate
FC	Forecast
GCM	Global Climate Model
GEOS	Geostationary Operational Environmental Satellite
GPP	Gross Primary Production
HRES	10-day High Resolution Forecast from ECMWF
HSS	Heidke Skill Score
ISBA	Interactions between Soil, Biosphere, and Atmosphere
LAI	Leaf Area Index
LDAS	Land Data Assimilation System
LPRM	Land Parameter Retrieval Model
LSM	Land Surface Model
LST	Land Surface Temperature
LSV	Land Surface Variable
MERRA	Modern-Era Retrospective Analysis for Research and Applications
MSG	Meteosat Second Generation
NARR	North American Regional Reanalysis

NCEI	National Centers for Environmental Information
NDVI	Normalized Difference Vegetation Index
NIC	Normalized Information Contribution
NOAA	National Oceanic and Atmospheric Administration
NWP	Numerical Weather Prediction
NWS	National Weather Service
OL	Open Loop
PDF	Probability Density Function
PODY	Probability of Detection "Yes" Event
RFI	Radio Frequency Interference
RMSD	Root Mean Squared Difference
RZSM	Root Zone Soil Moisture
SAR	Synthetic Aperture Radar
SEKF	Simplified Extended Kalman Filter
SiB	Simple Biosphere Model
SIF	Solar Induced Chlorophyll Fluorescence
SM	Soil Moisture
SMOSREX	Surface Monitoring of the Soil Reservoir EXperiment
SPI	Standardized Precipitation Index
SSM	Surface Soil Moisture
SURFEX	SURFace EXternalisée
SWI	Soil Wetness Index
TCI	Temperature Condition Index
TEB	Town Energy Balance
TER	Terrestrial Ecosystem Respiration
TIR	Thermal Infrared
TIROS-1	Television Infrared Observation Satellite
TOC	Top of Canopy
TRMM	Tropical Rainfall Measuring Mission
USCRN	United States Climate Reference Network
USDA	United States Department of Agriculture
VCI	Vegetation Condition Index
VOD	Vegetation Optical Depth
VODCA	Vegetation Optical Depth Climate Archive
WRF	Weather Research and Forecasting Model

Introduction Générale

A mesure que le climat évolue vers un état plus chaud et plus instable (GIEC, 2014), il devient d'autant plus important d'améliorer notre compréhension des événements climatiques extrêmes et de leurs impacts sur les variables des surfaces terrestres. Les événements extrêmes tels que les sécheresses, vagues de chaleur, inondations et feux de végétation devraient augmenter en fréquence et/ou en intensité en raison des changements causés par le réchauffement climatique d'origine humaine (GIEC, 2014; GIEC, 2018; Ionita et al., 2017). Les épisodes de sécheresses en particulier font partie des aléas les plus préjudiciables et les plus complexes (Bruce, 1994; Obasi, 1994; Cook et al., 2007; Hagman et al., 1984), avec de forts impacts sur l'agriculture et la sécurité alimentaire, à cause de la baisse du rendement des cultures.

Par conséquent, avoir un système de surveillance précis des sécheresses revêt une importance capitale afin de pouvoir apporter la meilleure réponse possible à leurs impacts. Le suivi des sécheresses inclut leur déclenchement, leur étendue et leur intensité. En plus d'une meilleure atténuation et planification, un meilleur suivi des sécheresses peut permettre de réduire les pertes liées aux ressources agricoles et en eau, et de minimiser les dommages causés aux filières agricoles et à la société (Wilhite, Hayes, and Svoboda, 2000).

L'importance grandissante du suivi des épisodes de sécheresse et de leurs impacts nécessite l'utilisation de toutes les ressources disponibles. Parmi elles, l'utilisation des modèles numérique des surfaces terrestres (LSM en anglais pour Land Surface Model) et des satellites d'observations de la Terre (EOs en anglais pour Earth Observations) est de plus en plus courante. Ces modèles numériques permettent de simuler les échanges complexes d'énergie et d'eau dans le continuum sol-végétation-atmosphère. Ils permettent également le suivi en temps quasi réel de l'évolution de la végétation et des ressources en eau à l'échelle mondiale, à des résolutions temporelles et spatiales cohérentes avec leur représentation des processus. Ces modèles peuvent également prévoir ces évolutions. Ils peuvent cependant souffrir d'une représentation trop simplifiée des processus, d'un forçage atmosphérique imparfait et de mauvaises conditions initiales.

Les données d'observations de la Terre depuis l'espace sont disponibles à l'échelle mondiale, souvent sur le long terme et de manière répétée dans le temps. Elles fournissent de nombreuses variables liées aux cycles de l'eau et du carbone, à la végétation, et au bilan d'énergie. Elles sont représentatives d'un instant précis dans le temps. Toutes les variables ne sont cependant pas observables depuis l'espace. Il est possible de combiner les modèles et les observations de la Terre depuis l'espace pour obtenir un résultat maximisant les avantages et minimisant les inconvénients de chacun. Cela s'opère au travers des systèmes d'assimilation de données pour les surfaces terrestres (LDAS en anglais pour Land Data Assimilation System) qui permettent d'assimiler ces observations dans les modèles de manière séquentielle, mettant à jour au fil de l'eau les variables d'états et incrémentant leur trajectoire.

Le Centre National de Recherche Météorologique (CNRM, UMR-3589, Météo-France/CNRS) a mis en place un tel système pour améliorer les sorties

du modèle ISBA (Interactions entre le Sol, la Biosphère et l'Atmosphère) au travers de l'assimilation d'observations satellitaire d'indice de surface foliaire et d'humidité superficielle du sol. Ce système global, LDAS-Monde (Albergel et al., 2017; Albergel et al., 2020), a la capacité de piloter l'humidité du sol grâce à l'assimilation de l'indice de surface foliaire, et inversement. De plus, il a été utilisé et validé dans diverses régions du monde. Il a été montré que LDAS-Monde peut identifier les sécheresses et suivre leurs impacts sur les variables des surfaces terrestres à différentes échelles spatiales.

L'objectif principal de ce travail de thèse est de mettre en place la prévision à court et moyen terme des variables terrestres dans LDAS-Monde. Pour cela, le forçage de LDAS-Monde par des prévisions de variables atmosphériques doit être mis en oeuvre. La qualité de l'estimation des variables terrestres qui caractérisent les sécheresses aux différentes échéances temporelles de la prévision est primordiale. Elle doit être évaluée dans la perspective d'un suivi précoce et d'une prévision des sécheresses. Cette nouvelle capacité de prévision dépend aussi de la qualité des prévisions atmosphériques. In fine, l'objectif est d'exploiter l'information provenant des prévisions des variables terrestres dans le but de détecter les épisodes de sécheresse avant qu'ils ne se développent, ce qui permet une meilleure anticipation et une meilleure atténuation de leurs effets.

Ce travail de thèse contribue à l'effort en cours pour développer l'utilisation de LDAS-Monde et de l'assimilation de données satellitaires afin de mettre en place un système d'alerte précoce des sécheresses. Les différents objectifs de cette thèse sont détaillés ci-dessous :

- Evaluer les performances du système LDAS-Monde, combinant des observations satellitaires et un LSM, pour une application dans le suivi et la caractérisation des sécheresses agricoles
- Faire passer LDAS-Monde d'un mode de suivi à un mode de prévision des variables terrestres
- Analyser l'impact des conditions initiales sur la qualité des prévisions
- Identifier de nouvelles observations permettant d'améliorer les conditions initiales dans LDAS-Monde
- Tirer profit des améliorations apportées au système LDAS-Monde pour mettre en place un système d'alerte des sécheresses fondé sur des variables terrestres utilisées comme indicateurs de sécheresse.

La Figure 1 illustre les questions scientifiques, ainsi que la logique et le contenu du travail réalisé pendant cette thèse.

Est d'abord présenté, dans le Chapitre 1, le contexte scientifique du travail et une revue de la littérature scientifique liée aux sécheresses, à la modélisation des surfaces terrestres et à l'observation de la Terre depuis l'espace. La méthodologie, y compris une description des données et des outils utilisés, les analyses

statistiques et les domaines analysés sont décrits dans le Chapitre 2. Les premières expériences avec LDAS-Monde pour la surveillance des sécheresses et la transition d'un système de suivi à un système de prévision sont présentées dans le Chapitre 3. Ce Chapitre comprend un article publié (Mucia et al., 2020). Le Chapitre 4 analyse les expériences conçues pour améliorer les conditions initiales du système LDAS-Monde grâce à l'assimilation de l'épaisseur optique de la végétation dans le domaine des micro-ondes (VOD). L'utilisation du VOD en tant que proxy de l'indice de surface foliaire est évaluée. Plusieurs expériences d'assimilation sont mises en œuvre afin de comparer l'assimilation séparée et conjointe de variables de la végétation et de l'humidité superficielle du sol. Enfin, le Chapitre 5 décrit une étude de cas utilisant un système d'alerte des sécheresses utilisant LDAS-Monde pour prévoir leurs impacts sur les variables des surfaces terrestres jusqu'à deux semaines d'échéance. Les premières étapes permettant d'aller vers la mise en œuvre de la méthodologie proposée sont appliquées aux expériences précédentes.

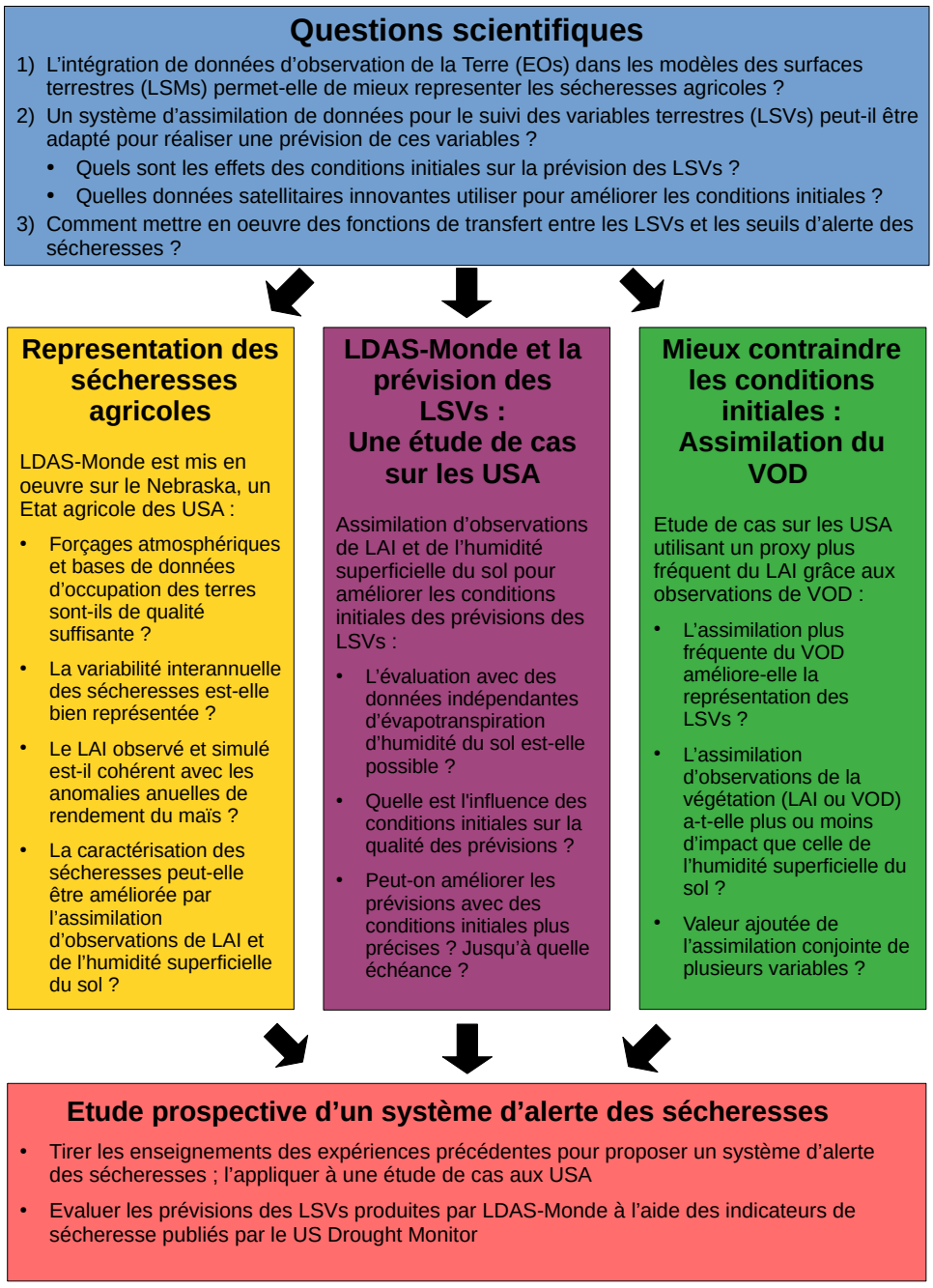

Figure 1: Un schéma de la motivation et du travail réalisé dans cette thèse.

General Introduction

As the climate changes towards a warmer and more volatile state (IPCC, 2014), the importance of understanding extreme climatic events, and their impacts on land surface variables (LSVs), mounts. These extreme events, such as droughts, heatwaves, floods, and fires, are projected to increase in frequency and/or magnitude from changes caused by anthropogenic warming (IPCC, 2014; IPCC, 2018; Ionita et al., 2017). Drought events in particular are among the most detrimental and complex hazards (Bruce, 1994; Obasi, 1994; Cook et al., 2007; Hagman et al., 1984), with impacts ranging from damaged crops to famine and human mortality.

Drought monitoring is then of great importance to track the timing, extent, and severity of ongoing drought events in order to better respond to their impacts. In addition to proper mitigation and planning, better drought monitoring can decrease agricultural and water resource losses and minimize damage to infrastructure and society (Wilhite, Hayes, and Svoboda, 2000).

With increased focus on monitoring drought events and their impacts, it is important to use all the available resources at our disposal. Included among those, and becoming ever more abundant, are land surface models (LSMs) and satellite Earth observations (EOs). LSMs are numerical models able to simulate components and fluxes of the surface vegetation, water, soils, and atmosphere through complex interactions and exchanges. These models allow for the near real-time tracking of the evolution of vegetation and water resources, globally and at consistent temporal and spatial resolutions. Moreover, LSMs have the ability to forecast LSVs. However, LSMs suffer from imperfect model physics and poor initial conditions.

Satellite EOs provide unparalleled, often global and long term, observations of many important variables, but also suffer from the inability to observe all LSVs, are representative of only an instant in time, and lack the capacity to forecast. However, through the combination of LSMs and EOs, a resulting product can be obtained maximizing the advantages, and minimizing the drawbacks of each by themselves. Land Data Assimilation Systems (LDAS) perform this task of sequentially assimilating and integrating observations into LSMs, and update the state variables to increment the model trajectory.

The French National Centre for Meteorological Research (CNRM) has implemented an LDAS in order to sequentially drive the ISBA (Interactions between Soil-Biosphere-Atmosphere) LSM through the assimilation of satellite-derived leaf area index (LAI) and surface soil moisture (SSM). This global LDAS-Monde (Albergel et al., 2017; Albergel et al., 2020) has the ability to drive soil moisture processes through the assimilation of LAI, and vice-versa. Additionally, it has been tested and analyzed over various regions and has been able to identify and monitor drought impacts on the land surface at various scales.

In this work, LDAS-Monde is enhanced with the capability to ingest medium range atmospheric forecasts hence producing daily forecasts of the land surface. The responses and quality of surface variables to different forecast time frames is critical for the monitoring and forecasting of drought impacts. With this new

forecast capability, and appropriate atmospheric forecasts, it is possible to analyze land surface forecasts with the intent to detect drought events before they fully develop, allowing for better preparation and mitigation of the disaster's effects.

This thesis works towards the ultimate goal of using LDAS-Monde, and the data assimilation therein, to create an objective drought alert system. The various objectives of this thesis are as follows:

- Assess the performance of the LDAS-Monde system, combining satellite Earth Observations (EOs) and Land Surface Models (LSMs), for the use of monitoring and representing agricultural droughts
- Transition LDAS-Monde from a purely monitoring mode to a forecast mode, enabling the prediction of Land Surface Variables (LSVs)
- Analyze the effects of system's initial conditions on the forecasts of LSVs
- Identify new observations of variables that can improve the initial conditions in LDAS-Monde
- Compile the combined improvements and knowledge of the LDAS to implement a drought alert and warning system based on selected LSVs as drought indicators

Figure 2 presents a diagram of the scientific questions, rationale, and work paths completed in this thesis.

This thesis first covers the scientific context for the work and a review of the scientific literature related to droughts, land surface modelling, and satellite Earth observations in Chapter 1. The methodology of this work, including a description of the data and tools used, the statistical analyses, and the domains analyzed are provided in Chapter 2. The initial experiments with LDAS-Monde for drought monitoring, and the transition from a monitoring to forecast system are shown in Chapter 3, including a published article (Mucia et al., 2020). Chapter 4 analyzes experiments designed to improve the initial conditions of LDAS-Monde through the assimilation of vegetation optical depth (VOD) as an LAI-proxy, and analyzes differences from the separate and joint assimilation of LAI or VOD and SSM. Finally, Chapter 5 proposes a drought warning case study using LDAS-Monde to predict drought impacts on LSVs. The first steps towards implementing the proposed drought alert methodology were also applied to previous experiments to provide a proof of concept.

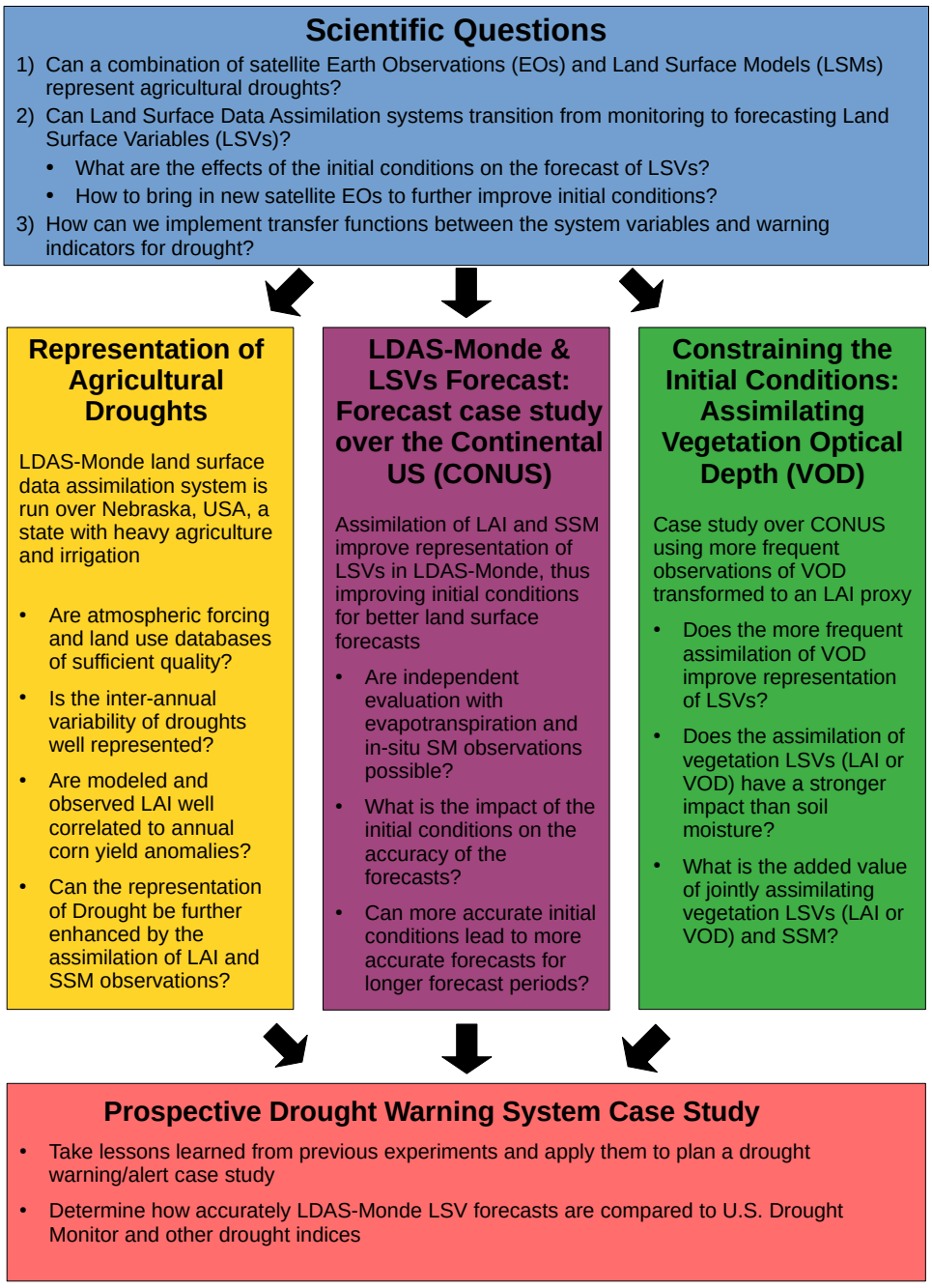

Figure 2: A diagram of the motivation and work completed in this thesis.

Chapter 1

Scientific Context

This chapter introduces the scientific context and motivation for the thesis. The importance of drought monitoring and forecasting is discussed, and a review of current drought monitoring efforts and the different definitions of drought is provided. Then, the significance of land-atmosphere interactions is detailed, focusing on the energy, water, and biogeochemical cycles. An overview of land surface modeling, its history and future directions is then presented. Satellite Earth observations are then examined, leading into the merging of Earth observations and land surface models. Finally, this chapter provides an overview of further chapters, and the work and objectives therein.

1.1 Importance of Monitoring and Forecasting Drought

Drought events have been found to be among the most detrimental natural hazards in existence in today's society (Bruce, 1994; Obasi, 1994; Cook et al., 2007). Billions of dollars are spent every year in the United States (Wilhite, 2000 citing FEMA, 1995) and the European Union (EC, 2007a; EC, 2007b) in effort to address drought damages and impacts. Drought is also the most complex natural hazard, leading to it being the least understood, and its effects reach more people than any other natural hazard (Hagman et al., 1984). In addition to the massive economic cost of damages, droughts can transcend monetary losses and cause famines and human mortality.

Agricultural regions everywhere can be especially vulnerable to drought based on their natural dependence on water resources. Impacts to agriculture can be felt from the relatively early stages of drought and can continue to ripple through communities long after the drought event itself has finished (Wilhite, 2000). Initial crop damage can lead to economic losses from low crop production, resulting in strained financial institutions and unemployment (Western Governors Policy Office, 1977). Some regions may face even more challenges associated with droughts, with impacts also damaging food security (FAO, 2003; Haile, 2005; Devereux, 2007).

While far from fully understood, advancing knowledge of how climate change affects drought brings additional worries. Extreme meteorological and climatic events are likely to increase in frequency and/or magnitude due to anthropogenic climate change (IPCC, 2014; Ionita et al., 2017). Specifically, droughts are projected to increase in the Southern Europe and Southern Africa as global surface temperatures approach 2°C above the pre-industrial average (IPCC, 2018). Projected changes to other regions of the globe are determined to have "Low Confidence" due to the high variability and longer duration of drought events. Because droughts tend to be longer lasting than other meteorological phenomenon, their frequency is low, and thus climate projections have less samples to analyze and any conclusions drawn are done so with less certainty (Vasiliades, Loukas, and Patsonas, 2009).

So-called flash droughts are an increasingly studied phenomenon in recent years. Defined as *"a subset of all droughts that are distinguished from more conventional slowly developing droughts by their unusually rapid rate of intensification"* (Otkin et al., 2018). Precipitation deficits alone are unlikely to lead to flash drought events. However, when those deficits occur in tandem with other anomalies such as high temperatures, strong winds, low humidity, and clear, sunny skies, the increased evaporative demand can quickly plunge vegetation into stress, depleting soil moisture reserves and striking the beginning of a flash drought event (Hunt et al., 2009; Hunt et al., 2014; Mozny et al., 2012; Otkin et al., 2013; Anderson et al., 2013; Ford et al., 2015; Ford and Labosier, 2017). With far more confident projections of surface temperature increases globally (IPCC, 2018), a warming planet provides a generally higher potential for rapid intensification and the appearance of flash droughts.

With looming impacts to agriculture, water resources, and food security, it is of the utmost importance to minimize loss of resources, infrastructure, and even life by preparing for, monitoring, and predicting drought events. Effective monitoring of droughts leads to reduced vulnerability and impacts (Wilhite, 2000). Drought monitoring is only a single part of a far larger strategy needed to manage the crisis and risk of drought, but it plays a key role. Providing timely, critical information regarding onset and severity of drought events to decision makers at all levels, enables local and national officials to enact responsive and mitigating measures (Wilhite, Hayes, and Svoboda, 2000), preventing more severe agricultural, hydrologic, and socio-economic losses.

Drought monitoring, defined as, "tracking the severity and location of drought" (Hayes et al., 2012), started out focusing specifically on precipitation and precipitation deficits (Heim, 2002). Various indicators (observed parameters such as temperature, precipitation, water levels, etc) and indices (computed numerical values representing drought magnitude or severity using indicators) (Hayes et al., 2012) were developed in order to attempt to track drought severity on a large scale. Popular and still commonly used examples include the Palmer Drought Severity Index (PDSI) (Palmer, 1965) and Standardized Precipitation Index (SPI) (McKee, 1995), among countless others. The SPI has since been adopted as the worldwide standard by the World Meteorological Organization used to track meteorological drought. And while these indices are simple, easy to calculate ways to measure aspects of drought, they are limited to meteorological assessments, can often have poor spatial representation due to limited observational coverage, and can also fail to adequately identify duration of events (Vasiliades and Loukas, 2009).

For current drought monitoring, it is well recognized that no single indicator or index fully describes the complex impacts of drought. It has therefore been a goal to merge indicators and indices into single products. These so-called hybrid approaches classify drought in simplistic systems for easy use by decision makers and the public (Hayes et al., 2012). Among these hybrid approaches, the United States Drought Monitor (USDM) (Svoboda et al., 2002), and North American Drought Monitor (NADM) (Lawrimore et al., 2002) are considered the current state-of-the-art drought monitoring tools. They take in a wide range of information into consideration, including local expertise, standard drought indicators, climatology, fire indices, satellite-based vegetation assessments, and hydrologic data. The USDM in particular then uses experts to synthesize all the available information from these sources, and ground truths this information with local observers across the country. The Drought Monitor works as a weekly summary and assessment of drought conditions and impacts, and has no forecast component. The severity of droughts is determined by a percentile ranking as seen in Figure 1.1. Analysis material and other objective drought indices have moved towards this same approach of percentile rank in order to better aid and compare to the USDM.

Other countries and regions have taken similar methodology and applied it to their region (such as The German Drought Monitor (Zink et al., 2016)), and others have created hybrid approaches with different metrics (Southeast

Category	Drought condition	Percentile chance
D0	Abnormally dry	20 to ≤30
D1	Drought—moderate	10 to ≤20
D2	Drought—severe	5 to ≤10
D3	Drought—extreme	2 to ≤5
D4	Drought—exceptional	≤ 2

Figure 1.1: Categories of drought magnitude use in the Drought Monitor. Each category is associated with its percentile chance of happening in any given year out of 100. (Source: Svoboda et al., 2002)

Asia Drought Monitor (IWMI, 2015), European Drought Observatory (Vogt et al., 2011), and African Drought Monitor (Sheffield et al., 2014) for example). Following advances in international and intergovernmental coordination, and solving scientific and technical issues, a Global Drought Monitor Portal has also been created, amassing a wide range of current drought indicators and impacts, as well as setting the stage for a Global Drought Early Warning System (Heim and Brewer, 2012).

While monitoring drought impacts and events provides necessary information to stakeholders regarding retroactive actions to take, it is also necessary to couple monitoring with early warning. A Drought Early Warning System (DEWS) is a tool designed to predict the occurrence and impact of a drought event as well as elicit a response (Buchanan Smith and Davies, 1995). Several shortcomings of DEWS have been noted (Wilhite, Sivakumar, and Wood, 2000), and specifically include the need to invest in research and increase the reliability of seasonal forecasts. Additionally, Hayes et al. (2011) identifies that successful DEWS may contain numerical models and Land Data Assimilation Systems (LDAS). With these tools, accurate numerical forecasts of land surface variables (LSVs) linked to drought are made possible.

1.1.1 Definitions of Drought

An event slow to develop and often identified in retrospect, drought has been called a *creeping phenomenon* (Gillette, 1950). While the definition of drought must be viewed from the perspective of each discipline (Wilhite and Glantz, 1985; Wilhite, 2000), the most general definition relates to a deficit or delay of precipitation over a region for some period of time. Clearly more specific definitions are required to better compartmentalize and eventually respond to drought events. This thesis uses the disciplinary views of drought from Wilhite and Glantz (1985). Droughts are separated into meteorological, hydrologic, agricultural, and socio-economic droughts. And they are defined as follows:

1. Meteorological Droughts are characterized by periods of prolonged and abnormal moisture deficiency, primarily measured by lack of solid or liquid precipitation (Palmer, 1965)

2. Hydrologic Droughts are characterized by deficiency in water supply on Earth's surface or deficiency in precipitation, runoff, or accumulated water in various storage capacities (Yevjevich, 1969)

3. Agricultural Droughts are characterized by deficiency in soil moisture that leads to hindrance of plant growth and a significant decline in crop yield (Boken, Cracknell, and Heathcote, 2005)

4. Socio-economic Droughts are characterized by conditions where water demand is greater than supply, leading to societal and economic impacts (Dinar and Mendelsohn, 2011)

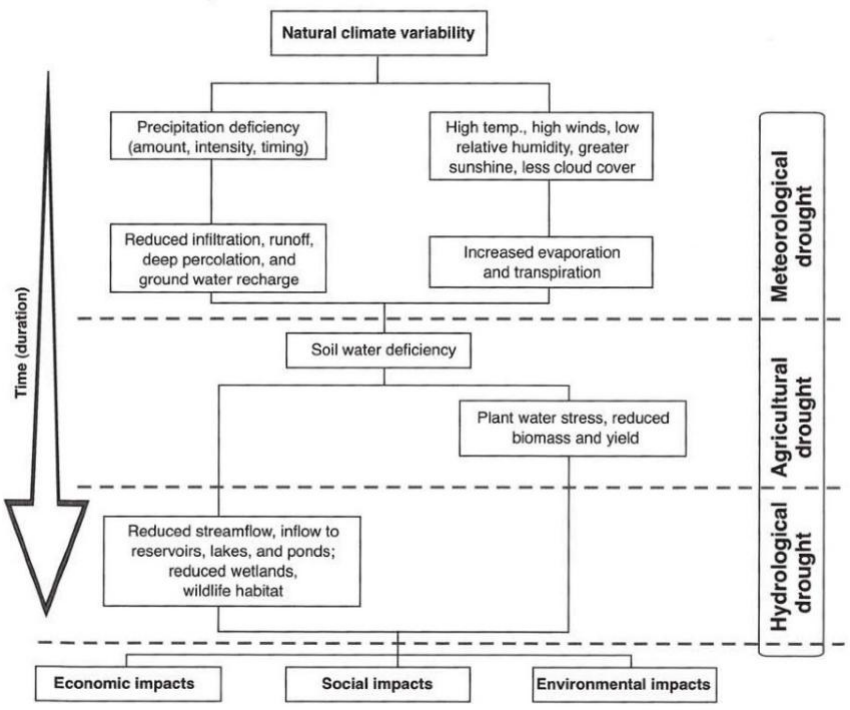

Figure 1.2: Relationship between various types of drought and duration of drought events. (Source: Wilhite, 2000)

Figure 1.2 displays the interconnected nature of the different types of drought, as well as characterizes them by duration. Meteorological, hydrologic, and agricultural droughts are more heavily dependant on various statistical indices to diagnose the onset, severity, and duration of drought events. Drought's impacts on agriculture are the most direct, devastating, and costly of the four types listed above (Drought.gov, 2021). Regional crop damage and

loss permeates globally, sharply increasing food prices, and hurting economies, particularly fledgling economies (Sternberg, 2011).

1.2 Land-Atmosphere Interactions

The weather and climate of the planet Earth are the results of a number of enormously complex, interconnected planetary systems. The main components of the Earth system are the atmosphere, the hydrosphere (including the cryosphere), pedosphere (soils), and the biosphere, (Bonan, 2016). The coupling of the biosphere and hydrosphere on the surface to the atmosphere are found to play a pivotal role in a better understanding of the Earth system (Bonan, 1995). Additionally, it has been found that land-atmosphere feedbacks, specifically soil-moisture-temperature and soil-moisture-precipitation feedbacks, are an important source variability when modelling future climates (Seneviratne et al., 2006).

Energy, water, and biogeophysical cycles link the aforementioned Earth systems. These cycles interact between the atmosphere, biosphere, pedosphere, and hydrosphere, driving changes and feedbacks, accentuating or mitigating changes to the greater Earth system (Bonan, 2016). Improving our understanding of the interconnections between vegetation and the atmosphere enables us to better describe, monitor, and predict the direct and indirect impacts of drought events.

1.2.1 Energy Balance

The sun is the driving force for the entire planet. The inbound energy drives the atmospheric and hydraulic circulations and most importantly, provides the energy for vegetation to grow. This vegetation base provides habitats and nutritional energy for a large array of animal life, as well as provides humans with fossil fuels, food, and materials such as wood.

Conservation of energy, described by the first law of thermodynamics, requires that a closed system may allow energy to change forms, but the total amount of energy must be conserved. We primarily talk about four types of energy in the global energy budget: shortwave solar radiation, longwave radiation, the turbulent surface fluxes of latent and sensible heat, and the ground heat flux (Kiehl and Trenberth, 1997; Trenberth, Fasullo, and Kiehl, 2009). Shortwave radiation from the sun enters Earth's atmosphere and is either absorbed or scattered (including scattering back out to space). In the context of this research, the energy balance of the surface is of the highest interest. When this incoming energy reaches the surface, its interactions with vegetation drive photosynthesis, as well as the latent and sensible heat fluxes that are critical to the coupling of the vegetation and the lower atmosphere. The most simplistic description of the surface energy balance can be stated as follows:

$$R_n = G + H + F + LE \tag{1.1}$$

Where R_n is the net radiation, G is the ground heat flux, H is sensible heat flux, F is energy absorbed in photosynthesis, and LE is the latent heat flux. Energy fluxes are expressed in Wm^{-2}. A transformation of the latent heat flux can often be useful when focused on this variable.

$$R_n = G + H + F + \lambda ET \tag{1.2}$$

Where λ is the latent heat of vaporization for water, and ET is the rate of evapotranspiration (Garratt, 1992). Sensible heat flux is the movement of long-wave radiation through conduction or convection into and from the lower atmosphere. Latent heat flux is the transformation of energy into and out of water through the processes of evaporation and condensation. The monitoring of LSVs, specifically regarding drought, are especially focused on this latent heat flux. Increased surface temperatures and low humidity can rapidly increase the rate of evapotranspiration, exhausting the soil water reserves and crippling vegetation.

1.2.2 The Water Cycle

Water dominates the surface of planet Earth. Seventy-one percent of the surface is covered by water, with oceans holding 96.5% of the world's supply. Freshwater only accounts for 2.5%, with nearly 70% of that trapped in glaciers and ice caps. Groundwater holds most of the remaining nearly 30% (Shiklomanov, 1993). This vast amount of water is continuously cycled between the surface, subsurface, and atmosphere, and it is the transition from one area and form to another that is of great concern to scientists. Figure 1.3 gives a simplified view of the water cycle (USGS, 2019).

In essence, the water cycle is a balance between water moving towards the surface by precipitation and deposition, and water moving towards the atmosphere via evapotranspiration and sublimation. Over the oceans, evaporation is greater than precipitation, resulting in a deficit of water. However, precipitation exceeds evaporation over land, with the surplus water flowing into the oceans and replenishes the net loss of water (Bonan, 2016). Water exists in three forms on the planet, and a significant amount of energy is required to change its state. In fact, over 50% of the net radiation at the surface is used to evaporate water (Trenberth, Fasullo, and Kiehl, 2009). This latent heat of vaporization is a key element in both the global energy and water cycles.

Water is the basis for all life on Earth, and therefore it is important to monitor this precious resource. With respect to drought, one of the most important sources of water is soil moisture. While only containing 0.05% of all freshwater (Shiklomanov, 1993), soil moisture controls a large number of interactions between the biosphere and atmosphere. Plant roots directly access this reservoir of water for photosynthesis, and lack of available soil moisture can lead to wilting and plant death.

The management of water resources is an ever growing field, and its importance has been highlighted in recent decades. Sustainability of water resources is also crucial, defined as *... systems designed and managed to fully contribute*

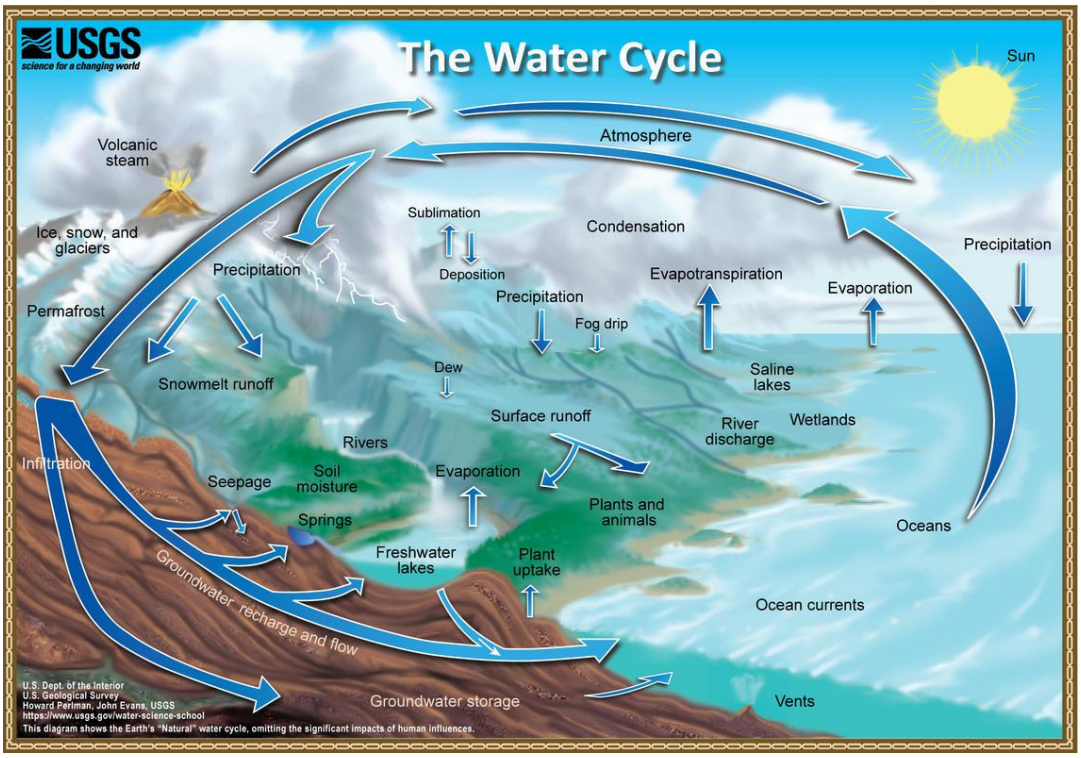

Figure 1.3: A simplified view of the global water cycle. (Source: USGS, 2019)

to the objectives of society, now and in the future, while maintaining their ecological, environmental, and hydrological integrity (ASCE, 1998). The delicate balance of sustainability makes water resource monitoring, as well as monitoring the impacts of divergence from water norms, a very important task.

1.2.3 Biogeochemical Cycles

Another major cycle describing land-atmosphere interactions is the biogeochemical cycle. This cycle includes the transition movement of major gases, namely carbon dioxide (CO_2), methane (CH_4), and nitrous oxide (N_2O). Each one of these are greenhouse gases, with significant perturbations in their relative amounts directly impacted by human behavior. This also makes the variability of sources and sinks of these gases less well understood, which poses a significant hurdle in projecting future changes.

Comprising a major portion of the biogeochemical cycle is the carbon cycle, illustrated in Figure 1.4. Comprised mainly of two domains in terms of the timeline of cycling, a fast flux reservoir, and slow flux reservoir. The fast domain consists of carbon rapidly exchanging between the atmosphere, the ocean, vegetation, and soils (IPCC, 2013). These transitions happen on the scale of years to millennia. On the other hand, the slowly fluxing domain consists of stored carbon in sediments and rocks, which are weathered or eroded on the scale of 10,000 years or more.

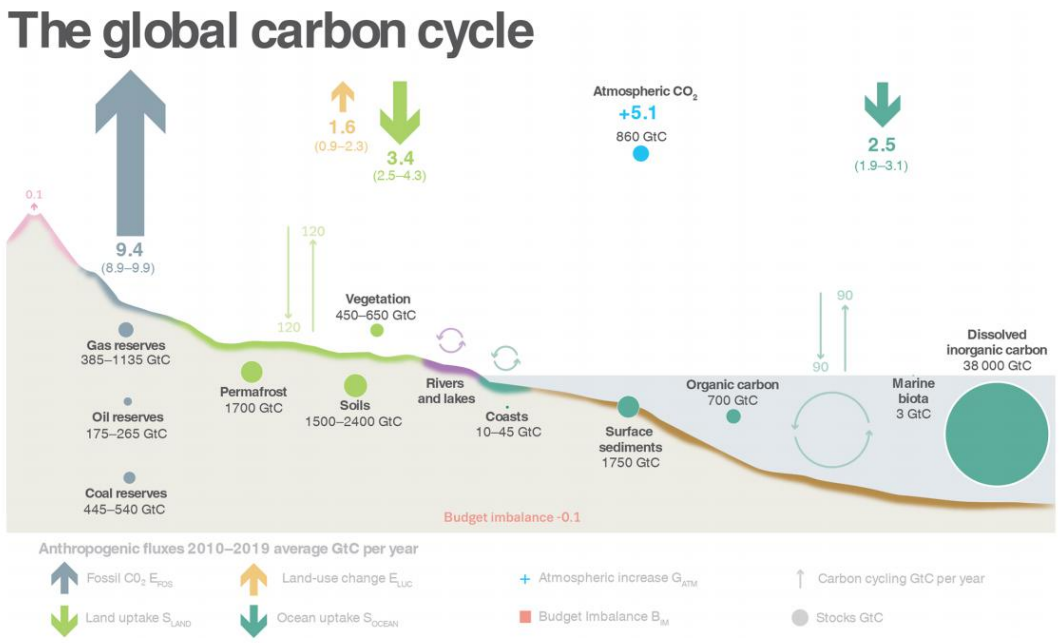

Figure 1.4: A simplified view of the global carbon cycle. (Source: Friedlingstein et al., 2020)

Of most immediate concern for this study is the carbon exchanges between the soils, atmosphere, and vegetation. At the canopy level and below, the process of carbon fixation and respiration is well understood (Schulze, 1970; Black, 1973; Poorter, Remkes, and Lambers, 1990). Carbon is fixed to plants from the atmosphere, and is cycled through the plants and soil through carbon respiration (Beer et al., 2010). Photosynthesis is the process by which plants absorb light energy to transform the carbon from the atmosphere into carbohydrates. This process occurs at the cellular level of leaves, and is given by the Equation 1.3:

$$nCO_2 + 2nH_2O \xrightarrow{\text{Light}} (CH_2O)_n + nO_2 + nH_2O \tag{1.3}$$

With n being the number of molecules of CO_2 combining with water to form $(CH_2O)_n$ and releasing n number of molecules of oxygen (Bonan, 2016). $(CH_2O)_n$ represents all carbohydrates, which contain many forms of sugars, starches, and other compounds of hydrogen, oxygen, and carbon.

Plants can maximize photosynthesis through several ways, namely reducing resistance to CO_2 diffusion by having many large and open stomata and/or increasing fixation rates by having higher concentrations of photosynthetic enzymes. However, these processes are not without drawbacks. Decreasing resistance to CO_2 diffusion also allows water to escape the plant through transpiration, and photosynthetic enzymes require the often limited resource of nitrogen. While the general photosynthetic reaction in Equation 1.3 is applicable to most plants, there are wildly differing pathways of exactly how plants efficiently fix

carbon, while minimizing water loss and conserving resources such as nitrogen. Primarily resulting from distinct environmental conditions, the C_3, C_4, and CAM photosynthetic pathways are important to understand for vegetation monitoring and modelling. The brief description of these pathways below are derived from Forseth (2010).

The C_3 pathway is one by which the initial products of photosynthesis are 3-carbon compounds known as phosphoglyceric acid or PGA. Carbon is fixed via Equation 1.4:

$$RuBP + CO_2 \rightarrow 2PGA \tag{1.4}$$

Where RuBP is a 5 carbon sugar, ribulose bisphosphate. This reaction is catalyzed by the rubisco enzyme (ribulose bisphosphate carboxylase-oxygenase), which also facilitates a secondary reaction with O_2 given in Equation 1.5.

$$RuBP + O_2 \rightarrow PGA + PG \tag{1.5}$$

With PG being 2-phosphoglycolate, which is later used in the photorespiration cycle (where CO_2 is released). This process of photorespiration acts to counter photosynthesis due to the rubisco being sought after by both CO_2 and O_2, as well as directly releasing CO_2 during the cycle. Often, warm and dry regions attain higher photorespiration after stomatal closure, reducing net photosynthesis up to 50% Forseth (2010). This detriment to the plant is generally thought of as the primary driver of the alternative C_4 and CAM pathways. The C_3 pathway is both the oldest evolutionarily, and the most environmentally and taxonomically diverse.

The C_4 pathway has different initial enzymes and initial products. To fix carbon, phosphoenolpyruvate or PEP is used at the outer stage. PEP prefers CO_2 more than rubisco, and does not have an oxygenase reaction. The CO_2 and PEP combine to form a 4-carbon acid. This new PEP reaction is used by the plant on the outer layers, while also having mostly separated C_3 biochemistry in the inner layers. The C_4 acid is moved inwards where photorespiration occurs, and CO_2 is released directly next to rubisco enzymes, thus providing a far higher advantage of CO_2 over O_2. This C_4 process becomes more efficient than C_3 when there is lower stomatal conductance (typically due to high temperatures and/or low humidity).

Lastly, CAM, or Crassulacean Acid Metabolism, is a pathway of photosynthesis that combines the PEP enzymes of the C_4 pathway with a rigid diurnal cycle. Stomata are open at night allowing the absorption and fixation of CO_2 via PEP, which is stored in malate. During the daytime, this malate undergoes photorespiration, and the released CO_2 is used in the C_3 pathway. Plants using the CAM pathway are generally succulents, although there are a number of species that have the ability to switch between C_3 and CAM pathways under stressful conditions, and even seen as a response to climate change (Lande, 2009; Nicotra et al., 2010).

In the context of this research, these differences in photosynthetic pathways

are important due to the often very different responses of LSVs and EOs. Environmental temperatures, humidity, soil composition, and even altitude play a large role determining the carbon fixation pathways, and are important aspects to consider during drought events. Additionally, many modern LSMs (including the ISBA model used in this thesis) explicitly model canopy carbon flux, intimately tied to the vegetation pathway type. Knowing these differences can help us understand different observed and modelled plant responses with and without drought. We can also analyze LSVs and EOs over regions primarily comprised of C_3 or C_4 plant types, which can be used as a way to check theory and analysis.

1.3 Land Surface Models

The energy balance, water cycle, and biogeochemical cycles play key roles in controlling land surface processes. Land surface models (LSMs) are numerical models that attempt to describe and represent these processes on both small and large scales. LSMs provide consistent temporal and spatial outputs of variables of interest called land surface variables (LSVs). These LSMs also provide the land surface component of global climate models (GCMs), thus directly influencing projections of future climate (McGuffie and Henderson-Sellers, 2001).

Land surface modelling first began with Manabe (1969) depicting soil-vegetation-atmosphere transfer of water with a simple bucket scheme. While becoming a major step forward for the representation of land surface processes, the simple bucket scheme was not able to well represent longer term hydrological variation, nor plant physiology. Various other simple transfer schemes emerged through a second generation of models including a soil heat and moisture flux scheme (Deardorff, 1978) and a general biosphere-atmosphere transfer scheme (BATS) (Dickinson, 1986). Soon after, Sellers et al. (1986) became the first model to explicitly model plant physiology simple biosphere transfers coupled to a GCM with the Simple Biosphere Model (SiB). Another model, Interactions between Soil, Biosphere, and Atmosphere (ISBA) was developed and incrementally improved, with the addition of high accuracy calibration of coefficients by other models and experimental data (Noilhan and Planton, 1989; Noilhan and Mahfouf, 1996). ISBA was also explicitly designed to be included in and coupled with meteorological models, which has been shown to improve boundary layer accuracy (Viterbo et al., 1999) and precipitation forecasts (Betts et al., 1996).

For use in climate models, Pitman (2003) postulates LSMs must accurately represent the surface energy budget, the surface water balance, surface carbon exchanges, and the climatic effect of snow. Third generation LSMs are moving to fulfil these requirements, and more. The focus of most of the improvements of third generation models focus on the carbon balance and representation of leaf stomatal conductance. Models continued to evolve, with several complex processes incorporated such as net photosynthetic rates, stomatal conductance (Sellers et al., 1992), and the exchange of CO_2 and H_2O gas at the leaf level

(Collatz et al., 1991). Advances to the ISBA model were also made by adding stomatal resistance and photosynthesis at the leaf level (Calvet et al., 1998).

According to (Blyth et al., 2021), modern LSMs contain seven basic components, as shown in Figure 1.5. These components include surface and canopy processes, snow and soil physics, water bodies, vegetation physiology and photosynthesis, soil biogeochemistry, vegetation dynamics, and land and water use. Figure 1.5 also describes the pre-2000 developments, recent advances, and future directions of each component. Figure 1.6 shows the same layout, taken from Blyth et al. (2021), but for LSM exchanges instead of the individual components. The exchanges include land-atmosphere, surface-subsurface, physics-biogeochemistry, vegetation-soil, vegetation-landscape, land-catchment, and water-humans.

Figure 1.5: Land Surface Model Component development for pre 2000, recent advances and future directions. (Source: Blyth et al., 2021)

As models include more and more processes, their complexity increases significantly. Current generation models face the delicate balancing act of attaining the most accurate modeling and forecasting, while keeping the model itself sufficiently simple to be understood and to be coupled with atmospheric and climate models. Figure 1.7 shows the rough timeline of parameters and processes added to LSMs (Fisher and Koven, 2020)

These modern LSMs play an important role in the monitoring of agriculture, particularly for extreme events. LSMs allow the selection of individual and collective LSVs to easily monitor through uniform temporal progression, and the inclusion of crop and crop types allows for a more dynamic approach

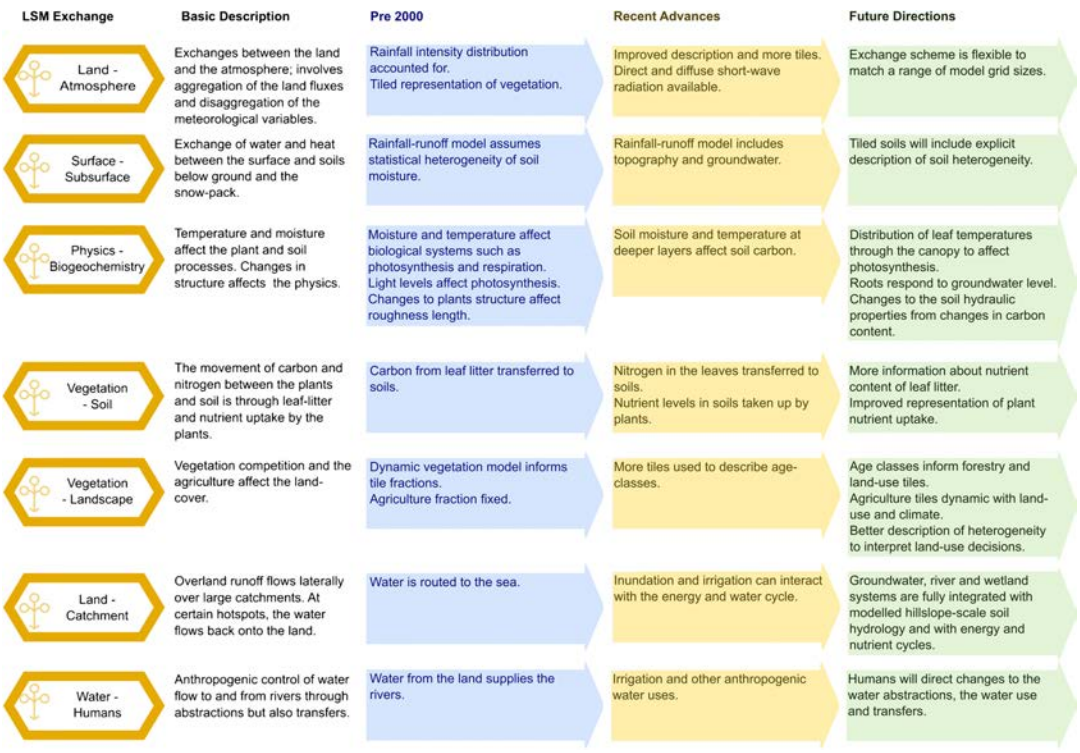

Figure 1.6: Land Surface Model exchanges, basic description, and development for pre 2000, recent advances and future directions. (Source: Blyth et al., 2021)

to monitoring droughts. Additionally, the coupling of LSMs to GCMs highlight the importance of land-atmosphere interactions, and contribute to the more accurate prediction of future climate conditions through more realistic photosynthesis and conduction schemes better representing the energy, water, and carbon fluxes.

While LSMs can have explicit or implicit crop schemes, they are different compared to traditional crop models. LSMs describe interactions and exchanges between the atmosphere, surface canopy processes, vegetation physiology and dynamics, land use, water use, soil chemistry, and water bodies. They are a broad look at all (or most) surface processes, able to represent land-atmosphere exchanges of heat and mass at fine temporal resolutions, and simulate the impact of the diurnal cycle. Crop models on the other hand, generally operate at a daily time step, focus specifically on predefined plant systems and the interactions between those crops and the environment, and simulate crop development, growth, nutrient and water uptake, and yield (Asseng et al., 2014; Nassiri Mahallati, 2020). Both crop models and LSMs have their uses in agricultural sciences, but this thesis focuses on the broader characterizations and applications of LSMs.

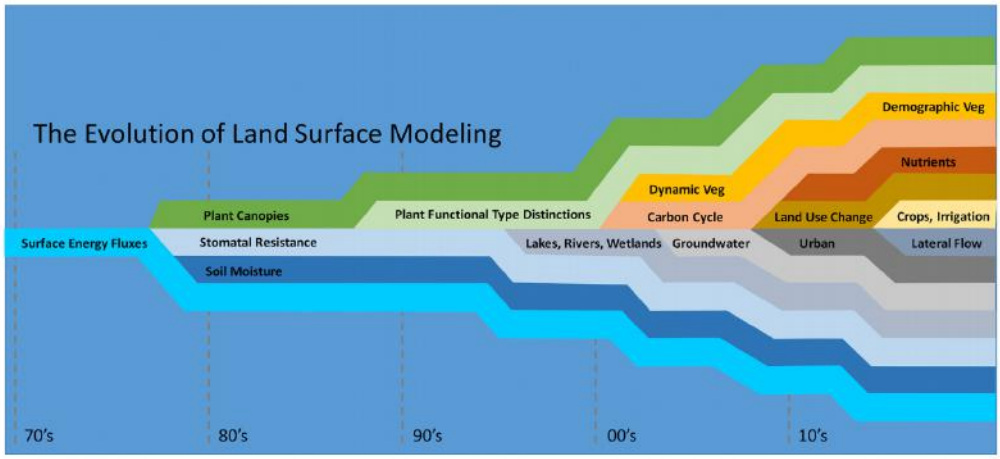

Figure 1.7: A schematic representing the approximate timing of new model components commonly employed in land surface models. (Source: Fisher and Koven, 2020)

1.4 Earth Observations

The launches of Sputnik 1 for the Soviet Union in 1957 and Explorer 1 for the United States in 1958 brought about the space age and soon after, ushered in a technological revolution for Earth sciences. TIROS-1 (Television Infrared Observation Satellite), launched in 1960 by the United States, became the first successful meteorological satellite, sending back the first images of the planet Earth, many of them used for meteorological analysis. But satellite remote sensing really took off in the early 1970s, with the launch of ERTS-1 (Earth Resources Technology Satellite), later renamed Landsat 1. This launch commenced the historic Landsat program which continues to this day.

The images and data observed by these early satellites began to grab the attention of more and more Earth scientists. With the scale and (more or less) consistent overpass rates, these satellite observations could provide detailed information on land use, water resources and hydrology, geology, soils, vegetation, forestry, and natural resource exploration (NASA Landsat, 2020; Lo, 1986). This technology offers the means to effectively study large areas in contrast to more traditional methods such as field surveys (Langley, Cheshire, and Humes, 2001). In fact, the importance of satellite Earth observations (EOs) have been so critical in Earth sciences that there have been calls to consider them "critical societal infrastructure" (Onoda, 2017).

Earth observations can be broadly separated into two types of remote sensing by the electromagnetic spectrum, optical (consisting of visible light, shortwave and intermediate infrared wavelengths), and microwave (consisting of microwave and radio wavelengths) (Onoda, 2017). The general tendency of reflectance of land, water, and vegetation in short wavelengths of the electromagnetic (EM) spectrum are seen in Figure 1.8.

Among the variables sensed in optical and infrared wavelengths are many observations of vegetation such as the Normalized Difference Vegetation Index

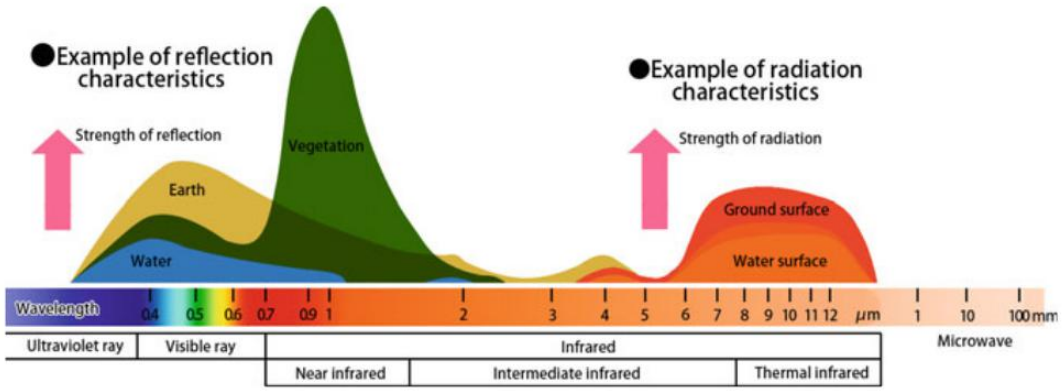

Figure 1.8: Reflectance of electromagnetic radiation by land, water, and vegetation as a function of wavelength. (Source: JAXA, 2016)

(NDVI) (Rouse et al., 1974; Tucker, 1979) or Leaf Area Index (LAI) (Watson, 1947; Price, 1993), ground cover such as land cover and land cover change (Lambin and Strahlers, 1994), and atmospheric phenomena such as clouds and water vapor (Fritz and Wexler, 1960; Inoue, 1985; Dalu, 1986). These wavelengths may be imaged with multispectral or hyperspectral sensors. Multispectral sensors only observe a handful of broad bands of the EM spectrum, while hyperspectral sensors observe up to hundreds of very narrow bands. Hyperspectral data can provide very useful information pertaining to vegetation and the atmosphere because of this high spectral resolution, but it can also be an overwhelming amount of data, and is not yet widely available at global scales.

Microwave sensors can be either active or passive. Active sensors emit microwave radiation themselves and receive the reflections (Dobson and Ulaby, 1986), while passive sensors only measure radiation emitted by the Earth's surface (Njoku and Entekhabi, 1996). Microwave remote sensing typically monitors parameters such as soil moisture, vegetation, and geography (altitude and land uplift/subsidence). Among the active sensors is Synthetic Aperture Radar (SAR) (Curlander and McDonough, 1991), which uses the motion of the satellite to construct an aperture far larger than its physical antenna, and thus increasing the spatial resolution of the system. Long microwave wavelengths of satellite sensors are used to provide important information on soil moisture properties, and even measure some subsurface moisture. By contrast, short wavelengths are easily scattered by vegetation, and are often used to gain valuable information about vegetation extent and health, as well as the soil moisture at the very top layers of soil (Onoda, 2017). While longer wavelengths of microwave radiation allow penetration and observations through vegetation, the signal is still attenuated. This attenuation by the vegetation is referred to as vegetation optical depth (VOD) (Jackson and Schmugge, 1991), and can be correlated to biomass as well as optical vegetation indices (Li et al., 2021). Many wavelengths of microwave observations also have a distinct advantage over optical indices:

all weather retrievals. High microwave frequencies become sensitive to atmospheric moisture, starting around the Ku-band. Lower microwave frequencies measured from space are not strongly impacted by clouds and precipitation.

Since the emergence of satellite remote sensing, these observations have been used to monitor drought conditions, typically through assessments of vegetation health. The first major application of this was with NDVI (Rouse et al., 1974; Tucker, 1979) applied to the AVHRR instruments (Tucker and Sellers, 1986; Gutman, 1990; Kogan, 1990; Hutchinson, 1991). The consistent and relatively high temporal resolution (compared to in situ monitoring stations with 10 day cumulative records) proved to be instrumental in monitoring droughts over the Sahel (Tucker et al., 1991), the United States (Burgan and Hartford, 1993), and Southern Africa (Unganai and Kogan, 1998) for just a few examples among many others. Several additional vegetation indices were created and widely used to monitor drought conditions such as the Vegetation Condition Index (VCI) (Liu and Kogan, 1996) which is based on NDVI, and the Temperature Condition Index (TCI) (Kogan, 1995) based on thermal vegetation and soil conditions. However, even with a wide range of observations and indices, it became well understood that no single index can sufficiently represent the complexity of drought (Hayes et al., 2005). Still, the continuous, high quality, large scale observations from satellites provide the ability to close the "information gap" and to improve drought monitoring capabilities (Hayes et al., 2012).

Future satellite missions promise to provide continuous observations in similar ways to the past decades. Mission continuations and follow-ons mean that the individual spectral bands will build long term databases, which can be used to study both short and long term environmental trends. Research also continues on how to best used these bands. Ever increasing spatial resolutions, both through scientific agency missions and private endeavors, are bringing subjects such as precision agriculture to the forefront.

1.5 Use of Earth Observations in Land Surface Models

Land surface models and Earth observations are good tools on their own to monitor LSVs. But it is also clear that both have weaknesses. LSMs provide the capacity of giving consistent, sub-daily output of dozens of important LSVs. These models help improve understanding of land surface interactions and feedbacks with the climate system. LSMs have gone beyond one of their initial purposes of providing land surface conditions for weather and climate models, and now can serve to monitor and forecast land surface conditions on their own rite (Balsamo et al., 2015; Schellekens et al., 2017) However, the biological and hydrological processes coded within the model are never perfect, and their accuracy can suffer from errors and uncertainties in model physics and parameterization, poor initialization, or errors in model forcing.

On the other hand, EOs provide high quality observations, often globally, of many important environmental parameters. But these observations too have

errors (calibration or otherwise), and are unable to observe all domains at all times. Autonomous satellite data analysis is constantly improving, yet there remains times when it is unclear exactly what the observations are of (fog vs low clouds, dew drops on vegetation vs soil moisture for example). They also are physically unable to directly observe some phenomena such as root zone soil moisture.

Through data assimilation (DA), we can integrate specific EOs into LSMs and are able to take full advantage of the major strengths of both, while reducing errors and uncertainties. Numerous studies have confirmed that DA results in better accuracy than either the model or observations alone (Reichle et al., 2007; Lahoz and De Lannoy, 2014; Albergel et al., 2017). Satellite observation products of soil moisture (Reichle, 2005; Barbu et al., 2011), snow cover (Rodell et al., 2004; Clark et al., 2006; Zhang et al., 2014), snow depth(De Lannoy et al., 2012; Liu et al., 2013), terrestrial water storage (Zaitchik, Rodell, and Reichle, 2008), land surface temperature (Reichle et al., 2010), and leaf area index (Demarty et al., 2007; Barbu et al., 2011; Kumar et al., 2019) have all been successfully assimilated using various assimilation schemes into various LSMs. The use of satellite DA has also proven to provide better initial conditions for numerical weather forecasting (Drusch and Viterbo, 2007; Mahfouf, 2010; Carrera et al., 2019), better monitoring of vegetation and crops (Albergel et al., 2010; Barbu et al., 2014; Sawada and Koike, 2016; McNally et al., 2017), floods (Kussul et al., 2008; Yucel et al., 2015; Chen et al., 2013), and droughts (Albergel et al., 2018b).

Integrating EOs into LSMs in an effective and consistent way can be a difficult task. The general overview of data assimilation provided in Lahoz and De Lannoy (2014) is summarized here. One way to approach data assimilation methodologies is to divide them into variational and sequential. Both can be associated to ensembles of model simulations. Many of the same statistical derivations apply throughout all data assimilation, and it is even possible to rewrite variational approaches as sequential as well as the inverse. Variational data assimilation is a way to minimize differences between the model and the background, as well as between predicted observations and real observations. An observation operator, H, is calculated to be able to merge the different spatial and temporal resolutions of the model and observations. Under linear conditions, the penalty function J, Equation 1.6, is sought to be minimized.

$$J = \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x^b})^T \boldsymbol{B}^{-1}(\boldsymbol{x} - \boldsymbol{x^b}) + \frac{1}{2}(\boldsymbol{y} - H(x))^T \boldsymbol{R}^{-1}(\boldsymbol{y} - H(x)) \tag{1.6}$$

Where $\boldsymbol{x}$ is the model state vector, $\boldsymbol{x^b}$ is the prior estimate of $\boldsymbol{x}$ (also called background), $\boldsymbol{B}$ is the background error covariance matrix, and $\boldsymbol{R}$ is the observation error covariance matrix.

For sequential data assimilation, the Kalman filter algorithm is used. This is computed by striving to find the best linear unbiased estimator, with the Equation 1.7.

$$\boldsymbol{x^a} = \boldsymbol{x^b} + \boldsymbol{K}(\boldsymbol{y} - H(\boldsymbol{x^b})) \tag{1.7}$$

Where $\boldsymbol{K}$ is the Kalman Gain Matrix defined in Equation 1.8, $\boldsymbol{x^a}$ is the result of the analysis, and $\boldsymbol{y}$ is the observation vector.

$$\boldsymbol{K} = \boldsymbol{B}H^T(H\boldsymbol{B}H^T + \boldsymbol{R})^{-1} \tag{1.8}$$

This Kalman Gain Matrix can be applied to nonlinear observation operators resulting in an Extended Kalman Filter (EKF) or a Simplified Extended Kalman Filter (SEKF), which is used in this thesis. The details of the SEKF data assimilation scheme used in this thesis are described in detail in Chapter 2.

Finally, both variational and sequential data assimilation can be used with an ensemble approach. Ensemble data assimilation casts a slew of short range forecasts, with increasing ensemble sizes increasing accuracy. One can do this with Monte Carlo simulations using the Kalman Gain Matrix, resulting in an Ensemble Kalman Filter (EnKF), or by using a particle filter (PF).

1.6 Objectives of this Thesis and Work Plan

This work tests the application of LDAS-Monde, through the SURFEX modelling platform over the Contiguous United States (CONUS), and several subdomains of interest. Various techniques were applied to the system and the assimilated observations to achieve the best potential for monitoring and forecasting drought. Statistical analyses compare the results of our simulations against independent observations and we examine applications where our analysis showed potential utility. This document is comprised of the consolidated results of all the work, to date, on the thesis titled "Assimilation of Satellite Data for the Monitoring and Prediction of Agricultural Droughts and Water Resources", and this work is presented in the following chapters.

Chapter 2 - Methodology

This chapter describes all the materials and methodologies employed throughout this thesis. The SURFEX modelling platform, along with all of its relevant components are described in detail. The parameters and settings of all of the experiments that were performed are also laid out. Then, a description of the data used to assess how the experiments performed is provided. Included in this data are also the observations that were assimilated in the experiments. Finally, a section is dedicated to describing the process of analysis and assessment, giving important details about the statistics and filtering used throughout many of the experiments.

Chapter 3 - Importance of Initial Conditions in Forecasting Land Surface Variables

Chapter 3 is the first section dedicated to results of the experiments and analyses. The forecast configuration of the system is described as well as discussion surrounding the different ways to use it. Next, the published article "From Monitoring to Forecasting Land Surface Conditions Using a Land Data Assimilation System: Application over the Contiguous United States", is attached, with expanded introduction and conclusion sections to fully encapsulate the work done on the forecast LDAS. Finally, a section is dedicated to several experiments over a particular state of the U.S., Nebraska, which is a good test bed for some analysis due to its heavy agriculture and significant amount of irrigation. These experiments also aided in the understanding of the model, data assimilation, land use database, and potential applications at higher resolutions.

Chapter 4 - Improving Initial Conditions of the LDAS

Chapter 4 builds on several of the conclusions of Mucia et al. (2020) in Chapter 3. Initial conditions of the system were found to be of great importance for forecast applications. Therefore, the use of assimilating VOD as a proxy to LAI was investigated to potentially improve initial conditions. Before the actual assimilation was performed, significant analysis was done comparing LAI, VOD, and VOD after linear re-scaling. These comparisons show the similarities of these two observation types over different domains and vegetation types. Several experiments were then performed with a wide range of assimilated observations. Analysis was done comparing the performance of assimilating matched VOD and LAI, as well as several variations of the matched VOD observations. The last section of the chapter analyses experiments dealing with the separate and joint assimilation of vegetation and soil moisture observations.

Chapter 5 - Towards Forecasting Impacts of Extreme Events on Land Surface Variables with LDAS-Monde

Chapter 5 applies the information learned in Chapter 3 and Chapter 4 to a potential real world application as a drought early warning system. A case study over CONUS is laid out for future work where the improvements seen through Chapter 3 and Chapter 4 are applied to LDAS-Monde. Several techniques for the analysis are also given as a way to assess the performance in this case study. Additionally, the methodology proposed for the case study is tested over a non-forecast experiment, with LSVs successfully converted to percentiles and drought levels, proving the utility of the method for future studies.

Chapters 6 and 7 - Conclusion and Prospects

Chapters 6 and 7 sum up the work done in this thesis (in French and English respectively) and provide concluding remarks. Prospects of future work with LDAS-Monde are also noted.

Chapter 2

Methodology

This chapter describes the methodology for the experiments and analyses performed throughout the chapters to follow. Details regarding the SURFEX modelling platform and its components are laid out, with specific emphasis on the descriptions of the ISBA LSM, the atmospheric data used as model forcing, and the ECOCLIMAP land use databases. Then, LDAS-Monde is presented, describing the Simplified Extended Kalman Filter data assimilation scheme, as well as a detailed definition and description of all the assimilated LSV observations. Vegetation optical depth in particular, along with the relationship between VOD and other vegetation LSVs is further explored.

As this thesis is concerned with the forecasting of LSVs, the transition of LDAS-Monde from a monitoring configuration to a forecasting configuration is presented. The data used in the assessment of the LDAS's performance are then detailed, including the USCRN in situ soil moisture observations, ALEXI evapotranspiration, and FLUXCOM gross primary production. Finally, the setups of the experiments performed in the thesis are outlined, as well as describing the statistical analyses used to measure and assess the performance of the system.

2.1 SURFEX Modelling Platform

The SURFEX (SURFace EXternalisée) modelling platform (Le Moigne et al., 2009; Le Moigne et al., 2012; Le Moigne et al., 2018) is a complete surface modelling package including the ISBA (Interactions between Soil, Biosphere, and Atmosphere) land surface model (Noilhan and Planton, 1989; Mahfouf et al., 1995; Noilhan and Mahfouf, 1996; Boone, Calvet, and Noilhan, 1999) for natural surfaces, the TEB (Town energy balance) model for urban surfaces (Masson, 2000), and a land use database, ECOCLIMAP (Champeaux, Masson, and Chauvin, 2005; Faroux et al., 2013), for use in defining tile surfaces. Lake surfaces can be treated in a simple way by SURFEX, or by the separate FLake model (Mironov, 2008). Likewise, sea and oceans are able to be simulated simply or by a more complex separate model. Aerosol emission and deposition can also be simulated. Alongside these models and inter-coupling, SURFEX provides the framework for data assimilation. In general, each tile type, nature, urban surfaces, ocean, and lakes, is modeled separately then the total flux for each grid cell is given by weighing each surface type's fraction of the overall tile area. A diagram of the SURFEX platform's interactions is shown in Figure 2.1. This thesis uses latest version of SURFEX, V8.1 (Le Moigne et al., 2018).

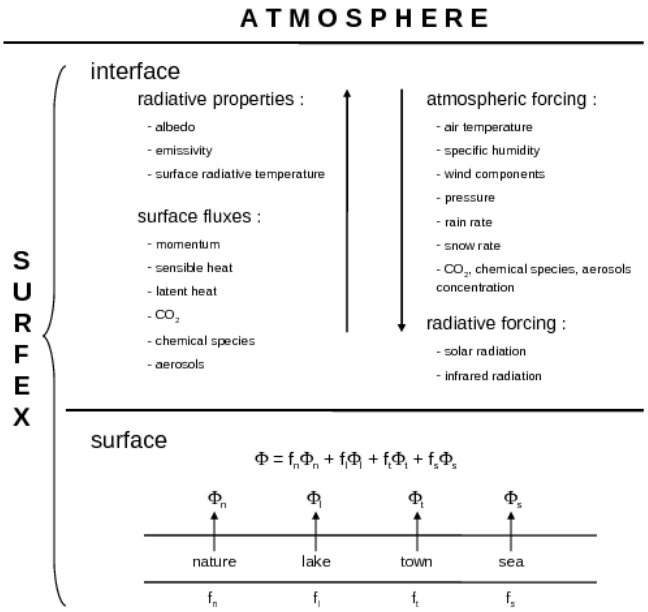

Figure 2.1: A diagram representing the components and interactions of the SURFEX modelling platform (Source: CNRM, 2020)

2.1.1 ISBA LSM

For nature tiles, the ISBA LSM simulates heat, water, carbon, and other surface fluxes. Included within ISBA are several components separately simulating

snow, hydrology, soil, and vegetation in the land surface system. The version of ISBA in this work uses the 12 layer snow parameterization scheme (Boone and Etchevers, 2001; Decharme et al., 2016), which better represents snow compaction, soil temperature, and surface albedo than previous snow schemes. Hydrologically, ISBA can also be coupled to the CNRM version of the Total Runoff Integrating Pathways (CTRIP) river-routing model (Oki and Sud, 1998; Decharme and Douville, 2007; Decharme et al., 2016), globally improving simulated discharge. Without this coupling enabled, ISBA simulates a simplified runoff routine, and this work contains experiments with and without CTRIP coupling.

The work in this thesis focuses on the evolution of vegetation, specifically in drought events. Therefore, ISBA is also required to accurately simulate vegetation and soil moisture dynamics. To that end, this thesis uses ISBA-A-gs (Calvet et al., 1998; Calvet, 2000; Calvet et al., 2004) which introduces the simulation of vegetation photosynthesis and stomatal conductance, as well as allowing for the calculation of CO_2 fluxes from photorespiration. Additionally, the *NIT* option is employed (Calvet and Soussana, 2001; Gibelin et al., 2006), which allows the modelling of above ground biomass, both leaf and structural, as well as transition the variable of LAI from being prescribed to diagnostic, based on the leaf biomass. ISBA also specifies minimum LAI thresholds, as modelled LAI that falls too low (notably in winter months) is unable to increase again in spring and summer. For evergreen forests, this threshold is 1 $m^2\ m^{-2}$, and for all other types of vegetation the threshold is 0.3 $m^2\ m^{-2}$. Gibelin et al. (2006) and Delire et al. (2020) have evolved the ISBA carbon cycle scheme introducing the NCB and NCB + CNT options. NCB is effectively the NIT option, but with conservation of carbon by using four discrete carbon reservoirs in grasses (leaves, stems, and fine roots) as well as six carbon reservoirs in trees (leaves, stems, wood, fine roots, and woody roots). Finally, NCB + CNT is the same as NCB but calculates fluxes of dead carbon in vegetation litter and soil carbon. While the NIT option is used in this thesis, NCB or NCB + CNT are clear advances which can be implemented in future evolutions of drought experiments.

The soil component of ISBA is ISBA-Diffusion (Boone et al., 2000), which uses a 14 layer grid with depths down to 12m, (0.01m, 0.04m, 0.1m, 0.2m, 0.4m, 0.6m, 0.8m, 1.0m, 1.5m, 2.0m, 3.0m, 5.0m, 8.0m, 12.0m). A mixed form of the Richards equation is used to describe water fluxes in the entire root zone. This multi-layer scheme also provides overall improved surface flux and temperature predictions, primarily due to better parameterization of latent heat from soil freezes. This Diffusion option is used as it represents the most layers of soil and has the most up to date soil physics of all the soil component options.

Atmospheric Forcing

ISBA has the capability to be coupled to atmospheric models and thus process atmospheric feedbacks (also known as "online"), but this thesis runs ISBA as

without this coupling or feedbacks enabled (also known as "offline"). Therefore, it is typically forced by static atmospheric reanalyses. The model requires the atmospheric variables at 2m above the surface of air temperature, wind speed, air specific humidity, atmospheric pressure, shortwave and longwave downwelling radiation, and liquid and solid precipitation that are ingested into ISBA, driving land surface processes. Through recent updates, LDAS-Monde now has the ability to run in forecast mode (Albergel et al., 2019; Albergel et al., 2020; Mucia et al., 2020) that is used in this thesis, where ISBA can accept daily forecasts and produce individual outputs for each of the forecast time steps.

When in the standard monitoring configuration, the system ingests atmospheric reanalyses from ECMWF's ERA5 (ERA5, 2018). ERA5 provides hourly data, globally over a 0.25° x 0.25° grid. This reanalysis is itself a product of data assimilation, combining model data and observations around the world to create this consistent dataset from 1950-present. ERA5 assimilates atmospheric observations every 12 hours, which updates to a new, more accurate forecast. Its uncertainty is measured by sampling a 10 member ensemble every 3 hours, and the mean and spread of the ensemble is pre-computed and provided to users. While not a real-time product, preliminary ERA5 data is available with an approximate 5 day delay, with a higher quality controlled release after 2-3 months. With this relatively low latency, ERA5 suits the needs of multi-annual or even multi-decadal monitoring studies very well. Before ERA5 was released, ERA-Interim provided similar products via the same data assimilation approach. Albergel et al. (2018a) compared ERA5 against ERA-Interim when driving ISBA, and found that ERA5 provided consistent improvement for nearly all land surface variables, specifically for variables linked to the terrestrial hydrological cycle. This also underscores the necessity of high quality atmospheric forcing for LSMs.

The forecast configuration requires the same atmospheric variables, and treats all the land processes the same. However, the atmospheric forcing is atmospheric forecasts instead of reanalyses. Twice daily, ECMWF runs a global 15-day ensemble forecast at 0.20° x 0.20° spatial resolution (ECMWF, 2018). This 51-member ensemble consists of a single control run (CTRL or ENS CTRL) as an unperturbed forecast, and 50 perturbed members. These perturbed members begin with the slightly perturbed initial conditions and model physics compared to the control member. These perturbations allow an exploration of the model uncertainty. These ensemble forecasts are also commonly used to investigate extreme weather such as tropical cyclones and heavy precipitation events, as well as being simple probability guides, often indicating warmer or colder than average temperatures or precipitation thresholds. This product has varying time-steps, with hourly data out to day 3, three-hourly out to day 6, and six-hourly out to day 15.

In addition to the 15-day ensemble forecast produced by ECMWF, a high resolution (HRES) 10-day forecast is produced at 0.10° x 0.10° spatial resolution using a single prediction model run instead of ensemble (ECMWF, 2018). This HRES product provides a more detailed description of future weather compared

to any individual ENS member, but contains the same physical and dynamic atmospheric representation. This HRES forcing was used over the domain of the U.S. state of Nebraska to test the higher resolution capabilities of LDAS-Monde. A comparison of the different forcings and forecasts used in this thesis are given in Table 2.1.

Table 2.1: Details of atmospheric forcing datasets used in this thesis

Forcing	Spatial Resolution	Temporal Availability	Forecast Length
ERA5	0.25° x 0.25°	1950-present	None
HRES	0.1° x 0.1°	April 2016- Present	10-Day
ENS CTRL	0.2° x 0.2°	April 2016- Present	15-Day

Land Use: ECOCLIMAP II & ECOCLIMAP SG

As previously mentioned, SURFEX uses the ECOCLIMAP land use database. Throughout several of the experiments in this thesis, two different versions of ECOCLIMAP are provided to the model, ECOCLIMAP-II (Faroux et al., 2013) and ECOCLIMAP Second Generation (SG) (CNRM, 2018; Calvet and Champeaux, 2020). Both versions of ECOCLIMAP include the 12 land surface types that are required for the version of ISBA used. These land surface types are divided into 9 predefined functional plant types (deciduous broadleaf forests, evergreen broadleaf trees, needleleaf forests, C3 crops, C4 crops, irrigated crops, C3 grasslands, C4 grasslands, and wetlands), as well as including bare soil, rock, and permanent snow and ice surfaces. For cases where the urban model TEB is not enabled, SURFEX converts urban areas defined by ECOCLIMAP into rock for use in ISBA.

ECOCLIMAP-II is a 1km resolution global scale database departing from previous land cover products by creating more classes which should ideally better represent regional environments. This database consists of a synthesis of existing land cover maps such as Corine Land Cover Map 2000 (CLC, 2000) and Global Land Cover 2000 (GLC, 2000), NDVI observations from SPOT VGT (Maisongrande, Duchemin, and Dedieu, 2004), LAI observations from MODIS (Yang et al., 2006), and CYCLOPES (Baret et al., 2007), and climatic datasets of the Koeppe classifications (Koeppe and De Long, 1958) and FIRS (Folving, Kennedy, and Megier, 1995). The merging of satellite sensor data strengthens the reliability of the product by making it less sensor dependant. ECOCLIMAP-II provides tiles or grid points that represent fractions of the four main surface types (nature, inland water, sea, or urban), and inside the nature tile fraction, are fractions of each plant functional type.

ECOCLIMAP-SG is an evolution of ECOCLIMAP-II, with an improved 300m resolution. At this native resolution, ECOCLIMAP-SG also abandons the notion of fractional vegetation cover, instead simplifying each grid point to represent a single functional vegetation type. The changes from ECOCLIMAP-II to ECOCLIMAP-SG do cause visible differences in dominant land cover over

the United States. Figure 2.2 demonstrates some of these differences. This figure shows the category of a patch where a single vegetation type is greater than 50%.

These differences are most visible over the Midwest and Southwest United States, as well as nearly all of Mexico seen in this domain. In the Midwest, strong changes between ECOCLIMAP-II and ECOCLIMAP-SG are seen reducing the land dominated by C4 crops which are primarily replace by areas dominated by C3 grasslands. In the US Southwest, much of the area dominated by bare soil or with no dominant vegetation in ECOCLIMAP-II are now dominated by deciduous forests in ECOCLIMAP-SG. Similarly, much of Mexico is dominated by bare soil, C4 grasslands, or has no dominant vegetation in ECOCLIMAP-II, but is almost entirely transformed to deciduous and coniferous forests in ECOCLIMAP-SG. Further analysis into the differences between these two products are possible, such as analyses comparing the fraction of all plant function types instead of simply visualizing differences where there is a single vegetation type covering more than 50% of the patch. However, these deeper analyses and further validation are outside the scope of this thesis.

2.1.2 LDAS-Monde

The land data assimilation system LDAS-Monde (Albergel et al., 2017) enables the integration of satellite observational products into the ISBA LSM. As its name suggests, LDAS-Monde is a global scale system. It can assimilate observations to directly update 8 control variables comprised of LAI and 7 soil moisture layers from 1cm-100cm depth. Additional variables are indirectly modified by the assimilation through their biophysical feedbacks. Because each observation directly updates LAI and soil moisture layers, even the assimilation of LAI by itself allows for an analysis of the root zone soil moisture (0-100cm), which is a unique capability of LDAS-Monde. Table 2.2 provides details about the LDAS-Monde parameters.

LDAS-Monde uses a Simplified Extended Kalman Filter (SEKF) as the default data assimilation scheme (described in the section below), but experiments have also allowed for an Ensemble Kalman Filter (EnKF) and an Ensemble Square Root Filter (EnSRF) (Bonan et al., 2020) schemes.

Table 2.2: Details of LDAS-Monde Assimilation Parameters

Model	Assimilated Observations	Model Equivalents of Observations	Control Variables
ISBA-A-gs, NIT, Diffusion	LAI (CGLS), VOD (VODCA), SSM (CGLS, ESA-CCI)	LAI (for LAI and VOD) Soil Layer WG2 (1-4cm)	LAI, Soil Layers WG2-WG8 (1-100cm)

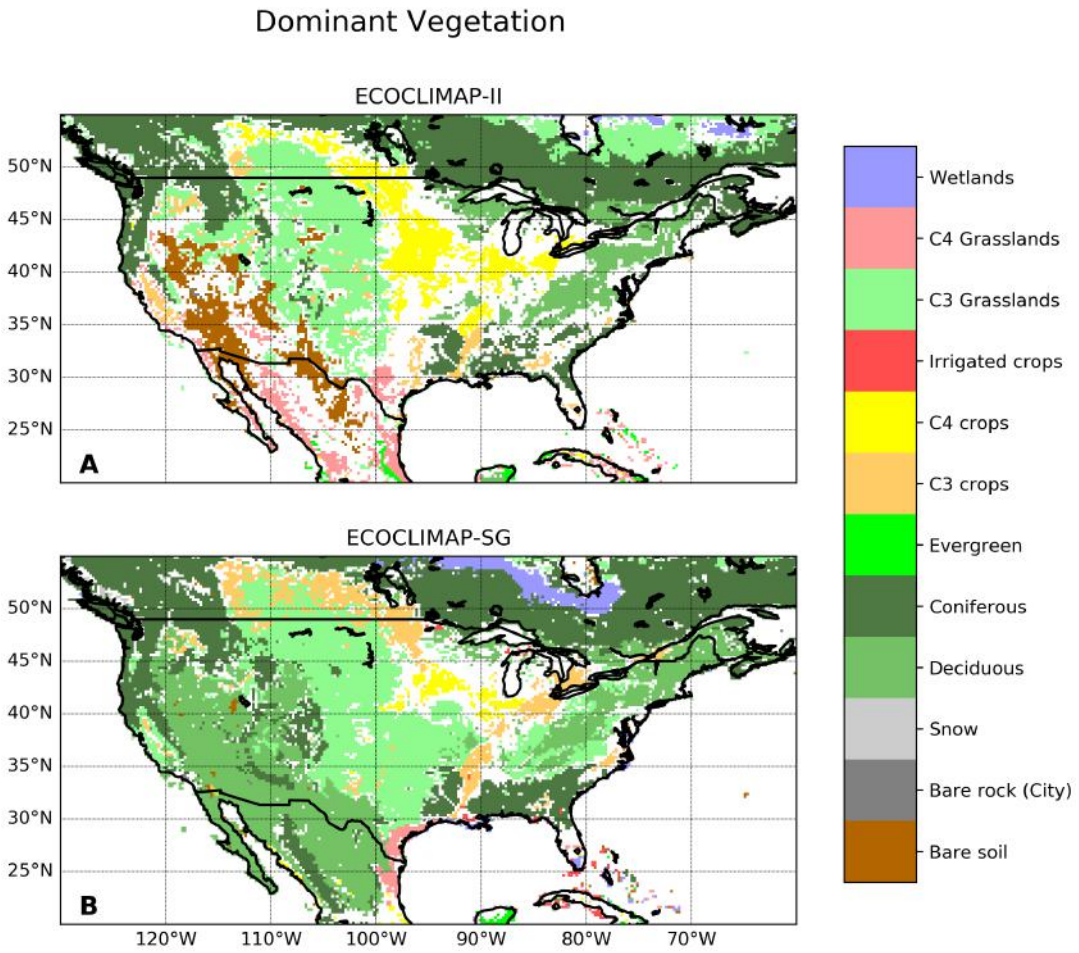

Figure 2.2: Map of dominant vegetation types (>50% single vegetation) according to a) ECOCLIMAP-II, and b) ECOCLIMAP-SG. White areas over land indicate no single vegetation type covers more than 50% of the tile.

Simplified Extended Kalman Filter

The Extended Kalman Filter (EKF) is computed using the $\boldsymbol{H}$ and $\boldsymbol{H^T}$ terms from Equation 1.8 which are the linearized version of the observation operator, also called the Jacobian matrix. Each element of this matrix is approximated via finite differences by perturbing each component (x_j) of the control vector $(\boldsymbol{x})$ by the amount (δx_j), as shown in Equation 2.1. The result is a column of the Jacobian matrix for each m integrations.

$$\boldsymbol{H}_{mj} = \frac{\delta y_m}{\delta x_j} \approx \frac{y_m(\boldsymbol{x} + \delta x_j) - y_m}{\delta x_j} \tag{2.1}$$

Prognostic equations representing the physical processes of the LSM then evolve the control vector to the end of the 24 hour assimilation window. Observations from the previous 24 hours are then assimilated forming the analyzed initial state of the next 24 hour period. Instead of calculating background error variances and covariances at the beginning of each cycle, fixed estimates are

used. This is the step that differentiates an EKF and SEKF. In LDAS-Monde, error values are fixed as 20% of observed LAI values and at a constant 0.05 m^3/m^3 for SSM. This simplification is continuing with previous studies (Mahfouf et al., 2009; Albergel et al., 2010; Barbu et al., 2011; Fairbairn et al., 2017) and Draper, Mahfouf, and Walker (2009) shows that the SEKF's 24hr assimilation window generates flow dependence and that the background error cycling of the EKF gave no significant additional benefit.

2.1.3 Assimilated Observations

LDAS-Monde can jointly assimilate observations as both LAI and SSM into the ISBA LSM. These observations are put into a simple text file format, with each file representing observations that day over every gridpoint in the domain. Observations sources are not limited by the model, but only by the workflow and acquisition of the data. The following sections describe the observational data that are assimilated in this thesis.

Surface Soil Moisture

Surface soil moisture (SSM) is a measure of the relative water content contained within the top few centimeters of soil (CGLS, 2020). The exact depth at which soil moisture ceases to be "Surface" soil moisture is dependant on discipline and domain. SSM is typically presented in terms of percent saturation or as a volumetric measure m^3/m^3.

The assimilation of surface soil moisture requires that the observations are re-scaled to fit consistently with the model climatology (Reichle and Koster, 2004; Drusch, Wood, and Gao, 2005). Otherwise, observation signals exhibit considerably different mean values as well as biased variability. This transition to the model-equivalent, volumetric SSM also addresses potential parameters such as porosity, field capacity and wilting point, but cannot account for errors in model physics or observation retrieval methods (Entin et al., 1999; Dirmeyer, Guo, and Gao, 2004). For LDAS-Monde, a linear re-scaling approach is used, as originally prescribed by (Scipal, Drusch, and Wagner, 2008), which simplifies a transformation by cumulative distribution function (CDF) matching as described in Equations 2.2 2.3 and 2.4. This simplification leads to removing differences in the first and second moments, the mean and variance respectively. Higher order moments produce stronger differences in dryer climates, and these differences are assimilated directly as uncorrected bias. Scipal, Drusch, and Wagner (2008) estimates these differences are still small and should scarcely reach values larger than 0.02 $m^3\ m^{-3}$.

$$SSM_r = a + b \cdot SSM_o \tag{2.2}$$

$$a = \overline{SSM_m} - b \cdot \overline{SSM_o} \tag{2.3}$$

$$b = \frac{\sigma_m}{\sigma_o} \tag{2.4}$$

SSM_r is the re-scaled SSM, SSM_o is the un-transformed observed SSM, and SSM_m is the model SSM. $\overline{SSM_m}$, $\overline{SSM_o}$, σ_m, and σ_o are the model and observation means and standard deviations respectively. The linear re-scaling parameters a and b are calculated monthly over a 3-month moving window for the period of each respective experiment, filtering out ice and altitudes over 1500m as well as pixels with urban surfaces exceeding 15% of the land cover fraction.

This thesis uses two separate sources of satellite observed SSM products. The first is from the Copernicus Global Land Services (CGLS) (Copernicus, 2021) derived from the Advanced Scatterometer (ASCAT) microwave instrument aboard the MetOp A and B polar orbiting satellites (Wagner, Lemoine, and Rott, 1999; Bartalis et al., 2007). The SSM product synthesizes both sensor's data daily, producing information based on the radar backscatter observations from 2007-present. These observations are then transformed into an estimated soil wetness index (SWI) by applying an exponential filter (Wagner, Lemoine, and Rott, 1999; Albergel et al., 2008) on a timescale parameter, T, varying between 1 and 100 days. This transformation is done over the T timescale to account for physical parameters such as soil type, thickness, texture, and density, as well as evaporation and runoff. Experiments in this study using this SWI product set the T value to 1 day. This T=1 is selected as it is the shortest time period, which effectively represents the most shallow layer of soil moisture assuming soil diffusivity is constant (Marschallinger, Paulik, and Jacobs, 2019). SWI is expressed as a saturation percentage, with 0 being completely dry and 100 as completely saturated. This CGLS SWI product is available globally at 0.1° x 0.1° spatial resolution.

The second source of SSM observations comes from the European Space Agency's (ESA) Climate Change Initiative (CCI) SM product (Dorigo et al., 2015; Gruber et al., 2017; Gruber et al., 2019). Version v04.5 of the COMBINED product is used, that merges active and passive microwave observations into a single product (Dorigo et al., 2017). Instruments merged in this product include four scatterometers (ERS-1/2 AMI, ERS-2 AMI, and MetOp-A & B ASCAT) and seven radiometers (SSMR, SSM/I, TMI, WindSat, AMSR-E, AMSR-2, and SMOS). Elevations greater than 1500m are filtered out of this product for assimilation, as soil moisture retrievals are difficult in complex terrain such as slopes, ridges, and valleys in close proximity often found at high elevations (Mätzler and Standley, 2000). Together, this product provides globally continuous coverage from November 1978-present, daily, at a spatial resolution of 0.25° x 0.25°.

Leaf Area Index

Leaf area index, or LAI, is the sum of the one-sided area of a leaf's surface per unit area of land (Watson, 1947). This index is a very useful metric, allowing for the comparison of natural vegetation and crops despite potentially different plant spacing. LAI has proven to be a key parameter when dealing with plant physiology (Breda, 2003), as well as being strongly linked to vegetation biomass (Friedl et al., 1994; Gitelson et al., 2003).

Assimilated LAI observations in this thesis come from the CGLS LAI V2 product (CGLS, 2019). The observations come from the SPOT/VGT and PROBA-V sensors. The top-of-canopy (TOC) reflectance is input into a neural network for instantaneous LAI estimates. The V2 algorithm then applies filtering, smoothing, gap filling, and temporal compositional techniques to derive consistent LAI estimates every 10 days (Verger, Baret, and Weiss, 2014). The product is also compared with various datasets following the CEOS Land Product Validation Group's guidelines to ensure consistency with other LAI datasets. The quality of the final product is highly dependant on the retrieval algorithm. As neural networks were trained using SPOT/VGT inputs, the application to PROBA-V data assumes sensor consistency between the two satellites. Additionally, pixels that are incorrectly categorized or oversimplified may produce incoherent or biased values, or potentially not capture seasonal cycles of vegetation. Some of these errors can be limited due to the higher spatial resolution of the product. It is important to note these errors, especially in the context of data assimilation, where it is assumed that neither the model nor the observations perfectly represent reality, but the acknowledgement and quantification of these errors is still necessary to successfully merge the two methods of monitoring LSVs. CGLS LAI V2 is available at 1km x 1km spatial resolution and from 1999-present.

Vegetation Optical Depth

What is Vegetation Optical Depth?

Up to this point, the assimilation of LAI has been of direct estimations of LAI from optical observations. For the testing of improving initial conditions of the model in Chapter 4, vegetation optical depth is used and transformed into an LAI-proxy. Kumar et al. (2020) has already shown that VOD assimilation as an LAI proxy is possible, with the linear re-scaling and assimilation into the Noah-MP LSM.

Vegetation optical depth, or VOD, is the measure of attenuation of microwave radiation passing through a vegetation canopy (Jackson and Schmugge, 1991). This attenuation is a function of microwave frequency, and can also be directly linked to vegetation water content (Jackson, Schmugge, and Wang, 1982; Wigneron et al., 1993; Owe, De Jeu, and Walker, 2001). Being the product of microwave radiation, VOD is a nearly all-weather parameter, passing through cloud cover almost unaffected. This allows for far more frequent VOD observations compared to LAI observations, which is demonstrated in Figure 2.3. For the same 2003-2018 period, LAI observations from CGLS are vastly outnumbered by VOD observations from VODCA by approximately a factor of 6.

VOD is comprised of attenuation from several sources, primarily standing vegetation (which itself is composed of green leaf biomass, non-green structural biomass such as stems, and wood), necromass (namely litter), and intercepted water from rain or dew. Recently, VOD has been more closely examined in regards to interacting effects of vegetation dynamics. A deeper look into L-band VOD by Konings et al. (2016) revealed that it is proportional to total

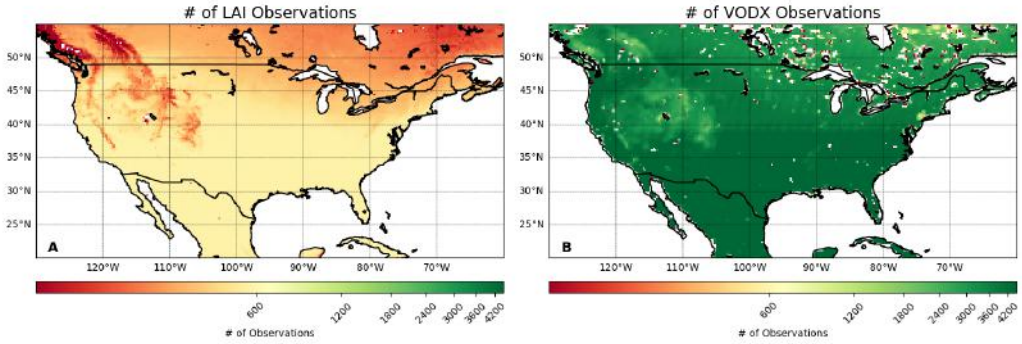

Figure 2.3: Maps showing the cumulative number of observations provided for A) CGLS LAI and B) VODCA VODX over CONUS between 2003-2018. The colorbar is on a log scale in order to show the vast differences between the two datasets.

vegetation water content and tweaks their retrieval algorithm to more accurately account for vegetation effects on soil moisture observations. Figure 2.4 shows an extreme example of the effect of the different components of VOD over a SMOSREX (Surface Monitoring of the Soil Reservoir EXperiment) site near Toulouse, France from Saleh et al. (2006). This is shown in order to visualized the individual impact of these components, namely the green vegetation (solid black line), ground litter (black dots), and leaf water interception (grey line).

Relationship between vegetation, VOD, and LAI

Momen et al. (2017) found that X-band VOD produced temporal dynamics closely resembling a new conception relationship between VOD and total biomass. This new framework provided significantly increased correlations between VOD, biomass, and leaf water potential. Then, Teubner et al. (2018) linked VOD and gross primary production (GPP), noting an similar strength or stronger relationship than between VOD and solar induced fluorescence (SIF). Teubner et al. (2021) also shows that X-band VOD provides improved average correlation coefficients to GPP estimates from in situ FLUXNET observations than L-band VOD, shown in Figure 2.5.

Additionally, Wigneron (2002) found the timing of peak vegetation water content and LAI can differ. When leaf biomass is the dominant signal from VOD, we assume:

$$VOD = B \times V_2 = B \times LAI \tag{2.5}$$

Where the B is a constant from the calculations of VOD based on land cover classification. V_2 is simply a factor representing a vegetation descriptor (Ulaby et al., 1984), which we choose to be approximated by LAI, as in Shamambo (2020). The Specific Leaf Area (SLA) is defined as the ratio of LAI/VOD, and is an indication of how much leaf area a plant makes for a given amount of leaf biomass. This also allows for:

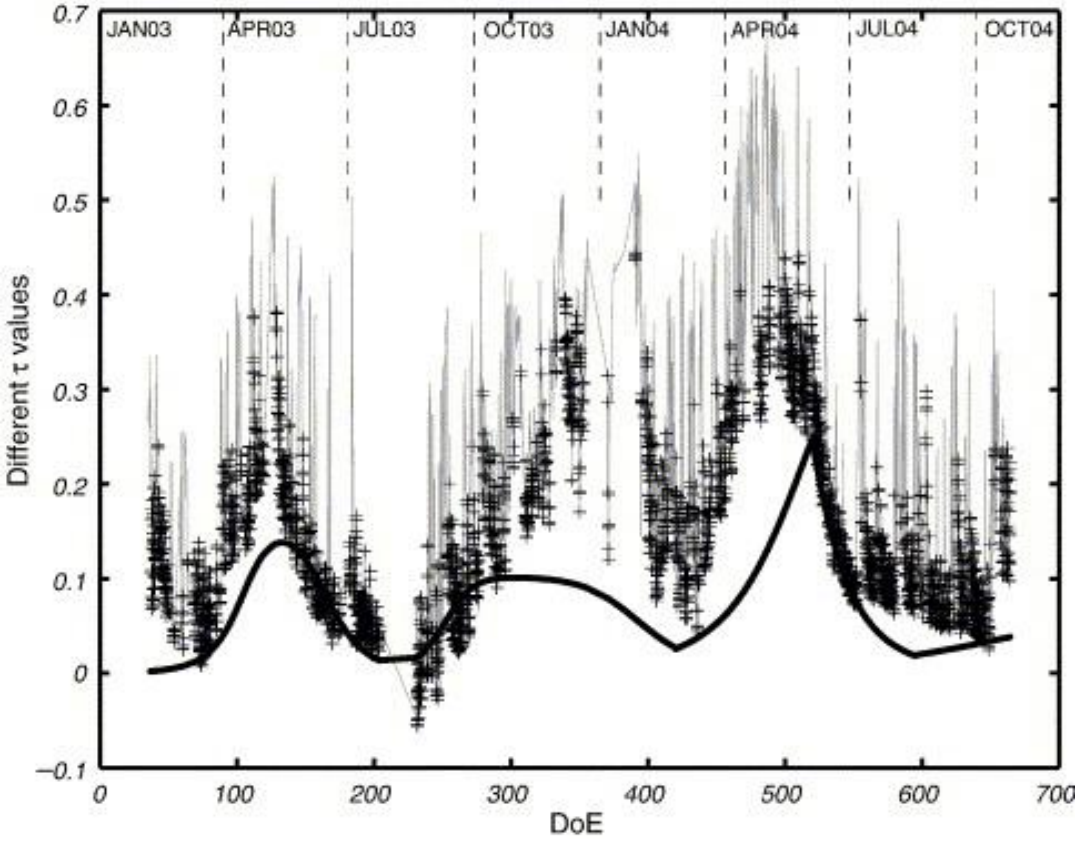

Figure 2.4: A time series of estimated VOD, with the solid black line representing green vegetation, black dots representing where there is no standing vegetation (i.e. litter effect), and light grey line representing estimates including immediately after rainfall events (Source: Saleh et al., 2006).

$$LAI/VOD = 1/B \tag{2.6}$$

Which shows an inverse relationship between SLA and B. In LSMs, SLA is a key parameter, and generally assumed to be constant. But this may not hold true, as Zakharova et al. (2012) found that the B parameter changes quite significantly between areas of low and high vegetation.

Figure 2.6 shows the time series response of CGLS LAI (green, solid), VODCA VODX (red, dashed), and VODCA VODC (blue, dotted) over Lincoln, Nebraska from 2003-2018. This pixel is composed primarily of C3 and C4 crops. LAI observations have a far more predictable and seasonal pattern. X-band VOD also is a stronger signal compared to C-band. The peaks are relatively close in timing in this case, but as discussed above, can also be offset due to the difference in peak vegetation water content.

This patch near Lincoln, Nebraska represents a mix between C3 and C4 crops, and does not include significant areas of irrigation. Shamambo et al. (2019) found over sites in France, that LAI observations from CGLS GEOV2 had two seasonal peaks, one in spring and one later in summer (Figure 10). The

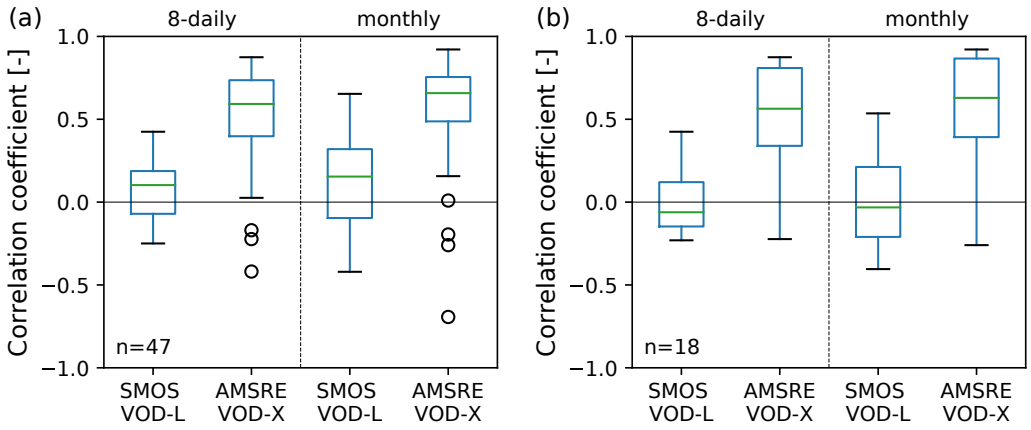

Figure 2.5: Correlations of VOD-L from SMOS and VOD-X from AMSRE to FLUXNET in situ observations of GPP. a) represents the stations where both VOD datasets were available, while b) represents only forest landcover. (Source: Adapted from Teubner et al., 2021).

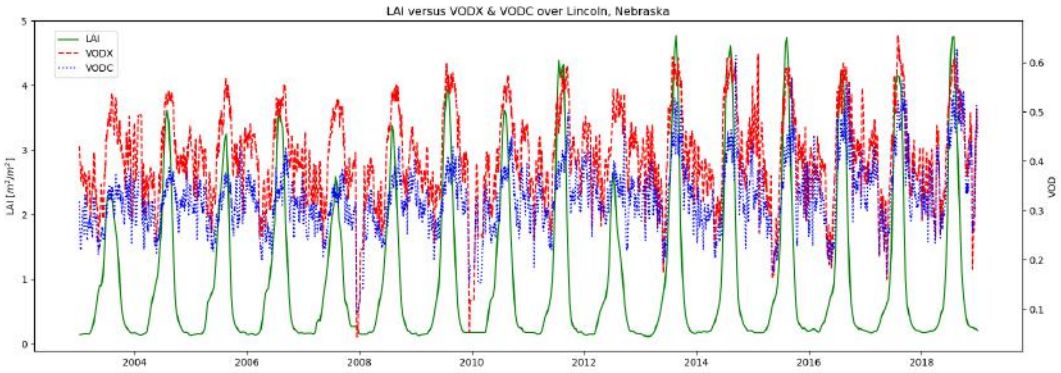

Figure 2.6: A time series showing the response of CGLS LAI (green, solid), VODCA VODX (red, dashed), and VODCA VODC (blue, dotted) over Lincoln Nebraska from 2003-2018.

later second peak was strongest in the area with more C4 crops. As Figure 2.6 shows, no such double peak can be seen over Nebraska. Similarly, in a study over a heavily irrigated patch of agriculture in Nebraska, Druel et al. (2021) found no such double peak from Observed LAI. These differences in results can simply be attributed to significantly different agricultural practices between Nebraska and France such as planting time, harvesting time, and crop rotations, as well a strongly different regional climates.

VODCA: Long-term VOD dataset

This work uses the X-Band of the newly created Vegetation Optical Depth Climate Archive (VODCA) (Moesinger et al., 2020) dataset. VODCA is a synthesis of various satellite sensors since 1987. It uses the Land Parameter Retrieval Model (LPRM) V6, which simultaneously retrieves and calculates

soil moisture and VOD from horizontally and vertically polarized microwave observations (Mo et al., 1982; Meesters, DeJeu, and Owe, 2005; Owe, Jeu, and Holmes, 2008; Schalie et al., 2017). The dataset is comprised from AMSR-E, AMSR-2, SSM/I, TMI, and WindSat sensors, and includes C, X, and Ku-Band VOD retrievals. The TMI sensor aboard the Tropical Rainfall Measuring Mission (TRMM) satellite is a notable exception in that its orbit is at a 35° inclination, and thus does not encompass the entirety of the domains in this thesis. For the X-Band of VODCA, TMI is the only sensor between 1998 and late 2002.

Each VODCA sensor source is first processed by removing locations influenced by radio frequency interference (RFI), removing observations where land surface temperature (LST) is below freezing (as the VOD cannot be accurately retrieved in frozen conditions due to the changing di-electric properties of water and ice), and removing negative values of VOD, which are data artifacts and not physically possible. Because daytime retrievals were found to have higher errors than their nighttime counterparts, only nighttime retrievals are used in VODCA. The sensor datasets are then individually matched based on the VOD band by using an improved CDF matching scheme to correct for systematic differences between the sensors (details of the improvements found in Moesinger et al. (2020)). The bands are then finally merged via arithmetic mean when multiple observations are available. This product provides a spatially continuous dataset, that is user ready. However, it is important to note that not only do errors exist in the observations themselves (such as those from the mischaracterization of the surface, interference, the retrieval model, or direct sensor errors), but these errors can be merged and blurred through the merging and combination of data (such as the mixing of observation times, incidence angles, and spatial resolutions). The evaluation of datasets such as this is non-trivial as independent and high quality datasets for comparison are rare or non-existent (Moesinger et al., 2020). As more products become available for comparison, independent evaluation products such as VODCA also can be performed.

This VODCA dataset is globally available at 0.25° x 0.25° spatial resolution. Because of the different number of sensors depending on each VOD band, and the timing of their overpasses, the merged product provides observations for at least 40% of all days with at least one sensor, and upwards of 70% with two or more. Figure 2.7 illustrates the temporal coverage for each microwave band in VODCA.

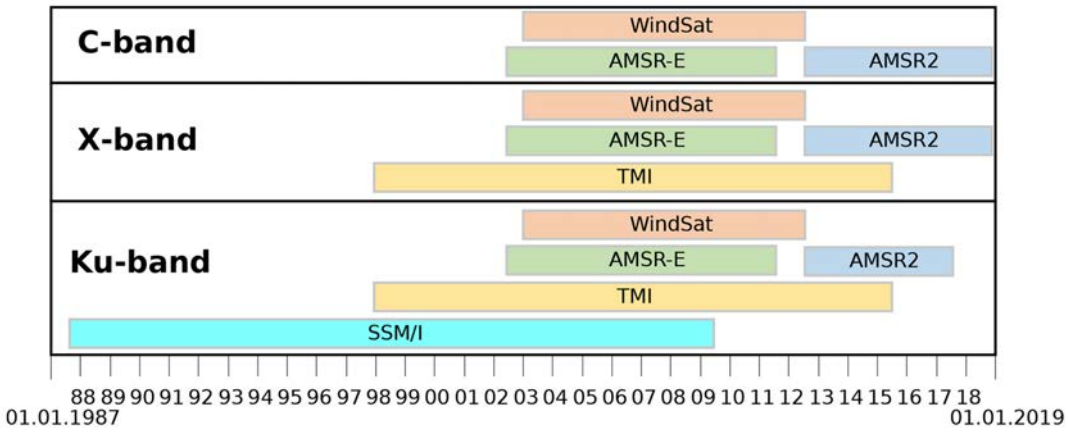

Figure 2.7: Time periods for each of the microwave sensors used in VODCA, for the C, X, and Ku Bands of VOD (Source: Moesinger et al., 2020)

Linear Re-scaling VOD to match LAI

As was done in Kumar et al. (2020), VOD observations are linearly re-scaled to match LAI observations over the same time period. A linear monthly re-scaling was performed using a 3-month moving window period to best match the two datasets over seasonal timescales. Over an entire year, this re-scaling is represented by 12 monthly equations each taking into account the climatologies of the months preceding and succeeding it. This re-scaling is applied because the ISBA LSM does not simulate VOD, and thus we cannot assimilate VOD data directly. As demonstrated in Albergel et al. (2018a) and as will be shown in Chapter 4, LAI and VOD observations are correlated. This assumption enables us to match the VOD to LAI observations and assimilate the resulting product in the place of LAI in the model. After re-scaling is performed, a 30-day rolling average is applied to act as a smoothing term, allowing for better performance of the assimilated data.

This method of re-scaling logically increases the match between LAI and the re-scaled VOD observations. It is a relatively simple method allow us to test the more frequent assimilation of satellite-derived observations in LDAS-Monde. However, weakness to this method exist, primarily in the merging of errors associated with both LAI and VOD observations (each briefly mentioned in the data descriptions), as well as errors stemming from the 3-month moving window and monthly re-scaling period. The re-scaling period could introduce errors due to the potential time lag between LAI and VOD peaks, as discussed above. Future experimentation and analysis could be performed using a daily re-scaling instead of monthly in order to better capture faster changes in seasonality.

Assimilating Level 1 Data

Recently, there has been a tendency towards directly assimilating lower level satellite observation data (level 1) in order to better preserve the contained information, whereas processing the data to level 2 or level 3 products introduces errors through cross correlations to required auxiliary information and removes

potentially valuable information related to vegetation and the surface (De Lannoy et al., 2012; Han et al., 2014; Lievens et al., 2015; Lievens et al., 2017). One such example of a level 1 product is radar backscatter observations, σ°. Directly assimilating such observations would require a coupling of the LSM to a radiative transfer model, predicting the σ°. Such work has been performed with the ISBA LSM in Shamambo et al. (2019) and Shamambo (2020), and shows neutral to moderate improvements, which can be strengthened by finding an optimal observation error covariance matrix. VOD and σ° are directly intertwined, and thus future work in the assimilation of VOD observations could be potentially improved by more direct assimilation into the LSM.

2.1.4 LDAS-Monde in Forecast Configuration

LDAS-Monde now has the ability to produce land surface forecasts when ingesting atmospheric forecasts as forcings. Uncoupled forecasts of LSVs and their agricultural impacts have been previously studied, as in Sawada et al. (2020). However, land surface forecasts using LDAS-Monde's capabilities to jointly assimilate LAI and SSM observations, and to analyze RZSM by use of assimilating LAI, are yet unparalleled. Sawada et al. (2020) finds that initial conditions, improved by the assimilation of brightness temperatures, improve the seasonal prediction of LAI. Using the new forecast configuration of LDAS-Monde, this thesis builds upon that idea and investigates joint assimilation of vegetation and soil moisture parameters towards improving the initial conditions.

The physical parameterization within the model works identically to that of a "normal" experiment using atmospheric reanalyses as forcings. The output variables are also identical. Figure 2.8 shows the different components of LDAS-Monde in both non-forecast and forecast modes. A typical configuration of LDAS-Monde produces three or four sets of daily netCDF files describing the output. These files include separated diagnostic and prognostic variables. In a forecast configuration, each day of the experimental time period creates those same files, instead 15 times, starting with the initial day, and moving to 15 days of forecast. Due to the timing mismatch between the data assimilation process occurring at 0900UTC, and the forecast data starting at 0000UTC, this allows for only 14 full days of forecast.

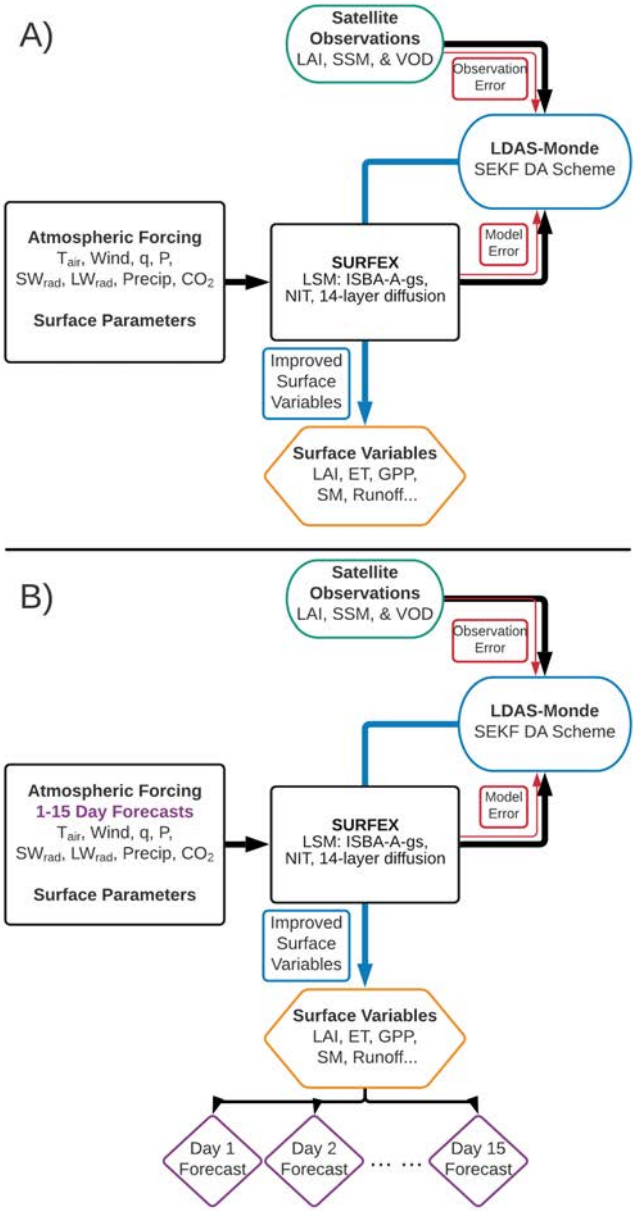

Figure 2.8: LDAS-Monde components in A) "normal", non-forecast configuration, and B) forecast configuration

As a data assimilation system, LDAS-Monde ingests observations into its SEKF scheme at the time of observations. In forecast mode, those observations are only assimilated at day 1, and are forced throughout the remaining forecast period exclusively by atmospheric forecasts. In this way, comparisons between the Open-Loop (OL), that is the model run by itself with no data assimilation, and SEKF runs at forecast days greater than 1 are, in fact, comparing how the OL and SEKF initial conditions perform at that forecast length.

2.2 Assessment Observations

This thesis uses various datasets of observations to quantify and assess the performance of the model and data assimilation. While there is no perfect reference data for any variable, the datasets listed below provide useful insights into the behavior of the model and of the assimilation. Satellite observations and satellite derived data are predominately used as they offer the best spatial coverage of the domains. In situ observations are also used, and are generally considered higher quality, but they only represent small points within a model grid cell.

2.2.1 Assimilated Observations

The first data that is compared to the model and analysis are generally the data that itself was assimilated. This comparison is a heavily biased and dependant one, but can give insights as to the effect of the assimilation. If the assimilation process is performing as expected, the assimilated values should strongly tend towards the observations and statistical scores compared to the observations should increase. Figure 2.9 shows the average values of the assimilated observations (VOD shown before linear re-scaling) over CONUS.

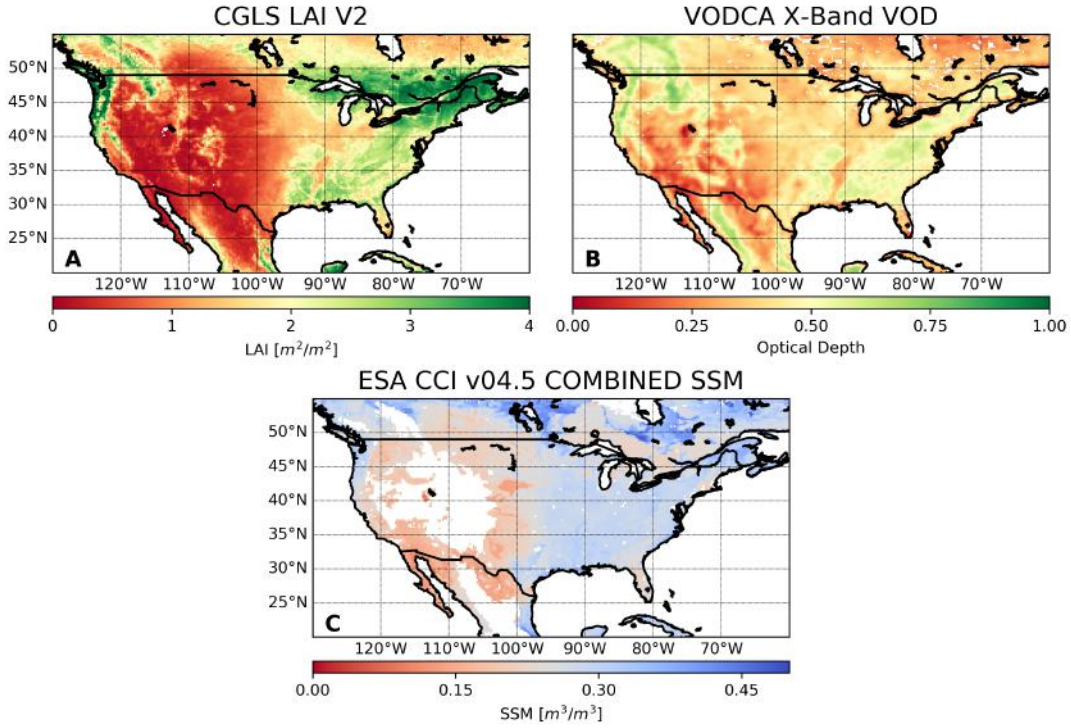

Figure 2.9: Map illustrating the average values of each of the assimilated observations datasets between 2003 and 2018. A) LAI V2 from CGLS B) X-Band VOD (before re-scaling) from VODCA C) SSM from ESA CCI v04.5 COMBINED

2.2.2 United States Climate Reference Network

The United States Climate Reference Network (USCRN) is a sustained network of climate monitoring stations maintained by the National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Information (NCEI) (Diamond et al., 2013; Bell et al., 2013). The network contains 114 stations in the contiguous U.S. and provides high quality, long-term temperature, precipitation, solar radiation, wind speed, humidity, soil moisture, and soil temperature observations. This study uses the soil temperature and soil moisture observations, which are provided sub-hourly.

At each site, USCRN places three plots of probe units at five different depths, 5, 10, 20, 50, and 100cm. The soil moisture probe measures the dielectric permittivity of the soil by observing reflected EM waves at 50MHz, which is then converted to volumetic soil moisture ($m^3\ m^{-3}$) via a calibration equation. Sensor calibration is also performed annually. A thermistor is also placed alongside the soil moisture sensor at all plots and depths. An average at each depth is calculated from the three plots every 5 minutes and output data is typically publicly available within an hour of the reading. Figure 2.10 shows the locations of the USCRN in situ observations.

This thesis compares USCRN data to LDAS-Monde soil moisture between the years of 2011 and 2018. While the network was operational as early as 2005, 2011 was selected as the start of the comparison in order to maximize the number of stations, and homogenize the results of comparisons between stations.

2.2.3 ALEXI Evapotranspiration

Evapotranspiration (ET) is a broad term including many individual components and sources of evaporation and transpiration. These components include leaf transpiration, bare-soil evaporation, interception loss, surface water evaporation, and sublimation. ET is also strongly coupled with ecosystem production (Law et al., 2002), which in turn is driven by water availability (Noy-Meir, 1973). Therefore, measuring and predicting ET can be a valuable asset in terms of monitoring and predicting agricultural droughts.

The Atmosphere-Land Exchange Inverse (ALEXI) is a surface energy balance model, which calculates evapotranspiration (ET) from a two-source land surface representation of the energy budget (Anderson et al., 1997; Anderson et al., 2007a; Anderson et al., 2007b; Anderson et al., 2011). The land surface is treated as a combination of soil and vegetation in the model, with each having unique temperatures, fluxes, and coupling with the atmosphere. Thermal infrared (TIR) bands from the Geostationary Operational Environmental Satellite (GOES) sensors estimate land surface temperature (LST) and provide the driving force for ALEXI over the United States, with Meteosat Second Generation (MSG) providing data over Europe and Africa. Global products use the Geoland2 land cover database (Lacaze et al., 2010) to estimate LST. Regional vegetation cover is estimated from MODIS-derived LAI products. Aerodynamic and atmospheric boundary layer conditions are derived from North American

Regional Reanalysis (NARR) (Mesinger et al., 2006), Weather Research and Forecasting model (WRF) (Skamarock et al., 2005), and the Modern-era Retrospective Analysis for Research and Applications (MERRA) (Bosilovich et al., 2006; Gelaro et al., 2017) for the US, European/African, and Global domains respectively. Finally, the University of Maryland's global landcover classification (Hansen et al., 2000) is used to define surface characteristics over all domains. The ALEXI ET product is available at a spatial resolution of 0.05° x 0.05° globally, and at 0.04° x 0.04° over CONUS.

While there are several other datasets that estimate ET, including GLEAM and FLUXCOM, ALEXI was chosen as in these studies as previous comparisons of ET estimations over CONUS showed it to be among the most consistent, especially over croplands (Kumar et al., 2019). Further analyses could very well compare results to other ET datasets such as GLEAM and FLUXCOM.

2.2.4 FLUXCOM Gross Primary Production

Gross primary production (GPP) is a measure of CO_2 assimilated into vegetation by photosynthesis. This sequestration of carbon plays an important role in the global carbon budget (Figure 1.4). GPP is indicative of vegetation health and photosynthetic activity, and is highly coupled to water, light, and soil nutrient availability. However, direct, global measures of GPP are not currently possible (Anav et al., 2015) and instead must be estimated by measurements of carbon exchange between the land surface and the atmosphere.

The global FLUXNET network is a vast organization of eddy covariance towers used to measure trace gas fluxes between the biosphere and atmosphere (Jung, Reichstein, and Bondeau, 2009; Pastorello et al., 2020). Machine learning algorithms are then applied to the energy and gas fluxes, as well as meteorological variables, to estimate fluxes in GPP and Terrestrial Ecosystem Respiration (TER) (Reichstein et al., 2005; Baldocchi, 2008; Lasslop et al., 2010). This network of in situ measurements are then taken and combined with MODIS imagery for quality control and feature selection, then put through several machine learning approaches, and finally combining with seasonal gridded satellite and meteorological observations to generate global carbon and energy flux products, FLUXCOM (Tramontana et al., 2016; Jung et al., 2019). This study uses the global GPP product from FLUXCOM to evaluate the performance of vegetation parameters independent from the LAI assimilated by LDAS-Monde. FLUXCOM GPP is available globally at 0.5° x 0.5° resolution, and from 1980 to present, however this study thesis only uses data up to 2013 due to lack of data access.

FLLUXCOM GPP was selected as an assessment dataset as it one of the few projects providing carbon fixation and photosynthesis, especially at a global scale. Furthermore, FLUXCOM GPP data was readily available for comparison in the LDAS-Monde ecosystem.

2.3 Experimental Setups

This thesis examines the results of three major experiments set over and within the contiguous United States, as shown in Table 2.3. The first is a long term (19-year) study over the U.S. State of Nebraska, using various spatial resolutions of LDAS-Monde as well experimenting with differences between ECOCLIMAP-II and ECOCLIMAP-SG. The purpose of this experiment is to link model variables to agricultural statistics and drought in an inter-annual comparison. The next experiment is the application of the LDAS-Monde forecast mode over CONUS. Using information and experience gathered from running the Nebraska experiments, this experiment uses 15-day deterministic forecasts from ECMWF to force 1-through-14 day land surface forecasts over 2017-2018. The reason a 15-day forecast produces only up to 14 days of land surface forecasts is due to LDAS-Monde using an assimilation window of 0900UTC to 0900UTC the next day, while the atmospheric forecasts start at 0000UTC. Thus only 14 full days of forecast data are produced. The 0900UTC assimilation window was initially implemented as that was the morning overpass time for ASCAT over France (Draper et al., 2011), and there was some evidence that morning observations were more accurate than evening ones (Wagner, Lemoine, and Rott, 1999; Albergel et al., 2009). Finally, another study was performed over the same CONUS domain using various assimilated observations in different combinations to study their effects. Each of these experiments used a 20 year spinup of their starting year to generate realistic initial conditions and for the model to reach equilibrium.

Table 2.3: Details and parameters of experiments used in this thesis

Domain	Time Period	Spatial Resolution	Atmospheric Forcing	Assimilated Observations	CTRIP Coupling	ECOCLIMAP Version
Nebraska (39.5-43.5°N 105-95°W)	Jan 2000-Aug 2018	0.25°, 0.10°, 0.01°	ERA5	SSM (CGLS), LAI (CGLS)	Yes	ECO-II ECO-SG
CONUS (20-55°N 130-60°W)	2017-2018	0.20°	ECMWF CTRL 15-day forecast	SSM (CGLS), LAI (CGLS)	No	ECO-SG
CONUS (20-55°N 130-60°W)	2003-2018	0.25°	ERA5	SSM (ESACCI), LAI (CGLS), VOD (VODCA)	No	ECO-SG

Figure 2.10 illustrates the main CONUS domain, with the blue box showing the domain of Nebraska. The four additional boxes indicate separate subdomains that are investigated more closely in Chapter 4. Finally, the red dots indicate the location of USCRN observation stations. CONUS was chosen as the primary domain of interest for several reason, notably the fact that it represents a varied geography and climate to test the model and analysis, as well

as the large availability of data and previous studies over this region. Nebraska was selected and analyzed as a subdomain as it is a large agricultural state, and agricultural yields can be linked to drought events. The four other subdomains, California, Midwest, Northeast, and Southern Plains, were selected from a larger global list of potential hotspots for droughts and heatwaves (Albergel et al., 2020). This also allows us to provide a more regional scale analyses of LDAS-Monde results. The analysis of these four subdomains are presented in the Appendices.

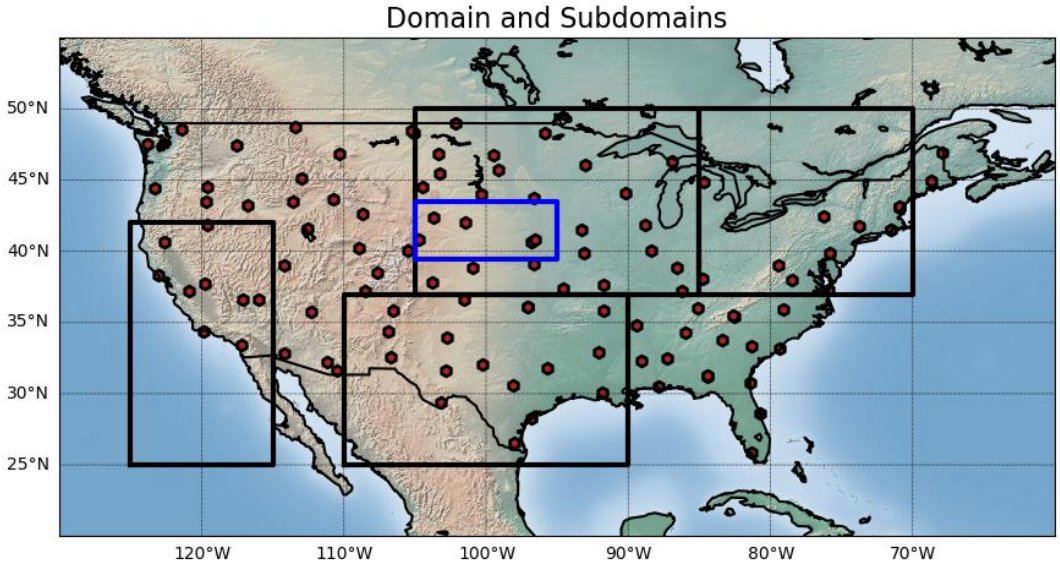

Figure 2.10: Map illustrating the CONUS domain, with boxes around selected subdomains of interest. The blue box represents the Nebraska domain. Red dots represent the locations of the USCRN SM stations.

2.4 Experiment Analysis and Assessment

The primary statistical scores used in this thesis are the Pearson's correlation coefficient (R) and root mean squared difference (RMSD). These two scores enable a quick assessment of improvement or degradation, and are consistent with previous studies of LSMs and LSVs. In addition to the correlation and RMSD, a normalized information contribution (NIC) is calculated for each as shown in Equations 2.7 and 2.8.

$$NIC_R = \frac{R_{Analysis} - R_{Model}}{1 - R_{Model}} X100 \tag{2.7}$$

$$NIC_{RMSD} = \frac{RMSD_{Analysis} - RMSD_{Model}}{RMSD_{Model}} X100 \tag{2.8}$$

The NIC in this context, as first shown in Kumar et al. (2009), provides a better quantitative metric for the improvement or degradation due to data

assimilation. The terms R_{Model} ($RMSD_{Model}$) are the correlation (RMSD) between the model OL and the observations, while $R_{Analysis}$ ($RMSD_{Analysis}$) are the correlation (RMSD) between a SEKF data assimilated run and the observations. For NIC_R, positive values indicate improvement from assimilation, while for NIC_{RMSD} negative values indicate improvement.

The next statistical approach used is bootstrapping. This approach is used to calculate confidence intervals and thus determine statistical significance between model and analysis. Essentially, bootstrapping is the repeated removal of random points in the dataset, and recalculation of the desired score or variable. This thesis uses a constant 10,000 repeats to calculate the confidence intervals in order to generate a sufficiently large number of samples.

For several analyses, probability distribution functions (PDFs) are estimated from the distribution of correlation scores of individual gridcells. These PDFs are derived using a Gaussian kernel density estimation, with Scott's Rule calculating the appropriate smoothing bandwidth. Scott's Rule minimizes the integrated mean squared error in each bin, and is a typically used approximation of optimal smoothing for PDFs (Scott, 1992). These PDFs give a far smoother and readable estimation of correlations when compared to simple histograms.

When analyzing the statistical scores of the USCRN, several conditions are applied. First, frozen soil conditions are avoided by only calculating scores based from observations when temperature measurements are above 4°C. As ISBA separately calculates frozen and liquid soil moisture, when conditions are close to freezing, there can be significant errors. Second, only stations with more than 100 observations (at the respective depths) are considered for a sufficient number of data points. Finally, p-values are calculated alongside the correlations, and stations without p-values of significance, where p-values $<$ 0.05, are screened out.

Additional processing steps were required for the ECMWF atmospheric forecast forcings. These forecasts are provided hourly up to day three, three-hourly up to day six, and six-hourly up to day fifteen. In order to match the 0900UTC assimilation window of LDAS-Monde, the three timesteps were transformed to three-hourly via linear interpolation. If the varying timesteps were left in their original formats, it would make a significant impact on the shape and amplitude of the diurnal cycle of incoming shortwave radiation. This impact is most evident on ET, due to the ISBA LSM directly changing stomatal properties based on shortwave radiation. With a consistent timestep, these problems do not appear to be found. Additionally, to better estimate hourly and sub-hourly shortwave radiation (in addition to the linear rescaling), the forecast experiment over CONUS uses a downward shortwave interpolation scheme in ISBA, which better captures peak hours of solar radiation after a calculation of the solar zenith angle based on global location and season (*SURFEX*).

2.5 Summary of Chapter 2

This chapter detailed the materials and methodologies used to perform experiments throughout this thesis.

- The SURFEX modelling platform was described in detail, along with its components including the land surface model ISBA-A-gs, the data assimilation system LDAS-Monde, as well as the assimilated observations of SSM, LAI, and VOD. VOD was described in depth, as well as a significant background given about the different bands and their interactions with vegetation. Included in the description of ISBA was the description of the atmospheric forcing and forecasts used to drive the model, as well as a description and discussion regarding the ECOCLIMAP land use database, and its different versions. The LDAS-Monde workflow was also outlined, highlighting the differences between a normal configuration and a forecast configuration.

- Additional observations are described that are used to assess the performance of the model and assimilation. This assessment data is comprised of the assimilated observations themselves, USCRN in situ soil moisture observations, ALEXI ET estimations, and FLUXCOM GPP estimations.

- The details and parameters of the experiments included in this thesis are also displayed, summarized in Table 2.3. This includes the domains, time periods, spatial resolutions, assimilated observations, and other important parameters. A map of the domains and certain subdomains of interest is presented.

- Finally, a description of the analysis itself is given, providing definitions of statistics used, as well as calculation conditions. Some necessary information about the processing of experiments is also provided.

Chapter 3

Importance of Initial Conditions in Forecasting Land Surface Variables

This chapter explores the work transitioning LDAS-Monde into a forecast configuration, and details the experiment using this configuration over the US. This work identifies having accurate initial conditions as being vital to providing accurate forecasts of LSVs. The experiment results and discussions are provided in a published article, Mucia et al. (2020). Supplementary work then briefly explores a way to potentially improve initial condition accuracy by using vegetation optical depth (VOD) observations in lieu of LAI. This introduction to VOD links the results of the land surface forecast system to results in Chapter 4.

3.1 Model Initial Conditions

Physical models, especially those representing nonlinear processes, are fundamentally limited in accuracy by their driving parameters and initial conditions (Rodell et al., 2005). The driving parameters of LSMs are primarily atmospheric forcings, as well as computed and assigned physical parameters. Initial conditions of models (LSMs, numerical weather prediction models, and other physical models) are an instantaneous representation of the physical states of the model variables at the time the simulation is started. For LSMs, these states include, but are not limited to, the surface and ground water levels, vegetation and canopy vegetation water content, snow cover and depth, and soil moisture and temperature at each prescribed layer. If a model run in a nonlinear system is started with faulty initial conditions, state variables can quickly diverge from realistic possibilities and can cause chaotic behavior in the system. Initial conditions can also depend on the specific model, as the climatology and land surface parameters, often necessary to compute relative values such as SWI, are determined by the model variables and physics (Koster and Milly, 1997). Improved land surface condition estimates are known to improve forecasts of weather patterns, temperature and precipitation at sub-seasonal scale, seasonal streamflow, agricultural productivity, droughts, floods, as well as carbon cycle fluxes (Bamzai and Shukla, 1999; Bierkens and Beek, 2009; Koster et al., 2010; Bauer, Thorpe, and Brunet, 2015; Massari et al., 2018)

Numerical weather prediction (NWP) models have much the same problems. During forecasts, any initial errors or inaccuracies in the initial conditions of the atmosphere can be amplified, resulting in lower accuracy forecasts. Areas with higher uncertainties in initial conditions, such as cloud-affected areas, also produce significant forecast error growth (McNally, 2002). NWP has then moved to use modern data assimilation techniques, with the ever increasing amount of satellite-based data, to correct poor initial conditions (Bauer et al., 2011). The same idea is steadily being worked on in LSMs, that is to say using the surface observations at our disposal to continuously correct known model biases and insufficient model physics when running the system over a longer historical period, as well as to create the most accurate initial conditions possible when focusing on immediate forecasts.

As this thesis works to transition to and enact the forecast configuration of LDAS-Monde, the data assimilation and thus the corrections to the initial conditions, become of utmost importance.

3.2 From Monitoring to Forecasting Land Surface Conditions Using a Land Data Assimilation System: Application over the Contiguous United States (Mucia et al., 2020)

3.2.1 Article Introduction and Context

Previous experiments with LDAS-Monde (Albergel et al., 2018b; Albergel et al., 2019) show that it can successfully monitor vegetation conditions, and that those indicators can be linked to droughts. These experiments also used the assimilation of satellite observed LAI and SSM which improves this ability. Work in this thesis is then performed to transition from the monitoring of conditions and experiment on the forecasting of them. With ERA5 reanalyses, only near-real time monitoring is possible, with delays on the order of a few days. Analysis in this mode provides useful information, but can rarely be a tool used to aide stakeholders in decision making. On the other hand, two week forecasts may prove useful for certain applications such as the timing of planting, irrigation, and harvests. Several weeks forecast of land surface conditions may also aide decision makers in domains such as agriculture, municipal water suppliers, wildfire managers, public health mangers, hydropower utilities, tourism, and local or regional resource management. Early warning in these domains can improve response quality and speed in order best mitigate the impact of drought events. The decisions themselves can range from dam operators deciding when to releasing water and how much to release, municipal water providers regulating or restricting water use, or even simply the scheduling of local tourism workers to account for changes in tourist and visitor influx. These land surface forecasts may also work towards providing a rough estimate of annual crop yields (although that would be an extension of this current work).

remote sensing

MDPI

Article

From Monitoring to Forecasting Land Surface Conditions Using a Land Data Assimilation System: Application over the Contiguous United States

Anthony Mucia [1], Bertrand Bonan [1], Yongjun Zheng [1], Clément Albergel [1,2,*] and Jean-Christophe Calvet [1]

[1] CNRM, Université de Toulouse, Météo-France, CNRS, 31057 Toulouse, France; anthony.mucia@meteo.fr (A.M.); bertrand.bonan@meteo.fr (B.B.); yongjun.zheng@meteo.fr (Y.Z.); jean-christophe.calvet@meteo.fr (J.-C.C.)

[2] Now at European Space Agency Climate Office, ECSAT, Harwell Campus, Oxfordshire, Didcot OX11 0FD, UK

* Correspondence: clement.albergel@meteo.fr

Received: 18 May 2020; Accepted: 23 June 2020; Published: 24 June 2020

Abstract: LDAS-Monde is a global land data assimilation system (LDAS) developed by Centre National de Recherches Météorologiques (CNRM) to monitor land surface variables (LSV) at various scales, from regional to global. With LDAS-Monde, it is possible to jointly assimilate satellite-derived observations of surface soil moisture (SSM) and leaf area index (LAI) into the interactions between soil biosphere and atmosphere (ISBA) land surface model (LSM) in order to analyze the soil moisture profile together with vegetation biomass. In this study, we investigate LDAS-Monde's ability to predict LSV states up to two weeks in the future using atmospheric forecasts. In particular, the impact of the initialization, and the evolution of the forecasted variables in the LSM are addressed. LDAS-Monde is an offline system normally driven by atmospheric reanalysis, but in this study is forced by atmospheric forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF) for the 2017–2018 period over the contiguous United States (CONUS) at a 0.2° × 0.2° spatial resolution. These LSV forecasts are initialized either by the model alone (LDAS-Monde open-loop, without assimilation) or by the analysis (assimilation of SSM and LAI). These two forecasts are then evaluated using satellite-derived observations of SSM and LAI, evapotranspiration (ET) estimates, as well as in situ measurements of soil moisture from the U.S. Climate Reference Network (USCRN). Results indicate that for the three evaluation variables (SSM, LAI, and ET), LDAS-Monde provides reasonably accurate and consistent predictions two weeks in advance. Additionally, the initial conditions after assimilation are shown to make a positive impact with respect to LAI and ET. This impact persists in time for these two vegetation-related variables. Many model variables, such as SSM, root zone soil moisture (RZSM), LAI, ET, and drainage, remain relatively consistent as the forecast lead time increases, while runoff is highly variable.

Keywords: ASCAT; data assimilation; soil moisture; leaf area index; evapotranspiration

1. Introduction

Extreme meteorological and climatic events, such as heatwaves and droughts, are predicted to increase in frequency and magnitude in future decades [1,2]. These events have significant ramifications on society, leading to environmental, economic, and societal damages. In particular, droughts have been found to be a detrimental and costly natural hazard [3–7], accounting for around 20% of all natural hazard damages [8] and costing society billions of dollars every year [7]. Due to this large impact, it is critical to monitor and predict the land surface variables (LSV) that link drought and

society [9]. The monitoring, prediction, and therefore warning of droughts, floods, famine, and other extreme events can be accomplished with land surface models (LSM) [10,11] that can simulate the LSV responses to these extreme events. These events and their responses also pose a significant scientific challenge for the adaptation to climate change [1]. Good knowledge of both land and lower atmosphere conditions is necessary to accurately monitor and predict LSV values. This understanding is also critical for the monitoring and prediction of drought. Specifically, modeling of surface soil moisture (SSM), leaf area index (LAI), and evapotranspiration (ET) is of high importance for agricultural producers in drought-prone areas [12–14]. While LSM simulations provide temporally and spatially continuous information about LSV, they are by no means perfect, and often lack complex interactions and physics that result in predictions differing from reality.

The availability of global LSV observations from satellite instruments enables the constraining of LSM simulations using data assimilation (DA) techniques. A number of DA systems are able to sequentially combine model variables and observations at regular and consistent intervals, as demonstrated by [15–20], among others. Many land data assimilation systems (LDAS) already exist, including the Global Land Data Assimilation System (GLDAS) [21], North American Land Data Assimilation System (NLDAS) [22,23], Coupled Land and Vegetation Data Assimilation System (CLVDAS) [24], the Famine Early Warming Systems Network (FEWS NET) Land Data Assimilation System (FLDAS) [25], and the National Climate Assessment – Land Data Assimilation System (NCA-LDAS) [26–28]. Barbu et al. [18] developed an offline regional LDAS over France in the research department of Météo-France (CNRM, Centre National de Recherches Météorologiques). This LDAS was able to sequentially assimilate LAI and SSM at the same time into the ISBA (interactions between soil biosphere and atmosphere) LSM [29–32]. The LDAS-Monde tool [33–35] is an upgraded version of this LDAS. LDAS-Monde is able to combine Earth observation data and LSM simulations at a global scale and over specific areas. It can be used for the monitoring, and possibly the prediction, of LSV conditions.

LDAS-Monde is forced by atmospheric variables and the sequential assimilation of satellite-derived LAI and SSM data produces an analysis of land surface conditions. The assimilation permits the analysis of the vegetation biomass and of the soil moisture profile. Albergel et al. [33] showed that the root zone soil moisture (RZSM) analysis benefits from the assimilation of SSM and also from the assimilation of LAI, especially in dry conditions. LDAS-Monde has been successful at representing LSV such as SSM, RZSM, LAI, gross primary production, and ET, by using atmospheric reanalysis data, such as those produced by the European Centre for Medium-Range Weather Forecast fifth generation reanalysis, ERA5. LDAS-Monde outputs result in good correlations when compared to satellite and/or in situ observations of land surface states [34–37]. So far, forecasting LSV conditions with LDAS-Monde is at a preliminary stage.

The objectives of this study are to assess to what extent (1) LSV conditions can be forecasted using an LSM, (2) LSV initial conditions influence the forecasts, (3) data assimilation can improve the accuracy of initial conditions of LSV forecasts, and (4) LSV forecasts can benefit to crop monitoring.

This study makes use of LDAS-Monde, towards the goal of moving from monitoring to predicting LSV conditions. Atmospheric forecasts are used to force the ISBA LSM to assess the forecast of LSV conditions. In situations when it is possible to predict atmospheric variables with reasonable accuracy days up to two weeks in advance, those same atmospheric predictions can be used to predict the vegetation and soil moisture variables used to characterize drought. The rationale underlying this study is that an LSM that can successfully monitor vegetation conditions forced by accurate atmospheric reanalyses would logically be able to forecast land surface conditions provided that accurate atmospheric forecasts are available. Running the LSV forecasts with atmospheric forecasts also underlines the importance of accurate initial conditions. Improving the accuracy of LSV initial conditions using data assimilation is a key issue.

The novel assessment in this study uses the LDAS-Monde data assimilation system with atmospheric forcing variables from the European Centre for Medium-Range Weather Forecasts (ECMWF) 15-day forecasts for the two-year 2017–2018 period over the contiguous United States (CONUS).

In both monitoring and forecasting modes, LDAS-Monde results are compared to several independent datasets for analyzing the accuracy of the LSV modeling and the impact of initial conditions (i.e., the impact of the data assimilation). To assess root zone and surface soil moisture simulations, an independent dataset of soil moisture measurements from the U.S. Climate Reference Network (USCRN) [38] across CONUS is used. The simulations of ET are compared to the satellite driven estimates of the Atmosphere-Land Exchange Inverse (ALEXI) [39] surface energy balance model. The LAI and SSM LDAS-Monde outputs are also compared to the matching satellite-derived products as a way to assess the correct behavior of the data assimilation impact.

Section 2 describes the materials and methods used in this study, including assimilated and verification datasets. Section 3 describes the results of the monitoring and forecast experiments by the comparison to verification datasets. Section 4 interprets and discusses these results in the context of known geographic and biophysical patterns. Finally, Section 5 summarizes and concludes this article.

2. Materials and Methods

2.1. Atmospheric Forcing

LDAS-Monde has recently been updated with the ability to also ingest forecasted atmospheric data to drive the ISBA model [35,36]. Twice daily, ECMWF runs global, 15-day, 51-member ensemble forecasts at a spatial resolution of $0.2° \times 0.2°$. This configuration consists of a control run (CTRL), which is an unperturbed forecast run, and 50 perturbed members. The perturbed members are similar to the control run, but with initial states and model physics slightly perturbed to explore the range of uncertainty in the model. The ensemble forecasts are commonly used as probability guides (e.g., warmer or colder than average, probability of more than a precipitation threshold, etc.) and to investigate extreme weather events (e.g., tropical cyclones, heavy precipitation). This study makes uses of the CTRL forecast with a 15-day lead time. The atmospheric forecast has an hourly timestep up to day three, a three-hourly timestep up to day six, and then a six-hourly timestep up to day fifteen. For our LSV forecast experiment, we only used the six-hourly timestep up to day 15 which we linearly interpolated to every three hours in order to match the 09:00 UTC hour start of the assimilation window.

The atmospheric variables required to force the ISBA LSM are air temperature, wind speed, air specific humidity, atmospheric pressure, shortwave and longwave downwelling radiation, and liquid and solid precipitation. All these variables are given by the ECMWF simulations, except for air specific humidity, which we calculated from air temperature and air dewpoint temperature. The unmodified atmospheric forcings were ingested into the ISBA LSM.

The different ECMWF atmospheric forcing datasets and temporal resolutions can impact the shape and amplitude of the diurnal cycle of shortwave solar radiation. These differences in solar radiation can have a large effect on ET, with stomatal changes responding directly to shortwave radiation in the ISBA LSM. To counteract these problems, the experiments were run using a downward shortwave interpolation scheme, which is intended to better capture the peak hours of solar radiation by using a calculation of the solar zenith angle at each latitude and longitude based on the time of year (http://www.umr-cnrm.fr/surfex/spip.php?article428, last accessed 12 May 2020).

2.2. Assimilated Satellite Observations

Figure 1 illustrates the domain as well as the assimilated land surface observations averaged over 2017–2018. The SSM and LAI satellite observations assimilated in this study are products of the Copernicus Global Land Service (CGLS) [40]. The SSM product is derived from the Advanced Scatterometer (ASCAT) microwave sensor aboard the polar orbiting MetOp A and B satellites [41,42]. This product is available from 2007 to present and is produced as a daily synthesis from both sensors.

The soil moisture information is based on radar backscatter observations. An exponential filter [43] is then applied to this product to estimate the soil wetness index (SWI) using a timescale parameter, T, that varies between 1 and 100 days. With this T parameter, SWI accounts for other physical parameters such as soil thickness and type, soil texture and density, evaporation, and runoff. This experiment uses the smallest possible T value, T = 1 day, to best estimate the surface soil moisture (SSM), which is expressed as a percentage of saturation between 0 (completely dry) and 100 (completely saturated) of the topsoil layer. This product is global at a 0.1° × 0.1° resolution, which is interpolated to an ISBA model grid of 0.2° × 0.2°.

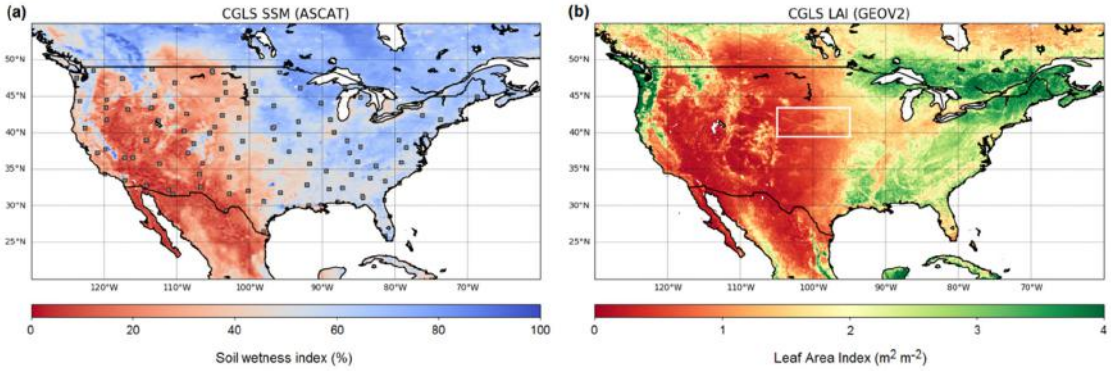

Figure 1. Average values of Copernicus Global Land Service (CGLS) soil wetness index (SWI) from combined (**a**) Advanced Scatterometer (ASCAT) observations and (**b**) CGLS leaf area index (LAI) observations from the GEOV2 dataset. Individual points on (**a**) are locations of U.S. Climate Reference Network (USCRN) in situ observations. The rectangle in (**b**) is the domain used to analyze LAI and corn yield over the state of Nebraska.

The SSM product that is assimilated must be rescaled to match the model climatology in units of m^3 m^{-3} in order to address a possible misclassification of land surface parameters such as soil porosity, wilting point, and field capacity [44,45]. The ASCAT SWI product is transformed into the model-equivalent SSM with the linear rescaling method as described in [46], where the observations' mean and variance are matched to the modeled surface soil moisture mean and variance. The second layer of soil in the ISBA model (WG2, 1–4 cm depth) is used to represent the topsoil properties. This rescaling gives, in practice, very similar results to cumulative distribution function (CDF) matching. The linear rescaling is performed on a seasonal basis (with a three-month moving window) as suggested by [16–18]. The rescaling parameters were derived monthly and screened for the presence of ice and urban surfaces. Average observed SWI over CONUS for 2017–2018 is displayed in Figure 1. The SSM observational dataset is filtered to exclude pixels whose average altitude is above 1500 m above sea level and pixels whose land cover fraction exceeds 15% of urban surfaces.

Over the considered period, 2017–2018, the assimilated LAI observations are derived from the PROBA-V (2014–present) mission. This study uses the GEOV2 dataset, which applies the observations to a 1 × 1 km spatial resolution with data steps every 10 days [47]. These observations are then also interpolated to the 0.20° ×0.2° model grid. Averaged LAI observations for 2017–2018 are displayed in Figure 1.

2.3. Independent Evaluation Datasets

USCRN is a National Oceanic and Atmospheric Administration (NOAA) network of climate-monitoring stations. USCRN contains 114 sites with hourly soil moisture and temperature measurements at 5, 10, 20, 50, and 100 cm depths, relatively well spread over CONUS. In order to effectively evaluate LDAS-Monde's ability to predict surface variables, the SSM simulations are compared to observed data at all sites with continuous quality-controlled data for all of 2018 in the USCRN over CONUS. The number of these sites differs based on the individual layer of soil moisture.

This study uses two layers of measurement from USCRN, the 5 and 20 cm observations. The 5 cm layer is compared to ISBA soil layer 3 (WG3) representing 4–10 cm depth, while the 20 cm observation is compared to an average of ISBA layers 4 and 5 (WG4 and WG5) representing 10–20 and 20–40 cm, respectively, weighted by each layer's respective thickness.

ALEXI [39,48–50] is a surface energy balance model calculating ET based on a two-source land surface representation of the energy budget. The model treats the land surface as a combination of soil and vegetation with unique fluxes, temperatures, and coupling with the atmosphere. This single model formulation can be applied to wide ranging moisture and vegetation conditions, including partially vegetated surfaces. ALEXI is produced at both the global scale and over CONUS at 0.05° × 0.05° and 0.04° × 0.04° spatial resolution, respectively. This independent dataset is used to evaluate LDAS-Monde's parameterization of ET processes.

2.4. LDAS-Monde

Within the SURFEX (Surface Externalisée Version 8.1) modeling system [51] developed by CNRM (http://www.umr-cnrm.fr/surfex), a package allows the assimilation of satellite-derived products into the ISBA land surface model using the LDAS-Monde data assimilation system. LDAS-Monde is a global scale LDAS, with options to couple to hydrological models. LDAS-Monde also provides statistics that can be used to monitor the impact of the assimilation of satellite observations. The specific components of the system are detailed below.

The LSM used in this study is the CO_2-responsive [30–32], multi-layer soil [52,53] version of the ISBA model. This version of the model allows the representation of vegetation biomass and LAI, in addition to exchanges in CO_2, energy, and water fluxes between the surface and the atmosphere. In our configuration, ISBA uses 12 predefined land surface patch types including nine functional plant types (needle leaf trees, evergreen broadleaf trees, deciduous broadleaf trees, C3 crops, C4 crops, C4 irrigated crops, herbaceous, tropical herbaceous, and wetlands) as well as bare soil, rocks, and permanent snow and ice surfaces. These patches are extracted from ECOCLIMAP Second Generation [54], an evolution of ECOCLIMAP-II [55]. Urban surfaces defined in ECOCLIMAP-SG are converted to bare rock when running ISBA alone. Atmospheric, climatic, and land use conditions create the evolution in the vegetation biomass through processes of plant growth and mortality. During growth phases, increased photosynthesis results in CO_2 assimilation, creating vegetation growth and increasing LAI from the minimum threshold (which is 1 m^2 m^{-2} for evergreen forests and 0.3 m^2 m^{-2} for all other types of vegetation). The experiments presented in this paper use a 14 layer, 12m depth soil diffusion scheme, with layers at 0.01, 0.04, 0.1, 0.2, 0.4, 0.6, 0.8, 1, 1.5, 2, 3, 5, 8, and 12 m depths [56].

In the LDAS-Monde system [33], observed satellite LAI and SSM data are assimilated into the ISBA LSM using a simplified extended Kalman filter (SEKF) DA technique [57]. The SEKF technique uses finite differences to compute the flow dependency between the assimilated observations, SSM and LAI, and the control variables (soil moisture from layer of soil 2 to 8 (1 cm to 100 cm) and LAI). The eight control variables are directly updated by the observed variables according to their sensitivity as given by the SEKF Jacobian matrices [19,33,58]. Other model variables are updated indirectly through feedbacks and other biophysical processes as related to the control variables. The SEKF technique used in this study is the most mature data assimilation scheme currently implemented in LDAS-Monde. Bonan et al. [59] recently implemented and validated an ensemble-based Kalman filter.

2.5. Experimental Setup

In this study, the LDAS-Monde offline system is used in a forecast mode, forced by ECMWF atmospheric forecasts over the CONUS domain. It produces one- through 14-day LSV forecasts for each day in the 2017–2018 period. The maximum 14-day lead time corresponds to potential changes in LSVs' conditions impacting agricultural decisions such as the timing of irrigation, planting, and harvesting. The LSV forecasts are initialized by either the open-loop (OL) and by the analysis resulting from the DA experiment based on the sequential use of the SEKF to assimilate SSM and LAI. The OL corresponds to

model-only simulations, without data assimilation. The OL- and SEKF-based forecast runs differ by changing their initialized conditions, with SEKF having what should be a more accurate initial state due to the assimilation of observations. The CONUS domain is defined as longitude 130° W to 60° W and latitude 55° N to 20° N. Table 1 presents the characteristics of the DA experiment. LDAS-Monde uses a 24 h assimilation window from 09:00 UTC to 09:00 UTC next day. The atmospheric forecast forcings begin at 00:00 UTC while ISBA begins running at 09:00 UTC. This discrepancy causes the 15-day atmospheric forecast to produce only 14 full days of data. For this forecast experiment to reach an equilibrium able to generate physically realistic initial conditions, the first year (2017) was spun up 20 times using the earliest time period of the CTRL member of the ECMWF ensemble forecast (FC1, corresponding to the first day of forecast) as atmospheric forcing.

Table 1. Description of the LDAS-Monde data assimilation (DA) experiment used in this study.

Model	Domain	Time Scale and Model Resolution	Atmospheric Forcing	Deterministic Atmospheric Forecast	Assimilated Observations	Model Equivalent of Observations	Control Variables
ISBA Multi-layer soil Plant growth ("NIT" option in SURFEX)	CONUS (20N–55N, 130W–60W)	2017–2018, 0.20° × 0.20°	CTRL first 24 h (3–hourly)	Up to 15 days	SSM (ASCAT) LAI (GEOV2)	Rescaled WG2 (1–4 cm) LAI	Layers 2 to 8 (1–100 cm) LAI

Since previous studies [33–37] have demonstrated that LDAS-Monde analyses can accurately represent LSVs, this study uses a novel approach regarding the assessment of the forecasted variables. The SEKF at FC1 is used as a reference. If the variables do not change greatly between later forecast periods and FC1, it can be said that forecast variables are persistent. This persistence is important to find as highly different values should encourage more scrutiny for that variable, whereas highly similar values can be treated like the reanalysis (i.e., no forecast).

2.6. Assessment

Two scores are considered in this study: the Pearson's correlation coefficient (R) and the root mean squared difference (RMSD).

For surface soil moisture, score values between in situ observations and modeled soil moisture are only calculated when soil temperatures (as given by USCRN measurements) are above 4 °C to avoid frozen conditions. The ISBA LSM separates liquid and solid soil moisture, and the simulated SSM variability can be affected by soil freezing when the modeled soil temperature is close to the freezing level. Correlations are only calculated when there are more than 100 days of soil moisture measurements in the 2017–2018 timeframe above this 4 °C threshold. Correlations are only retained when the p-value is less than 0.05 (i.e., scores are significant at P values < 0.05).

The normalized information contribution (NIC) is defined by Equations (1) and (2) as in [60]. The NIC provides a metric to quantify the improvement or degradation from the analysis compared to the model. It is computed for the correlations, and for RMSD (Equations (1) and (2), respectively): $R_{Analysis}$ ($RMSD_{Analysis}$) is the correlation coefficient (RMSD) between the SEKF experiment and observations, while R_{Model} ($RMSD_{Model}$) is the correlation coefficient (RMSD) between the OL experiment and observations. While the average NIC_R and NIC_{RMSD} values over all the stations are simple and good indications of overall trends, the effect of the assimilation on individual stations can be highly variable. This is important to note if work is done at a more local or regional level as the CONUS average may not represent each individual station. These normalized scores are used in assessments using satellite observations of LAI, SSM, and ET. For NIC_R, positive values indicate improvement from the assimilation, while negative values indicate degradation. For NIC_{RMSD}, this pattern is

reversed, with negative values denoting improvement from the assimilation, and positive values denoting deterioration.

$$NIC_R = \frac{R_{Analysis} - R_{Model}}{1 - R_{Model}} \times 100 \quad (1)$$

$$NIC_{RMSD} = \frac{RMSD_{Analysis} - RMSD_{Model}}{RMSD_{Model}} \times 100 \quad (2)$$

Finally, the bootstrapping statistical approach is used in order to calculate 99% confidence intervals and to determine statistically significant differences between the model and analysis with regards to forecast time period. The bootstrapping method can be summarized by the repeated random removal of points and recalculation of the desired variable. In this analysis, this removal process is repeated 10,000 times in order to generate a sufficiently large number of samples in order to find 99% confidence intervals. Bootstrapping was applied to the correlation values of all evaluation datasets (CGLS SSM, CGLS LAI, ALEXI ET, and USCRN in situ soil moisture) versus matching OL and SEKF variables.

3. Results

3.1. Impact of the Analysis

Analysis increments, given in Figure 2, are the amount of daily mean change in each variable from the assimilation of observations. These demonstrate the effect that the assimilation of LAI and SSM has on five of the control variables (LAI, WG2–WG5 representing soil moisture layers 1–40 cm). Specifically, it gives the average LAI (m^2 m^{-2}) or volumetric soil moisture (m^3 m^{-3}) added or removed from the model run due to the assimilation over the considered 2017–2018 period. LAI experiences the strongest effect, with negative LAI increment values being observed over most of the domain, most strongly in the Great Plains, Midwest, and Eastern U.S. One distinct pattern is the Mississippi River valley, where LAI is more markedly decreased. The coastal Pacific Northwest is the main region with the opposite reaction, instead adding to the LAI. The model-equivalent of SSM, at 1–4 cm depth (WG2), has a more diverse reaction to the assimilation, with significant areas where water is added, mostly in the southwest and southern Texas, and where it is removed, predominantly in the Washington, Oregon, Idaho, and Nevada areas. Much of the domain also has a weak or neutral change. Increments at 10 cm (WG3) (Figure 2c) are generally still removing soil moisture, with some areas strongly removing, and scattered, weaker areas of adding soil moisture. Soil layers at the 20 and 40 cm depths (WG4 and WG5, respectively) have increments (Figure 2d–e) in a more uniform pattern, with nearly all regions having soil moisture removed.

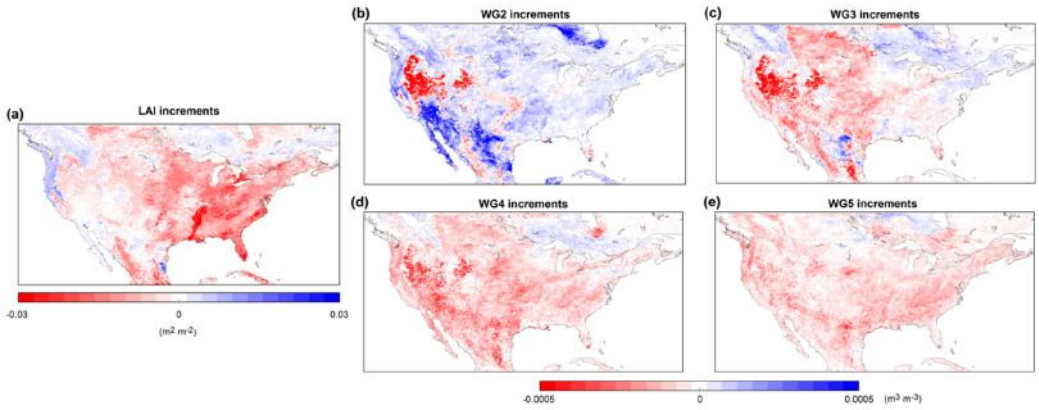

Figure 2. LDAS-Monde daily mean analysis increments for (**a**) LAI (m^2 m^{-2}), (**b**) WG2 (1–4 cm) (m^3 m^{-3}), (**c**) WG3 (4–10 cm) (m^3 m^{-3}), (**d**) WG4 (10–20 cm) (m^3 m^{-3}), and (**e**) WG5 (20–40 cm) (m^3 m^{-3}). These increments represent if the assimilation of observations increases or decreases the amount of leaf area or water for each individual pixel.

3.2. Persistence: Forecast Versus Initial Conditions

In this section, all SEKF-initialized forecast lead times are compared to the initial condition of the analysis produced by the assimilation (i.e., the SEKF) in order to assess each variable's persistence in time. Understanding this persistence was done as an initial step towards understanding the persistence of the system as in a forecast mode. This analysis provides a standardized comparison to determine the persistence of each variable in the LSM. Calculations of average correlation and RMSD for five LSVs (LAI, WG2, RZSM, Runoff, Drainage, and ET) are given in Figure 3. As these correlations and RMSD are compared to the initial conditions at FC1, the statistics would be in perfect agreement at that time step (comparing FC1 against itself). Due to this, RMSD in Figure 3 is shown as a percentage change from the FC1 initial condition. The individual RMSD for the LSVs follow the same general pattern. The LAI, RZSM, WG2, and drainage variables all remain within 20% of the initial condition at FC14. Unlike correlation, ET quickly increases RMSD and at FC14 is ~88% higher than its initial condition. Finally, runoff changes the quickest and finishes at over 210% changed from the initial condition. All correlations decrease as the forecast period increases, but at different rates per variable. For instance, the LAI and RZSM correlations are nearly identical in their decrease, barely having diminished at the fourteen-day forecast time. Drainage decreases in correlation slightly more with values dropping 0.1 by forecast day 14. The WG2 and ET correlations diminish more rapidly, both decreasing by approximately 0.2 over the fourteen days of lead time. Finally, the runoff correlation quickly falls, of about 0.3, between forecast day 11 and day 14.

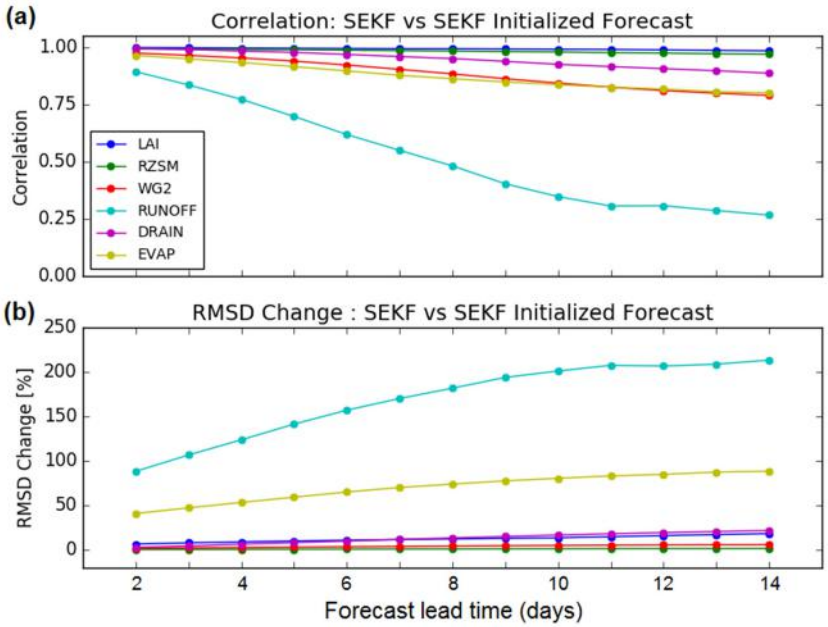

Figure 3. Comparison of forecasted LAI, root zone soil moisture (RZSM), WG2, runoff, drainage, and evapotranspiration (ET) variables initialized by the simplified extended Kalman filter (SEKF), with their values on the first day of forecast. The (**a**) correlation and (**b**) root mean squared difference (RMSD) relative changes with respect to FC1 score values are presented. These comparisons show the persistence of each variable with the use of atmospheric forecasts in the LDAS-Monde system over the contiguous United States (CONUS) for the 2017–2018 time period.

3.3. Evaluation Using Satellite-Derived Products

LDAS-Monde has the ability to ingest atmospheric forecasts in order to output forecasts of land surface conditions. Presented here are the results of comparing LDAS-Monde-forecasted SSM, LAI, and ET to the satellite-derived evaluation datasets of the same variables. This analysis is not only focused on how the forecasts perform in a general sense (i.e., how accurate are these predictions 14 days in advance),

but also focuses on the impact of the initialization, that is, if the forecast is initialized by the OL or SEKF. The correlations and RSMD values for SSM, LAI, and ET are summarized in Table 2. The bootstrapped 99% confidence intervals of correlation (not shown) range from ± 0.001 to ± 0.003 for SSM (either OL or SEKF), ± 0.003 to ± 0.004 for LAI (either OL or SEKF), and ± 0.002 to ± 0.003 for ET (either OL or SEKF). The temporal correlations of SSM, LAI, and ET to the satellite-derived observations for FC1 are given in Figure 4 together with NIC_R, with red colors representing improvement from assimilation, and blue representing degradation. The RMSD score and NIC_{RMSD} are given in Figure 5 with blue colors for NIC_{RMSD} representing improvement from the assimilation, and red representing degradation. Results for SSM, LAI, and ET are further described in Section 3.3.1, Section 3.3.2, and Section 3.3.3, respectively.

Table 2. Impact of forecast lead time and initialization on the prediction of surface soil moisture (SSM), LAI, and ET by LDAS-Monde over the CONUS, during the 2017–2018 time period. Only even values of forecast lead times are indicated, from 2 to 14 days (FC2 and FC14, respectively). The *R* values smaller than 0.5 and the SSM RSMD values larger than 0.05 m^3 m^{-3} are in bold.

LSV	Initialization	Score	Forecast Lead Time						
			FC2	FC4	FC6	FC8	FC10	FC12	FC14
SSM	OL	*R*	0.62	0.58	0.52	**0.46**	**0.41**	**0.36**	**0.35**
		RMSD (m^3 m^{-3})	0.044	0.046	0.050	**0.053**	**0.056**	**0.059**	**0.060**
	SEKF	*R*	0.64	0.59	0.53	**0.46**	**0.41**	**0.36**	**0.34**
		RMSD (m^3 m^{-3})	0.042	0.046	0.049	**0.053**	**0.056**	**0.059**	**0.060**
LAI	OL	*R*	0.56	0.55	0.55	0.57	0.56	0.56	0.55
		RMSD (m^2 m^{-2})	1.02	1.02	1.01	1.01	1.01	1.01	1.01
	SEKF	*R*	0.69	0.69	0.69	0.71	0.71	0.65	0.64
		RMSD (m^2 m^{-2})	0.73	0.73	0.73	0.73	0.73	0.82	0.82
ET	OL	*R*	0.57	0.57	0.57	0.56	0.55	0.54	0.54
		RMSD (mm day^{-1})	1.37	1.37	1.37	1.39	1.40	1.40	1.42
	SEKF	*R*	0.58	0.58	0.58	0.57	0.56	0.55	0.55
		RMSD (mm day^{-1})	1.35	1.35	1.36	1.38	1.39	1.39	1.40

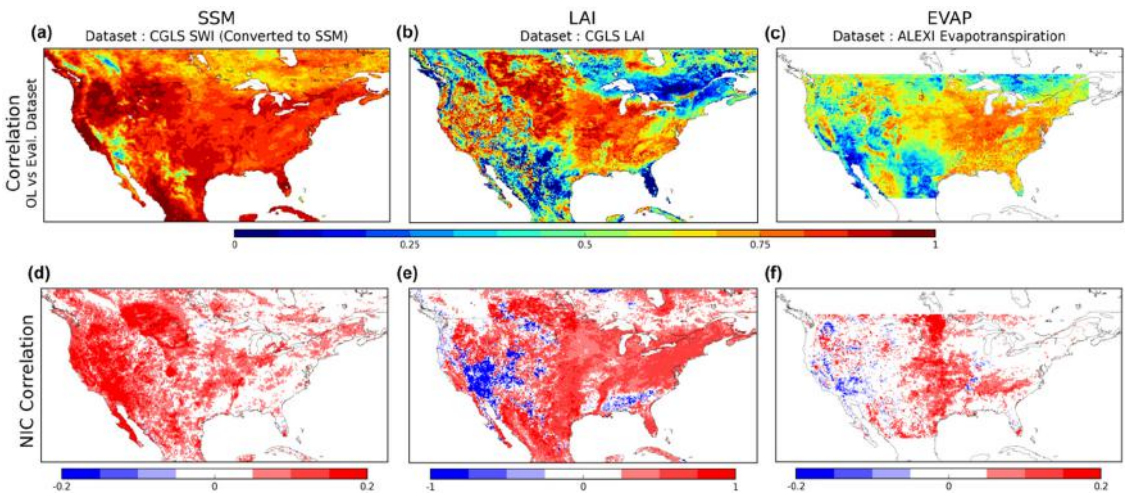

Figure 4. Upper row maps (**a**–**c**) are of average temporal correlation between LDAS-Monde open-loop (OL) and satellite-derived products for SSM (**a**), LAI (**b**), and ET (**c**) at forecast day 1. Bottom row maps are average normalized information contribution (NIC) correlation impact from the SEKF assimilation for the same variables ((**d**–**f**), respectively).

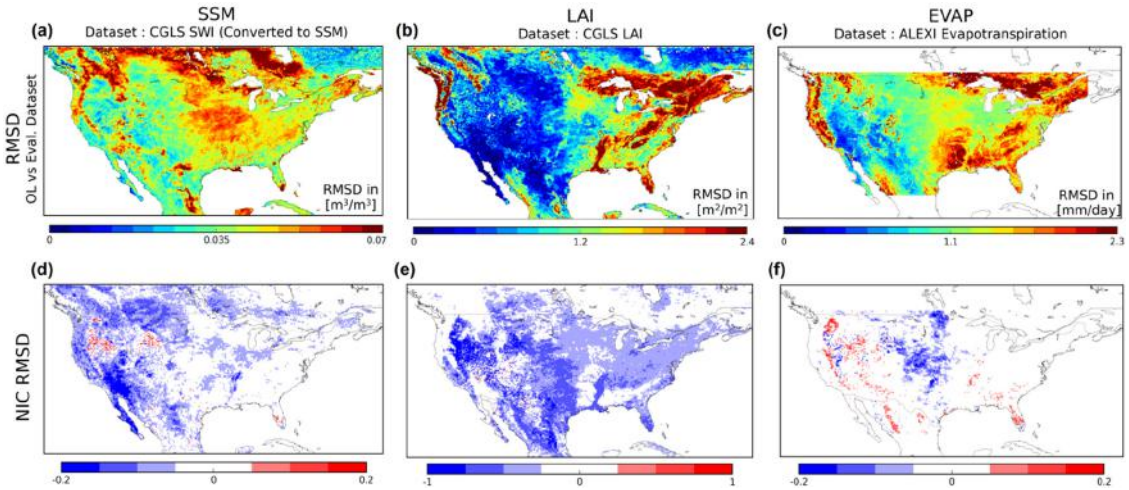

Figure 5. As in Figure 4, except for RMSD. Upper row maps (**a**–**c**) are of average temporal correlation between LDAS-Monde open-loop (OL) and satellite-derived products for SSM (**a**), LAI (**b**), and ET (**c**) at forecast day 1. Bottom row maps are average normalized information contribution (NIC) correlation impact from the SEKF assimilation for the same variables ((**d**–**f**), respectively).

3.3.1. Surface Soil Moisture

Figure 4 shows that SSM temporal correlations are generally very high across much of the continent, with some particularly dry areas, specifically in the U.S. Southwest and southern Texas, having far lower values. The NIC_R values show the visible improvement of correlation values by the assimilation in much of the domain, including the entire Western U.S., much of the central Great Plains, and a large portion of the Midwest. Few, if any, parts of the domain show any degradation after assimilation. The RMSD values of SSM are shown in Figure 5. There are some regions of locally higher RMSD, specifically in the Midwest and eastern Great Plains including some southern states, and additional pockets of high error in portions of the Pacific Northwest, some of the Rocky Mountains, and much of Canada. The data assimilation lowers this error in much of the Southwest, Pacific coast, and Montana, with a more moderate reduction in the Midwest. The lower Mississippi River valley is also outlined by a reduction in error. Scattered patterns of higher RMSD can be seen in some of the mountainous Northwest and Wyoming areas. The average correlation and RMSD for SSM for each forecast period using both model (OL) and analysis (SEKF) initializations are given in Table 2 and in Figure 6. Both OL- and SEKF-initialized correlations experience a consistent drop as the forecast period increases, with correlations at forecast day 14 around 50% lower than the first day of forecast. At the FC1 lead time, the OL correlations start at 0.63 (± 0.003), while SEKF correlations begin at 0.66 (± 0.0025) and are significantly higher than OL until FC5, where the analysis and model are nearly identical. From FC6 to FC14, the two initializations have similar values, with correlations of both ending at FC14 of about 0.35 (± 0.003). Likewise, the RMSD shows a consistent increase (degradation) as the forecast moves forward, with the model having either a slightly higher value or exactly the same error, as far as the output precision allows. As with correlation, the first few days of forecast see RMSD values that have a larger difference between the OL- and SEKF-initialized states compared with the mid length and later forecast days. From FC8 to FC14, the mean SSM R (RMSD) values are smaller (larger) than 0.5 (0.05 m^3m^{-3}). The OL and SEKF initialization modes give very similar R and RMSD values from FC8 to FC14. The 99% confidence intervals determine that the OL and SEKF initializations from FC1 to FC4 are significantly different, whereas the rest of the forecast period do not see statistically different values based on the initializations. These results denote a poor predictability of SSM beyond less than a week forecast lead time.

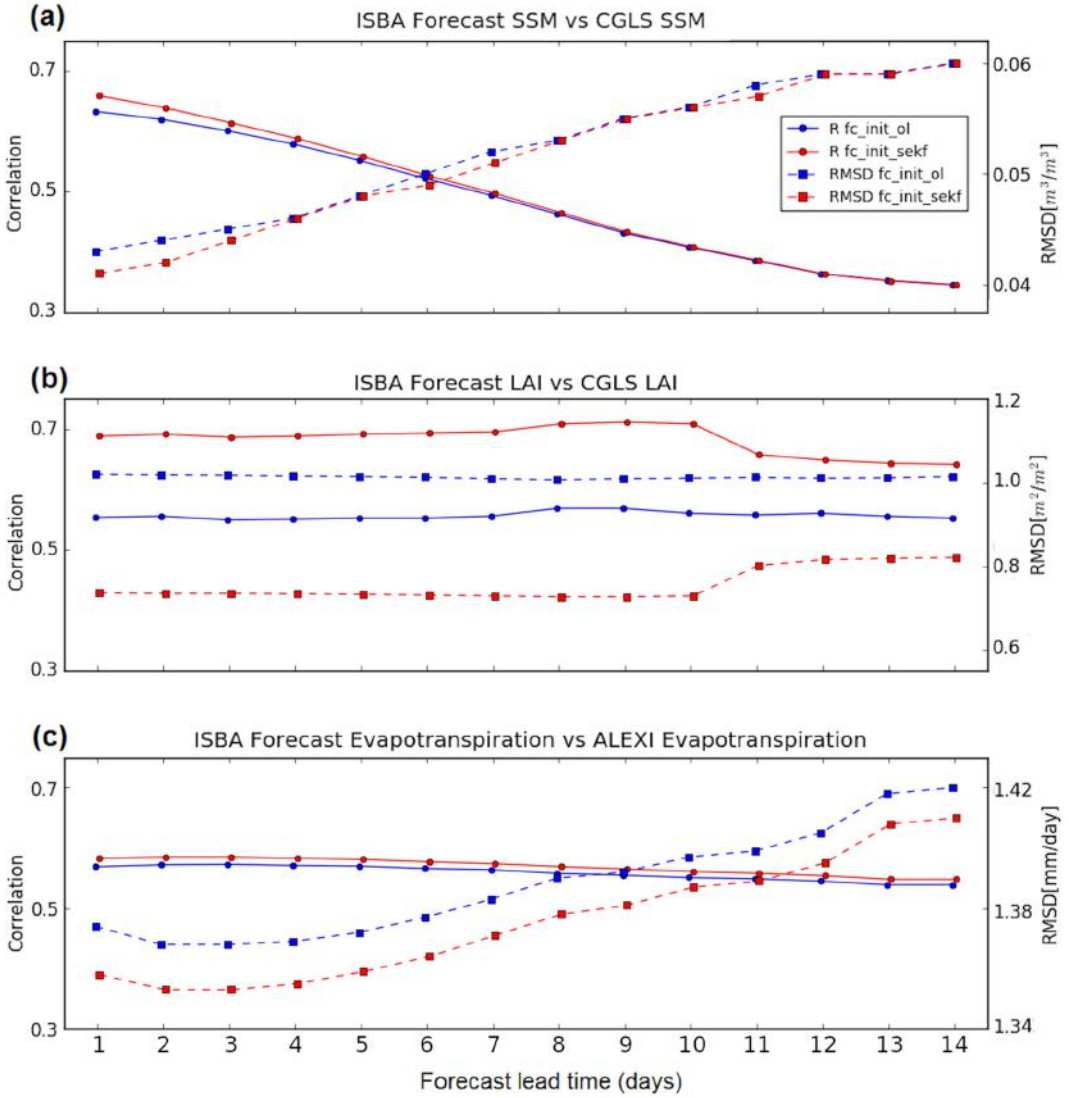

Figure 6. Correlations (solid lines, circles) and RMSD (dashed lines, squares) given for the variables of (**a**) SSM, (**b**) LAI, and (**c**) ET. Both statistics are presented for both OL (blue) and SEKF (red) initializations, and for each day of forecast lead time.

In order to see the geographic patterns of the SEKF and OL initialization differences, maps of NIC_R and NIC_{RMSD} are presented for FC2, FC8, and FC14 in Figure 7. The SSM NIC_R shows how over much of the US, the early forecast lead times see strong improvement (positive, red) across the domain, with strongest improvements in the dry Southwest as well as in the northern Great Plains. As the forecast lead time increases, some areas still show strong improvement, with the Southwest remaining strongly influenced by the initialization out to FC14. Additionally, after the eight-day lead time, there is a small pocket of reduced correlation found at the Washington–Oregon state border, which is not present in earlier lead times, becoming more negative at FC14. Over the entire domain, the magnitude of improvement markedly decreases with the forecast lead time, reducing it to almost equally low positive and negative impact at 14 days. Further shown in Figure 7 is the NIC_{RMSD} for SSM presenting a similar pattern of improvement (negative, blue) and degradation (positive, red). The Washington–Oregon border shows degradation at earlier lead times than for correlation, while much of the other domain indicates the analysis initialization improves error. The Southwest patch also stands out with particularly strong and lasting improvements. As the lead time increases, these effects are reduced in magnitude, but even at 14–days lead time, those two patches are distinguishable.

As with correlation, a general reduction in error over the domain is most prominent only in the two-day lead time, and after it is reduced to near equal improvement and degradation.

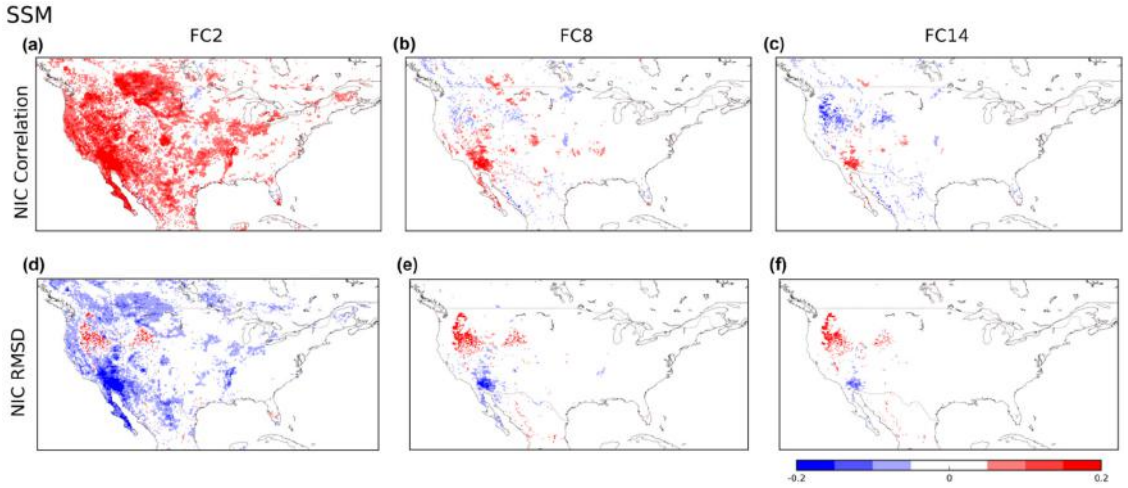

Figure 7. Spatial patterns of persistence of SEKF impact on simulated SSM as compared to CGLS SSM as forecast lead time increases. Upper row of maps shows NIC_R values at FC2 (**a**), FC8 (**b**), and FC14 (**c**). Bottom row of maps (**d**–**f**) is the NIC_{RMSD} values at the same forecast lead times. Note that red color for correlation is improvement from assimilation, whereas blue is improvement for RMSD.

3.3.2. Leaf Area Index

Temporal correlation maps for LAI at the FC1 lead time are given in Figure 4 with impacts from the assimilation given in Figure 5. Correlations are generally very positive over much of the central East Coast, Midwest, and Great Plains. In the mountainous and Western U.S., the correlations are more sporadic, highly variable between very high and low correlations. The impact of the assimilation is overall positive, increasing the correlation significantly in most areas, except for degradation that occurs strongly in the Southwest, and in some mountainous areas of Colorado and Utah. Figure 5 shows that RMSD has particularly high values around the lower Mississippi River, much of Appalachia, New England, and Florida. Unlike correlation, impacts on the RMSD from the assimilation are almost entirely improvements. The biggest changes occur in the Great Plains, mountainous Pacific Northwest, Florida, and again the lower Mississippi River. Smaller improvements are made in more error-prone areas such as Appalachia and New England. It is important to note the scale on the LAI improvements in correlation and RMSD from the assimilation is far larger than both SSM and ET due to the larger differences. The maximum error is also large, reaching up to 2.4 m^2 m^{-2}. Importantly, the Midwest and Great Plains, strongly agricultural regions, show very strong improvement from the assimilation, particularly in correlation, along with a distinct patch of the lower Mississippi River.

Spatially averaged LAI correlations and RMSD are given in Table 2 and Figure 6. The SEKF statistics are significantly improved compared with the OL-initialized states at all forecast steps. The OL correlations stay more or less constant at all forecast periods. The LAI RMSD values for the model indicate a similar pattern of little change. The analysis correlations do experience a strong decrease between forecast day 10 and day 11. The same pattern is seen in reverse with an analysis RMSD increase during this same period. With regards to spatial statistics for different forecast lead times (Figure 8), most of the domain experiences improvement, with NIC_R values often greater than 0.5 and NIC_{RMSD} often less than -0.25. Particularly, the Northeast, Midwest, and Great Plains show consistent and strong improvement at all forecast lead times. There is also weaker degradation in a belt across Mississippi, Alabama, and Georgia that becomes stronger degradation as the lead time increases. Generally, these impacts persist well through the forecast lead time, although like all other variables, they do decrease intensity somewhat. Even at 14–days lead time, the initialization by the analysis shows strong improvement in LAI correlations in much of the agricultural areas of the U.S. A similar

trend is seen with NIC_{RMSD}. The strong improvement in the Mississippi River valley is still easy to see. The degradation in correlation seen in the Southwest is closer to no change in NIC_{RMSD}. Finally, the improvements in RMSD appear to fade quicker than NIC_R as the forecast lead time increases.

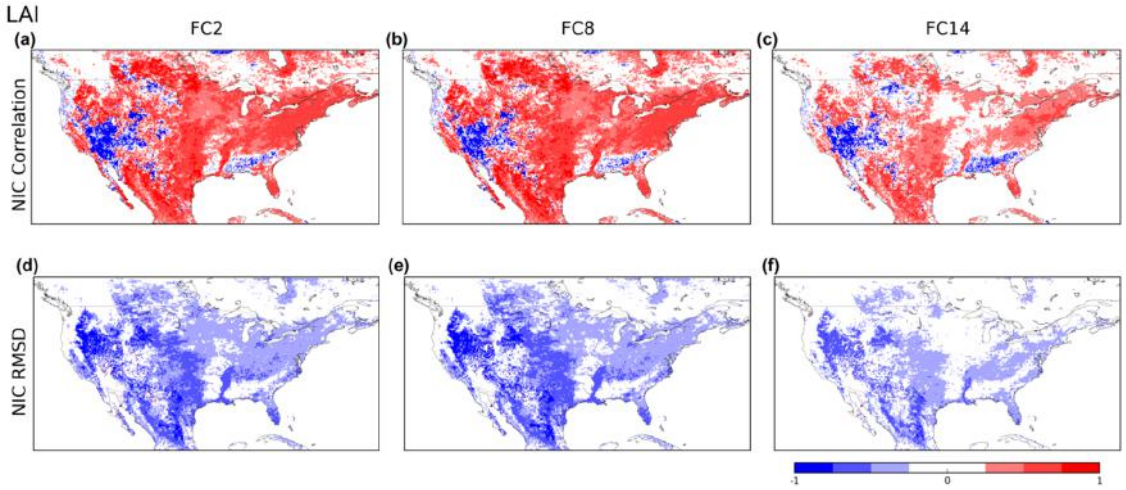

Figure 8. Same as Figure 7, except for simulated LAI and CGLS LAI. Spatial patterns of persistence of SEKF impact on simulated SSM as compared to CGLS SSM as forecast lead time increases. Upper row of maps shows NIC_R values at FC2 (**a**), FC8 (**b**), and FC14 (**c**). Bottom row of maps (**d**–**f**) is the NIC_{RMSD} values at the same forecast lead times. Note that red color for correlation is improvement from assimilation, whereas blue is improvement for RMSD.

3.3.3. Evapotranspiration

Spatial correlations with ALEXI ET at FC1 (Figure 4) show strong agreement over the Midwest and some of the Great Plains. Much of the East Coast and South also demonstrated relatively strong agreement. Impacts from the assimilation indicate improvement over almost the entire Central U.S., with another area of strong improvement in the Kentucky and Tennessee regions. Some scattered degradation is seen in the West, most notably at the California, Arizona, and Nevada border. The ET RMSD and NIC_{RMSD} impacts are given in Figure 5. The error is largest in the South, Pacific Coast, and New England, with most of the Midwest and Great Plains having a mild RMSD. A very low RMSD occurs in the dry Southwest, areas where correlations are also very low. The NIC_{RMSD} demonstrates that the impacts are not as widespread as they were with correlation. The strongest improvement is in the Great Plains states of Kansas, Nebraska, South Dakota, and North Dakota. Degradation up to $NIC_{RMSD} = 0.15$ is most notably seen in the Sierra Nevada and Cascade Mountains.

With the spatially averaged correlations and RMSD of ET (Table 2, Figure 6), both the OL- and SEKF-initialized experiments show a consistent, but small, decrease in correlation as the forecast lead time increases, beginning at 0.57 (± 0.003) and 0.58 (± 0.002), respectively, and ending at the 14–day lead time at 0.54 (± 0.0025) and 0.55 (± 0.002), respectively. The RMSD values of ET forecasts initialized with OL and SEKF increase as the forecast period increases, beginning with 1.37 and 1.35 (mm day^{-1}) and ending at the 14-day lead time with 1.42 and 1.40, respectively. At all forecast lead times, SEKF-initialized runs always have significantly improved correlations and RMSD compared with the OL-initialized runs.

Figure 9 shows that the geographic patterns in the forecast time periods show that impacts on ET correlation from the assimilation are most persistent in the Central U.S. While the magnitude of these improvements does decrease, they are still evident, and in the same pattern, at FC14. With NIC_{RMSD}, most patterns are similar. The area of strongly improved RMSD in the central and northern Great Plains can be said to persist through the forecast lead time, but diminishes far more at FC8 and FC14 compared with the improvements in correlation.

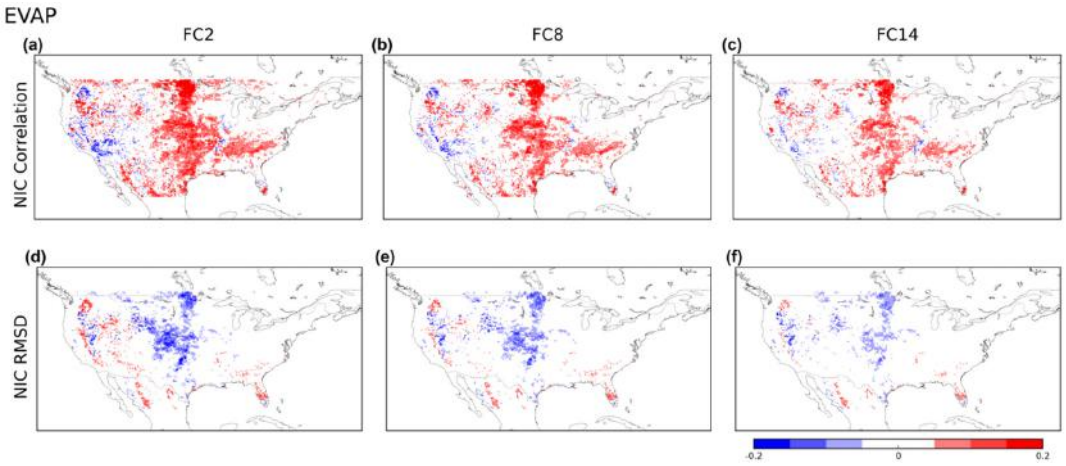

Figure 9. Same as Figure 7, except for simulated ET and Atmosphere-Land Exchange Inverse (ALEXI) ET. Spatial patterns of persistence of SEKF impact on simulated SSM as compared to CGLS SSM as forecast lead time increases. Upper row of maps shows NIC_R values at FC2 (**a**), FC8 (**b**), and FC14 (**c**). Bottom row of maps (**d**–**f**) is the NIC_{RMSD} values at the same forecast lead times. Note that red color for correlation is improvement from assimilation, whereas blue is improvement for RMSD.

3.4. Evaluation Using In Situ Soil Moisture Observations

At 5 cm, with 106 USCRN stations of data, the initialization makes little impact, with OL- and SEKF-initialized correlations to in situ measurements showing no significant differences. Increasing the forecast lead time logically decreases the correlation with in situ measurements, with the lowest correlation at the longest lead time as shown in Figure 10. Correlations for both initializations begin at approximately 0.74 and diminish as the forecast lead time increases, reaching a minimum below 0.50 at FC14. At the 20 cm depth, with 84 stations worth of data, the initializations do cause a visible systematic difference at early lead times, with correlations of OL and SEKF at FC1 of 0.68 and 0.69, respectively. There is also a trend of the OL- and SEKF-initialized correlations becoming closer as the lead time increases (i.e., impact of the initialization decreases with the forecast lead time). These visual differences, however, are not significant as determined by the bootstrap calculations. The 20 cm depth begins with lower correlations than the 5 cm depth but also decreases slower as the lead time increases. However, past FC8, the correlations at 20 cm are better than their 5 cm counterparts, ending at FC14 with R = 0.51 for the OL and SEKF initializations.

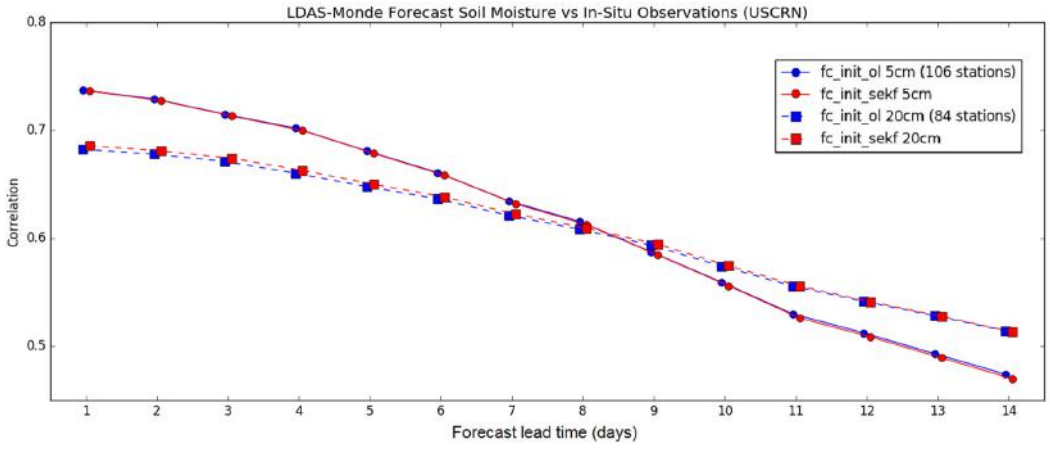

Figure 10. Correlations of OL- (blue) and SEKF (red)-initialized forecasts of SSM to USCRN in situ observations at 5 cm (circles) and 20 cm (squares). Due to the removal of freezing conditions and removal stations with less than 100 days of observations, the 5 cm correlations are averaged over 106 stations, whereas the 20 cm correlations are averaged over 84 stations.

4. Discussion

4.1. Can LSV Conditions Be Forecasted Using a LSM?

The correlations of evaluating the forecast periods against their initial SEKF reference conditions suggest that most LSV (except for runoff) are rather persistent in time in the model. LAI and RZSM had minimal changes from forecast lead time day 2 to 14 (FC2 to FC14), indicating that, working together, the model and forecasts limit the variability of these variables. ET, WG2 (1–4 cm soil moisture), and drainage have slightly more variability, but still had very similar results between the monitoring mode and 14–day forecast. Runoff, on the other hand, varied greatly as the forecast period increased. This behavior tends to demonstrate the variable's high degree of coupling with precipitation [61], which is a highly variable forecast variable.

From FC1, RMSD had a far greater relative change than correlation. The LAI, WG2, RZSM, and drainage all have very small changes out to FC14. ET RMSD is at least 50% higher than the previous variables at all forecast lead times. Further, like correlation, the runoff RMSD quickly decouples from the SEKF initial condition. These results show that forecasted LAI, SSM, RZSM, as well as ET and drainage can be considered rather similar to their non-forecast, SEKF, counterparts. However, run-off differs greatly and its variability indicates that we should not treat the forecasted values the same as their non-forecast counterparts.

4.2. Do LSV Initial Conditions Influence Their Forecasts?

The results of this experiment demonstrate that the influence of the initial condition on the LSV forecast depends on exactly what the LSV in question is, as well as the evaluation dataset. For example, the evaluation of SSM using USCRN in situ observations resulted in no discernible difference between the OL and SEKF initial conditions at 5cm, and a small, but not significant difference at 20cm.

At 5 cm depth, averaged LDAS-Monde OL- and SEKF-initialized correlations to the in situ observations did not significantly differ at 5 cm depth for any forecast period. Still, the correlations at FC1 generally agree with other similar comparisons such as [34], with the main difference being that this study only uses a two-year time period. This could be caused by a few different factors. First is the relatively short time period. Due to the ensemble forecast data only being used from 2017 onward, a longer period of study is not possible in this forecast capacity. The two-year analysis may not be enough time for any statistical differences to appear. Another reason is the relatively small impact of the assimilation of SSM, especially when compared with the impact of the LAI assimilation, which is discussed in the next paragraph. Notably, at FC4, the forecast of SSM is still higher than 0.70, indicating some persistence of model skill as the forecast lead time increases.

At the 20 cm layer, the initialization shows a visible, however not significant, difference. This can be caused by the impact from the assimilation of LAI, which strongly affects deeper layers of soil moisture (20 cm or deeper, ISBA equivalent WG4 and deeper) compared with SSM (1–10 cm, WG2 and WG3 ISBA equivalent) as demonstrated by [33,62]. The correlation at deeper layers is still less than at the surface, which may be explained by the lower amount of variability in the observations, leading to a lower correlation score. As seen in other instances of assimilating soil moisture [63], the neutral impact of the assimilation of ASCAT data can be partially explained by the generally high quality of LDAS-Monde OL soil moisture, compared with the quality of ASCAT. When compared directly with the ASCAT-based SWI product, observations are far less correlated with the same in situ observations (R = 0.56), while LDAS-Monde OL and SEKF initializations at FC1 show higher values (R = 0.74 for both). More research is planned to investigate the individual effects of assimilating SSM and LAI separately in order to quantify their added value independent of one another.

For forecasts in general, as the forecast period increases, the forecast skill of the LSM decreases. With a fast-evolving variable such as SSM, the model decreases in skill in a similar manner as the period increases. SSM is shown to be strongly sensitive to atmospheric conditions, and as the quality of those forcings decreases (as lead time increases), the correlations with measurements decrease. Deeper

layers of soil moisture are still strongly connected to precipitation, but have less variability, leading to a slower decrease in correlation with the forecast period.

Using the CGLS SSM, differences were shown at early forecast periods, but this influence became smaller and smaller until disappearing at longer time forecast periods. The other variables of LAI and ET always resulted in differences corresponding to the initial conditions.

4.3. Can Data Assimilation Improve the Accuracy of Initial Conditions of LSV Forecasts?

Previous experiments using LDAS-Monde [33,34,62] showed that assimilated LAI and SSM satellite-derived products increase their correlations and decrease RMSD with LDAS-Monde LAI and SSM after data assimilation, averaged over the domain. This is a good check to be sure that the data assimilation is working as intended. The response of ET to the data assimilation is of special importance due to its independence from the assimilated datasets. Furthermore, the analysis of how these statistics evolve as a function of forecast lead time is a question of interest (i.e., how long do these variables provide useful, usable information). Several previous studies have looked at the impact of initial conditions on LSVs. Wood and Lettenmaier [64] found that accurate initial conditions and future atmospheric forcings are critical for good streamflow predictions. Shukla et al. [65] demonstrated similar results linking better initial conditions to improved hydrologic forecasts. Sawada et al. [66] have used CLVDAS, assimilated microwave brightness temperature and general circulation model (GCM)-based seasonal meteorological forcings, to predict LAI and agricultural drought with up to a three-month lead time. That study also found that initial conditions play an important role in the prediction of LAI. This study with LDAS-Monde differentiates itself by operating on smaller temporal scales, with operational 15–day meteorological forecasts to ensure the most accurate atmospheric forecasts available. Shorter, but more accurate atmospheric forecasts may provide useful information to stakeholders and influence their decision-making, notably for planting, irrigation, and harvest, which may require higher certainty to take action. Even on the shorter timescales, the results corroborate the idea that the initial conditions are improved after data assimilation, and that these better initial conditions contribute to more accurate forecasts for LSVs. As in [66], the strongest improvement is seen in LAI. In models coupled with the atmosphere, more accurate initial conditions through the use of data assimilation of Earth observations can also feed-back and create better forecasts of atmospheric weather patterns [67,68]. An application where LDAS-Monde would run coupled with an atmospheric model would prove useful, in particular to assess the role of vegetation.

Some geographic patterns of NIC_R and NIC_{RMSD} for LAI and SSM at FC1 can also be seen in the maps of analysis increments (Figure 2). Stronger improvements in correlation are generally associated with stronger increments. For example, with SSM NIC_R and NIC_{RMSD} (Figure 7), we see degradation in some areas of the Pacific Northwest that worsen with the forecast time. This same area is where the analysis increments show that water is removed in the model-equivalent soil layer, WG2 (1-4 cm). This may suggest that the removal of water from the assimilation may be an oversensitivity to the SSM observations. This large change, which may result in slightly improved monitoring at FC1, cascades to become less correlated and with more error at later forecast times. Additionally, average SSM correlations and RMSD (Figure 6) worsen quickly with forecast time, demonstrating the fast-evolving nature of this variable. As accurate atmospheric reanalyses are important to successfully monitor LSVs, specifically those linked to the terrestrial water cycle such as SSM [34], the same reasoning can be extended to include the importance of accurate atmospheric forecasts, in this case, up to 14–days lead time. Finally, the SEKF assimilation does prove to provide significantly more accurate initial conditions, and thus more accurate forecasts for at least some days before the OL- and SEKF-initialized SSM converge.

Correlations and RMSD to the satellite-derived LAI product from the SEKF-initialized forecasts are consistently and significantly better than the OL-initialized counterpart. This strong change driven by the assimilation corrects for model biases in the seasonal timing of vegetation peak and senescence, as well as accounting for known shortcomings such as not considering any human impacts

on vegetation, such as irrigation and harvests. As with SSM, LAI using the SEKF-initialized state proves to be a more accurate initial condition. A notable behavior of the correlations and RMSD between SEKF and the satellite-derived LAI is a strong change between the day 10 and day 11 forecasts (Figure 6). This drop in correlation and rise in RMSD can be explained by examining the frequency of the LAI observations, which are approximately every 10 days. The initial state, day 1, is the only result that has directly changed due to the assimilation, and all forecast days following it only see the difference in initial conditions. These first ten days are compared directly to the most recent satellite observation, which is unchanging across those ten days. At day 11, another LAI observation is available, which is then compared to the day 11 forecast for the purposes of correlation and RMSD, but the forecast results have not seen this updated observation, which leads to a stronger disagreement. The plateau of steady correlations and RMSD between the first and tenth day, and after the eleventh through fourteenth day of forecasts, can be explained by the slow evolving nature of LAI. This shows that the 10–day sampling time for the LAI estimates is a limitation of the data assimilation impact. Such a low sampling time is used to mitigate the lack of data due to cloud coverage. Merging these data with more frequent all-weather vegetation products derived from microwave observations could be a way to increase the sampling time. Microwave vegetation optical depth (VOD) products have potential to improve the analyses [69]. In particular, VOD estimates from ASCAT backscatter observations can present very good (daily or better) sampling times at mid-latitudes [70]. Many works have addressed the relationship between LAI and VOD. Following the work of [34], a comparison between the SEKF and OL FC1 simulations and a VOD product is presented in Figure S1 (see the Supplementary).

As with LAI and the first several forecast days of SSM, ET forecasts initialized by the SEKF state are significantly better than their OL-initialized counterparts, once again showing the utility of the data assimilation to provide more accurate initial conditions. ET correlations and RMSD are very good metrics that demonstrate the potential value in this new predictive capacity due to the completely independent ALEXI evaluation dataset. Additionally, ET shows stronger persistence from the more accurate SEKF initial conditions than SSM, which is likely due to the impact on ET of the relatively slow LAI and root zone soil moisture dynamics. ET continues to have very good correlations even with two–weeks lead time. This independent variable can give stakeholders an advanced look at how much potential evaporation stress their crops may be exposed to in the near future and can provide necessary information on the timing of irrigation.

Finally, biases in atmospheric forcing are inevitable, and can cause errors when monitoring or predicting LSV conditions. No changes were made in this experiment to minimize those errors. While this bias is inevitable, it provides even more reason to assimilate satellite observations, which can work to correct the impact of these biases and move towards more accurate LSV conditions. While data assimilation has been shown to give better initial conditions averaged over the domain, specific regions may see poorer responses. The most notable examples are regions of decreased correlation in LAI after assimilation (Figure 4) without forecasts and appearing in SSM RMSD, mostly notably at FC8 and FC14 (Figure 7). These areas where the initial conditions provide worse scores after assimilation persist through the forecast time. While the exact cause of this degradation is not yet known, a possible issue could be with the fraction of bare ground or vegetation cover, as a misrepresentation of these parameters can affect the partitioning of water fluxes. Investigating these regions of declined statistical scores after assimilation is also the subject of future work.

4.4. Can LSV Forecasts Benefit Crop Monitoring?

Previous work with LDAS-Monde compared grain yields over France with OL and SEKF above-ground biomass [33,71], which, similarly to LAI, is a useful indicator of vegetation health and development [72]. Those findings include significant added value from the data assimilation, and thus improve the representation of agriculture and agricultural droughts on a global scale. Towards this goal of analyzing and improving LDAS-Monde's capacity to monitor agricultural droughts, CGLS LAI observations were extracted from 2000–2018 over the U.S. state of Nebraska, which is a significant

producer of maize. The United States Department of Agriculture (USDA) has yearly harvest yields for many crops at the county and state level [73]. Annual satellite-observed LAI and annual corn yield anomalies over Nebraska are illustrated in Figure 11. The correlation coefficient of these annual values is 0.92.

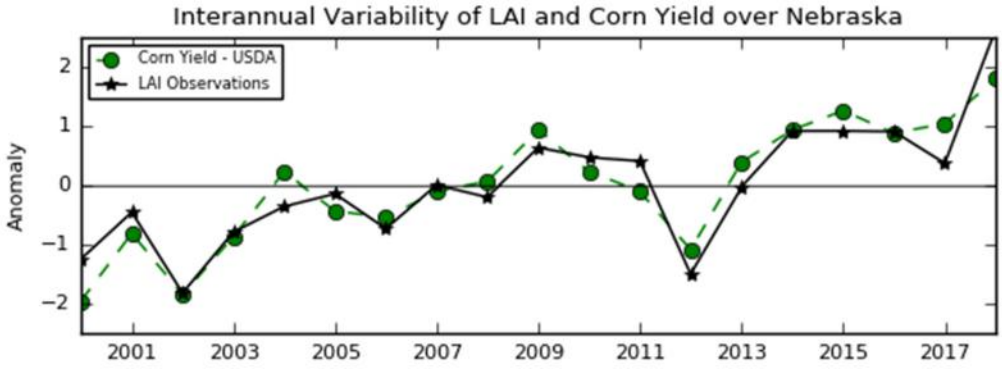

Figure 11. Inter-annual variability of CGLS LAI anomalies compared to USDA recorded corn yield anomalies over the U.S. state of Nebraska. Correlation coefficient of the anomalies is 0.92.

This comparison provides an example of the ability to use LAI as a proxy for crop yield, which, when coupled with the forecast capacity when integrated in LDAS-Monde, could be a useful tool for decision-makers and stakeholders. Inter-annual variability is adequately displayed by the observed LAI, and the two major drought events in this time span (2002 and 2012) are well captured.

5. Conclusions

LDAS-Monde has repeatedly proven to be a valuable tool for monitoring LSVs on the regional and continental scale. The process of data assimilation has also previously been shown to increase this value. This research advances these advantages by ingesting atmospheric forecasts allowing the prediction of future land surface conditions up to two weeks in advance with reasonable accuracy, particularly for LAI and ET. This study of only a two–year period did not show significant differences in SSM between the OL- and SEKF-initialized states when compared with in situ observations, possibly caused by the small time period. However, this difference is seen and is significant when comparing to satellite-derived LAI, SSM, and ET. Just as numerical weather forecasts are dependent on their initial conditions, forecasts of LSV are also strongly linked to their initial states. A consistent improvement of these initializations will continue to improve forecast accuracy.

This new capability of forecasting LSVs can be used by agricultural stakeholders in certain decision-making processes. One to two–week forecasts of SSM and ET can be a deciding factor for planting, harvesting, and irrigation timing. LAI forecasts can be used to estimate crop yield as demonstrated by the correlation of inter-annual variability between LAI and corn yield over Nebraska. The limiting factor in the capacity to provide the most accurate results is the accuracy of atmospheric forecasts themselves. Further work will investigate using longer range forecasts (at decreased spatial resolution) to provide a longer forecast period. This longer outlook may be helpful for seasonal instead of sub-seasonal trends and potentially a better prediction of annual crop yield. Additionally, future experiments will test the potential improvement from the assimilation of the more frequent VOD observations. These results will be the basis to build an agricultural drought monitoring and warning system based on LDAS-Monde for the purpose of adequately warning agricultural producers and decision-makers of the future states of the land surface.

Supplementary Materials: The following are available online at http://www.mdpi.com/2072-4292/12/12/2020/s1, Supplement – Microwave Vegetation Optical Depth. Figure S1: A) Correlation between LDAS-Monde SEKF LAI at FC1 vs. X-Band VOD from VODCA B) NICR illustrating the improvement or degradation in correlation between the VOD dataset and the SEKF and OL LAI C) Average monthly correlation scores between the VOD dataset and the SEKF and OL LAI.

Author Contributions: Conceptualization, A.M., C.A., B.B., Y.Z., and J.-C.C.; methodology, A.M., B.B., Y.Z., C.A., and J.-C.C.; investigation, A.M., C.A., B.B., Y.Z., and J.-C.C.; formal analysis, A.M..; supervision, C.A. and J.-C.C.; writing—original draft preparation, A.M..; writing—review and editing, A.M., C.A., B.B., Y.Z., and J.-C.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the French Make Our Plant Great Again programme, IRT Antoine de Saint-Exupéry Foundation, grant number CDT-R056-L00-T00 (POMME-V project), and by Météo-France.

Acknowledgments: The authors would like to thank the Copernicus Global Land Service for providing the satellite-derived LAI products. The authors would like to thank Christopher Hain for his help in accessing and processing the ALEXI ET dataset.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ASCAT	Advanced Scatterometer
CDF	Cumulative distribution function
CNRM	Centre National de Recherches Météorologiques
CGLS	Copernicus Global Land Service
CONUS	Contiguous United States
CTRL	Control run of ECMWF atmospheric forecasts
DA	Data Assimilation
ECMWF	European Centre for Medium-Range Weather Forecasts
ERA5	ECMWF Reanalysis 5th generation
ET	Evapotranspiration
FC	Forecast
ISBA	Interactions between Soil, Biosphere, and Atmosphere
LAI	Leaf Area Index
LDAS	Land Data Assimilation System
LSM	Land Surface Model
LSV	Land Surface Variable
NIC	Normalized contribution index
NOAA	National Oceanic and Atmospheric Administration
OL	Open-loop (simulation without assimilation)
PROBA-V	Project for On-Board Autonomy – Vegetation
RMSD	Root-Mean-Square Deviation
RZSM	Root-zone soil moisture
SEKF	Simplified Extended Kalman Filter
SSM	Surface Soil Moisture
SWI	Soil Wetness Index
SURFEX	Surface Externalisée (externalized surface models)
USCRN	U.S. Climate Reference Network
VOD	Vegetation Optical Depth

References

1. Pachauri, R.K.; Allen, M.R.; Barros, V.R.; Broome, J.; Cramer, W.; Christ, R.; Church, J.A.; Clarke, L.; Dahe, Q.; Dasgupta, P.; et al. *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; IPCC: Geneva, Switzerland, 2014; 151p, ISBN 9789291691432.
2. Ionita, M.; Tallaksen, L.M.; Kingston, D.G.; Stagge, J.H.; Laaha, G.; Van Lanen, H.A.J.; Scholz, P.; Chelcea, S.M.; Haslinger, K. The European 2015 drought from a climatological perspective. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 1397–1419. [CrossRef]
3. Bruce, J.P. Natural Disaster Reduction and Global Change. *Bull. Am. Meteorol. Soc.* **1994**, *75*, 1831–1835. [CrossRef]

4. Obasi, G.O.P. WMO's role in the international decade for natural disaster reduction. *Bull. Am. Meteorol. Soc.* **1994**, *75*, 1655–1662. [CrossRef]
5. Cook, E.R.; Seager, R.; Cane, M.A.; Stahle, D.W. North American drought: Reconstructions, causes, and consequences. *Earth-Sci. Rev.* **2007**, *81*, 93–134. [CrossRef]
6. Mishra, A.K.; Singh, V.P. A review of drought concepts. *J. Hydrol.* **2010**, *391*, 202–216. [CrossRef]
7. Gerber, N.; Mirzabaev, A. *Benefits of Action and Costs of Inaction: Drought Mitigation and Preparedness–a Literature Review*; Integrated Drought Management Programme (IDMP) Working Paper 1; WMO: Geneva, Switzerland; GWP: Stockholm, Sweden, 2017; p. 24.
8. Wilhite, D.A. Drought as a Natural Hazard: Concepts and Definitions. In *Drought: A Global Assessment*; Wilhite, D.A., Ed.; Routledge: London, UK, 2000; Volume 1, pp. 3–18. ISBN 9781315830896.
9. Di Napoli, C.; Pappenberger, F.; Cloke, H.L. Verification of Heat Stress Thresholds for a Health-Based Heat-Wave Definition. *J. Appl. Meteorol. Climatol.* **2019**, *58*, 1177–1194. [CrossRef]
10. Svoboda, M.; LeComte, D.; Hayes, M.; Heim, R.; Gleason, K.; Angel, J.; Rippey, B.; Tinker, R.; Palecki, M.; Stooksbury, D.; et al. The drought monitor. *Bull. Am. Meteorol. Soc.* **2002**, *83*, 1181–1190. [CrossRef]
11. Luo, L.; Wood, E.F. Monitoring and predicting the 2007 U.S. drought. *Geophys. Res. Lett.* **2007**, *34*, L22702. [CrossRef]
12. Dai, A.; Trenberth, K.E.; Qian, T. A Global Dataset of Palmer Drought Severity Index for 1870–2002: Relationship with Soil Moisture and Effects of Surface Warming. *J. Hydrometeorol.* **2004**, *5*, 1117–1130. [CrossRef]
13. Huang, L.; McDonald-Buller, E.C.; McGaughey, G.; Kimura, Y.; Allen, D.T. Annual variability in leaf area index and isoprene and monoterpene emissions during drought years in Texas. *Atmos. Environ.* **2014**, *92*, 240–249. [CrossRef]
14. Hanson, R.L. Evapotranspiration and droughts. In *National Water Summary 1988-89: Hydrologic Events and Floods and Droughts: U.S. Geological Survey Water-Supply Paper 2375*; Paulson, R.W., Chase, E.B., Roberts, R.S., Moody, D.W., Eds.; U.S. Geological Survey: Denver, CO, USA, 1991; pp. 99–104. [CrossRef]
15. Reichle, R.H.; McLaughlin, D.B.; Entekhabi, D. Hydrologic Data Assimilation with the Ensemble Kalman Filter. *Mon. Weather Rev.* **2002**, *130*, 103–114. [CrossRef]
16. Draper, C.; Mahfouf, J.-F.; Calvet, J.-C.; Martin, E.; Wagner, W. Assimilation of ASCAT near-surface soil moisture into the SIM hydrological model over France. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 3829–3841. [CrossRef]
17. Draper, C.S.; Reichle, R.H.; De Lannoy, G.J.M.; Liu, Q. Assimilation of passive and active microwave soil moisture retrievals. *Geophys. Res. Lett.* **2012**, *39*, L04401. [CrossRef]
18. Barbu, A.L.; Calvet, J.-C.; Mahfouf, J.-F.; Lafont, S. Integrating ASCAT surface soil moisture and GEOV1 leaf area index into the SURFEX modelling platform: A land data assimilation application over France. *Hydrol. Earth Syst. Sci.* **2014**, *18*, 173–192. [CrossRef]
19. Fairbairn, D.; Barbu, A.L.; Napoly, A.; Albergel, C.; Mahfouf, J.-F.; Calvet, J.-C. The effect of satellite-derived surface soil moisture and leaf area index land data assimilation on streamflow simulations over France. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 2015–2033. [CrossRef]
20. Blyverket, J.; Hamer, P.; Bertino, L.; Albergel, C.; Fairbairn, D.; Lahoz, W. An evaluation of the EnKF vs. EnOI and the assimilation of SMAP, SMOS and ESA CCI soil moisture data over the contiguous US. *Remote Sens.* **2019**, *11*, 478. [CrossRef]
21. Rodell, M.; Houser, P.R.; Jambor, U.; Gottschalck, J.; Mitchell, K.; Meng, C.-J.; Arsenault, K.; Cosgrove, B.; Radakovich, J.; Bosilovich, M.; et al. The global land data assimilation system. *Bull. Am. Meteorol. Soc.* **2004**, *85*, 381–394. [CrossRef]
22. Xia, Y.; Mitchell, K.; Ek, M.; Sheffield, J.; Cosgrove, B.; Wood, E.; Luo, L.; Alonge, C.; Wei, H.; Meng, J.; et al. Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *J. Geophys. Res. Atmos.* **2012**, *117*, D03109. [CrossRef]
23. Xia, Y.; Mitchell, K.; Ek, M.; Cosgrove, B.; Sheffield, J.; Luo, L.; Alonge, C.; Wei, H.; Meng, J.; Livneh, B.; et al. Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow. *J. Geophys. Res. Atmos.* **2012**, *117*, D03110. [CrossRef]

24. Sawada, Y.; Koike, T.; Walker, J.P. A land data assimilation system for simultaneous simulation of soil moisture and vegetation dynamics. *J. Geophys. Res. Atmos.* **2015**, *120*, 5910–5930. [CrossRef]
25. McNally, A.; Arsenault, K.; Kumar, S.; Shukla, S.; Peterson, P.; Wang, S.; Funk, C.; Peters-Lidard, C.D.; Verdin, J.P. A land data assimilation system for sub-Saharan Africa food and water security applications. *Sci. Data* **2017**, *4*, 170012. [CrossRef] [PubMed]
26. Jasinski, M.F.; Borak, J.S.; Kumar, S.V.; Mocko, D.M.; Peters-Lidard, C.D.; Rodell, M.; Rui, H.; Beaudoing, H.K.; Vollmer, B.E.; Arsenault, K.R.; et al. NCA-LDAS: Overview and analysis of hydrologic trends for the national climate assessment. *J. Hydrometeorol.* **2019**, *20*, 1595–1617. [CrossRef]
27. Kumar, S.V.; Zaitchik, B.F.; Peters-Lidard, C.D.; Rodell, M.; Reichle, R.; Li, B.; Jasinski, M.; Mocko, D.; Getirana, A.; De Lannoy, G.; et al. Assimilation of gridded GRACE terrestrial water storage estimates in the North American land data assimilation system. *J. Hydrometeorol.* **2016**, *17*, 1951–1972. [CrossRef]
28. Kumar, S.V.; Jasinski, M.; Mocko, D.M.; Rodell, M.; Borak, J.; Li, B.; Beaudoing, H.K.; Peters-Lidard, C.D. NCA-LDAS land analysis: Development and performance of a multisensor, multivariate land data assimilation system for the national climate assessment. *J. Hydrometeorol.* **2019**, *20*, 1571–1593. [CrossRef]
29. Noilhan, J.; Mahfouf, J.-F. The ISBA land surface parameterisation scheme. *Glob. Planet. Change* **1996**, *13*, 145–159. [CrossRef]
30. Calvet, J.-C.; Noilhan, J.; Roujean, J.-L.; Bessemoulin, P.; Cabelguenne, M.; Olioso, A.; Wigneron, J.-P. An interactive vegetation SVAT model tested against data from six contrasting sites. *Agric. For. Meteorol.* **1998**, *92*, 73–95. [CrossRef]
31. Calvet, J.-C.; Rivalland, V.; Picon-Cochard, C.; Guehl, J.-M. Modelling forest transpiration and CO_2 fluxes—response to soil moisture stress. *Agric. For. Meteorol.* **2004**, *124*, 143–156. [CrossRef]
32. Gibelin, A.-L.; Calvet, J.-C.; Roujean, J.-L.; Jarlan, L.; Los, S.O. Ability of the land surface model ISBA-A-gs to simulate leaf area index at the global scale: Comparison with satellites products. *J. Geophys. Res.* **2006**, *111*, D18102. [CrossRef]
33. Albergel, C.; Munier, S.; Leroux, D.J.; Dewaele, H.; Fairbairn, D.; Barbu, A.L.; Gelati, E.; Dorigo, W.; Faroux, S.; Meurey, C.; et al. Sequential assimilation of satellite-derived vegetation and soil moisture products using SURFEX_v8.0: LDAS-Monde assessment over the Euro-Mediterranean area. *Geosci. Model Dev.* **2017**, *10*, 3889–3912. [CrossRef]
34. Albergel, C.; Munier, S.; Bocher, A.; Bonan, B.; Zheng, Y.; Draper, C.; Leroux, D.; Calvet, J.-C. LDAS-Monde sequential assimilation of satellite derived observations applied to the contiguous US: An ERA-5 driven reanalysis of the land surface variables. *Remote Sens.* **2018**, *10*, 1627. [CrossRef]
35. Albergel, C.; Dutra, E.; Bonan, B.; Zheng, Y.; Munier, S.; Balsamo, G.; de Rosnay, P.; Muñoz-Sabater, J.; Calvet, J.-C. Monitoring and forecasting the impact of the 2018 summer heatwave on vegetation. *Remote Sens.* **2019**, *11*, 520. [CrossRef]
36. Albergel, C.; Zheng, Y.; Bonan, B.; Dutra, E.; Rodríguez-Fernández, N.; Munier, S.; Draper, C.; de Rosnay, P.; Muñoz-Sabater, J.; Balsamo, G.; et al. Data assimilation for continuous global assessment of severe conditions over terrestrial surfaces. *Hydrol. Earth Syst. Sci. Discuss.* 2019, in review. [CrossRef]
37. Leroux, D.; Calvet, J.-C.; Munier, S.; Albergel, C. Using satellite-derived vegetation products to evaluate LDAS-Monde over the Euro-Mediterranean area. *Remote Sens.* **2018**, *10*, 1199. [CrossRef]
38. Bell, J.E.; Palecki, M.A.; Baker, C.B.; Collins, W.G.; Lawrimore, J.H.; Leeper, R.D.; Hall, M.E.; Kochendorfer, J.; Meyers, T.P.; Wilson, T.; et al. U.S. Climate reference network soil moisture and temperature observations. *J. Hydrometeorol.* **2013**, *14*, 977–988. [CrossRef]
39. Anderson, M.; Norman, J.M.; Diak, G.R.; Kustas, W.P.; Mecikalski, J.R. A two-source time-integrated model for estimating surface fluxes using thermal infrared remote sensing. *Remote Sens. Environ.* **1997**, *60*, 195–216. [CrossRef]
40. Copernicus Global Land Service. Available online: https://land.copernicus.eu/global/ (accessed on 12 May 2020).
41. Wagner, W.; Lemoine, G.; Rott, H. A method for estimating soil moisture from ERS scatterometer and soil data. *Remote Sens. Environ.* **1999**, *70*, 191–207. [CrossRef]
42. Bartalis, Z.; Wagner, W.; Naeimi, V.; Hasenauer, S.; Scipal, K.; Bonekamp, H.; Figa, J.; Anderson, C. Initial soil moisture retrievals from the METOP-A Advanced Scatterometer (ASCAT). *Geophys. Res. Lett.* **2007**, *34*, L20401. [CrossRef]

43. Albergel, C.; Rüdiger, C.; Pellarin, T.; Calvet, J.-C.; Fritz, N.; Froissard, F.; Suquia, D.; Petitpa, A.; Piguet, B.; Martin, E. From near-surface to root-zone soil moisture using an exponential filter: An assessment of the method based on in-situ observations and model simulations. *Hydrol. Earth Syst. Sci.* **2008**, *12*, 1323–1337. [CrossRef]
44. Reichle, R.H.; Koster, R.D. Bias reduction in short records of satellite soil moisture. *Geophys. Res. Lett.* **2004**, *31*, L19501. [CrossRef]
45. Drusch, M.; Wood, E.F.; Gao, H. Observation operators for the direct assimilation of TRMM microwave imager retrieved soil moisture. *Geophys. Res. Lett.* **2005**, *32*, L15403. [CrossRef]
46. Scipal, K.; Holmes, T.; de Jeu, R.; Naeimi, V.; Wagner, W. A possible solution for the problem of estimating the error structure of global soil moisture data sets. *Geophys. Res. Lett.* **2008**, *35*, L24403. [CrossRef]
47. Verger, A.; Baret, F.; Weiss, M. Near real-time vegetation monitoring at global scale. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 3473–3481. [CrossRef]
48. Anderson, M.C.; Norman, J.M.; Mecikalski, J.R.; Otkin, J.A.; Kustas, W.P. A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing: 1. Model formulation. *J. Geophys. Res. Atmos.* **2007**, *112*. [CrossRef]
49. Anderson, M.C.; Norman, J.M.; Mecikalski, J.R.; Otkin, J.A.; Kustas, W.P. A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing: 2. Surface moisture climatology. *J. Geophys. Res.* **2007**, *112*, D11112. [CrossRef]
50. Anderson, M.C.; Hain, C.; Wardlow, B.; Pimstein, A.; Mecikalski, J.R.; Kustas, W.P. Evaluation of drought indices based on thermal remote sensing of evapotranspiration over the continental United States. *J. Clim.* **2011**, *24*, 2025–2044. [CrossRef]
51. Masson, V.; Le Moigne, P.; Martin, E.; Faroux, S.; Alias, A.; Alkama, R.; Belamari, S.; Barbu, A.; Boone, A.; Bouyssel, F.; et al. The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of Earth surface variables and fluxes. *Geosci. Model Dev.* **2013**, *6*, 929–960. [CrossRef]
52. Boone, A.; Masson, V.; Meyers, T.; Noilhan, J. The influence of the inclusion of soil freezing on simulations by a soil–vegetation–atmosphere transfer scheme. *J. Appl. Meteorol.* **2000**, *39*, 1544–1569. [CrossRef]
53. Decharme, B.; Boone, A.; Delire, C.; Noilhan, J. Local evaluation of the Interaction between Soil Biosphere Atmosphere soil multilayer diffusion scheme using four pedotransfer functions. *J. Geophys. Res.* **2011**, *116*, D20126. [CrossRef]
54. Calvet, J.-C.; Champeaux, J.-L. L'apport de la télédétection spatiale à la modélisation des surfaces continentales. *La Météorologie* **2020**, *108*, 52–58.
55. Faroux, S.; Kaptué Tchuenté, A.T.; Roujean, J.-L.; Masson, V.; Martin, E.; Le Moigne, P. ECOCLIMAP-II/Europe: A twofold database of ecosystems and surface parameters at 1 km resolution based on satellite information for use in land surface, meteorological and climate models. *Geosci. Model Dev.* **2013**, *6*, 563–582. [CrossRef]
56. Decharme, B.; Martin, E.; Faroux, S. Reconciling soil thermal and hydrological lower boundary conditions in land surface models. *J. Geophys. Res. Atmos.* **2013**, *118*, 7819–7834. [CrossRef]
57. Mahfouf, J.-F.; Bergaoui, K.; Draper, C.; Bouyssel, F.; Taillefer, F.; Taseva, L. A comparison of two off-line soil analysis schemes for assimilation of screen level observations. *J. Geophys. Res.* **2009**, *114*, D08105. [CrossRef]
58. Barbu, A.L.; Calvet, J.-C.; Mahfouf, J.-F.; Albergel, C.; Lafont, S. Assimilation of soil wetness index and leaf area index into the ISBA-A-gs land surface model: Grassland case study. *Biogeosciences* **2011**, *8*, 1971–1986. [CrossRef]
59. Bonan, B.; Albergel, C.; Zheng, Y.; Barbu, A.L.; Fairbairn, D.; Munier, S.; Calvet, J.-C. An ensemble square root filter for the joint assimilation of surface soil moisture and leaf area index within the Land Data Assimilation System LDAS-Monde: Application over the Euro-Mediterranean region. *Hydrol. Earth Syst. Sci.* **2020**, *24*, 325–347. [CrossRef]
60. Kumar, S.V.; Reichle, R.H.; Koster, R.D.; Crow, W.T.; Peters-Lidard, C.D. Role of subsurface physics in the assimilation of surface soil moisture observations. *J. Hydrometeorol.* **2009**, *10*, 1534–1547. [CrossRef]
61. Linsley, R.K. The relation between rainfall and runoff. *J. Hydrol.* **1967**, *5*, 297–311. [CrossRef]
62. Tall, M.; Albergel, C.; Bonan, B.; Zheng, Y.; Guichard, F.; Dramé, M.; Gaye, A.; Sintondji, L.; Hountondji, F.; Nikiema, P.; et al. Towards a long-term reanalysis of land surface variables over Western Africa: LDAS-Monde applied over Burkina Faso from 2001 to 2018. *Remote Sens.* **2019**, *11*, 735. [CrossRef]

63. Martens, B.; Miralles, D.G.; Lievens, H.; van der Schalie, R.; de Jeu, R.A.M.; Fernández-Prieto, D.; Beck, H.E.; Dorigo, W.A.; Verhoest, N.E.C. GLEAM v3: Satellite-based land evaporation and root-zone soil moisture. *Geosci. Model Dev.* **2017**, *10*, 1903–1925. [CrossRef]
64. Wood, A.W.; Lettenmaier, D.P. An ensemble approach for attribution of hydrologic prediction uncertainty. *Geophys. Res. Lett.* **2008**, *35*, 1–5. [CrossRef]
65. Shukla, S.; Sheffield, J.; Wood, E.F.; Lettenmaier, D.P. On the sources of global land surface hydrologic predictability. *Hydrol. Earth Syst. Sci.* **2013**, *17*, 2781–2796. [CrossRef]
66. Sawada, Y.; Koike, T.; Ikoma, E.; Kitsuregawa, M. Monitoring and predicting agricultural droughts for a water-limited sub-continental region by integrating a land surface model and microwave remote sensing. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 14–33. [CrossRef]
67. de Rosnay, P. A simplified Extended Kalman Filter for the global operational soil moisture analysis at ECMWF. *Q. J. R. Meteorol. Soc.* **2013**, *139*, 1199–1213. [CrossRef]
68. de Rosnay, P.; Balsamo, G.; Albergel, C.; Muñoz-Sabater, J.; Isaksen, L. Initialisation of land surface variables for numerical weather prediction. *Surv. Geophys.* **2014**, *35*, 607–621. [CrossRef]
69. Shamambo, D.; Bonan, B.; Calvet, J.-C.; Albergel, C.; Hahn, S. Interpretation of ASCAT radar scatterometer observations over land: A case study over southwestern France. *Remote Sens.* **2019**, *11*, 2842. [CrossRef]
70. Vreugdenhil, M.; Hahn, S.; Melzer, T.; Bauer-Marschallinger, B.; Reimer, C.; Dorigo, W.; Wagner, W. Characteristing vegetation dynamics over Australia with ASCAT. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 2240–2248. [CrossRef]
71. Dewaele, H.; Munier, S.; Albergel, C.; Planque, C.; Laanaia, N.; Carrer, D.; Calvet, J.-C. Parameter optimisation for a better representation of drought by LSMs: Inverse modelling vs. sequential data assimilation. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 4861–4878. [CrossRef]
72. Kross, A.; McNairn, H.; Lapen, D.; Sunohara, M.; Champagne, C. Assessment of RapidEye vegetation indices for estimation of leaf area index and biomass in corn and soybean crops. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *34*, 235–248. [CrossRef]
73. USDA (United States Department of Agriculture) National Agricultural Statistics Service. Available online: https://www.nass.usda.gov/Data_and_Statistics/ (accessed on 12 May 2020).

3.2.2 Towards Assimilating VOD

As the conclusions of Mucia et al. (2020) discussed, more frequent assimilation constraining vegetation in the model could improve LSV correlations. VOD, being an all weather observation unaffected by clouds, provides far more frequent observations than LAI. Therefore, supplementary and complementary work to Mucia et al. (2020) explored the direct statistical relationship between the LDAS-Monde LAI before and after data assimilation, and observed VOD observations.

Using a similar LDAS-Monde system configuration over CONUS, Albergel et al. (2018b) have compared modeled LAI and microwave-derived VOD from radar backscatter measurements of ASCAT for 2010-2016. They found high correlation values in large parts of the domain, with a median value of 0.57. The northern part of the CONUS domain showed R values greater than 0.7, while smaller R values (and even negative R values) were found in the southern part of the domain. They suggested that over dry soils, sub-surface scattering from the microwave signal may have affected the VOD estimates. They also showed that the same VOD dataset had a higher median R value with the observed CGLS LAI data, that is, 0.88. Consequently it was better correlated with the analysis (median R value of 0.61) than with the model.

An additional analysis of the link between observed VOD and modeled LAI was performed using VODCA VODX over the United States domain, showing similar trends. Figure 3.1 demonstrates A) the spatial distribution of correlations of SEKF day 1 forecasts of LDAS-Monde LAI versus VODX from VODCA, B) the NICR showing the improvement or degradation in correlation between the VOD dataset and the SEKF and OL LAI, and C) average monthly correlations between VODX and the SEKF and OL LAI. Strong correlations are seen in the East Coast, and Great Plains, while some Southern and Southwestern regions see very low or negative correlations. As in Albergel et al. (2018b), the SEKF gave stronger average correlations than the OL, as shown in most of the domain in panel B and over most months in panel C. The improvement in correlation by using the SEKF LAI over the OL LAI also indicates that the assimilated LAI observations and VODX observations contain much of the same vegetation-related information.

This direct correlation of VOD and LAI shows that these values are strongly linked, and is worthy to be studied further. As previously stated, assimilating vegetation-related information at the far higher frequency of VOD observations has the potential to improve the capability of LDAS-Monde to monitor LSVs. This improvement directly translates to improved initial conditions of the LDAS, which has just been demonstrated to be vitally important to accurate land surface forecasts for many LSVs. Additionally, more comparisons can be performed regarding the impact of the individual and joint assimilation of these LAI, VOD, and/or SSM observations, which can reaffirm their respective strengths and overall value. Chapter 4 explores and analyzes these ideas.

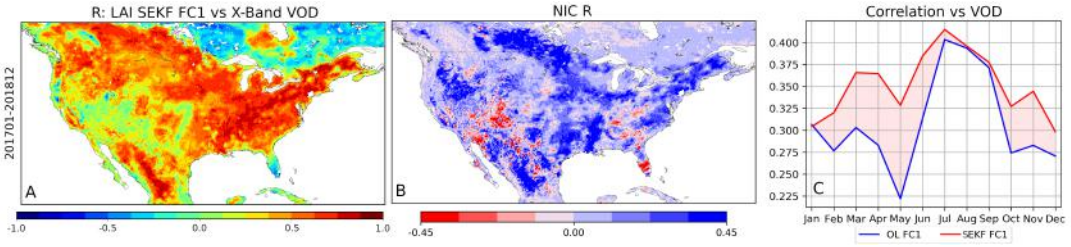

Figure 3.1: A) Correlation between LDAS-Monde SEKF LAI at FC1 vs. X-Band VOD from VODCA B) NIC R illustrating the improvement or degradation in correlation between the VOD dataset and the SEKF and OL LAI C) Average monthly correlation scores between the VOD dataset and the SEKF and OL LAI.

3.2.3 Mucia et al., 2020 Summary and Conclusions

This article successfully demonstrates the ability of LDAS-Monde to forecast LSVs linked to drought and water resources over CONUS. In particular, LAI and ET variables are shown to be reasonably accurate out to two weeks in advance. While this configuration of the ISBA LSM does not simulate irrigation processes, the assimilation of the satellite observations does drive corrections of the initial conditions from these human impacts. The practicality of these forecasts for stakeholders is also quickly addressed. Finally, this article states that these land surface forecasts can and will be used as the basis for an agricultural drought early warning system. The next steps are laid out, detailing the experimentation with assimilating VOD to potentially improve the initial conditions, which as proven in this article, greatly impact the performance of the forecasts.

The supplementary material also addresses this VOD assimilation, taking the first step in the assessment of the VOD dataset. A direct correlation was taken between VOD and LAI from LDAS-Monde SEKF on the first day of forecast, showing variable correlations across CONUS, but generally high mean and median scores. This material also describes the precise VOD data that would be used in the assimilation, demonstrating that this specific VOD and LAI are highly linked, justifying further assimilation studies.

3.3 Summary of Chapter 3

This chapter detailed the first experiments tested and run, initially over the US state of Nebraska, then the major forecast experiment over CONUS, which resulted in a published article. Information includes:

- Initial conditions of physical models are described, with an emphasis on how they can be improved via data assimilation.
- An article on the LDAS-Monde transition from monitoring to forecasting over CONUS was presented, with an expanded introduction and conclusion. This article demonstrates that LDAS-Monde can transition to a forecast mode, that the forecasts of LSVs can provide useful information

up to two weeks in advance, and that the initial conditions of the model play a large role in forecast accuracy.

– Supplementary material was also analyzed which looks at a basic relationship between VOD and LAI, which serves as a link to future work assimilating VOD to potentially improve initial conditions.

Chapter 4

Improving Initial Conditions Using the LDAS

This chapter explores the relationship between Vegetation Optical Depth (VOD) and Leaf Area Index (LAI). Analysis is done examining the regressions between both VODC and VODX against LAI observations (along with LAI from the ISBA LSM OL) over the entire CONUS domain. Additional analysis is performed over six dominant vegetation covers as defined by both ECOCLIMAP-II and ECOCLIMAP-SG comparing first unmatched VODX then matched VODX to LAI observations over those patches. A brief analysis is also given on the general tendency of VOD anomalies compared to LAI and corn yield anomalies seen previously in Chapter 3.

The remainder of the chapter describes and analyzes the effect of assimilating matched VODX into LDAS-Monde as an LAI proxy. Various other assimilation scenarios are also explored, including the assimilation of matched VODX only when an LAI observation is present, as well as matching VODX to LAI observations that have been interpolated to daily availability. The variables of LAI, GPP, ET, and SSM are compared to satellite derived observations, and probability distribution functions of the correlations are compared. Correlation scores are also compared to in situ observations of soil moisture through the USCRN. In addition to assessing the impact of assimilating matched VODX, several comparisons are made describing the improvement seen by assimilating joint vegetation (LAI or matched VODX) as well as SSM.

4.1 Vegetation Optical Depth as an LAI Proxy

Chapter 2 has previously discussed VOD, its definition, uses, strengths, and weaknesses. It has also touched on some key questions, such as the effects of different VOD bands, and VOD responses over different vegetation types. This section helps answer some of those questions.

Presented in this section are numerous comparisons between VOD and LAI over different vegetation types. Both X-band and C-band VOD are compared in several cases. This is because, while X-band was chosen for transformation and assimilation, C-band is a perfectly fine alternative, and investigating the properties and responses of two bands of VOD has the potential to provoke discoveries regarding model and vegetation behavior.

4.1.1 Transforming VOD into an LAI Proxy

While VOD can work as a proxy to vegetation, and even be transformed into a proxy for LAI, it is important to note that VOD is not LAI. This transformation will produce some error, and the model treating the matched VOD as LAI in its assimilation will produce additional errors.

As mentioned in Chapter 2, a simple linear re-scaling method was employed, similar to CDF matching, to match the VOD observations to the LAI observations. A 3-month re-scaling length was selected, and then a 30-day rolling mean was applied to the result to smooth it.

4.1.2 Comparison over CONUS

This section will compare the data from the overall CONUS domain. Appendix C provides this same comparison over the subdomains shown in Chapter 2 and Figure 2.10.

These density scatter plots represent all times and points when there is both VOD and LAI data. A linear regression and correlation score have been plotted over the data. Table 4.1 gives these correlations in a single table. Logarithmic transformations to the VOD data were also applied, but with no significant increase in regression correlation or regression shape. Furthermore, a linear regression and re-scaling was ultimately selected as it is supported by previous experiments (Kumar et al., 2020), and it is more robust over varying vegetation types. As will be shown in the VOD and LAI comparisons over patches dominated by different vegetation, some vegetation has more of a logarithmic shape, while other types have a far more linear shape. These comparisons only looks at the months of April-September, which are typically considered the "growing season" in much of the CONUS domain. All density scatter plots are set to have the same x and y axes for better visual comparison. The colorbar is logarithmically scaled in order to emphasize the distributions, and bins with under 5 count are eliminated. Figure 2.2 can be used to suppose certain VOD or LAI responses over different regions of the US based on dominant vegetation type in ECOCLIMAP-II or ECOCLIMAP-SG.

Table 4.1: Correlation coefficients between VODCA VODX and VODC versus CGLS LAI and ISBA LAI over the various subdomains

Domain	CGLS LAI vs VODC	CGLS LAI vs VODX	ISBA LAI vs VODC	ISBA LAI vs VODX
CONUS	0.482	0.660	0.659	0.795
California	0.184	0.181	0.539	0.557
Midwest	0.610	0.717	0.700	0.761
Northeast	0.007	0.413	0.329	0.667
Southern Plains	0.346	0.317	0.595	0.587
Nebraska	0.517	0.637	0.548	0.608

On average over the CONUS domain, there is a moderately strong relationship between LAI and VOD, for both VODC in Figure 4.1 and VODX in Figure 4.2. Panel A is LAI from observations in the CGLS dataset plotted against VOD from the VODCA dataset. Panel B is LAI from the ISBA LSM (with no data assimilation), using ECOCLIMAP-SG land surface parameterization plotted against VOD from VODCA. A pattern observed here, and in the subdomains analyzed in Appendix C, is that VODX values are consistently higher than VODC values, as well the VODC comparison containing more noise. This may be explained by C-band radar's longer wavelengths better penetrating larger canopies, and retrieving more ground or low vegetation backscatter (Frappart et al., 2020; Jiao et al., 2010). The higher noise from VODC is part of why over CONUS, the correlations between VODX and LAI are higher. Over CONUS, the VODC-LAI observation correlation is 0.48, while the VODX-LAI observation correlation is 0.66.

While the LAI observations are moderately well correlated to VOD, the relationship of VOD with LAI from ISBA (Panel B) shows consistently stronger values. The VODC-LAI ISBA correlation is 0.66 while VODX-LAI ISBA has a 0.80 correlation value. Visually, we can see at higher LAI values, there is a more dramatic increase in VOD in Panel B, whereas in Panel A, the VOD values seem almost as if they flatten out. This is partly seen in Panel B, but is also slightly compensated by progressively higher VOD, while eliminating low VOD values at high LAI. Additionally, Panel B clearly shows certain artificial thresholds from ISBA, including the 0.3 m^2/m^2 lower limit LAI for most vegetation and the 1 m^2/m^2 lower limit for evergreen forests. These artifacts can be seen at a wide range of VOD values, and are also visible in these graphics for the sub-domains (Appendix C). As explained in Chapter 2, these lower limits in ISBA are often reached in Winter months when vegetation activity and LAI are lowest. Because this comparison, as well as those of the subdomains to follow, compare VOD to LAI at all points in time, these limited minimum values are seen.

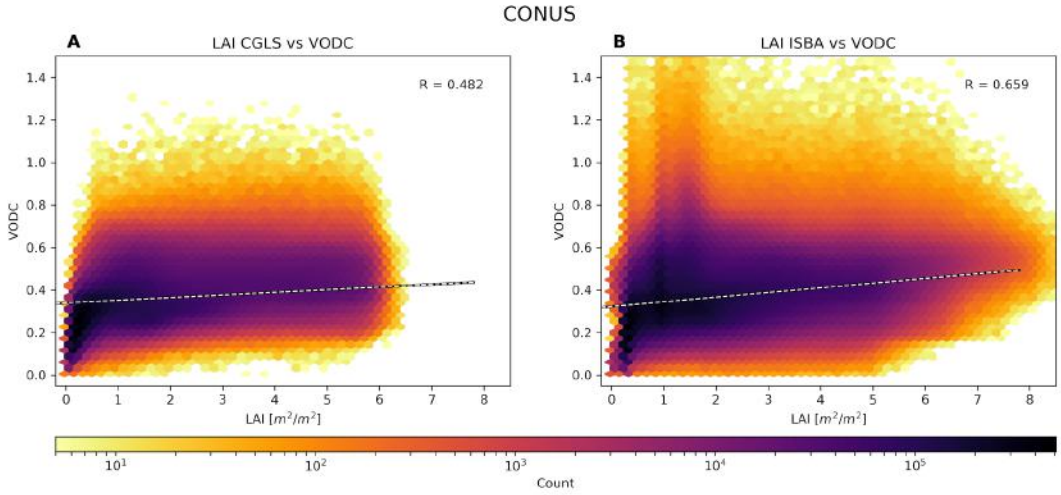

Figure 4.1: A density scatter plot detailing the relationship between LAI and VODC from VODCA. This comparison only analyzes where and when there are both LAI and VOD observations. Warm colors represent more counts of points in the hexagonal bins. The colorbar is logarithmically scaled in order to emphasize the distributions, and bins with under 5 count are eliminated. A regression line and correlation score are added. This comparison only looks at points during the growing season months (April-September). Panel A) compares LAI from CGLS observations while panel B) compares to LAI from ISBA OL.

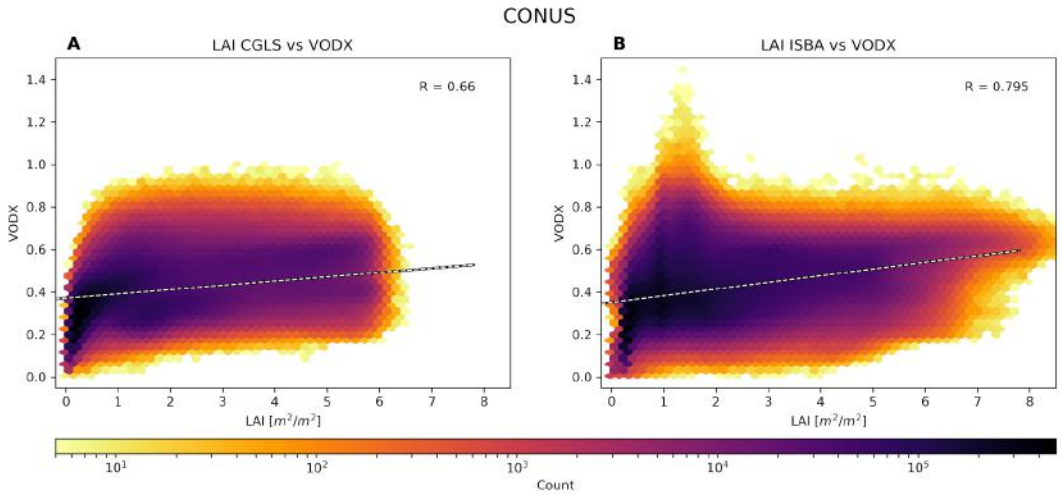

Figure 4.2: Same as Figure 4.1 but with VODX instead of VODC

4.1.3 Anomaly Comparison of Corn Yield to VOD and LAI over Nebraska

It is clear that the VOD observations from VODCA share a significant amount of information with LAI, even before being linearly rescaled, as just demonstrated over the various vegetation types. In order to take a slightly deeper look into the relationship between VOD and C4 crops, the same comparison as in Figure A.2 was performed, but adding VODC and VODX anomalies. As seen in Table A.1, the OL LAI anomaly had a 0.80 correlation to the annual corn yield anomaly, while the SEKF (assimilating both LAI and SSM) saw an improved score of 0.85. The VOD anomalies also provide relatively high correlation scores with corn yield, at 0.77 for VODX and 0.80 for VODC.

Overall, the variation between VOD observations and corn yield is not perfect, but the general tendency is similar. Most notably, the 2012 drought is not strongly seen in VOD, but from 2013 to 2018, the anomalies match well. This is just another indicator that shows that VOD observations have utility in monitoring vegetation. And while they do not perfectly match any direct drought or vegetation health index, they provide useful information regardless, and in the context of this thesis, they provide far more frequent observations.

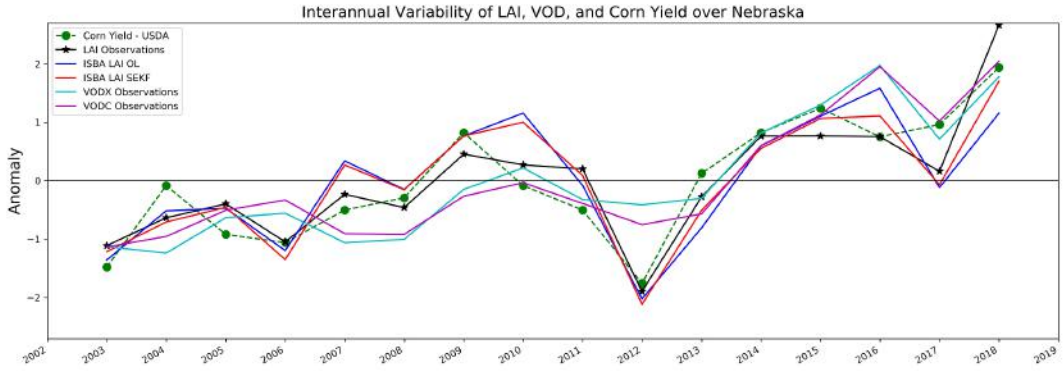

Figure 4.3: A time series representing the inter-annual anomalies of observed, modelled, and analyzed LAI, and corn yield over Nebraska. The black line and stars represent the mean annual observed LAI anomalies from the CGLS LAI V2 dataset. The green dashed line represents annual corn yield anomalies as reported by the USDA. The blue and red lines are the LDAS-Monde OL and SEKF LAI products respectively. The cyan and magenta lines are VODX and VODC observations from VODCA respectively.

4.1.4 Relationship Between VOD and LAI over Dominant Vegetation Types

As a more direct way to look at VOD and LAI responses over different vegetation types, their relationship is compared over six different dominant vegetation types in the following section. Dominant vegetation is defined as 50% or more of a single type in a patch as defined by the ECOCLIMAP land use database(s). The six patches investigated are A) deciduous forests, B) coniferous forests, C)

C3 crops, D) C4 crops, E) C3 grasslands, and F) irrigated crops, and are labeled as such in the figures to follow. This section looks at the response for both ECOCLIMAP-II and ECOCLIMAP-SG databases over the patches. The tables also present the correlations over all types of vegetation, where there is less than 10% of non-vegetated surfaces in a patch defined by each ECOCLIMAP version respectively. As there are changes to these non-vegetated surfaces, between the two versions, the correlations themselves over all vegetation can differ. The contrast between the two ECOCLIMAP versions can also help shed light on large changes or mischaracterizations of vegetation that may be highlighted in model results. The primary focus is on the VODX product from VODCA, and all the tables and figures in the chapter are dealing with the X-Band, as this band was chosen to assimilate in the LDAS-Monde experiments. However, this same analysis was performed on the C-Band from VODCA and the results are laid out in Appendix B.

This section will first look at ECOCLIMAP-II, and how the relationship between LAI observations from CGLS and VODX observations from VODCA change with vegetation type. The same analysis is then done, but substituting the matched VODX. This process is repeated in the same way for ECOCLIMAP-SG. Each of the following density scatter plots records any point and time when there are both VODX and LAI observations for the patches dominated by the specific vegetation type of interest, with darker colors representing a higher concentration of points. A seasonal analysis is also shown using the meteorological seasons (Winter:DJF, Spring:MAM, Summer:JJA, and Autumn:SON), with each season being a spatial average of the days with observations. Those dots are color coded as follows: cyan is Winter, green is Spring, red is Summer, and yellow is Autumn. After each density scatter plot, the correlation scores are given in a table, for the entire time period, as well as split up into the same meteorological seasons.

LAI versus VODX

Beginning with the earlier ECOCLIMAP version, ECOCLIMAP-II, it is immediately apparent that vegetation type plays huge role in the LAI and VOD response. Figure 4.4 shows the density scatter plot of the patches along with the average seasonal correlations in the top left, and Table 4.2 gives all the correlation values along with the number of points in each for the seasonal calculations. Of all the vegetation types, coniferous forests were the only to have a negative all season correlation at -0.09. All of the other vegetation types show a moderately strong correlation between 0.43 for irrigated crops and 0.68 for C4 crops.

The seasonality does play a strong part as well, and the figures show a clear separation of values according to seasons over the patches. Winter correlations are typically low for all vegetation types, and the cyan seen in the graphs are often clumped at low LAI values. Spring scores are on average increased, and contain a far wider range of LAI values, but similar range of VOD values. Summer and Autumn see the highest correlation scores and are characterized by a wider range of LAI and VOD values. Notably, deciduous forests have

a negative correlation during Summer months, but the Autumn correlation is strongly positive.

As previously mentioned, these same tables and figures are given for C-band VOD in Appendix B. In general, the same conclusions can be gained from looking at the C-band. Correlations are generally slightly degraded, and there are certainly more outliers in the high VOD values. Additionally, the correlations between VODC and LAI observations over forests are weaker than with VODX.

Using different land use databases does show some different LAI and VOD responses, while the same overall trends are still present. ECOCLIMAP-SG patches show almost identical shapes in the density scatter plots, shown in Figure 4.5, but do see some correlation strength changes as well as slightly altered seasonal responses, shown in Table 4.3. The shape of the distribution of deciduous forests is slightly changed as well, with a significantly higher concentration of low LAI and low VOD points, which in turn, skew the seasonal means lower. The density scatter plot colors, which are fixed between all dominant patch plots, indicate that ECOCLIMAP-SG has significantly more patches dominated by deciduous forests compared to ECOCLIMAP-II. Irrigated crops also see stronger changes, with ECOCLIMAP-SG strongly reducing the number of these patches, and the seasonal averages are far more variable. C3 and C4 crops are mostly similar, with the starkest change being a wider range of seasonal mean averages in ECOCLIMAP-SG. In the case of C3 crops, this is likely caused by a visible increase in the range of LAI values over C3 patches from ECOCLIMAP-SG.

Considering all vegetation, ECOCLIMAP-SG allows for better correlations between LAI and VODX observations, but individual vegetation types have more variability. The correlations over coniferous forests are lowered the most dramatically from ECOCLIMAP-II to ECOCLIMAP-SG, while deciduous, C3 crops, C4 crops, and C3 grasslands vegetation are overall similar between the two databases. Irrigated crops also show a dramatic decrease, going from weak or moderately positive, to almost zero correlation. This response is likely due to the far fewer patches of irrigated crops included in ECOCLIMAP-SG. Seasonal patterns are quite similar, with Winter having lowest correlations, and spring having the highest, but using ECOCLIMAP-SG reduces the positive summer correlations for most vegetation types except C3 and C4 crops.

These differences do suggest that the selection of land use parameters used in the LSM is important and can change the interpretation of result, especially if looking at small domains. Even the relatively minor changes between ECOCLIMAP-II and ECOCLIMAP-SG are visible in this comparison over a very large domain. Smaller domains that see a greater relative change between the two versions would amplify these differences. Additionally, the relatively straightforward and high correlations of LAI and VODX over C3, C4 crops, and C3 grasslands patches signal that the Midwest and Great Plains regions of the United States may be the most appropriate regions where VOD may act as an LAI proxy.

Table 4.2: Seasonal and total correlations of VODX versus LAI for ECOCLIMAP-II patches

VODX ECOII					
Veg Type	All Seasons (575)	Winter (144)	Spring (143)	Summer (144)	Autumn (144)
All Vegetation	0.58	0.57	0.70	0.68	0.56
Deciduous	0.56	-0.16	0.22	-0.23	0.78
Coniferous	-0.09	0.23	-0.19	0.16	-0.04
C3 Crops	0.67	-0.04	0.46	0.47	0.61
C4 Crops	0.68	-0.04	-0.34	0.85	0.74
C3 Grasslands	0.58	0.34	0.59	0.62	0.30
Irrigated Crops	0.43	0.18	0.31	0.57	0.34

Table 4.3: Seasonal and total correlations of VODX versus LAI for ECOCLIMAP-SG patches

VODX ECOSG					
Veg Type	All Seasons (575)	Winter (144)	Spring (143)	Summer (144)	Autumn (144)
All Vegetation	0.76	0.37	0.74	0.32	0.61
Deciduous	0.53	0.57	0.39	0.67	0.56
Coniferous	-0.46	0.17	-0.42	-0.06	-0.23
C3 Crops	0.73	0.07	0.02	0.76	0.56
C4 Crops	0.68	-0.33	-0.21	0.89	0.60
C3 Grasslands	0.68	0.42	0.61	0.58	0.57
Irrigated Crops	0.14	-0.02	0.21	0.10	0.03

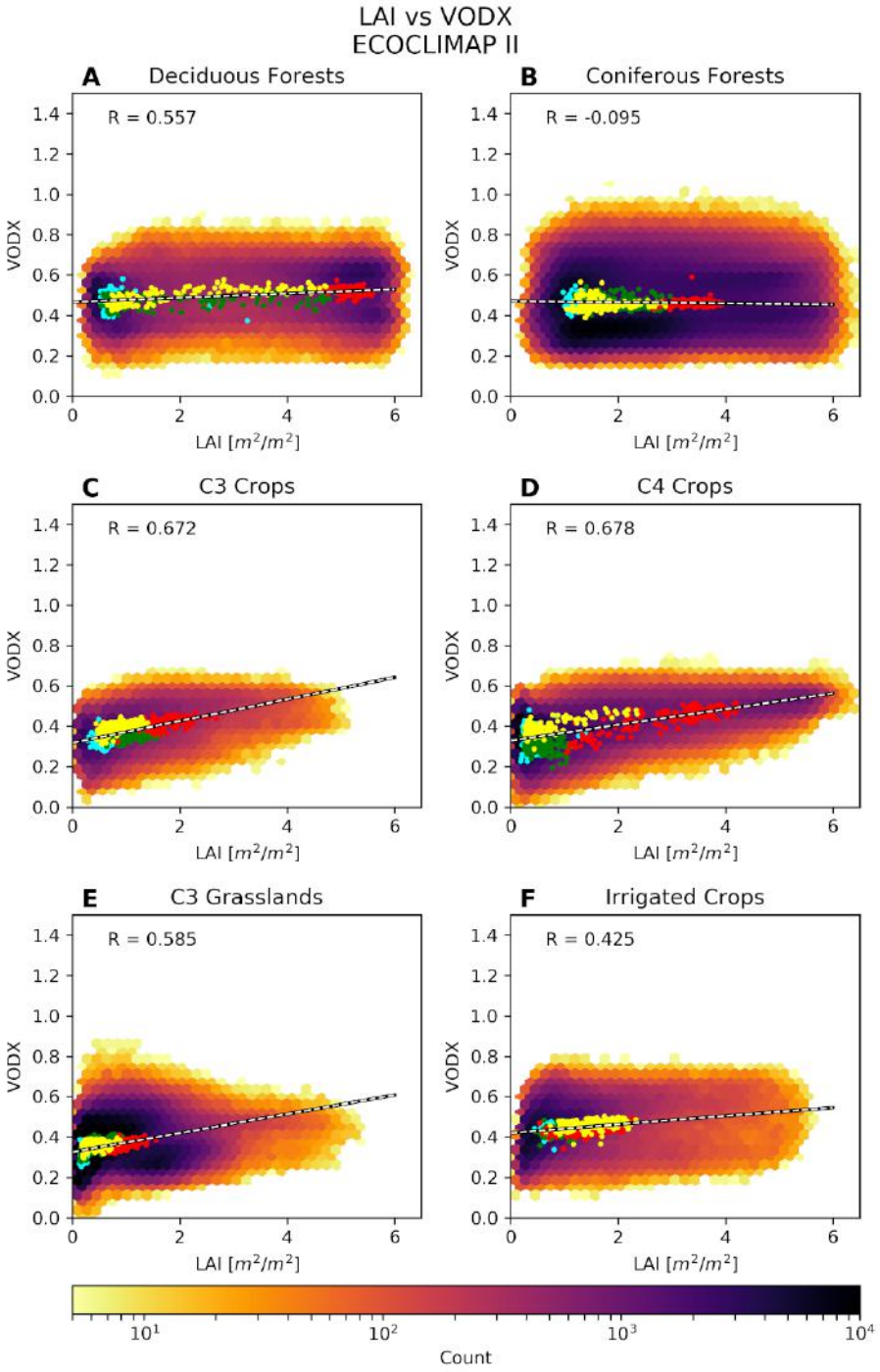

Figure 4.4: A density scatter plot detailing the relationship between LAI and VODX from VODCA over six dominant vegetation types, A) Deciduous Forests, B) Coniferous Forests, C) C3 Crops, D) C4 Crops, E) C3 Grasslands, and F) Irrigated Crops. Dominant vegetation is defined as where 50% or more of a patch containing a single vegetation type. Higher concentrations of points trend towards black. Colored dots represent the spatial average over the four seasons, where cyan is Winter, green is Spring, red is Summer, and yellow is Autumn. Black and white dashed lines represent the linear regression of the seasonal scores.

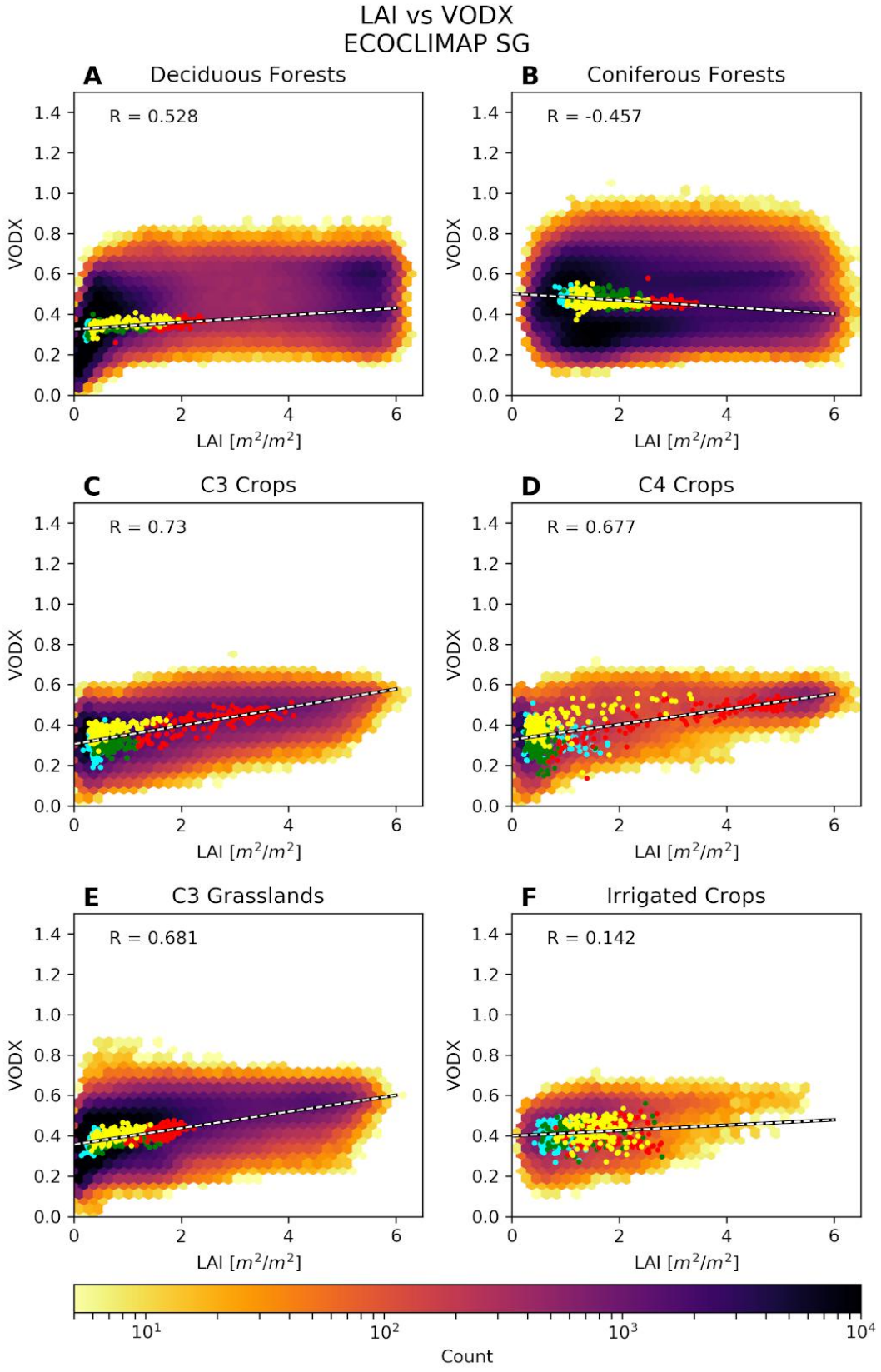

Figure 4.5: Same as Figure 4.4, but using ECOCLIMAP-SG instead of ECOCLIMAP-II.

LAI versus Matched VODX

Now using matched VODX to compare to LAI observations, it is shown that that the correlations between the two datasets increase dramatically. Visually, the data are strongly aligned (Figure 4.6), and the correlations in Table 4.4 have scores closely approaching the maximum of 1. Deciduous and coniferous forest along with C4 crops share a relatively wide variation, while C3 crops, C3 grasslands and irrigated crops have more compact distributions. These distributions are likely due to the natural tendency of forests to have higher LAI values, while most C3 plants are far lower. In the United States, the primary C4 crops are corn (maize), millet, sorghum, and in a few parts, sugar cane. C4 plants have higher optimal leaf areas compared to C3 (Anten et al., 1995), and seasonality plays a clear role in the distributions of LAI and matched VOD. The seasonal variations generally widen slightly compared to non-matched VOD. For all vegetation types except for irrigated crops, Summer months provide clearly higher coupled LAI and VOD values. Irrigated crops instead show that Autumn and Summer compete for the highest levels. This is a response from the far less water restricted environment of irrigated fields, leading to a longer growth phase and thus delaying the timing of maximum leaf area, even when the region is under light to moderate water stress.

Additional things to note are that in deciduous forests, C3 and C4 crops, there is a clear indication that matched VOD in Autumn months are higher than expected from the LAI observations. In fact, both deciduous forests (Figure 4.6 Panel A) and C4 crops (Figure 4.6 Panel D) show a signal of hysteresis, where the timing of the LAI or matched VOD observation changes the pattern. In the case of C4 crops, I hypothesize that heavier ground litter during and after harvesting produces a stronger VOD response while at lower LAI. This effect is more apparent in C4 crops potentially due to their overall higher LAI and VOD values. Another potential cause of this hysteresis could be from leaf water interception, both on growing plants and ground litter, which would also disproportionately increase the VOD observations, leading to an offset. Yet another explanation could be linked to the surface roughness. Fernandez-Moran et al. (2015) and Hornbuckle et al. (2016) found that satellite microwave retrievals are impacted by surface roughness as well as vegetation water content and soil moisture. As management decisions for agricultural producers can impact a field's roughness, (i.e by plowing or tilling), it is possible higher values of microwave retrievals, and ultimately VOD, may be at least partially explained by a higher roughness. Further analysis of LAI and VOD over agricultural lands of varying roughness could be conducted towards better understanding this phenomenon.

This mismatch in seasonality between matched VOD and LAI does pose the question: is the linear re-scaling, as implemented, sufficient to account for the differences in values and trends between these two datasets? The 3-month re-scaling period is meant to account for this type of response, but clearly does not perfectly transform the VOD product into a matching LAI proxy. While that is somewhat expected, these results can be further analyzed and can lead to finding even more solutions.

When comparing ECOCLIMAP-II to ECOCLIMAP-SG for LAI versus matched VODX, the overall shapes of the plots are similar. Figure 4.7 provides the same figure comparisons, but for ECOCLIMAP-SG. Some of the notable differences include the far more concentrated seasonal means in deciduous forests, the overall increase in LAI and matched VODX values for C3 crops, the enhanced hysteresis seen in C3 (Figure 4.7 Panel C), and especially C4 crops (Figure 4.7 Panel D), and the skewed regression over irrigated crops. The seasonal variation in deciduous forests strongly shrinks from ECOCLIMAP-II to ECOCLIMAP-SG, but the general pattern of highest values in Summer, Autumn and Spring sharing the transition period, and Winter values being lowest, remains. This shift actually increases the seasonal correlation in Winter, but lowers it in Summer. C3 crop patches show a general expansion of LAI and matched VOD values, widening the shape of the plot, and their seasonal means. This change could be due to ECOCLIMAP-SG increasing the amount of C3 crops in patches, which in turn would widen the range of observed LAI values. It is likely that this expansion directly leads to the enhanced hysteresis pattern seen in C3 crops, with the Autumn averages trending consistently above the regression. The C4 crop patches see the same widening of the hysteresis, but with an overall loss of number of C4 crop patches. One potential explanation to this response is that ECOCLIMAP-SG shifted a number of C4 crop patch fractions to C3 crops, resulting in an overall smaller selection of C4 crop dominated patches. If this change moves the land cover closer to reality, it suggests that C4 crops are more susceptible to this hysteresis phenomenon. This also again calls into question if the linear re-scaling is sufficient to account for these seasonal changes.

The use of linear re-scaling to produce matched VOD observations dramatically increases correlation over all vegetation types. While it is important to keep in mind VOD is not LAI, and the two datasets, even after re-scaling will never be perfectly matched, the strong agreement after re-scaling is a promising sign for VOD's use in data assimilation, discussed later in this chapter.

Table 4.4: Seasonal and total correlations of Matched VODX versus LAI for ECOCLIMAP-II patches

	Matched VODX ECOII				
Veg Type	All Seasons (575)	Winter (144)	Spring (143)	Summer (144)	Autumn (144)
All Vegetation	0.93	0.72	0.89	0.72	0.90
Deciduous	0.97	0.09	0.96	0.29	0.94
Coniferous	0.95	0.20	0.72	0.67	0.90
C3 Crops	0.95	0.49	0.83	0.69	0.91
C4 Crops	0.94	0.04	0.89	0.87	0.94
C3 Grasslands	0.95	0.48	0.90	0.87	0.93
Irrigated Crops	0.93	0.71	0.63	0.88	0.90

Table 4.5: Seasonal and total correlations of Matched VODX versus LAI for ECOCLIMAP-SG patches

	Matched VODX ECOSG				
Veg Type	All Seasons (575)	Winter (144)	Spring (143)	Summer (144)	Autumn (144)
All Vegetation	0.96	0.29	0.92	0.35	0.96
Deciduous	0.96	0.61	0.91	0.14	0.95
Coniferous	0.97	0.27	0.82	0.63	0.94
C3 Crops	0.95	0.26	0.74	0.80	0.95
C4 Crops	0.92	0.41	0.46	0.92	0.89
C3 Grasslands	0.96	0.30	0.93	0.72	0.95
Irrigated Crops	0.83	0.47	0.76	0.62	0.67

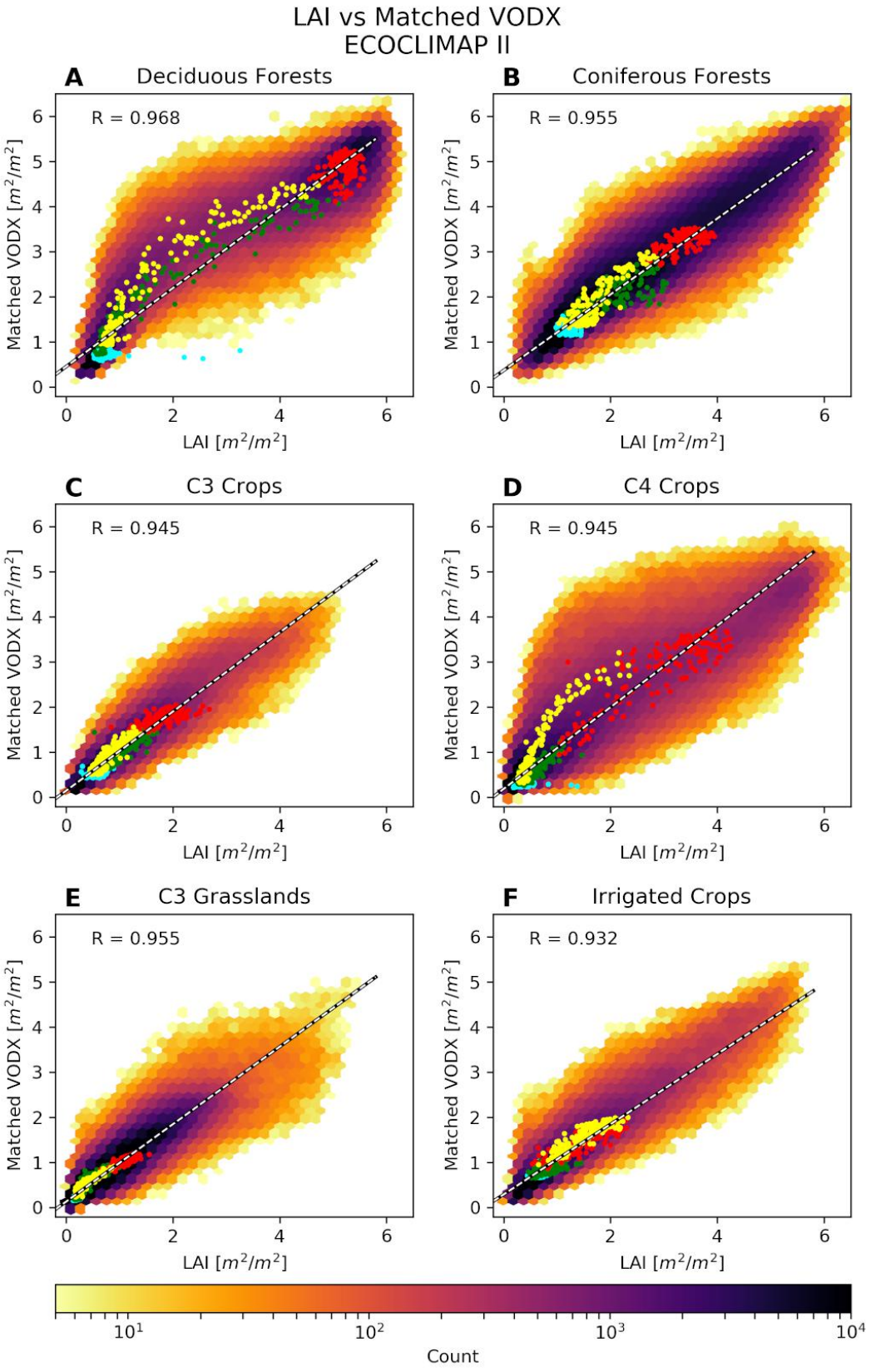

Figure 4.6: Same as Figure 4.4, but with matched VODX instead of raw.

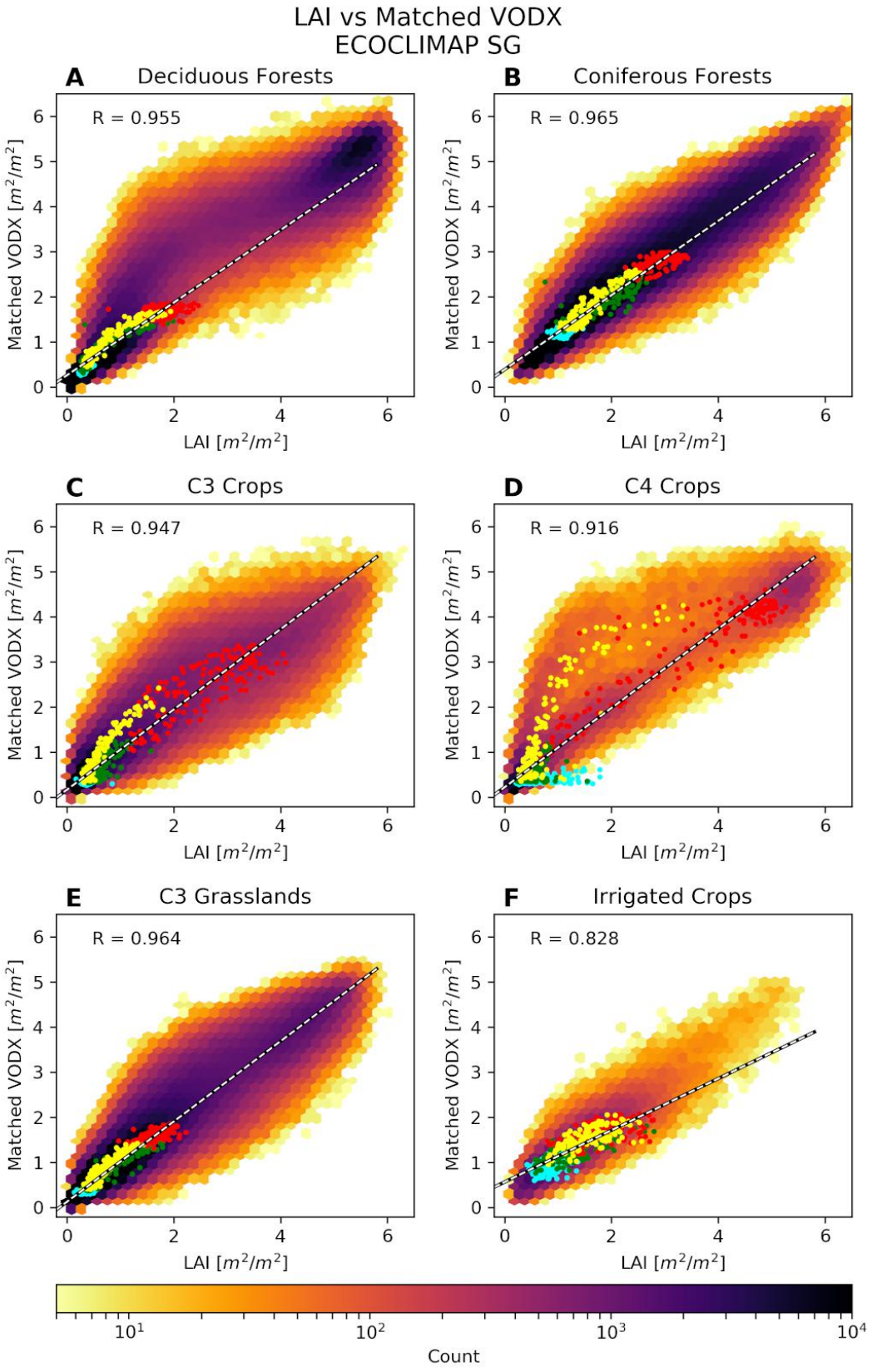

Figure 4.7: Same as Figure 4.6, but using ECOCLIMAP-SG instead of ECOCLIMAP-II.

4.2 Impact of Assimilating VOD as an LAI proxy

This section of Chapter 4 analyzes the impact of using and assimilating VODX as an LAI proxy in the LDAS-Monde system. A total of 10 experimental assimilation setups were run over the CONUS domain from 2003-2018. 2003 was chosen as the start date, as this is approximately when the TRRM mission ceases to be the only functioning VOD dataset included in VODCA. Because of the limited geographic extent of this mission (only up to 35°N), the analysis would be skewed. Table 4.6 provides the experiment names used throughout to reference the assimilation setups, and briefly describes what data is assimilated for each one. Besides the OL, SEKF LAI, SEKF VODX, SEKF SSM, SEKF LAI SSM, and SEKF VODX SSM, several experiments were run at the same time, but using modified observations of VODX. SEKF VODX10 and the joint assimilation with SSM (SEKF VODX10 SSM), uses VODX observations from VODCA as before, but has filtered those observations to coincide only where and when LAI observations from CGLS exist. This is used to test whether the changes produced between SEKF LAI and SEKF VODX are truly from the more frequent assimilation, or from the quantifiable differences between matched VODX and LAI. If the SEKF VODX10 results are closely resembling SEKF LAI results, but SEKF VODX are far different, this indicates the frequency of assimilated observations is the primary cause of those differences.

Additional experiments of SEKF VODX_Int and SEKF VODX_Int SSM slightly change the processing of linear re-scaling the VODX by linearly interpolating the CGLS LAI observations from their native temporal frequency of every 10 days at best, to daily values. The VODX is then linearly re-scaled as before, but to these values provided every day. It is important to note that the SEKF VODX_Int and SEKF VODX_Int SSM experiments were ran to make sure that the LAI observations approximately every 10 days were sufficient for linear re-scaling. The results of these experiments are shown, but the focus stays with the other experiments.

This section specifically analyzes the assimilation of the various matched VOD data (VODX, VODX10 and VODX_Int) in place of LAI, as well as some assessment of SEKF SSM by itself. The following section then looks at the impact of jointly assimilating SSM with LAI and the matched VOD proxies of LAI.

This section will go through the results of these experiments over CONUS. Appendix D details this same analysis over the subdomains shown in Chapter 2 and Figure 2.10. As described in Chapter 2, the primary variables of interest are LAI, GPP, ET, and SSM. GPP, ET, and SSM (depending on the experiments analyzed) are truly independent datasets to compare to, and the main conclusions are drawn from these variables. Comparisons to LAI observations are still presented, but as the LAI was used in the assimilation itself (except for SEKF SSM), or to re-scale the VOD, it acts more as a benchmark for the assimilation.

The primary focus of this analysis is done using the statistical score of correlation coefficient. The LDAS-Monde workflow includes two ways of calculating the correlation over a domain. First is what is termed the "Per Point" calculation, which is a simple and intuitive average of correlations over the gridcells of interest. Each gridcell's correlation to the observations through the time period is separately calculated, and the arithmetic mean is then taken. While this score is the most commonly used, and is easy to calculate and visualize, there are concerns that an averaged correlation looses strict meaning, primarily when the distribution of the scores are not Gaussian. Critical information about the variability and range of the scores is lost. Depending on time scale, the significance of this score can also be weak if there are fewer observations over certain points.

The second option is termed the "For All Points" calculation, which combines all points in the domain into a single, long, time series, which is then computed against the observations processed in the same way. This provides only one correlation score over the domain. However, the significance of the score is strengthened due to the far larger sample length.

This analysis uses both these scores in combination against the satellite derived observations of LAI, GPP, ET, and SSM. For each domain, correlations by month are provided using the "For All Points" calculation. Then probability distribution functions (PDF), derived from the histogram of individual gridcell correlations, are calculated using a Gaussian kernel density estimation, and use the "Scott's Rule" to calculate an appropriate smoothing bandwidth, as justified in Chapter 2. This dual calculation approach allows us to both describe seasonal variations of correlation for each experiment, while also showing the distribution of scores throughout the domain.

Table 4.6: List of experiment names and their assimilated observations analyzed in Chapter 4

#	**Experiment Name**	**Assimilated Observations**
1	Open Loop (OL)	No Assimilation
2	SEKF LAI	CGLS LAI
3	SEKF VODX	VODCA Matched VODX
4	SEKF SSM	ESA CCI SSM
5	SEKF VODX10	VODCA Matched VODX only when there is an associated LAI observation (every 10 days)
6	SEKF VODX_Int	VODCA Matched VODX linearly rescaled using LAI observations that were interpolated to daily data
7	SEKF LAI SSM	Joint CGLS LAI + ESA CCI SSM
8	SEKF VODX SSM	Joint VODCA Matched VODX + ESA CCI SSM
9	SEKF VODX10 SSM	Joint VODCA Matched VODX10 + ESA CCI SSM
10	SEKF VODX_Int SSM	Joint VODCA Matched VODX_Int + ESA CCI SSM

4.2.1 Analysis over CONUS using Satellite-Derived Observations

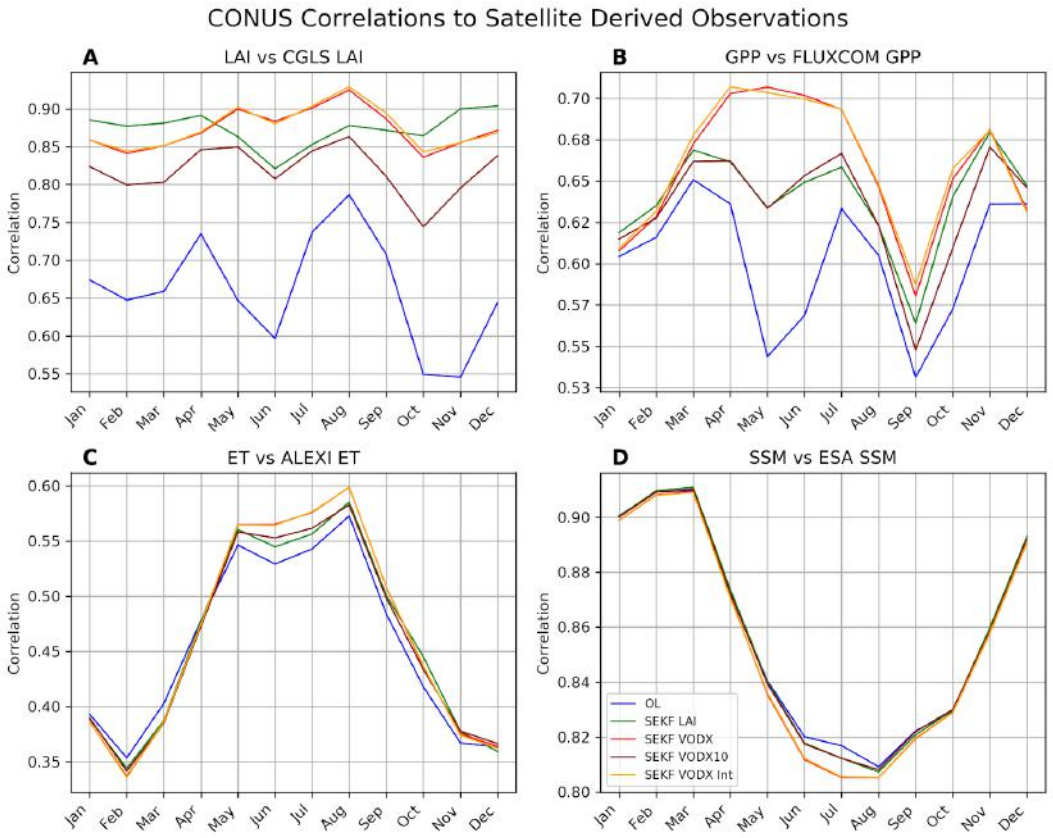

Figure 4.8: Graphs of monthly correlations over CONUS between LDAS-Monde OL (blue), SEKF LAI (green), SEKF VODX (red), SEKF VODX10 (maroon), and SEKF VODX Int (yellow) and satellite derived observations of A) LAI, B) GPP, C) ET, and D) SSM.

Over CONUS, the experiments of OL, SEKF LAI, SEKF VODX, SEKF VODX, and SEKF VODX_Int are analyzed. Figure 4.8 gives the monthly correlation scores throughout the year of the experiments compared to satellite derived observations of A) LAI, B) GPP, C) ET, and D) SSM.

LAI: First looking at LAI, the whole CONUS domain sees added value during the months of May through September when assimilating matched VODX in place of LAI. The rest of the year, the scores for SEKF VODX are slightly below that of SEKF LAI. SEKF VODX_Int matches VODX almost perfectly through the year, and this trend is seen in all the other variables as well. The improvement in LAI correlation from assimilating VODX comes as a slight surprise, as this is comparing to the CGLS LAI observations that themselves were assimilated in SEKF LAI. Some potential explanations include the far

more frequent assimilation of VOD during the summer months when LAI is most rapidly changing. The results of SEKF VODX10 also show definitively that it is the more frequent observations, and not the differences between LAI and matched VOD that are causing some improvement, as SEKF VODX10 is consistently lower than both SEKF LAI and SEKF VODX. Another reason this improvement of SEKF VODX is seen compared to SEKF LAI, is possibly related to the "For All Points" calculation of correlation. Additionally, this panel shows that any assimilation of VODX or LAI significantly improves correlations compared to the model by itself (OL).

GPP: For the variable of GPP, some similar trends are seen. During the months of March through July, the assimilation of VODX (or VODX_Int) performs far better than SEKF LAI or SEKF VODX10. From July to October, there is also some improvement, but not as strongly as in the Spring and early Summer. Interestingly, for the OL, SEKF LAI, and SEKF VODX10, there is a visible dip in correlation scores during the month of May, while the SEKF VODX and the interpolated counterpart see near constant, or even slightly higher correlations compared to previous and future months. I hypothesize that May sees some of the fastest vegetation change of the year for CONUS, and all the model, and even all the observations assimilated at best every 10 days do not provide sufficient constraint to the vegetation. The near daily VODX products do provide that, and thus prove immediately their utility in use as LAI proxies for data assimilation. The changes in correlations between experiments and GPP observations are not as drastic as seen in LAI, but they do show the same overall trends. And importantly, this GPP observation is an independent evaluation of vegetation health and production, where the large improvements from VOD assimilation are observed in the Spring and Summer months, when droughts and heatwaves are most likely to damage agricultural production.

ET: The ET variable is a slightly more tricky one to unpack, as the correlation scores for all the experiments seen here are relatively close. The only easily distinguishable differences arise in the months of May to August. During these months, like with LAI and GPP, SEKF VODX and SEKF VODX_Int are nearly indistinguishable and have the highest correlations. They are followed by SEKF LAI and SEKF VODX10, which are close to one another, then finally followed by the OL. In general, these correlation scores are lower than for LAI or GPP, but also are at their peak during much of the Summer, when evaporative demand is highest, and when it is critical for agricultural production to account for hot and dry conditions.

SSM: With SSM, as with ET, correlations between our experiments are nearly indistinguishable except for some months in the Summer, namely June through August in this case. Overall correlations are very high, consistently higher than 0.80, and contrary to all the other variables, provide the best correlations in Winter months. During these months of differences in the Summer, we

can see that the OL actually performs the best, followed by SEKF LAI and SEKF VODX10, and finally with the lowest scores given to SEKF VODX and SEKF VODX_Int. While in absolute terms these differences are small, a logical explanation can be had that supports these rankings: any data assimilation in LDAS-Monde, whether it is of vegetation such as LAI or VOD, or SSM, directly changes the 8 control variables. Seven of these variables are soil moisture, with 6 of them deeper than the 5cm WG3 layer used to compare against these ESA SSM observations. The assimilation, of LAI or VODX in this case, impacts all these layers and can adjust the uppermost layer used here to coincide with higher LAI values. In these experiments, only the vegetation variable is assimilated, and thus there is no secondary compensation at the upper soil levels done by assimilating SSM observations. We will see more evidence in later sections to support this idea.

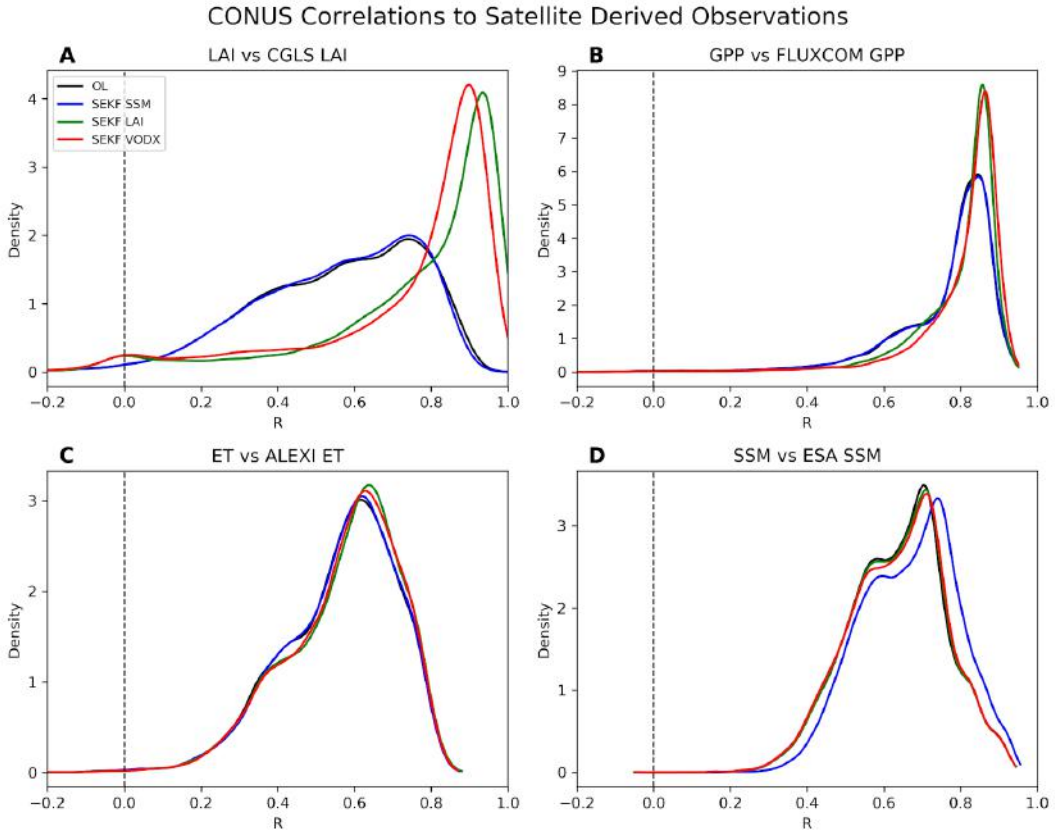

Figure 4.9: Graphs of Probability Distribution Function (PDF) correlation distributions over CONUS between LDAS-Monde OL (black), SEKF SSM (blue), SEKF LAI (green), SEKF VODX (red) and satellite derived observations of A) LAI, B) GPP, C) ET, and D) SSM. The PDFs were calculated using a Gaussian kernel density estimation of the scores.

The PDFs (derived from the histogram of correlation scores) are given in

Figure 4.9 for CONUS. In order to better view certain differences, these histograms analyze the experiments of OL, SEKF SSM, SEKF LAI, and SEKF VODX. The x axes are fixed to the same range for all the variables.

LAI: LAI provides a very clear indication that the assimilation of vegetation variables, whether LAI or VODX heavily shifts the distribution of correlations higher. For CONUS, the assimilation of SSM changes the distribution very little compared the OL, while both SEKF LAI and SEKF VODX share a lower amount of points at the low and moderate correlations, while beginning a strong upward trend at 0.7 and higher. It is also of note, that there is a small "bump" in correlations for SEKF LAI and SEKF VODX at and around 0.0, not seen in the OL or SEKF SSM. This is seen in more detail in the California domain (Appendix D, and is likely related to land use and the amount of bare ground. From correlations of 0.1 to 0.45, SEKF VODX has slightly more points than SEKF LAI, and this is reversed from 0.45 to 0.75. SEKF VODX then quickly spikes at a correlation value of 0.88, with SEKF LAI shifted higher, with a peak closer to 0.9. These strongly changed values from the OL and similarities between SEKF LAI and SEKF VODX demonstrate even more that VODX can properly act as an LAI proxy.

GPP: The distribution of GPP correlations are not as widespread as LAI, however, a clear pattern still emerges. Starting at 0.4, the SEKF LAI and SEKF VODX have less gridcells than the OL, which lasts until around 0.8. It is around this point that the shift towards more higher correlation points with SEKF LAI and SEKF VODX is strongly apparent. While similar, SEKF VODX does slightly outperform SEKF LAI in this case as well, having more higher correlation values. OL and SEKF SSM are indistinguishable for the entire range of correlations.

ET: Like with the monthly correlations presented before, this distribution of ET scores are very similar between all the experiments. At around 0.4, there is a noticeable difference where SEKF VODX and SEKF LAI begin containing less points, which is then made up with those experiments having more higher values consistently between 0.55 and 0.7. At their peak densities, SEKF LAI slightly outperforms SEKF VODX, but SEKF VODX still improves over the OL and SEKF SSM, which, again, are nearly indistinguishable.

SSM: For all experiments in the figure, the distribution of SSM correlations cores are almost bimodal. There is one peak at 0.55 and another between 0.7 and 0.8. In this case, the SEKF SSM, which has assimilated ESA CCI SSM observations, is not compared to an independent dataset, but exactly that which was assimilated. It is then comforting to see that the assimilation increases the score distributions across the board, outperforming all other scenarios. Other SSM datasets would be needed for an independent evaluation, and may be performed in follow up studies. The other experiments of SEKF LAI and SEKF VODX are only slightly changed, with both edging out better performance than

the OL. SEKF VODX also edges out a better performance than SEKF LAI, but just barely. This is due to a lower amount of scores in the first peak, while consistently having very slightly more scores from 0.7 and higher.

4.2.2 USCRN Soil Moisture

The comparison of LAI, GPP, ET, and even SSM against satellite derived observations serves an important purpose, as those observations are spatially continuous. However, errors in the sensors or processing of the data still exist, and relatively large spatial resolutions mean losses of more localized information. With this in mind, I compare all of the experiments listed in Table 4.6 to soil moisture observations from the United States Climate Reference Network (USCRN) in situ soil moisture monitoring stations.

The comparison is done at four depths, which are matched to ISBA soil layers, 5cm (WG3), 20cm (WG_20), 50cm (WG6), and 100cm (WG8). The USCRN observations are point measures at each of those depths. This comparison uses the ISBA soil layers to directly compare against these point measures, but it is important to keep in mind that WG3, WG_20, WG6, and WG8 are just that, layers of soil. WG3 is from 5cm to 10cm, WG_20 is a weighted average of WG4 and WG5 (as performed in Mucia et al., 2020) representing 10cm to 40cm, WG6 is a layer from 40cm to 60cm, and WG8 is the 80cm to 100cm layer. This comparison between points and layers will certainly result in some error, but the overall scores and trends seen will still be valid.

As discussed in Chapter 2 and Mucia et al., 2020, this comparison uses USCRN data between the years of 2011 and 2018. While the network was operational as early as 2005, 2011 was selected as the start of the comparison in order to maximize the number of stations, and homogenize the results of comparisons between stations. The station observations are processed and filtered, with the most notable filters including the removal of any data with corresponding soil temperature observations at or below 4°C. Stations with less than 100 days of observations are also removed, as the scores proved too variable. Finally, only correlation scores with associated p-values less than or equal to 0.05 are retained ($\alpha = 95\%$).

Table 4.7 provides the mean calculated correlations for each of the experiments and at each of the depths. As previously seen in (Mucia et al., 2020), correlations strongly drop as the depths become lower. With 2 significant figures, the correlations at 5cm are all identical at 0.75, except for SEKF VODX_Int SSM, which is at 0.76. There is very slightly more variability at lower depths with 20cm scores ranging from 0.68 for the OL, to 0.70 for all the experiments jointly assimilating vegetation and soil moisture observations. Similar variations are present for the 50 and 100cm depths. Notably, the 100cm scores are all the same at 0.48 except for the OL and SEKF SSM experiments with 0.46.

Statistical bootstrapping (as described in Chapter 2) was performed on all the calculated values, before rounding significant figures, to calculate the upper and lower bounds of the 95% confidence intervals (CIs). The resulting CIs showed that at every depth, all the experiment's mean correlations were within

every CI, meaning no experiment could be said to be statistically different than any other at the same depth.

While the differences may not be statistically significant, the small trends seen between the mean correlations of different assimilation scenarios still do point to some changes. Figure 4.10 displays in graph form the figures from Table 4.7, limiting the number of experiments shown. Figure 4.11 is the same, but including all experiments. The experiments displayed in the figures are sorted from lowest correlation at 100cm on the left to highest correlation at 100cm on the right. These figures demonstrate that the differences in mean correlations between experiments seem to be amplified at depths. Very little changes are seen at 5 and 20cm, but at 50 and especially at 100cm, there is a clear improvement from the OL and SEKF SSM to assimilation using vegetation and joint vegetation and soil moisture. This result follows the logic that the assimilation of vegetation, with LAI or VODX and its similarly processed counterparts, more strongly influences deeper soil moisture layers, and, in general, improves the correlation to these in situ observations. And again, while not statistically significant, the assimilation of VOD and joint VOD and SSM assimilation provides better scores at the 100cm level.

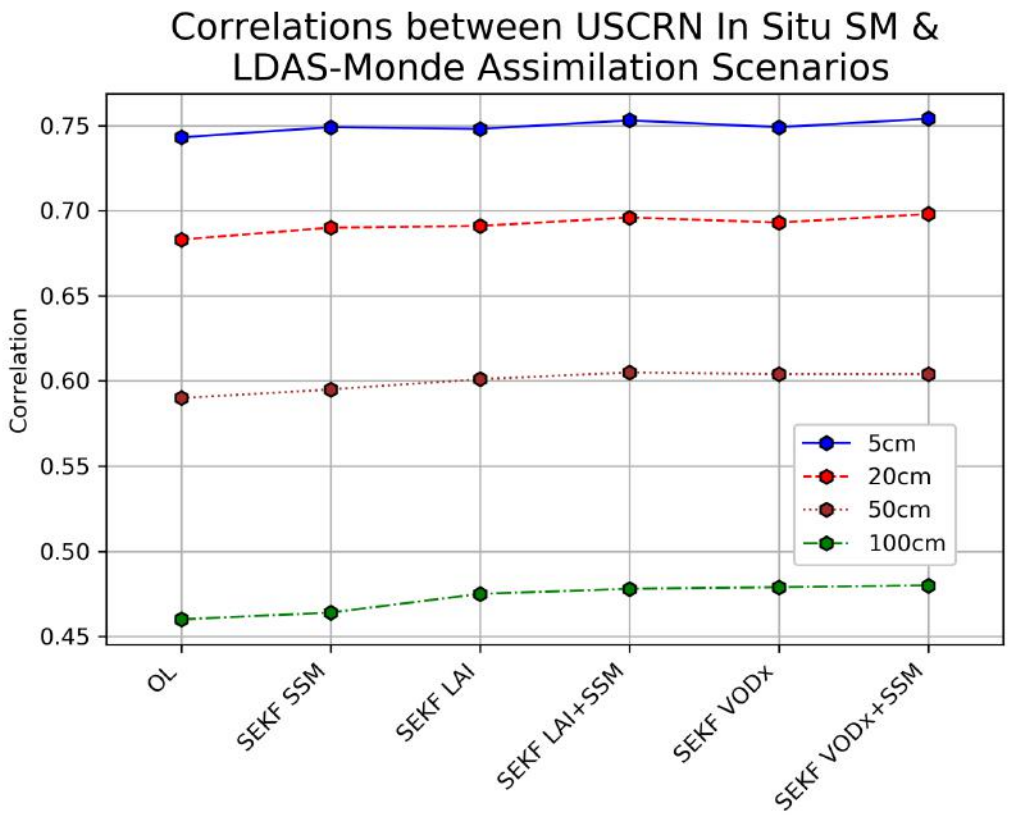

Figure 4.10: Average correlation scores between USCRN stations and a limited selection of LDAS-Monde Experiments at 5 (blue, solid), 20 (red, dashed), 50 (maroon, dotted), and 100cm (green, dash/dot).

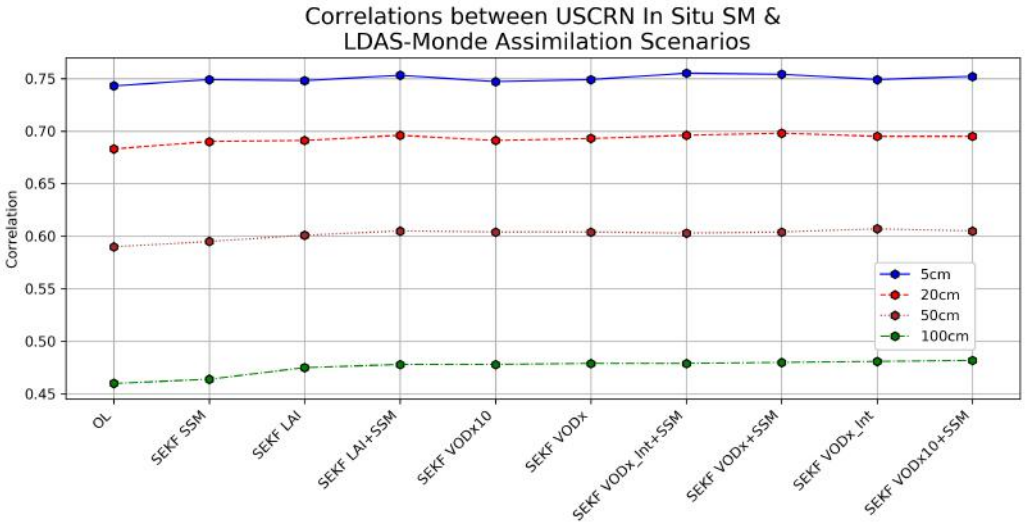

Figure 4.11: Average correlation scores between USCRN stations and all LDAS-Monde experiments at 5 (blue, solid), 20 (red, dashed), 50 (maroon, dotted), and 100cm (green, dash/dot).

Table 4.7: Average correlations scores between USCRN in situ soil moisture observations and LDAS-Monde soil moisture at 5, 20, 50, and 100cm depths. Bolded values indicate the highest score at each depth. *WG_20 is a weighted average of WG4 and WG5 in order to directly compare to 20cm observations from USCRN.

Experiment	WG3 (5cm) (n=110)	WG_20* (20cm) (n=87)	WG6 (50cm) (n=85)	WG8 (100cm) (n=84)
OL	0.75	0.68	0.59	0.46
SEKF SSM	0.75	0.69	0.60	0.46
SEKF LAI	0.75	0.69	0.60	**0.48**
SEKF VODX	0.75	0.69	0.60	**0.48**
SEKF VODX10	0.75	0.69	0.60	**0.48**
SEKF VODX_Int	0.75	**0.70**	**0.61**	**0.48**
SEKF LAI SSM	0.75	**0.70**	**0.61**	**0.48**
SEKF VODX SSM	0.75	**0.70**	0.60	**0.48**
SEKF VODX10 SSM	0.75	**0.70**	**0.61**	**0.48**
SEKF VODX_Int SSM	**0.76**	**0.70**	0.60	**0.48**

Seen throughout these results is the changing number of stations (n) used in the analyses. 5cm comparisons have 110 stations, 20cm uses 87 stations, 50cm uses 85 stations, and 100cm uses 84 stations. This is simply due to the fact that the USCRN network cannot install soil moisture or soil temperature probes in hard or rocky ground layers, as stated in the USCRN soil climate observations documentation (*USCRN Soil Climate Observations Documentation*). In all cases, the 5cm and 10cm probes are installed, but deeper layers depend on the regolith type.

As previously discussed with the "Per Point" and "For All Point" correlation calculation, the averaging of correlation scores does not have much meaning

and looses information. To better analyze differences on a more individual scale, the normalized information contribution of the correlations (NICR) were calculated for each experiment and each depth in comparison to the OL. These NICR values tell us by how much the assimilation experiments improved or degraded scores in respect to the OL. Table 4.8 displays each experiment and depth, and the number of stations that were degraded (red), neutral (black), and improved (green). This approach avoids averaging scores, while still providing a performance overview of the whole domain.

In a similar manner, Figure 4.12 displays the PDF of the distribution of differences in correlation for each of the four depths compared to the OL, and looks at the responses of SEKF SSM, SEKF LAI, SEKF VODX, SEKF LAI SSM, and SEKF VODX SSM experiments.

In both the NICR table and PDFs, the 5cm changes are the strongest seen, with SEKF VODX SSM, SEKF VODX_Int SSM providing the highest number of improved stations, while having some of the lowest number of degraded stations. SEKF LAI SSM and SEKF VODX10 SSM perform similarly, but with a reduction in improved stations. Assimilating just LAI, VODX, VODX10, or VODX_Int consistently under-perform their matching joint assimilation experiments by increasing the number of degraded stations, while having fewer improved stations. And finally, SEKF SSM, while showing the fewest degraded stations, also has the fewest improved. As depths get lower, the numbers become closer. It is generally still seen that the joint assimilation of vegetation and soil moisture improve more stations than the individual assimilation, and the number of stations degraded stays similar. These trends go to show that the joint assimilation has distinct added value in soil moisture monitoring, and will be discussed in more detail in the next section.

Table 4.8: Number of degraded (red), neutral (black), and improved (green) USCRN stations after assimilation using NIC R between OL and various LDAS-Monde experiments at 5, 20, 50, and 100cm depths. Stations are considered improved if the NICR is greater than 3, degraded if the score is less than 3, and neutral if it is between -3 and 3. *WG_20 is a weighted average of WG4 and WG5 in order to directly compare to 20cm observations from USCRN.

Experiment	WG3 (5cm) (n=110)	WG_20* (20cm) (n=87)	WG6 (50cm) (n=85)	WG8 (100cm) (n=84)
SEKF SSM	3/79/28	4/59/24	8/59/18	15/52/17
SEKF LAI	10/69/31	10/51/26	8/49/28	14/43/27
SEKF VODX	13/55/42	10/40/37	14/36/35	17/35/32
SEKF VODX10	9/68/33	7/53/27	9/52/24	13/44/27
SEKF VODX_Int	15/54/41	12/38/37	14/37/34	16/36/32
SEKF LAI SSM	7/57/46	6/41/40	12/41/32	13/36/35
SEKF VODX SSM	8/45/57	10/34/43	15/36/34	17/35/32
SEKF VODX10 SSM	7/56/47	7/44/36	13/38/34	17/37/30
SEKF VODX_Int SSM	8/48/54	10/35/42	14/35/36	18/31/35

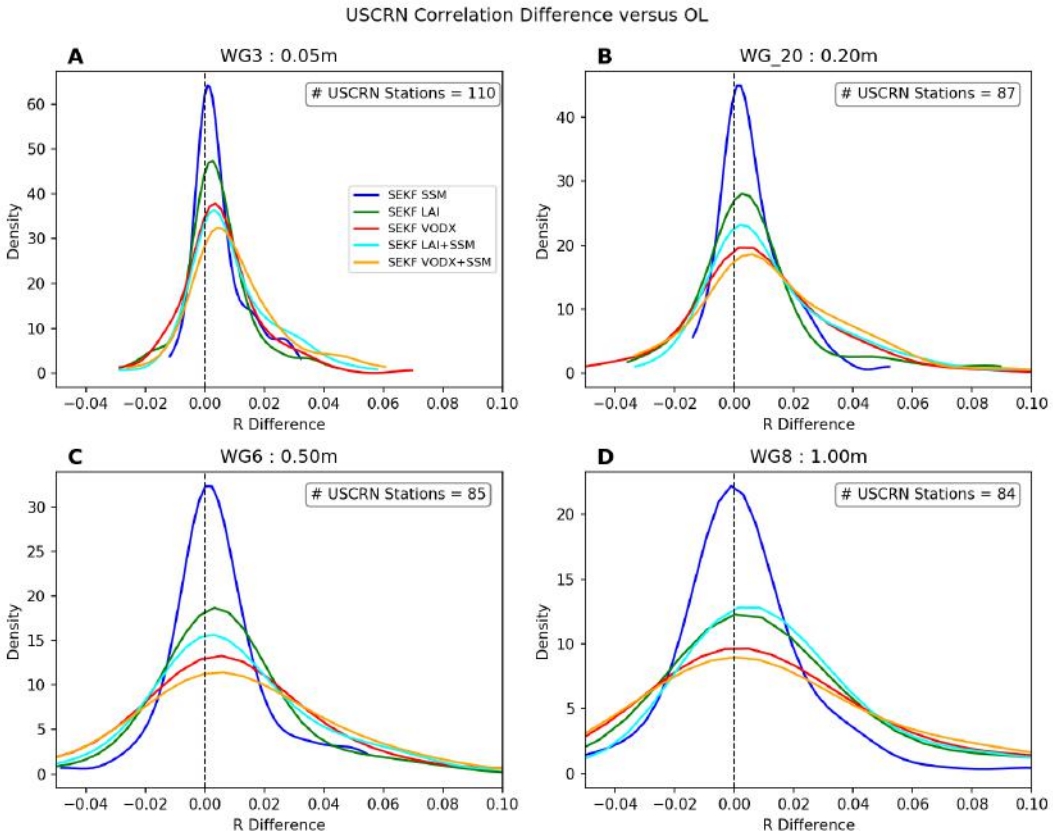

Figure 4.12: Probability Distribution Functions of the Distribution of Correlation Differences between OL and SEKF SSM (blue), SEKF LAI (green), SEKF VODX (red), SEKF LAI SSM (cyan), and SEKF VODX SSM (orange) for USCRN at A) WG3 (5cm), B) WG_20 (20cm), C) WG6 (50cm), and D) WG8 (100cm).

In order to assess any geographic patterns of the USCRN correlations, maps of the NICR of each station are plotted for the four depths looking at the SEKF SSM, SEKF LAI SSM, and SEKF VODX SSM experiments. Figure 4.13 displays WG3 maps in Panels A-C and WG_20 maps in Panels D-F. Figure 4.14 provides WG6 maps in Panels A-C and WG8 maps in Panels D-F. It is clear from these figures that stations are strongly improved by using either LAI SSM or VODX SSM assimilation, compared to just SSM. Additionally, a geographic pattern does emerge, primarily at 5 and 20cm. This pattern is that much of the Great Plains and Midwest show strong improvement due to the joint assimilation, while the south and east coasts show little change. Stations in the western US are more sporadic, with improved, degraded, or neutral stations spread throughout. This western region is, however, where most of the stations experiencing degradation are located. At 50 and 100cm, the SEKF SSM show minimal concrete changes compared to the OL, with near-even numbers of improved and degraded sites. The joint assimilation improves upon these values, and while the degraded stations are typically still degraded in SEKF LAI SSM and SEKF VODX SSM, more stations move from neutral to improved.

Overall, the use of the USCRN in situ observations strongly agrees with the hypothesis that the assimilation of vegetation has stronger effects on soil moisture than the assimilation of SSM. It also goes to show that the assimilation of VODX is on par or even an improvement from the assimilation of LAI.

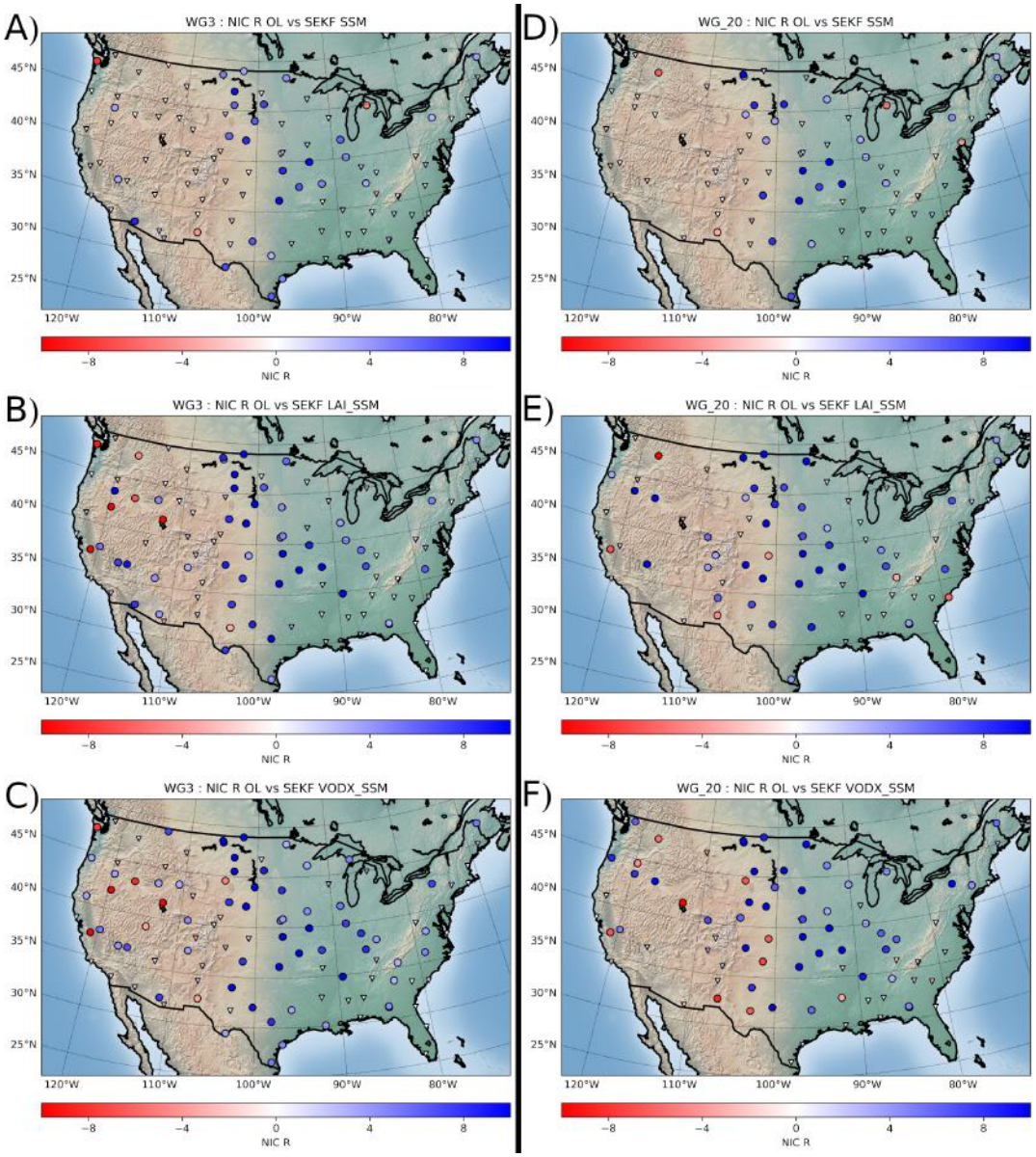

Figure 4.13: Maps of Normalized Information Contribution (NIC) correlation for A-C) WG3 (5cm depth) and D-F) WG_20 (20cm) between the OL and SEKF SSM, SEKF LAI SSM, and SEKF VODX SSM. Circles represent a change greater than 3%, while triangles indicate changes less than 3%. Blue indicates correlation improvement from the assimilation with respect to the OL, while red indicates degradation.

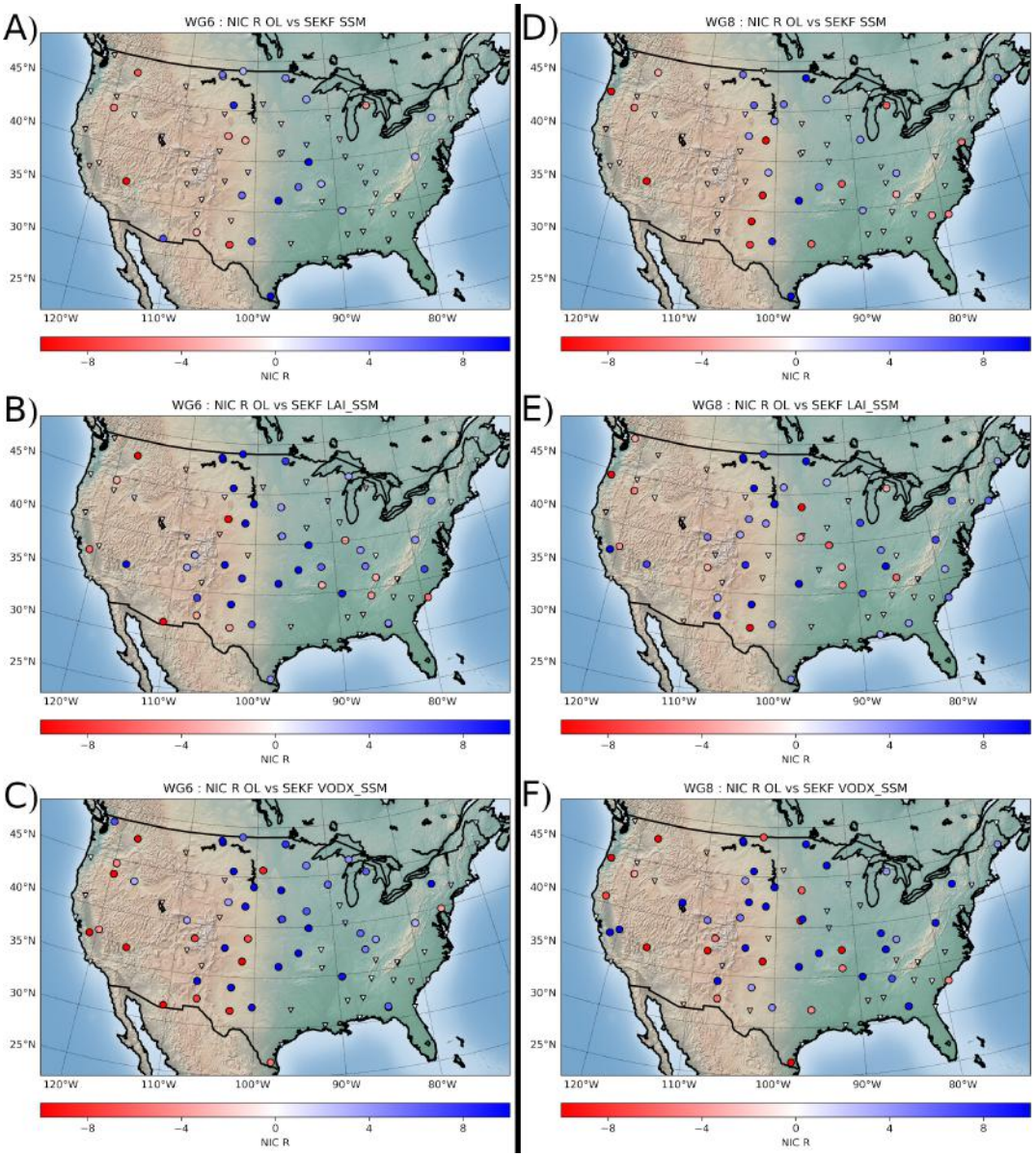

Figure 4.14: Same as Figure 4.13, but for WG6 (50cm) and WG8 (100cm).

4.3 Impact of Jointly Assimilating Vegetation Variables and SSM

While previously discussed above when assessing correlations and NICR for the USCRN, this section will go into more detail regarding the effects of jointly assimilating variables of vegetation (LAI or Matched VODX) and SSM in LDAS-Monde. We have already seen that the joint assimilation provides a noticeable increase in improved USCRN stations relative to the OL, over the single assimilation of vegetation variables or SSM. This section analyzes the joint assimilation of observations over CONUS, while Appendix D details this same analysis over the subdomains shown in Chapter 2 and Figure 2.10.

Figure 4.15 presents the same type of figure as previously, looking at the monthly scores of four LSVs of interest over CONUS, but including the joint assimilation experiments of SEKF LAI SSM and SEKF VODX SSM. The main concern in this figure and the following, is determining the improvement, if any, between the dashed (single assimilation) and solid (joint assimilation) red and green lines. As seen in Panel A, there is no discernible difference between the single and joint assimilation for LAI. But as Panel B and C shows, jointly assimilating vegetation and SSM produces slightly improved monthly correlations over the whole CONUS domain for GPP and ET respectively. These improvements are primarily seen in the months of June through August, and are quite small. Regarding the variable of SSM, the joint assimilation strongly improves the monthly correlations from LAI to LAI SSM and from VODX to VODX SSM. To reiterate, because the SSM observations assimilated are the same used to compare in Panel D, it is merely an indication that the data assimilation is truly shifting model soil moisture values closer to that of observations.

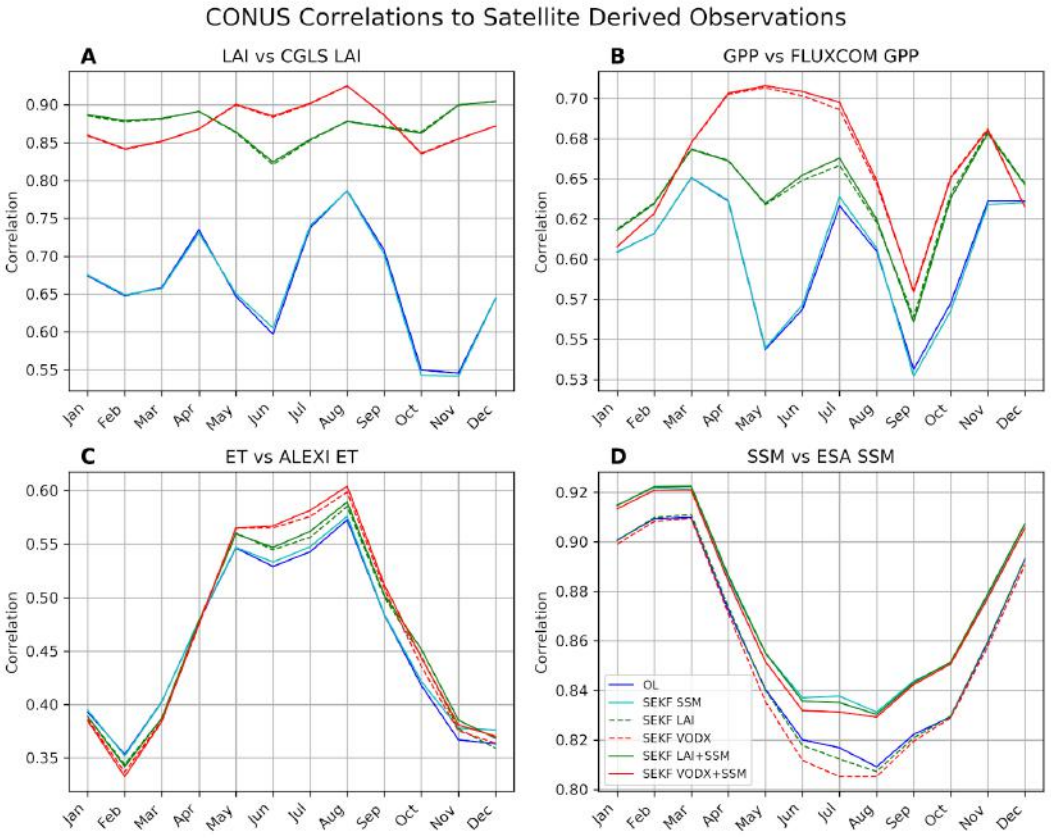

Figure 4.15: Graphs of monthly correlations over CONUS between LDAS-Monde OL (blue), SEKF SSM (cyan), SEKF LAI (green, dashed), SEKF VODX (red, dashed), SEKF LAI SSM (green, solid), and SEKF VODX SSM (red, solid) and satellite derived observations of A) LAI, B) GPP, C) ET, and D) SSM.

4.4 Summary of Chapter 4

This chapter explores the relationship between LAI and VOD, both X and C-Band over CONUS, and shows agreement with previous research that these two variables do share useful vegetation-related information. A deeper dive into this relationship is performed, focusing on the differences over six vegetation types. The linear re-scaling of VODX was described, and analysis was done regarding the responses of LAI versus VODX and LAI versus matched VODX over the two versions of the land use database, ECOCLIMAP-II and ECOCLIMAP-SG. A brief comparison was also performed regarding the VODX and VODC anomalies compared to annual corn yield and observed LAI anomalies. Then the last sections deal with the analysis resulting from 10 different LDAS-Monde experiments focused on the assimilation of matched VODX as an LAI proxy. Included in these experiments were those designed to test the differences between single assimilation of vegetation (LAI or matched VOD) or SSM, and the joint vegetation-SSM assimilation.

This is one of the very few attempts at assimilating VOD. To the best of our knowledge, only Kumar et al. (2020) have, to date, developed a similar methodology. Unlike Kumar et al. (2020), in our study, assimilating VOD has a direct impact, not only on LAI, but also on the root zone soil moisture analysis (as SM layers 2 to 8 are also control variables). In addition to the successful results linked to a stronger constraint on vegetation within LDAS-Monde, this pioneering work shed light on additional research needed on (1) how to best re-scale VOD to LAI, (2) how to best set up observational errors to the re-scaled VOD in the data assimilation system, (3) how to deal with areas where a weaker relationship between VOD and LAI was obtained.

The major individual conclusions are as follows:

- Analysis of VOD versus LAI
  - Both X and C-band VOD observations from VODCA contain a significant amount of information related to vegetation.
  - VOD is not LAI, even while direct correlations between LAI and VOD observations were positive and moderately strong.
  - C-band VOD was shown to contain far more variation and noise when compared to LAI, and resulted in overall weaker correlations than X-band VOD. Therefore, X-band VOD from VODCA was selected to be assimilated in the LDAS-Monde experiments.
  - The overall tendency between annual VOD, LAI, and corn yield anomalies are well matched, indicating the utility of VOD in vegetation monitoring.
- Role of Dominant Vegetation
  - Dominant vegetation plays a strong role in the relationship between LAI and X-band VOD. Coniferous forests consistently had the weakest matching, while C3 and C4 crops typically performed the best.

  - Seasonality also dramatically changed the LAI-VODX relationship, with Winter scores typically the lowest, and Summer and Autumn correlations typically the strongest.
  - Matched VODX was far more strongly correlated to LAI observations than non-matched VODX, but still exhibited significant variation.
  - While both ECOCLIMAP versions showed similar results, some of the differences point to mischaracterization of land surfaces, specifically irrigated crops.
  - A hysteresis pattern was observed in both LAI versus matched and non-matched VODX comparisons, and in both ECOCLIMAP versions. The VOD values in Autumn specifically tend to indicate a possibility that harvest produced ground litter increases VOD responses. This also may indicate the linear re-scaling is not sufficient to correct for these differences.

- Assimilating VOD as an LAI Proxy

  - Assimilating VODX in place of LAI generally improved month to month correlations of GPP and ET. The distribution of GPP correlations were also strongly and consistently shifted to higher values.
  - The improvements seen in GPP and ET correlations by assimilating VODX in place of LAI are almost entirely due to the more frequent observations of VOD. This is shown because the experiment SEKF VODX10, which assimilated matched VODX at the same frequency as LAI observations performs considerably worse than the natural VODX observation frequency.
  - Correlations to the USCRN in situ observations strongly support the assertion that matched VODX can act as an assimilation replacement for VOD, with scores staying near, or even improving upon those of assimilated LAI.
  - The counting of degraded, neutral or improved NICR for USCRN stations indicate the most definitive effect of data assimilation is seen at the upper layers of 5 and 20cm, even when the averaged scores show no significant differences.
  - The spatial distribution of the USCRN NICR scores demonstrate the Great Plains and Midwest as benefiting the most from assimilation.

- Joint Assimilation of Vegetation and SSM

  - The counting of degraded, neutral or improved NICR for USCRN stations tend to show that the joint assimilation of vegetation and SSM regularly performs better, with fewer degraded and more improved stations at upper soil layers, and simply more improved stations at deeper layers.

- Correlations of LAI show no strong trend of improvement or degradation from single to joint vegetation-SSM assimilation.
- GPP and ET correlation scores exhibit a very small improvement due to the joint assimilation over the broader CONUS domain.
- The comparisons of all the experiments definitively reveal that the assimilation of SSM has a far weaker effect on the model and system.

Chapter 5

Towards Forecasting Impacts of Extreme Events on Land Surface Variables with LDAS-Monde

This chapter describes a prospective experiment as a follow up to work presented in Chapters 3 and 4. The purpose of this experiment is to analyze the effectiveness of the LDAS-Monde system in a forecast configuration to provide drought forecast alerts. The experiment takes lessons learned from the initial forecast experiment over CONUS (Chapter 3; Mucia et al. (2020)) and combines it with assimilation of matched VODX (Chapter 4) to provide better initial conditions. After details of the experiment requirements are laid out, the next section provides details of how the analysis will be done, and how the assessment of the system's drought forecast and alert capability will be conducted.

A proof of concept analysis is performed using the 2003-2018 experiment in Chapter 4 and a 19 year OL baseline. This analysis demonstrates the capability to transform LDAS-Monde LSVs into percentile ranks, that is to say that for each selected time window (week or month) the LSV in question is percentile ranked as a function of the distribution of that LSV in the same window, for every year in the time period. This percentile rank is then subsequently converted into drought categories, as defined by the US Drought Monitor (USDM). And even with the relatively short comparison period, the RZSM and ET LSVs correctly identify the major 2012-2013 Great Plains drought. These results of LDAS-Monde in a monitoring configuration set the stage for the proposed experiment in forecast configuration.

Primary datasets used in the comparison include the USDM over the selected area, as well as other potential drought indices (DI) such as the Vegetation Drought Response Index (VegDRI) or the Normalized Difference Vegetation Index (NDVI). An OL experiment similar to the one used in Chapter 4, but instead from 1970-2020, forced by ERA5, will serve as a baseline of LSVs in order to compute percentiles. Certain LSVs such as ET, RZSM, and LAI are selected to be blended with variable weighting so that during the case study period, times when the blends drop below the Drought Monitor thresholds (Figure 1.1) will be noted and analyzed. The timing of these alerts will be compared to onset of drought determined by the USDM, and the LSV percentiles can then be directly compared to both USDM drought levels (which are defined with percentiles) or other DIs that are percentiles or standardized anomalies.

5.1 Drought Forecast Alert System Case Study

In this section, requirements and parameters of the prospective CONUS drought case study will be presented and discussed. While LDAS-Monde is a global system, ancillary data over CONUS provide the capability to correlate and analyze results against a large amount of continuous drought monitoring datasets, most importantly the United States Drought Monitor, which is not available outside the US (or North America at reduced frequency).

5.1.1 Requirements of the Case Study

Any forecast case study used as a follow up to the work presented in Chapter 3 and Chapter 4 must meet a number of basic requirements in order to truly continue the work and draw meaningful conclusions. As discussed above, CONUS is the primary working domain, and thus, it is required that any drought event must have taken place over the same domain as previously assessed. Additionally, while drought is a widespread event, the resolution of LDAS-Monde must be high enough to still see regional details, and thus 0.5° x 0.5° will be imposed as the upper resolution limit.

Time Frame

The availability of datasets used either as forcing (atmospheric forecasts) or analysis (USDM) are the most immediate limiting factors of the time frame of this experiment. A wider limit is forced by the USDM, with drought monitoring started in the year 2000, and available continuously until present. As described in Chapter 1, the USDM is a weekly synthesis of objective drought indicators and indices made by experts and coordinated with ground truth information and local observers across the country. And being the default tool for drought monitoring in the United States, this data must be used in this future assessment. The USDM provides a near real time (2 day time lag) look at current drought extent and impacts by categorizing into drought levels defined with specific percentile rank ranges. As shown in Figure 1.1, drought levels are defined as: D0 (Abnormally Dry) = 30-21%; D1 (Moderate Drought) = 20-11%; D2 (Severe Drought) = 10-6%; D3 (Extreme Drought) = 5-3%; D4 (Exceptional Drought) = 2-0%.

However, the atmospheric forecasts are a far stronger limiting factor. Table 5.1 provides the basic details of some atmospheric forecast forcings available for use in LDAS-Monde. ECMWF forecasts used in Chapter 3 and Mucia et al. (2020) only provide data from April 2016 onward, and thus limit our case selection to only a few years.

The inclusion of the ECMWF ENS Extended forecast (ECMWF, 2018), which provides 0.4° x 0.4° resolution forecasts from 16 to 46 days in advance, significantly lengthens the potential lead time of any alert system. The ENS Extended forecast has had significant changes since its first operational real-time forecast in October 2004, including the extension from 32 to 46 days maximum lead time in May 2015. The data availability and procurement of this forecast

would ultimately be the determining factor of drought event selection. If even a 32 day surface forecast that is quantitatively consistent was procured for years earlier than 2016, it would significantly help choosing a specific case study. However, with forecast periods this long, the quality of the specific, daily (or 6-hourly in this case) information is constantly challenged.

Additionally, this ENS Extended forecast is run only twice per week, Mondays and Thursdays. As will be discussed in section 2 of this chapter, most drought indices, including the Drought Monitor, are weekly products valid on one day (Tuesdays for the DM), but accounting for the developments earlier in the week. Because the ENS Extended run starts are fixed, it means that only 2/7 of the individual forecast lead times will ever be directly comparable to USDM (for example there will never be a 3-day forecast valid at the same time as the DM). However, as the forecasts are continuous and initiated in the same manor, it is expected that the comparison of every 2/7 forecast days will still give a reasonable approximation of the forecast LDAS's capabilities.

Finally, as the ENS Extended only produces forecasts beginning from a 16-day lead time, the ENS CTRL would be used as the forecast forcing for days 1-15. This transition can present some problems, most notably related to the changing spatial resolution and temporal frequency. Regarding the spatial resolution, the ENS CTRL is produced at 0.2° x 0.2° while ENS Extended is at 0.4° x 0.4°. The most straightforward approach would be to either up-scale the ENS CTRL to 0.4° or to down-scale the ENS Extended to 0.2°. Albergel et al. (2019) has previously tested down-scaling of ERA5 reanalyses to match the higher resolution of HRES, and found the new resolution performed similarly and nearly as well as the first period of these operational forcings. This can justify the use of the higher resolution, down-scaled, forecasts, providing more detailed regional analyses. The differences in temporal frequency (ENS CTRL: 3-hourly up to day six, 6-hourly up to day 15, ENS Extended: 6-hourly through end), can be solved in the same manor as in Mucia et al. (2020), with taking only the six hourly time steps, and linearly interpolating them to 3-hourly. This allows for the matching to the 0900UTC assimilation window in LDAS-Monde, and also avoids variable temporal frequencies as forecasts progress.

Table 5.1: Atmospheric Forecast Options for Case Study

Atmospheric Forecast	Spatial Resolution	Temporal Availability	Forecast Length
HRES	0.1° x 0.1°	April 2016 - Present	1 to 10-Day
ENS CTRL	0.2° x 0.2°	April 2016 - Present	1 to 15-Day
ENS Extended	0.4° x 0.4°	October 2004 - Present	16 to 32 or 16 to 46-Day

Selection of Drought Event

A major step in this prospective case study is the selection of a drought event corresponding to the requirements laid out above.

Over CONUS the most ideal regions for study are over the Great Plains, Midwest, Northeast, and some of the Southeast. LDAS-Monde over these locations demonstrated strong performance in monitoring vegetation and soil moisture conditions, while certain drier regions, most notably the California/Arizona/Nevada border, identified monitoring problems.

While the selection of the atmospheric forecast forcing will primarily drive the drought event selection, the seasonal timing can also be important. Spring through Autumn months typically show the best LDAS-Monde performance in monitoring LSVs. These months are also when some of the most extreme causes and effects of drought emerge, such as heat waves, plant stress, and plant death.

Using the USDM database of archived drought monitor maps, the selection of a case study event can begin by selecting a geographic region, and reviewing a time series of percent of that area in each of the drought categories (D0-D4), as is depicted for the North Central region in Figure 5.1.This method of selection will give a general overview of a drought's severity and extent, which will then be further investigated by reviewing the weekly drought maps for that region.

This North Central region's percent area in drought time series for 2003-2018 clearly shows a significant drought event beginning mid-2012. The percent of this region in every drought level very quickly spikes, with a nearly 20% of this region in D4 drought, a 2% or 1 in 50 year event. This 2012-2013 drought event, which most strongly affected the Great Plains and Midwest United States, is the most extreme, widespread drought event since the beginning of the USDM. This event will also serve as an important sanity check when comparing to LDAS-Monde derived drought indicator time series.

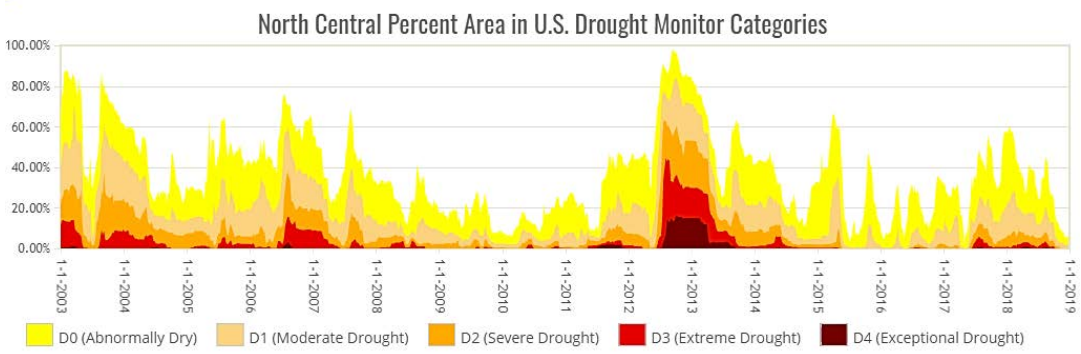

Figure 5.1: Time series of percent area in drought over the North Central region of the US according to the US Drought Monitor for 2003-2018. The y-axis represents the percent of the region in each level of drought (D0-D4). Time series such as these over US regions can help identify and select case study areas and time periods for the prospective experiment in this chapter. This time series was produced via the website of the National Drought Mitigation Center: `https://droughtmonitor.unl.edu/DmData/TimeSeries.aspx`

5.1.2 LDAS-Monde Experimental Parameters

Once selected, the LDAS-Monde parameters can be set and the experiment run. The following subsections will detail the characteristics of the SEKF assimilation analysis, but there will also be a matching OL run along side. The same statistics and LSVs parameters will be analyzed for both the OL and SEKF runs, and thus this OL experiment will be a good tool to use as a baseline alert system. The comparison of the alerts between the OL and SEKF runs will provide more information on the utility and performance of the assimilation in the direct context of drought monitoring and warning.

Improving Initial Conditions with VOD assimilation

As demonstrated in the previous chapter (Chapter 4), the assimilation of matched VODX as an LAI proxy provided improved correlations of GPP and ET over CONUS to satellite derived observations, while also improving aspects of soil moisture monitoring compared to the USCRN. These improvements were caused almost exclusively by the more frequent availability, and thus more frequent assimilation, of the VOD observations compared to LAI observations. The improvements seen in this non-forecast configuration would directly impact the initial conditions of the forecast experiments. And as shown in Chapter 3, the initial conditions have a generally strong and long-lasting impact on the quality of the forecasts. Therefore, this case study will assimilate matched VODX as an LAI proxy.

Joint Assimilation of Vegetation and Soil Moisture Observations

The joint vegetation-soil moisture assimilation demonstrated minor improvements in certain domains concerning ET and GPP monitoring, and more drastically improved soil moisture correlations to the USCRN. Therefore, ESA CCI SSM observations will also be assimilated in LDAS-Monde for this case study.

5.2 Linking LDAS-Monde Forecasts to a Drought Indicator

This experiment follows a similar methodology as Mocko et al. (2021), which reviewed OL and LAI assimilated experiments with Noah-MP LSM in order to improve the representation of drought for agricultural areas. The prospective study described in this chapter will use many of the same analysis techniques and statistics later presented in this section, but is targeted towards the LDAS's potential capacity for forecast and alerts of future drought events. This study also differentiates itself by creating the blend of LSVs for use as a drought index instead of just soil moisture percentiles.

One of the first steps will be converting LDAS-Monde LSVs of interest to a percentile, which can be directly compared to USDM drought levels, as well as

standardized anomaly indices. These LDAS-Monde percentiles are calculated as a percentile of a long term OL experiment over the CONUS domain.

Among the statistics used in the analysis will be a Spearman rank correlation for each gridcell and over the entire time frame. The Spearman rank correlation is used to determine a monotonic relationship between data, which better suits this ranked data. After the immediate post-processing, our LSV data from LDAS-Monde, the USDM, and other DI data are percentiles (or can be converted to percentiles). As such, the value difference between any percentile is entirely dependant on the immediate neighbor percentiles and cannot be assumed to be linear. In other words, the values associated with each percentile are not necessarily distributed evenly.

Additionally, statistics and scores designed specifically for the verification of forecast will be used for the full drought forecast case study. Starting off, calculations of the Probability of Detection "Yes" Events (PODY) and False Alarm Rate (FAR) will be performed for each gridcell. PODY is defined as the number of correctly simulated drought level "hits", divided by the sum total of both hits and misses. In this case, a "hit" is a week when the gridcell correctly characterizes the USDM drought level, while a miss is when the gridcell shows a drought level not matched by the USDM. In the reverse way, FAR is calculated as a ratio of the number of false alarms, meaning a drought level is simulated from LDAS-Monde but the USDM shows no drought, over the sum total of false alarms and hits. These two statistics provide a useful metric in determining if the LDAS-Monde drought levels are over-sensitive or under-sensitive or both, compared to the USDM standard. Together, PODY and FAR used in conjunction allow for a significant analysis of the system's performance.

More scores, such as Brier's score, continuous ranked probability score (CRPS), and Heidke skill score (HSS) can also be calculated in order to provide further in depth analysis, as well as to note this system's scores for comparisons to other drought forecasts as well as to potentially improved LDAS-Monde forecasts. The Brier's score (Brier, 1950) measures the probability of errors associated in forecasts. While a useful quantity in terms of the system's mistakes, a larger sample of extreme events would be required to function as a robust score. The CRPS is another quantitative measure of performance where the cumulative distribution functions (CDFs) of the forecast and of empirical reality are compared, and has a similar decomposition to the Brier score (Hersbach, 2000). Unlike the Brier score, the CRPS allows for an evaluation of more than one aspect of the forecast. In the use with the LDAS-Monde forecast system, this means the CRPS can evaluate the LSV percentiles at drought and non-drought levels. Finally, the Heidke skill score (Heidke, 1926) is a simple score of categorical forecasts in a "proportional correct" form, using hits, misses, false alarms, and correct rejections. This score provides a simple way to encompass the four forecast outcomes, but is relative to random chance and would also require a large sample for comparison. For all of the described metrics, improvements or degradation between the OL and assimilated experiment can also prove useful.

5.2.1 Indicators Used in Drought Alerts

Certain LSVs that are directly output from LDAS-Monde will be taken and converted to percentile ranks for the whole period of the case study. The primary LSVs concerned, as an initial step, will be RZSM, ET, and LAI. LDAS-Monde RZSM will be calculated averaging layers WG2-WG8 (2-100cm) by the weights of their respective layer thicknesses. Once in percentile form, the three LSVs will then be blended using the different proposed weighted average schemes as follows:

$$Blend\,1 = (0.33)LAI + (0.33)RZSM + (0.33)ET \tag{5.1}$$

$$Blend\,2 = (0.50)RZSM + (0.50)ET \tag{5.2}$$

Blend 1 in Equation 5.1 is simply the normal arithmetic mean (with rounded weights) of the three LSVs. Blend 2 in Equation 5.2 is the simple mean of RZSM and ET, in order to isolate potential biases with LAI. These blends are formulated in this way so that additional weights can be developed during the analysis as well in order to arrive at a best match. For example, it may be found that RZSM should be weighted far more than ET in order to best match known events. As discussed and shown below, individual LDAS-Monde LSV percentiles do not perfectly match historic observations and drought categories, and so the given blends will be used simply as an arbitrary starting point for a best initial guess to best match to USDM levels.

The duration of the drought in the case study is also of strong importance. All three of the selected variables are strongly indicative of short to medium length drought events, but do not consider long-term impacts such as hydrological variables (runoff, drainage, groundwater). However, additional blends can be quickly made later during the analysis using the same process with percentiles of alternative hydrologic or vegetation-related variables. Additionally, as these variables in percentile form are normalized, this combination of variables can help represent a wider array of drought conditions including vegetation and water availability, rather than looking at each variable individually.

5.2.2 Proof of Methodology

To test and advance the techniques that will be used in the full case study, the SEKF VODX SSM analysis from Chapter 4, for the Midwest domain, has been processed according to the above description. This experiment was run from 2003 to 2018, strictly in a monitoring configuration, that is to say not producing any land surface forecasts. In this way, this comparison acts as a proof of concept for the methodology, and shows that the processing and analysis can be successfully performed on the data, resulting in similar results to the USDM. Further work will expand this same type of processing and analysis to forecast experiments.

For this initial analysis, the comparison period of the OL for this test is only 19 years (2000-2018), and thus the percentile rankings of the 16 year 2003-2018

SEKF VODX SSM are rather weak. A longer time period for the comparison period, such as the proposed 50 years, would strongly increase the significance, and likely increase the similarities between LDAS-Monde and the USDM. In this initial comparison, the LSVs of LAI, RZSM, and ET were computed and then compared in a percentile rank compared to a slightly longer OL experiment that was run over CONUS from 2000-2018.

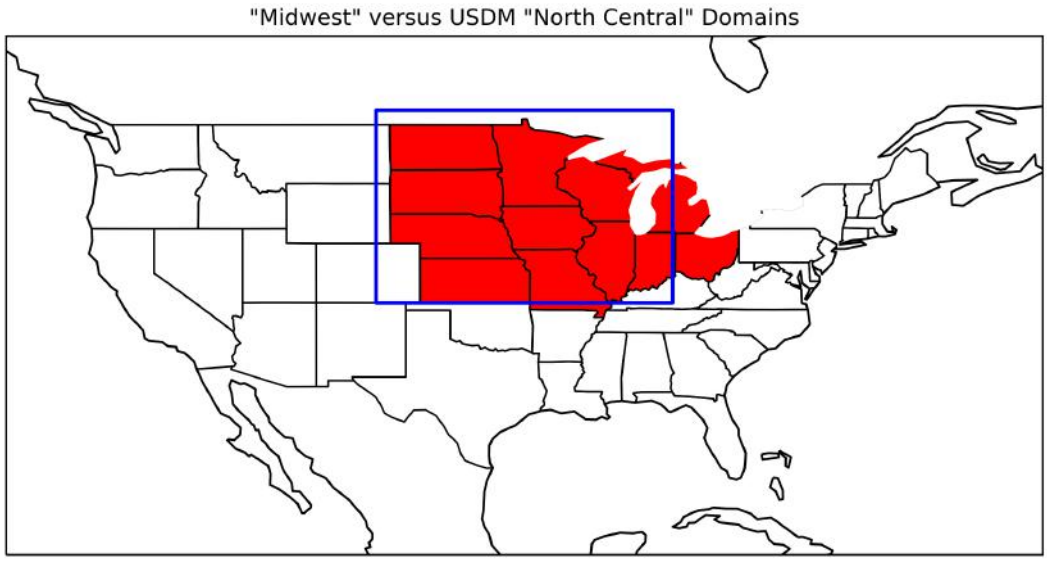

Figure 5.2: Map of the two different domains used in the USDM comparison. The blue box represents the region defined in Figure 2.10 as "Midwest" for LDAS-Monde purposes. The US states shaded in red are a part of the "North Central" region according to the USDM. These are the closest matching pre-defined domains, and thus used in this initial comparison.

Figure 5.2 presents a map of the domains used in this proof of concept analysis. The blue box is the same "Midwest" domain as in Figure 2.10 used in LDAS-Monde, and the USDM "North Central" region is comprised of the red shaded states (North Dakota, South Dakota, Nebraska, Kansas, Minnesota, Iowa, Missouri, Wisconsin, Illinois, Michigan, Indiana, and Ohio).

Figure 5.1 shows the time series of the percent of this area in each drought level. The following analysis attempts to recreate this type of figure with LDAS-Monde LSVs and blends. The first section discusses RZSM and ET percentiles. Figures 5.3 and 5.4 present the time series of percent of area in each drought category as determined by the SEKF VODX SSM RZSM, and ET respectively. The next section looks at LAI percentiles Figures 5.5, and associated problems and solutions. These percentiles are taken weekly, comparing each week (Monday-Sunday) of the year throughout the 2003-2018 time period, to the same week of the year from the 2000-2018 OL experiment time period. Note that due to the small comparison time period, it is not possible for non-averaged/non-blended percentile ranks to fall between 5 and 2 percent, which is why D3 droughts are not seen in the time series to follow of individual LSV percentile ranks.

RZSM and ET

The RZSM and ET drought indicators present a relatively realistic perspective drought events. These two variables show high variability, possibly due to a small comparison period, resulting in a less smooth time series compared to Figure 5.1. A longer comparison period would include a larger number of percentiles, allowing for smaller week to week changes, instead of a minimum 5% increment with the 19 year period. Still, the general patterns and events can be easily seen in both LDAS-Monde RZSM and ET as well as USDM. The most notable and easily seen event is the 2012-2013 drought that struck this region. This event is visible and strong with the RZSM drought indicator, and can still be seen in the ET drought indicator. Also, the relatively lull in drought from 2010 to 2012 matches well.

Primary differences between the two variables include the overall larger amount of area in more severe droughts with RZSM compared to ET, and the higher variability or "noise" of ET drought levels compared to RZSM. These are just some reasons to merge the variables into blends, which would still include the data provided by each variable, while also eliminating some of the noise.

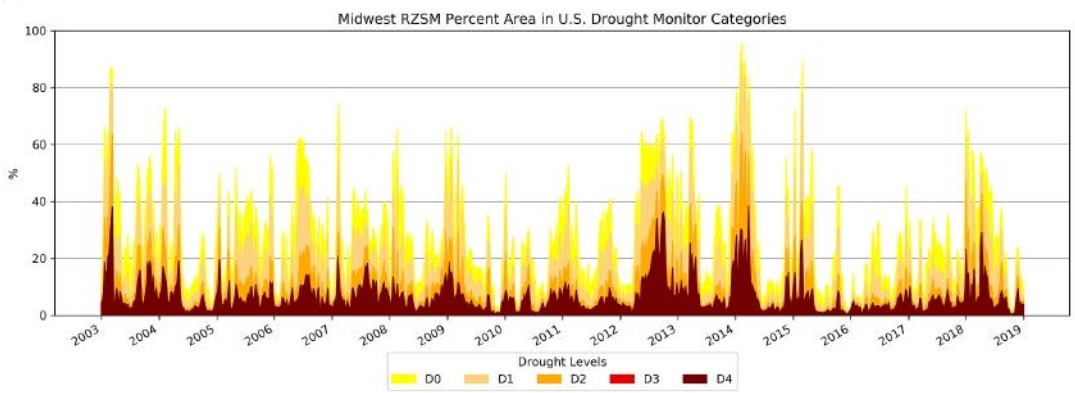

Figure 5.3: A time series of the percent area in each drought category, using RZSM percentile rank in drought category determination.

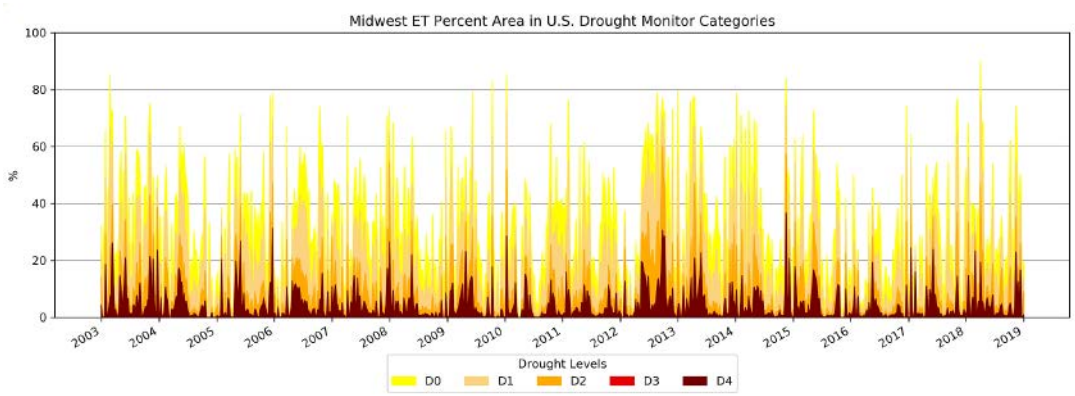

Figure 5.4: A time series of the percent area in each drought category, using ET percentile rank in drought category determination.

LAI

The LAI drought percentile responses in Figure 5.5 are striking, and immediately indicate bias with this LSV. There is a bi-annual cycle of D4 droughts, with more area covered in the summer event. This response is because the SEKF VODX SSM LSVs are compared to that of the OL for a longer time period, as done in similar studies (Houborg et al., 2012; Mocko et al., 2021). LAI is a control variable strongly impacted by the VODX assimilation, and in this case, there is a significant bias where the OL consistently overestimates LAI. After assimilation, these values are constrained and reduced, and thus when comparing the two, it seems as if there are strong droughts every summer. Additionally, as ISBA has employed minimum LAI values (1 m^2/m^2 for evergreen forests and 0.3 m^2/m^2 for all other vegetation), these percentiles will very often tie in value and in rank during winter or low vegetation months when these minimums are consistently reached in both the OL comparison period and the SEKF, leading to little to no drought information being retrieved. Together, these responses

identify weaknesses with using LAI as a variable in this methodology for an LDAS-Monde alert system.

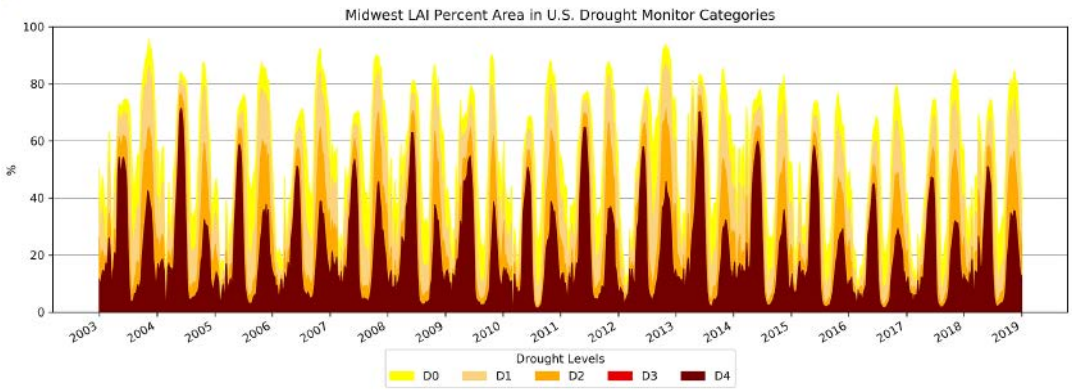

Figure 5.5: A time series of the percent area in each drought category, using LAI percentile rank in drought category.

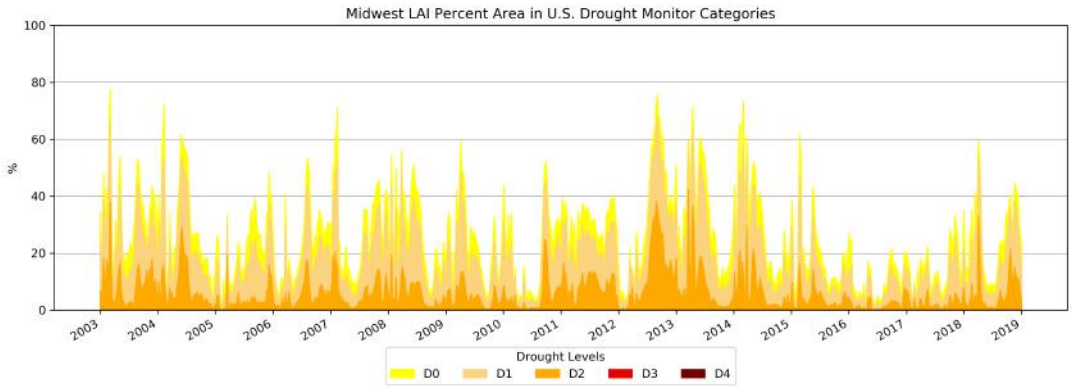

Figure 5.6: A time series of the percent area in each drought category, using LAI percentile rank, with the SEKF VODX SSM as compared against itself, as a drought category. The SEKF comparing against itself removes the bias due to the assimilation of the LAI-proxy, but cannot scale to large climatologies.

Figure 5.6 demonstrates the fact that this is in fact an artifact of the bias due to assimilation. This figure makes the same calculation of percentile rank as the figure above, but compares to its own 2003-2018 period including the LAI-proxy assimilation, instead of the 2000-2018 period OL with no assimilation. In this setup, we see similar responses to that of RZSM and ET, with no bi-annual events as seen in the normal LAI drought indicator. The 2012-2013 drought event is still visible, as are the inter-drought periods of 2016 to 2017. Because of the calculation of percentile rank, and only the 16 year comparison period, these LAI SEKF-SEKF percentiles do not drop below 5%, and thus no D3 or D4 drought events are seen.

While this strategy eliminates the bias associated with the LAI-proxy assimilation, it presents other challenges such as a significantly limited comparison

period (limited by satellite observation availability, for CGLS LAI available from 2000 to present), which weakens the significance of the percentiles. As stated in Houborg et al. (2012), which performs a similar percentile rank calculation on their GRACE DA system, "Calculating meaningful, current drought indicator percentiles requires a consistent long-term data record that extends to present time". Additional development can be made regarding selections of new, less biased, drought-related LSVs from LDAS-Monde, such as biomass and drainage, or statistical bias-correcting can be performed on the LAI variable. Additional analysis can also look at specific vegetation types such as was done in Chapter 4, with more focus on crops and grasslands over forests, as this system is targeted towards identifying and predicting agricultural droughts.

While LAI, and other vegetation-related LSVs are an important aspect of LDAS-Monde, their effectiveness as drought monitoring indicators tends to weaken in wintertime. Wintertime droughts, at least for much of CONUS, can strongly affect hydrology as well as certain lower biomass crops such as winter wheat, but much of the other vegetation is dormant. Snow cover can also impact the usefulness of these LSVs, as optical or microwave estimations of vegetation will be intermixed or contaminated by the snow. Thus, LAI would not be a relevant observation or indicator during this time over regions with this type of seasonality of vegetation. This is especially true when using LDAS-Monde LAI, as wintertime values often hit the defined minimum thresholds and give no more information. Other regions with smaller seasonal changes and little to no snow, would still find LAI and vegetation LSVs to be useful and impactful for drought monitoring. Therefore, careful attention must be paid to domain's climate zone and season when studying drought events with this system and these LSVs.

Blends

Taking a further step, two different blends were also calculated and plotted. Figure 5.7 shows the Blend 1, with equal weighting between LAI, RZSM, and ET variables, and Figure 5.8 presents an equal weighting of just RZSM and ET, as we have seen that LDAS-Monde LAI can be a biased indicator. These blends are the first to show significant area covered by D3 drought category, due to the weighing of different LSV percentiles. There is also a slight limit of some variability of the more extreme D3 and D4 drought categories. The blends also can provide a more holistic view of drought events, and can easily incorporate new variables and quickly change weighting for rapid prototyping of a drought index. The Blend 1 in Figure 5.7 includes the biased LAI response as discussed above, but does show that the weighing and merging can still provide a reasonable product. Without using LAI, as in Blend 2 and Figure 5.8, we do see some minor reduced variability.

Overall, these comparisons using the 19 year comparison period do show promise, and show that the methodology is sound for RZSM and ET LSVs. But the extreme variability (spikiness) and low number of extreme events in the period highlight that a longer period would greatly aid in a more in depth comparison.

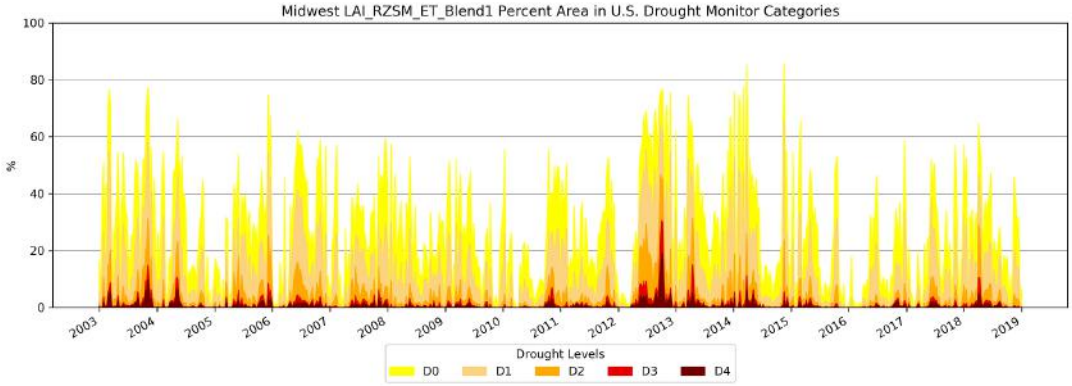

Figure 5.7: A time series of the percent area in each drought category, using a blend of LAI, RZSM, and ET percentile rank in drought category determination.

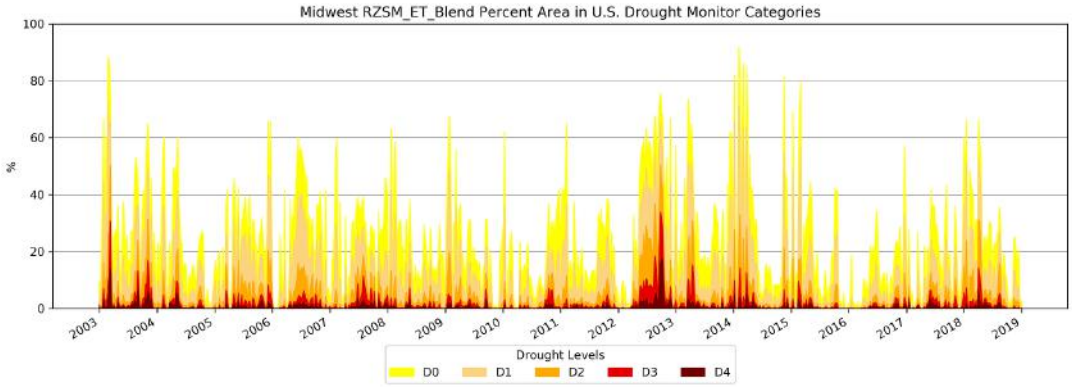

Figure 5.8: A time series of the percent area in each drought category, using a blend of RZSM and ET percentile rank in drought category determination.

5.2.3 Comparison to the United States Drought Monitor and other Drought Indices

USDM data, which is produced weekly and valid on Tuesdays, will be matched to the domain grid and time frame of the experiment. Initial analysis will look at the different drought index blends from the OL and SEKF valid on the same day as the USDM, converted to USDM equivalent drought levels, where the Spearman rank correlation, PODY, and FAR will be calculated. The resulting correlations and scores, along with the difference in these scores between the OL and SEKF, will be visualized on maps, potentially identifying more localized troublesome regions. Theses scores can also be plotted on charts and time series to focus on the timing of drought onset. For additional analysis, weekly averages of the LDAS-Monde drought blends could be compared to the USDM.

There may also be use to evaluate the performance of LDAS-Monde drought indices compared to only certainly levels of drought on the USDM. For example, the lower levels of drought (D0-D3) may be significantly better matched than if they included D4. In this analysis, the D4 levels of LDAS-Monde indices and USDM would just be converted to D3 levels, and the same scores and analyses re-run. As this is intended to work as a potential early warning system, the warning at earlier stages of drought is more important than perfectly matching the timing and extent of the slow developing extremes.

The same procedure described to compare LDAS-Monde forecast results to the USDM can be similarly applied to many other DIs. One such example is VegDRI, or Vegetation Drought Response Index, which is a composite map of satellite-based observations related to vegetation stress. Comparisons to other DIs can enhance the understanding of the system's performance, specifically related to the timescale of drought. For example, the Palmer Drought Severity Index (PDSI) is typically targeted towards long-term droughts, while the Standardized Precipitation Index (SPI) can be adjusted to account for short or long-term trends. Other DIs such as VegDRI target seasonal variability, and even others, such as QuickDRI were developed with short-term, rapid changes in mind.

5.2.4 Impact of Forecast LSVs on the LDAS-Monde Drought Indicators

While the presented proof of concept and methodology demonstrates the analysis can be performed on this data, the ultimate goal of the proposed experiment is to study the effect of forecasting LSVs on the LDAS-Monde drought indicators. We know from Chapter 3 that LSV correlations to observations decrease as forecast lead time increases, as a logical progression from decreasing atmospheric forecast accuracy. I hypothesize that the same will hold true for LDAS-Monde drought indicators and their associated drought levels. As forecast lead time increases, the Spearman correlation will decrease, the PODY will decrease, and the FAR will increase. When using 6 week atmospheric forecasts, a sufficiently long time period will also be required to build a large enough sample size of each forecast length to analyze changes in these statistical scores.

5.3 Summary of Chapter 5

This chapter explores the details of a prospective experiment to continue and follow on the work presented in Chapter 3 and Chapter 4. It takes lessons learned from both the two year forecast experiment over CONUS and the experiments with various assimilation scenarios, combining the knowledge in order to evaluate how the LDAS in forecast mode can perform as a drought alert system. The requirements of the case study (time period, domain) are also presented, which are primarily linked to data availability of the atmospheric forecasts and drought analysis data. The selection of LDAS-Monde configuration is then given and explained, with the joint assimilation of matched VODX and SSM

as the chosen setup in order to best improve the land surface initial conditions, primarily over the months of April through October.

Next, details are provided regarding the proposed analysis of the results of this prospective experiment. The methodology of this analysis follows Mocko et al. (2021), using the Spearman correlation, PODY, and FAR statistics, but instead focuses on the system's forecast and alert potential. Then, ET, RZSM, and LAI are selected as important LSVs which can be merged into a single blend to work as a drought indicator for LDAS-Monde. An analysis using existing experiments was performed using the proposed methodology, allowing for the processing and conversion of LDAS-Monde LSVs into percentile rank and USDM drought categories over the Midwest. These new drought indicators were also compared to a USDM time series of the area in each drought category. The RZSM and ET LSVs, along with the two blends, were able to match many details of USDM product over a similar area, most notably the 2012-2013 drought, while LAI was found to be heavily biased by the assimilation, limiting its use in this methodology. Finally, an explanation is given regarding how the comparison to USDM and other drought indicators will be conducted, and a discussion is had regarding the impact of forecasts on the LDAS-Monde drought indicators.

Chapter 6

Conclusions et Perspectives

Ce travail de thèse a eu pour objet d'étude l'évaluation de la capacité du système d'assimilation de données satellitaires pour les surfaces terrestres LDAS-Monde à suivre et à prévoir les sécheresses agricoles. De manière générale, les modèles des surfaces terrestres et les données d'observation de la Terre depuis l'espace sont affectés par des incertitudes qui limitent leur capacité à suivre et à prévoir les sécheresses et leurs impacts. Ce travail a montré que l'assimilation d'observations satellitaires dans le modèle ISBA permet de réduire ces incertitudes et d'améliorer les conditions initiales du modèle pour réaliser des prévisions. La configuration de LDAS-Monde a été modifiée afin de pouvoir utiliser des prévisions des variables atmosphériques pour forcer le système. Cela a rendu cet outil capable de fournir une prévision des variables des surfaces terrestres. Après avoir été mise en œuvre, cette nouvelle capacité de LDAS-Monde a été évaluée sur les USA. Une fois les limitations du système connues, des solutions permettant d'améliorer la précision des prévisions ont été proposées et testées dans la perspective de créer un système d'alerte des sécheresses agricoles et de leurs impacts.

6.1 Résumé des résultats

Les premières étapes de cette thèse se sont concentrées sur l'analyse des performances du système LDAS-Monde pour le suivi des impacts de la sécheresse sur la végétation et l'humidité du sol, et sur la manière dont les données auxiliaires affectent les résultats (par exemple les paramètres statiques du modèle ISBA gérés par ECOCLIMAP). Deux simulations ont été effectuées sur l'état du Nébraska, aux USA : une simulation « open-loop » (OL, une simulation sans assimilation de données) et une analyse (SEKF, une simulation avec assimilation d'observations d'humidité superficielle du sol et de l'indice de surface foliaire - LAI). Les anomalies annuelles moyennes de LAI simulées par OL et SEKF ont été comparées aux anomalies de rendement agricole de maïs, montrant de fortes correspondances, en particulier avec SEKF (les corrélations obtenues avec SEKF sont meilleures que celles d'OL). L'utilisation des valeurs annuelles maximales de LAI (au lieu de la moyenne annuelle) pour calculer les anomalies a donné des résultats presque identiques. Les données de précipitation de la ré-analyse atmosphérique ERA5 utilisées comme forçage atmosphérique ont été

évaluées en utilisant les mesures réalisées in situ à deux stations météorologiques du Nebraska. La comparaison montre une bonne correspondance générale.

A suivi ensuite le développement et l'utilisation de la capacité de LDAS-Monde a prévoir les variables des surfaces terrestres. Une expérience numérique a été mise en place sur les USA, couvrant la période de temps 2017-2018. Les simulations OL et SEKF ont été réalisées en étant forcées par le membre de contrôle de la prévision d'ensemble du ECMWF (ENS-CTRL). Cela a produit une prévision de 15 jours pour chaque journée de la période considérée. La persistance de diverses variables (LAI, RZSM, SSM, ruissellement, drainage et ET) a été analysée. Les variables LAI, ET et SSM ont été comparées à des observations satellitaires et in situ. Les résultats indiquent que le ruissellement est la variable la plus dynamique et la moins persistante dans le temps. La qualité des prévisions du ruissellement et du SSM dépend fortement de la qualité des prévisions atmosphériques. LAI et RZSM sont plus persistants dans le temps et varient moins vite à mesure que la prévision avance dans le temps. Pour les simulations de LAI, SSM et ET qui ont été comparées aux observations satellitaires, il apparaît clairement qu'à mesure que la période de prévision augmente, la précision du système diminue, mais fournit toujours des informations utiles sur toute la longueur de la période de prévision. On constate aussi que le choix des conditions initiales, c'est-à-dire les prévisions initialisées avec OL ou SEKF, a un fort impact sur les performances. Les conditions initiales plus précises fournies par SEKF sont plus performantes et pour des durées de prévision plus longues, notamment pour ET et LAI. Cette étude a été publiée dans Mucia et al., 2020.

Ayant identifié un moyen d'améliorer la précision de la prévision des variables des surfaces terrestres, c'est-à-dire l'amélioration des conditions initiales du système, et l'assimilation des observations liées à la végétation (comme le LAI dans notre cas), mon travail s'est orienté vers la recherche d'un moyen de mieux contraindre le LDAS en assimilant plus fréquemment des variables liées à la végétation. Contrairement au LAI qui provient d'observations réalisées dans le domaine du visible, du proche infrarouge et de l'infrarouge moyen, l'épaisseur optique de la végétation dans le domaine des micro-ondes (VOD en anglais pour « Vegetation Optical Depth ») est dérivé d'observations micro-ondes disponibles dans presque toutes les conditions météorologiques, de jour comme de nuit. Le VOD a été sélectionné comme variable à utiliser comme proxy du LAI et à assimiler par le LDAS car cette quantité contient de l'information pertinente relative à la végétation. Afin de mieux caractériser la relation entre LAI et VOD, des analyses ont été menées sur plusieurs zones géographiques en comparant les observations de LAI satellitaires et les simulations du modèle ISBA, aux observations de VOD en bande X et C (VOD-X et VOD-C, base de données VODCA) sur l'ensemble du domaine USA et sur des sous-domaines : la Californie, le Midwest, le Nord-Est, les plaines du Sud et le Nebraska. Ces comparaisons ont montré une corrélation généralement élevée entre LAI et VOD, mais avec des variations liées au type de végétation dominant dans chaque sous-domaine. Il a également été constaté que le VOD en bande X présentait moins de bruit et des corrélations plus fortes que celui en bande C. Les anomalies de VOD en bande X

et C ont également été évaluées, comme celles concernant les observations LAI, pour leur capacité à représenter la variabilité interannuelle des rendements du maïs. Les résultats sont comparables. Comme la végétation dominante semble jouer un rôle majeur dans la force de la relation entre LAI et VOD, six types de végétation (forêts de feuillus, forêts de conifères, cultures en C3, cultures en C4, prairies en C3 et cultures irriguées) ont été sélectionnés pour approfondir l'analyse. La relation entre LAI et VOD a été analysée sur les USA pour les pixels contenant plus de 50% de ces types de végétation selon les bases de données d'occupation des sols ECOCLIMAP-II et ECOCLIMAP-SG. Les forêts de conifères présentent les corrélations les plus faibles, tandis que les autres types de végétation présentent des corrélations modérées à fortes. La saisonnalité joue également un rôle important, de meilleures corrélations entre VOD et LAI étant généralement observées au printemps et en automne qu'en été et (surtout) qu'en hiver. La même analyse a été effectuée avec le VOD redimensionné linéairement aux observations de LAI via une fonction de distribution cumulative sur une fenêtre glissante de 3 mois. Cela permet d'obtenir des corrélations beaucoup plus fortes. Un phénomène d'hystérésis de l'évolution du VOD par rapport au LAI a été observé. Il se manifeste surtout pour les cultures en C4 en automne, avec des valeurs élevées de VOD. Ce phénomène semble indiquer que l'utilisation d'une méthode linéaire pour passer du VOD au LAI n'est pas suffisante pour tenir compte des tendances saisonnières. Cette analyse a également montré de fortes différences entre ECOCLIMAP-II et ECOCLIMAP-SG. En particulier, ECOCLIMAP-SG réduit le nombre de pixel dominés par les cultures en C4 et les cultures irriguées sur l'ensemble du domaine USA.

Après avoir comparé LAI et VOD-X (avant et après linéarisation), une série d'expériences d'assimilation a été réalisée. VOD-X, une fois redimensionné linéairement aux observations de LAI, a été assimilé à la place du LAI dans LDAS-Monde. Les résultats ont été comparés à quatre jeux de données d'observation satellitaires. Parmi eux, la production primaire de la végétation (GPP) et l'évapotranspiration (ET) sont indépendants des expériences. Sur la base des corrélations mensuelles moyennes, ainsi que des distributions des corrélations sur chaque point de grille, les résultats montrent que l'assimilation de VOD-X (mis à l'échelle du LAI) a amélioré la représentation de la GPP et de l'ET par rapport à l'assimilation directe du LAI sur la plus grande partie du domaine. Ces améliorations se produisent principalement pendant la saison de croissance de la végétation. Elles sont dues aux observations plus fréquentes du VOD par rapport au LAI qui permettent de mieux contraindre le modèle ISBA. Une évaluation complémentaire de ces résultats par rapport à des mesures in situ de l'humidité du sol indique que l'utilisation de VOD-X comme proxy du LAI dans l'assimilation permet d'obtenir des corrélations inchangées ou légèrement améliorées.

C'est l'une des premières études à assimiler le VOD dans un LSM. Kumar et al. (2020) a réalisé des expériences avec une méthodologie similaire, mais notre étude fournit l'utilité de l'assimilation VOD ayant un impact direct, non seulement sur le LAI, mais également sur les variables de contrôle de l'humidité du sol dans la zone racinaire. Les résultats ont démontré une contrainte plus

forte sur la végétation au sein de LDAS-Monde, et ce travail pionnier a mis en lumière des domaines où des recherches supplémentaires sont nécessaires, en particulier (1) comment redimensionner au mieux le VOD en LAI, (2) comment configurer au mieux les erreurs d'observation pour le VOD redimensionnée dans le système d'assimilation de données, (3) comment traiter les zones où une relation plus faible entre VOD et LAI a été obtenue.

Des expériences supplémentaires ont permis d'analyser les effets de l'assimilation individuelle et/ou conjointe des différentes observations. Aucune amélioration notable des scores de GPP et ET n'a été obtenue par assimilation conjointe des variables, par rapport à l'assimilation du LAI ou du VOD-X seuls. Cependant les comparaisons avec les observations in situ ont montré moins d'impacts négatifs et une plus grande amélioration des corrélations via l'assimilation conjointe. Ces expériences ont également montré que l'assimilation du SSM a un effet beaucoup plus faible sur l'évolution des variables des surfaces terrestres par rapport à l'assimilation du LAI ou du VOD-X.

Ayant acquis la certitude que l'assimilation du VOD-X en tant que proxy du LAI permet d'améliorer les conditions initiales, on peut recommander d'utiliser cette configuration de LDAS-Monde dans un contexte de prévision. Une méthodologie a été proposée pour comparer les variables produites par LDAS-Monde et un système existant de suivi des sécheresses. Afin de tester cette méthodologie, l'expérience SEKF assimilant VOD-X (en tant que proxy du LAI) et SSM de 2003-2018 a été comparée à une simulation OL de 2000 à 2018, et des classes de percentiles ont été produites sur la zone du Midwest. Une série chronologique sur cette zone dans chaque catégorie de sécheresses a été comparée aux produits du système de surveillance US Drought Monitor. L'utilisation des variables RZSM et ET issus de l'analyse ont montré certaines similitudes avec le US Drought Monitor, notamment en ce qui concerne la sécheresse majeure de 2012-2013.

La variable LAI produite par le SEKF présente des valeurs plus faibles que le LAI issu de OL. L'assimilation du LAI ou de VOD-X en proxy LAI réduit systématiquement le LAI simulé, conduisant souvent à des classes de percentiles artificiellement basses, c'est-à-dire à des sécheresses de forte intensité. Cependant, cette comparaison montre que cette méthodologie est applicable à LDAS-Monde, et que même avec une période de comparaison courte, des épisodes de sécheresse peuvent être détectés.

6.2 Perspectives

Bien que cette thèse ait fait une comparaison approfondie du VOD en tant que proxy du LAI, plusieurs questions restent en suspens. Le VOD n'est pas une quantité équivalente au LAI. Certains comportements, comme l'hystérésis observée au Chapitre 4 sur les forêts de feuillus, et surtout sur les cultures en C4, indiquent que leur relation n'est pas linéaire. Il serait utile d'approfondir ces résultats afin de déterminer la cause précise de ces phénomènes, qu'il s'agisse de la présence de litière au sol, de l'interception de l'eau précipitée par les feuilles ou

d'un autre processus. Effectuer la même analyse, mais filtrer les résultats par la présence d'un réservoir d'interception non nul (une variable de sortie d'ISBA), puis analyser les différences, serait un premier pas dans cette direction.

Une autre piste de travail consisterait à assimiler directement des données de niveau 1 dans LDAS-Monde, par exemple les coefficients de rétrodiffusion radar ASCAT ou des températures de brillance micro-ondes issues d'instruments passifs. Cela permettrait d'utiliser toute l'information contenue dans le signal, à la fois sur l'humidité du sol et sur la végétation, tout en gérant mieux les erreurs d'observation dans l'assimilation de données (Shamambo et al., 2019; Shamambo, 2020). Une étape majeure dans ce développement est la mise en place d'opérateurs d'observation. Cela peut se faire par l'utilisation de l'apprentissage automatique.

Concernant le système d'alerte des sécheresses et les comparaisons avec le système US Drought Monitor sur les USA, les prochains travaux pourraient concerner l'application de LDAS-Monde à l'alerte précoce des sécheresses, ainsi que sur l'utilisation des prévisions saisonnières pour générer des prévisions de variables des surfaces terrestres à plus longue échéance. Comme décrit au chapitre 5, une étape pourrait consister à effectuer un test des capacités d'alerte des sécheresses du système en utilisant les prévisions atmosphériques à moyen et à long terme du ECMWF. Des concepts similaires ont été appliqués à d'autres systèmes d'assimilation de données dans le but de produire des indicateurs de sécheresse correspondant au US Drought Monitor. Par exemple, Houborg et al., 2012 ont assimilé des valeurs d'anomalies gravimétriques de GRACE dans le modèle de surface « Catchment » pour produire des indicateurs de sécheresse hydrologiques. Getirana et al., 2020 ont ensuite utilisé ce système dans une configuration de prévision en utilisant des prévisions saisonnières à 30, 60 et 90 jours du système de prévision saisonnière à interannuelle du système d'observation de la Terre Goddard (GEOS) de la NASA (Borovikov et al., 2019), pour produire des indicateurs de sécheresse. Une approche similaire pourrait être mise en œuvre dans LDAS-Monde, mais avec l'avantage d'assimiler des variables de la végétation afin de mieux caractériser les sécheresses agricoles. Un tel système offrirait la possibilité de customiser des produits plus proches des besoins des acteurs du monde agricole.

Cependant, l'utilisation de prévisions saisonnières présenterait plusieurs défis à surmonter. Par exemple, la base de données du projet de prévision subsaisonnière à saisonnière (S2S) (Vitart et al., 2017) fournit des prévisions jusqu'à 60 jours provenant de nombreux centres opérationnels. Cependant, ces prévisions ne comportent pas toujours toutes les variables atmosphériques requises pour forcer ISBA. Par ailleurs, pour les durées de prévision plus longues, en particulier pour les durées saisonnières, les pas de temps sont souvent quotidiens. Par conséquent, une réduction d'échelle statistique devrait être utilisée pour obtenir un pas de temps suffisamment fin pour forcer ISBA (horaire, trihoraire). Une autre option serait d'initialiser un système de prévision saisonnière en utilisant les variables analysées par LDAS-Monde.

A partir des méthodes mises en œuvre dans le Chapitre 5, la capacité à caractériser les sécheresses de diverses variables ou combinaison de variables

provenant du LDAS ou pouvant être intégrées dans l'assimilation de données pourrait être explorée. Parmi ces variables supplémentaires, la fluorescence chlorophyllienne (SIF en anglais pour Solar-Induced chlorophyll Fluorescence) et la biomasse de la végétation pourraient être étudiées plus en détail. Lors de la photosynthèse, les plantes émettent un rayonnement dans une bande spectrale, qui change instantanément lorsqu'elle réagit aux conditions environnementales telles que l'éclairement ou la disponibilité en eau. Cela fait du SIF un proxy de l'activité photosynthétique pouvant être observé depuis l'espace (Meroni et al., 2009). Au fur et à mesure que les observations satellitaires du SIF se développent et qu'une période d'observation plus longue devient disponible, le SIF a tout pour devenir une donnée de plus en plus utile pour suivre l'état de la végétation. De même, la biomasse aérienne de la végétation, c'est à dire la masse sèche de bois par unité de surface, peut fournir une valeur de référence pour la production et la santé des forêts. Elle peut être utilisée avec d'autres variables de surface pour déterminer les impacts à long terme des sécheresses. Des liens étroits ont également été montrés entre le VOD en bande L et la biomasse aérienne de la végétation (par exemple Rodríguez-Fernández et al., 2018). Ces liens pourraient être analysés plus en détail dans le contexte de l'assimilation du VOD en tant que proxy du LAI dans LDAS-Monde.

Malgré les défis inhérents aux indicateurs individuels et les biais qui en découlent, la méthodologie présentée au Chapitre 5 fournit une base pour prévoir le déclenchement et la sévérité des épisodes de sécheresse avec LDAS-Monde. Ce dernier pourrait encore être amélioré avec de nouvelles variables simulées et de nouvelles observations à assimiler. Les résultats de prévisions à de plus longues échéances temporelles, par exemple saisonnière, permettraient également d'apporter un éclairage sur la précision et, par conséquent, l'utilité d'un tel système de prévision des sécheresses. Les étapes suivantes pourraient alors être définies, sur la base de ces résultats, de la disponibilité des données et des besoins des utilisateurs. Finalement, les futurs systèmes d'alerte et de prévision des sécheresses, tels que celui proposé dans cette thèse, pourraient fournir des informations vitales pour la prise de décisions en agriculture (dates d'irrigation, de semis, et de récolte par exemple), afin d'atténuer l'impact des sécheresses.

Chapter 7

Conclusions and Prospects

This thesis has investigated the ability and utility of LDAS-Monde to monitor and forecast agricultural droughts and water resources. Land surface models or satellite Earth observations alone are not sufficient to accurately monitor and predict drought events and impacts. However, the assimilation of satellite observations into LSMs provides reduced error and improved initial conditions. LDAS-Monde has been upgraded to allow for the ability to ingest atmospheric forecasts and produce land surface forecasts. This thesis focused on the use of this new capability, its limits, and analyzed ways to improve the accuracy of the land surface forecasts for the purpose of creating a drought alert warning system targeted towards agricultural impacts of drought.

7.1 Summary of Results

The initial steps of this thesis focused on analyzing how LDAS-Monde performs in monitoring impacts of drought on vegetation and soil moisture, and how the required supporting data (e.g. ECOCLIMAP) affect the results. Over the US State of Nebraska, the system was run in both OL (no assimilation) and SEKF (assimilating LAI and SSM) modes. These simulated mean annual LAI anomalies of the OL and SEKF were compared to reported corn yield anomalies, with strong matches between the experiments and agricultural yields. The SEKF also improved upon the OL correlations. Using yearly maximum LAI values (instead of yearly mean) to calculate the anomalies provided nearly identical results. The precipitation data from the ERA5 atmospheric reanalysis used as atmospheric forcing was also analyzed at two weather stations over Nebraska, overall providing a good match.

Work then moved to implementing the forecast configuration of LDAS-Monde and applying this new functionality over CONUS. An experiment was conducted, running the OL and SEKF over CONUS for 2017-2018, with the ECMWF ENS CTRL atmospheric forecast acting as the atmospheric forcing. This produced a 15-day forecast for each day of the time period. The persistence of LAI, RZSM, SSM, Runoff, Drainage, and ET variables were analyzed. LAI, ET, and SSM were compared to satellite derived and in situ observations. Results indicate that Runoff is the most dynamic and least persistent variable, and along with SSM, heavily dependent on accurate forecasts. LAI and RZSM, however, are more persistent in time and change far less strongly as the forecast

length increases. For the LAI, SSM, and ET that were compared to satellite derived observations, it is clear that as the forecast period increases, the accuracy of system decreases, but still provides useful information out to the full forecast length. It is also clear that the initial conditions, that is the forecasts initialized with the OL or SEKF, make a strong impact on the performance, with the more accurate initial conditions of the SEKF performing better and for longer forecast lengths, particularly for ET and LAI. This analysis was published in Mucia et al. (2020).

Now having identified a way to improve land surface forecast accuracy, that is improving the initial conditions of the system, and that the assimilation of vegetation-related observations (i.e. LAI) have the strongest impact on the system, the focus of the thesis shifted towards a way to better constrain the LDAS by assimilating an LAI-like variable more frequently. VOD from the VODCA dataset was selected as the variable to transform to an LAI proxy and assimilate into the LDAS, as VOD is a microwave observation producing nearly all weather and close to daily observations in contrast to optical observations such as LAI, and it contains pertinent information related to vegetation. To better characterize the LAI-VOD relationship over various domains, analyses were conducted comparing LAI observations and ISBA simulated LAI to X and C-band VOD observations over CONUS, California, the Midwest, the Northeast, the Southern Plains, and Nebraska. These comparisons found a generally good correlation between the two variables, but with significant variation strongly linked to dominant vegetation type in each sub-domain. It also found X-Band VOD had significantly less noise and stronger correlations than C-Band. The two VOD bands were also plotted as anomalies next to LAI and corn yield anomalies, providing a moderately good match. As the dominant vegetation seemed to play a key role in the strength of the LAI-VOD relationship, six vegetation types (deciduous forests, coniferous forests, C3 crops, C4 crops, C3 grasslands, and irrigated crops) were selected. The LAI-VOD relationship was analyzed over CONUS for pixels with more than 50% of these vegetation types according to the ECOCLIMAP-II and ECOCLIMAP-SG land use databases. Coniferous forests consistently had the lowest correlations, while other vegetation types had moderately strong correlations. Seasonality played a large role as well, with spring and autumn months typically outperforming summer, and strongly outperforming winter. The same analysis was performed with VOD that had been linearly re-scaled to LAI observations through a 3-month moving window cumulative distribution function, instead of using the raw observations, resulting in very strong correlations throughout. A hysteresis pattern was observed in the autumn season of the LAI to VOD comparison where C4 crops had higher VOD at lower LAI. This pattern potentially indicates that the linear re-scaling is not sufficient to account for seasonal trends. This analysis also found some strong differences between ECOCLIMAP-II and ECOCLIMAP-SG, with a ECOCLIMAP-SG reducing patches dominated by C4 and irrigated crops over CONUS.

After comparing LAI and both matched and unmatched VODX, a series of assimilation experiments were run, where X-band VOD, linearly re-scaled to

LAI observations, was assimilated in place of LAI in LDAS-Monde. The results were compared to four satellite-derived observation datasets, with GPP and ET being completely independent evaluations datasets. Through averaged monthly correlations, as well as the distributions of the individual point correlations, the results demonstrate the assimilation of matched VODX improved representation of GPP and ET over that of assimilating LAI over most of the US. These improvements primarily occurred during the growing season, when the vegetation dynamics are at their peak, and the improvements are due to the more frequent observations of VOD compared to LAI. Additional evaluation of these results against in situ soil moisture observations support that matched VODX can replace LAI in assimilation with correlations being either unchanged or slightly improved.

This is one of the first studies to assimilate VOD in an LSM. Kumar et al. (2020) has performed experiments with similar methodology, but our study provides the utility of VOD assimilation directly impacting not only on LAI, but the root zone soil moisture control variables as well. The results demonstrated stronger constraint on vegetation within LDAS-Monde, and this pioneering work illuminated areas where additional research is needed, specifically (1) how to best re-scale VOD to LAI, (2) how to best set up observational errors to the re-scaled VOD in the data assimilation system, (3) how to deal with areas where a weaker relationship between VOD and LAI was obtained.

Additional experiments analyzed the effects of individual and joint assimilation of vegetation related LSVs and SSM. No notable differences were seen in the independent GPP or ET observations when assimilating vegetation related information by itself or jointly with SSM. However, in situ comparisons demonstrated fewer negative impacts, and more improved correlations via joint assimilation. These experiments also proved that the assimilation of SSM has a far weaker effect on the evolution of the land surface compared to the assimilation of LAI or matched VODX.

With evidence that shows assimilating matched VOD provides some advantages improving initial conditions, we can recommend using this LDAS-Monde forecast configuration. A methodology was proposed to compare LDAS-Monde LSVs to an existing drought monitoring system. In order to test this methodology, the SEKF experiment assimilating VODX (as an LAI-proxy) and SSM from 2003 to 2018 was compared to an OL simulation from 2000 to 2018, and the percentile ranks were produced over the Midwest. A time series over the area in each drought category was compared to a similar region from the USDM. Variables of RZSM and ET demonstrated some similarities, specifically picking up on the major 2012-2013 drought event. LAI was found to be significantly biased by the SEKF to OL comparison, as the assimilation of LAI or LAI-proxies systematically reduces the simulated LAI, often leading to artificially low percentiles, i.e. severe droughts. However, this comparison goes to show that the percentile rank methodology can be performed on LDAS-Monde results, and that even with a short comparison period, drought events can be detected.

7.2 Prospects

While this thesis has made an in depth comparison of VOD as an LAI-proxy, there remain several important questions. VOD is not equivalent to LAI. Certain responses, namely the hysteresis seen in Chapter 4 over deciduous forests and especially over C4 crops, indicate that VOD and LAI do not have a purely linear relationship. Future analysis continuing this work can look deeper into these results, and determine the precise cause, whether ground litter, leaf-water interception, or some other process. Performing the same analysis, but filtering out any point and time when there is a non-zero interception reservoir (an output variable of ISBA), and analyzing the differences would be the first step in that direction.

Another step in that direction consists of directly assimilating level 1 data into LDAS-Monde, ASCAT radar backscatter or microwave brightness temperature from passive instruments for example. This would allow the use of all the information contained in the signal, including soil moisture and vegetation, and at the same time reduce observations errors in the data assimilation (Shamambo et al., 2019; Shamambo, 2020). A major step in this development is the solution of the observation operator. Further steps towards developing this operator will focus on using machine learning approaches to find the optimal term.

Regarding the drought alert system and comparisons to the US Drought Monitor, follow up work to this thesis could focus on the application of the LDAS-Monde forecast system towards drought early warning, as well using seasonal forecasts to drive longer land surface forecasts. As described in Chapter 5, a next step could be to perform a test of the system's drought alert capabilities using medium to long range atmospheric forecasts from ECMWF. Similar concepts have been applied to other data assimilation systems with the goal of producing objective drought indicators matching the Drought Monitor. Houborg et al. (2012) uses assimilated GRACE gravity anomaly data in the Catchment LSM to produce hydrological drought indicators in percentile form. Getirana et al. (2020) then takes the GRACE-DA system into a forecast configuration using 30, 60, and 90-day seasonal forecasts from NASA's Goddard Earth Observing System (GEOS) Seasonal-to-Interannual Forecast System (Borovikov et al., 2019), producing drought indicator outlooks in the same percentile form. A similar implementation can be imagined for LDAS-Monde, but with the advantage of assimilating vegetation variables and a focus on agricultural drought. This gives the ability to experiment with different variables and blends of variables, potentially better tuning the system to the needs of agricultural producers.

However, using seasonal forecasts would present several challenges needing to be overcome. The Subseasonal to Seasonal (S2S) Prediction Project Database (Vitart et al., 2017) provides seasonal forecasts, up to 60 day, from numerous operational centers. A significant first hurdle would be to find a forecast that outputs all the atmospheric variables required by ISBA at the surface. Variables such as wind speed, or shortwave and longwave radiation are not guaranteed to be available products. For longer forecast lengths, especially at seasonal length, the time steps are often daily. Therefore, statistical down-scaling would need

to be employed to obtain the sub-daily evolution of the atmosphere needed to drive ISBA. Another option would be to initialize a seasonal forecast system using the LDAS-Monde analyzed variables.

Using the methodology of Chapter 5, additional LSVs from the LDAS can be reviewed and processed to determine their utility as individual or a part of blended indicators. Among these additional LSVs, Solar-Induced chlorophyll Fluorescence (SIF) and vegetation biomass can be studied in more detail. As plants photosynthesize, they emit radiation in a narrow spectral band, which instantly changes as the plant responds to environmental conditions such as water and light availability. This makes SIF a direct proxy for photosynthetic activity, and can be actively monitored via remote sensing (Meroni et al., 2009). As these satellite observations of SIF become more widely available and a longer observation period is built up, SIF becomes a more and more useful tool in monitoring vegetation conditions. Similarly, plant biomass, the total mass of vegetation in a unit area, provides a reference LSV for vegetation production and health, which can be used together with other surface variables to determine drought impacts. There have also been strong links found between L-Band VOD and above ground biomass (Rodríguez-Fernández et al., 2018), which can be further analyzed in the context of assimilating VOD as an LAI-proxy in LDAS-Monde.

Despite the challenges associated with individual indicators and the biases found therein, the methodology shown in Chapter 5 provides a solid base for predicting drought onset timing and severity with LDAS-Monde that can be incrementally improved with new variables and as new observations become available. The results of longer, seasonal scale, forecasts will also shed light on the accuracy, and consequently the utility, of this drought forecast system. Steps beyond that will be determined by those results, data availability, and the needs of potential users. Ultimately, future drought warning and prediction systems, such as the one proposed in this thesis, can provide stakeholders with vital information to make crop and timing decisions (irrigation, planting, and harvesting for example), as well as better mitigate drought impacts.

Appendix A

Appendix A

A.1 LDAS Experimentation over Nebraska

The U.S. state of Nebraska was chosen as a domain for initial case studies in which LDAS-Monde runs in a normal configuration, as this state is a significant agricultural producer, as well as having a very strong presence of irrigation. Being a significant producer of crops allows for good comparisons between LSVs and crop production on an annual basis. The smaller domain of the state (compared to CONUS) also allows for the experimentation of LDAS-Monde in a higher spatial resolution. The atmospheric forcing can also be evaluated against in situ weather observations stations. Differences in land surface properties between ECOCLIMAP-II and ECOCLIMAP-SG are also investigated.

A.1.1 Potential High Resolution LDAS

LDAS-Monde can be set to any spatial resolution, and is limited primarily by the resolutions of atmospheric forcing, land cover, observations, and coupled models (e.g. CTRIP). Higher resolution monitoring with LDAS-Monde has been shown to improve representation of vegetation conditions (Albergel et al., 2019). As ERA5 resolution is natively 0.25° x 0.25°, most LDAS-Monde experiments have been run with equal spatial resolution. However, there is some evidence from Albergel et al. (2019) that simply using a bilinear interpolation to downscale ERA5 forcing to 0.1° x 0.1° spatial resolution improves the monitoring of vegetation almost as much as using the native 0.1° IFS HRES forcing.

Some initial experiments with LDAS-Monde over Nebraska included a run from 2017-2018 at 0.1° spatial resolution and forced by the ECMWF IFS HRES atmospheric forecasts (see Table 2.1). This experiment was also a first test of the forecast configuration of LDAS-Monde, with 10 day forecasts being produced. With the increased spatial resolution, coupling to CTRIP becomes impractical and introduces errors. As CTRIP is run at 0.5°, resolutions smaller than half of that produce visible artifacts in output data.

Even without a full run experiment, some insight can be gained from looking at parameters at high resolutions. Figure A.1 shows maps of Nebraska's C4 fraction as given by ECOCLIMAP-SG for 0.25°, 0.1°, and 0.01° spatial resolutions. ECOCLIMAP has an effective spatial resolution of 1km, but is interpolated to the model grid resolution. In this figure, the heavily pixelized large spatial resolutions progressively become clearer, to the point where, at

the 0.01° level, an large amount of detail can be gleaned from the land use. Geographic features such as rivers, lakes, urban centers, and topography all can be easily viewed at these higher resolutions. Field scale or smaller detail is possible, and applications of this scale can be made even to precision agriculture or more targeted monitoring and forecasting.

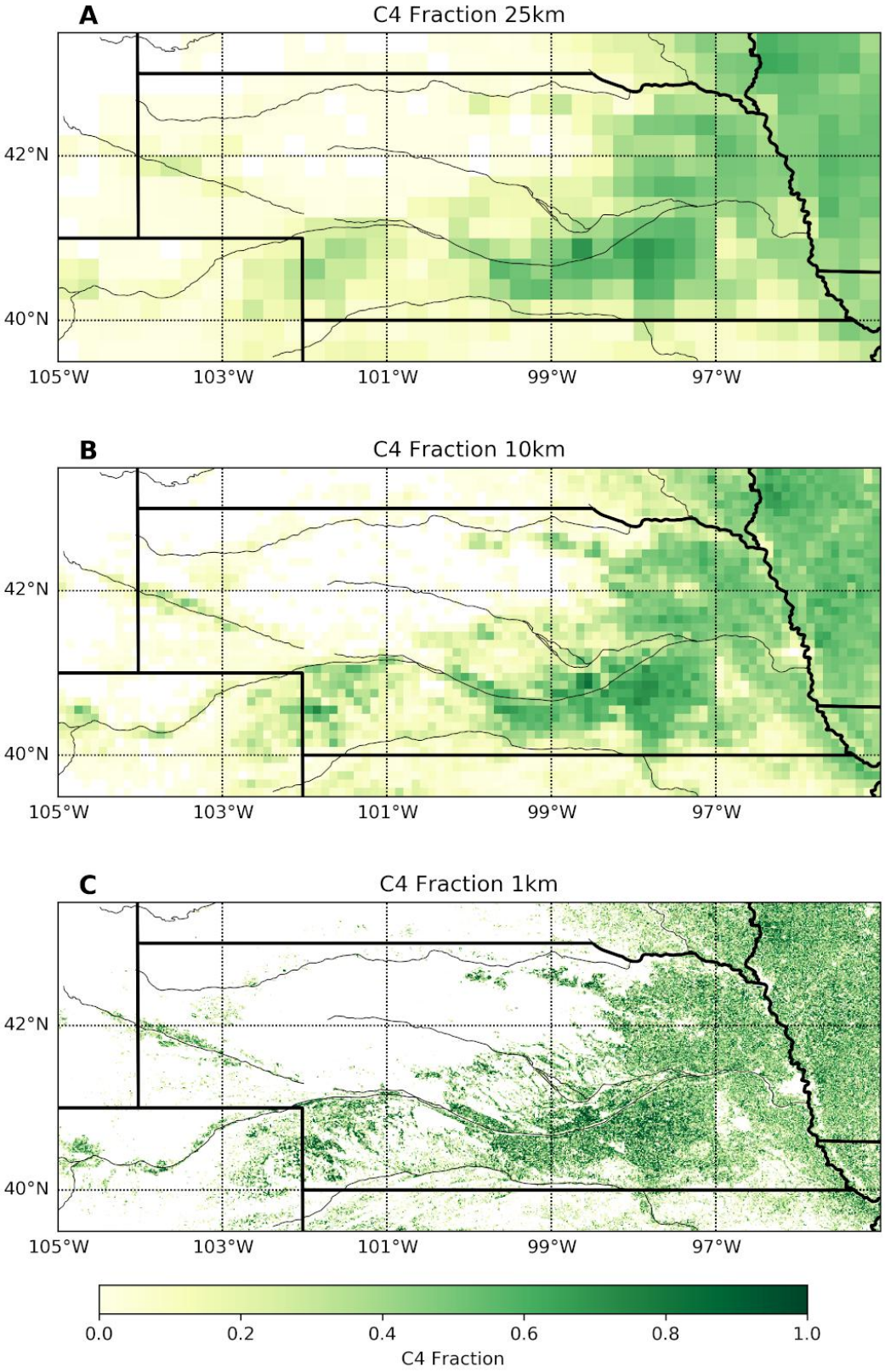

Figure A.1: Maps of the fraction of C4 crops as given by ECOCLIMAP-SG at A) 0.25°, B) 0.1°, and C) 0.01° spatial resolutions.

The downside to using this high resolution is primarily the cost (both time and energy) of running the system for such a domain. Such detailed resolution of a small domain on the scale of a state could be equal in computing cost to that of a more traditional resolution experiment over the scale of a continent. Additionally, the atmospheric forcing or forecasts are typically the bottleneck in terms of spatial resolutions of the input datasets.

A.1.2 Assessment of LDAS-Monde, ERA5, and ECO-CLIMAP over Nebraska

LDAS-Monde

Mucia et al. (2020), shown later in this chapter, analyzed and compared LAI observations and corn yield on an annual basis. Figure A.2 demonstrates this same comparison, but with the addition of LDAS-Monde OL and SEKF results. This figure, and subsequent anomaly figures over Nebraska, use the mean annual LAI for observations and OL/SEKF results to calculate the anomalies. When the maximum LAI was use to calculate the anomalies, little to no change was seen. It is seen that both the model and analysis closely match the observed inter-annual variability of LAI as well as of corn yield. Table A.1 provides the correlation scores between the observed, modelled, and analyzed LAI and corn yield.

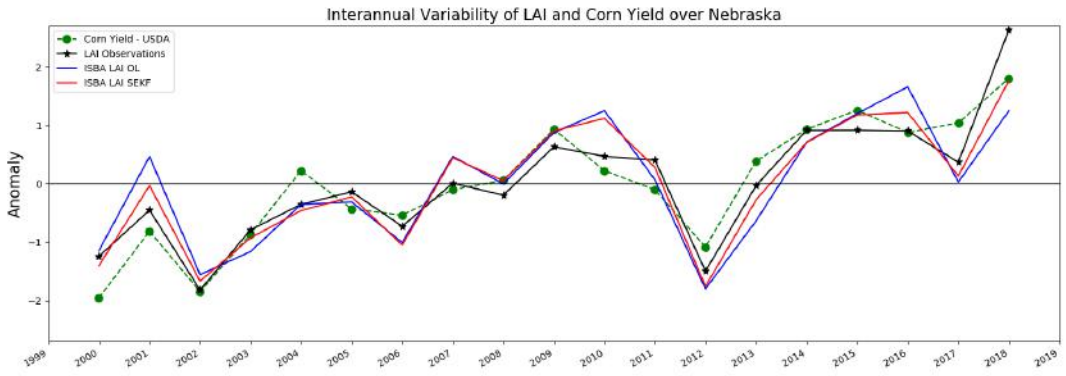

Figure A.2: A time series representing the inter-annual anomalies of mean observed, modelled, and analyzed LAI, and corn yield over Nebraska. The black line and stars represent the mean annual observed LAI anomalies from the CGLS LAI V2 dataset. The green dashed line represents annual corn yield anomalies as reported by the USDA. The blue and red lines are the LDAS-Monde OL and SEKF LAI products respectively.

Table A.1: Correlation Coefficients between Observed, Model, and Analysis LAI and Corn Yield over Nebraska

	Yield	Model	Analysis	Observations
Yield	1	0.79	0.87	0.92
Model		1	0.97	0.86
Analysis			1	0.94
Observations				1

As stated in Mucia et al. (2020), the LAI observations closely match the annual corn yield anomalies, with a correlation of 0.92. We can extrapolate that annual LAI anomalies can be used as a yield proxy. LDAS-Monde also proved to be able to follow these inter-annual variations of corn yield with its simulated LAI. The model alone provided a correlation of 0.79, while the analysis improved that to a correlation of 0.87 thanks to the assimilation of LAI and SSM observations.

The results of this test can be broadly categorized as showing that the LDAS-Monde system has the capability to monitor land surface variables and conditions related to drought events in an agricultural setting. Coupled with previous studies (Albergel et al., 2018a; Albergel et al., 2019; Tall et al., 2019), LDAS-Monde has shown that it can successfully monitor more general drought conditions.

ERA5

Analysis was performed over Nebraska to assess the accuracy of ERA5 atmospheric forcing compared to monitoring stations. While previous studies have shown ERA5 provides accurate atmospheric conditions, and even improves the performance of LDAS-Monde compared to ERA-Interim (Albergel et al., 2018a), over a more localized domain, this analysis is also necessary to confirm the accuracy.

In particular, the variable of precipitation was chosen as the variable itself is an important driver of land surface properties, and the comparison and validation is straightforward. Two sites in Nebraska with National Weather Service (NWS) weather stations were also chosen, that of Lincoln (40.83N,96.76W) and Grand Island (40.96N,98.31W). To encompass a wide range of precipitation conditions, the three years of 2009 (wet), 2011 (normal), and 2012 (dry) were selected and investigated. Figures A.3 and A.4 represent the monthly cumulative precipitation totals between the station observations and ERA5 at grid cell containing the station.

In general, one can see a good correspondence for most months between the two datasets. The seasonality is well captured, and the inter-annual changes are also seen. There are sporadic months of disconnect between ERA5 and the observations, but the biggest consistent mismatch is for 2012 in Grand

Island (Figure A.4C). In this specific case, ERA5 consistently overestimates precipitation throughout the year. While Lincoln has a slight mismatch between the two for August and October, this same general pattern is not seen.

One must note that this comparison is to test the general accuracy of the ERA5 dataset, but this analysis is also a simple one. The observations are a very high quality point measurement of precipitation, while ERA5 is an estimation over a 0.25° x 0.25° grid cell. Precipitation events are often very localized, especially in thunderstorms and extreme severe weather, as is often seen by both Grand Island and Lincoln. In fact, the localized precipitation events, often seen in Spring, could be able to explain the individual large discrepancy between the two during the month of April in 2009 for Lincoln, and 2012 for Grand Island.

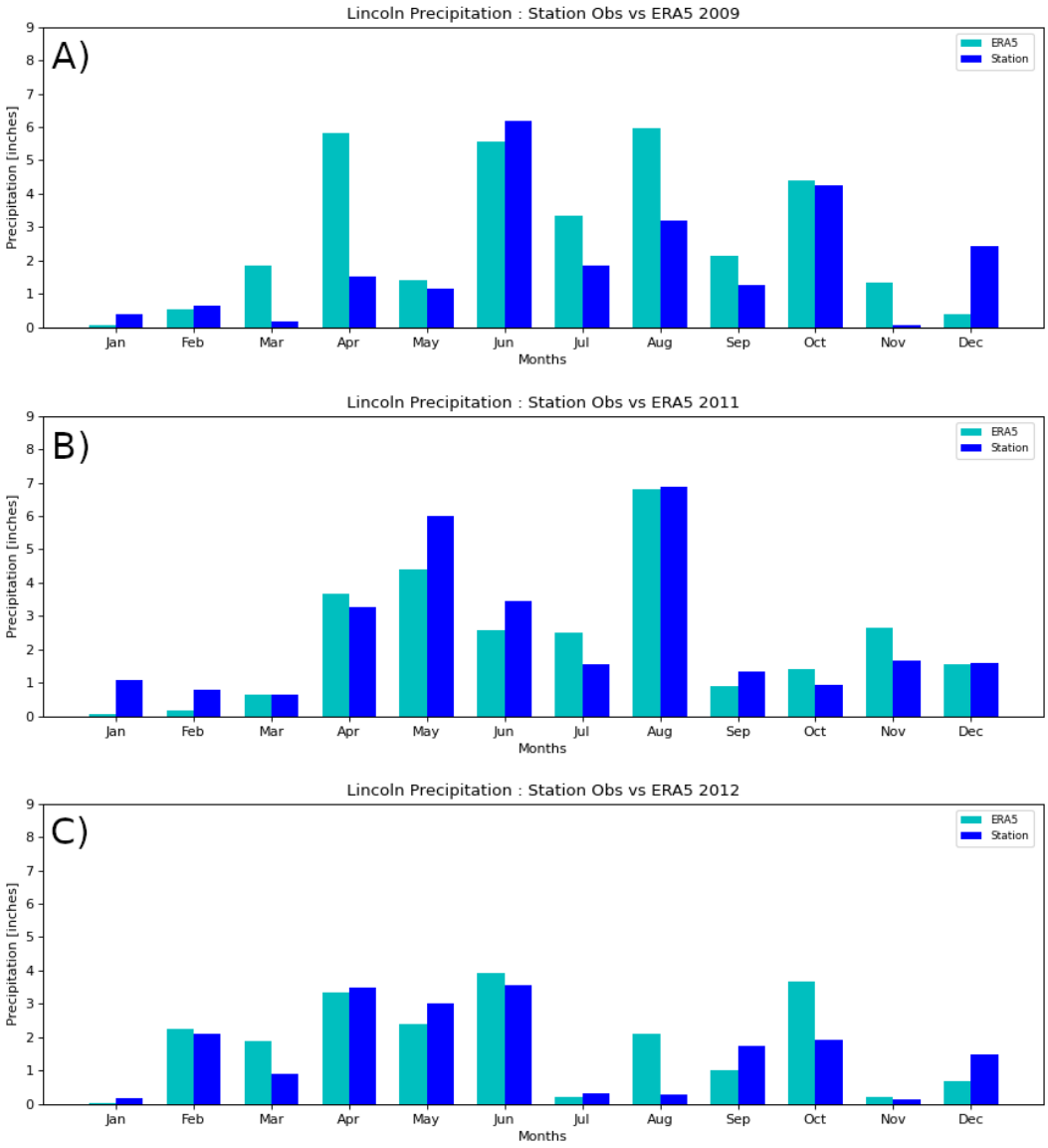

Figure A.3: Comparison of monthly cumulative precipitation over Lincoln, Nebraska between NWS weather station (KLNK) observations and ERA5 for A) 2009, B) 2011, and C) 2012

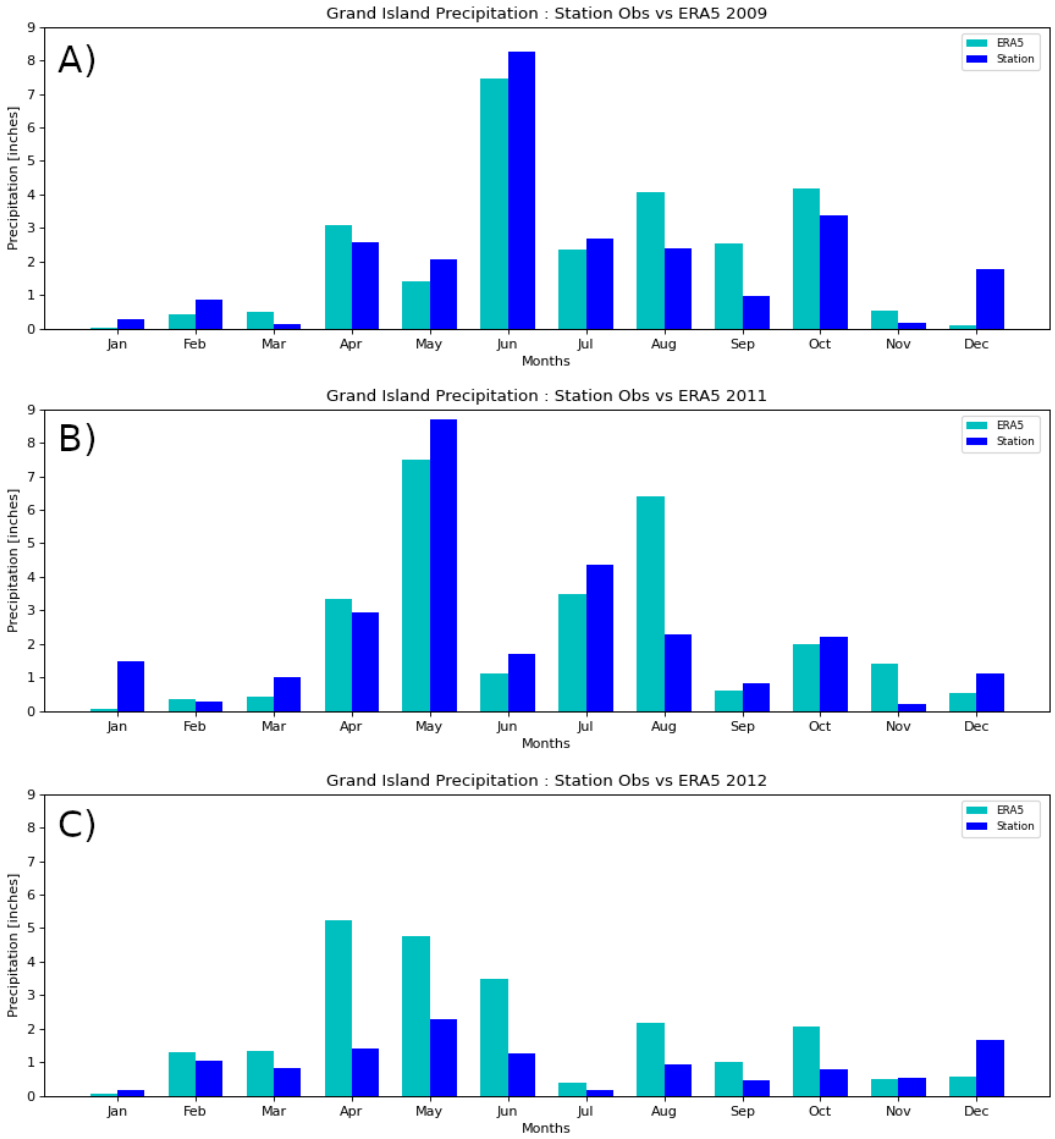

Figure A.4: Comparison of monthly cumulative precipitation over Grand Island, Nebraska between NWS weather station (KGRI) observations and ERA5 for A) 2009, B) 2011, and C) 2012

ECOCLIMAP

As demonstrated by Figure 2.2, ECOCLIMAP-II and ECOCLIMAP-SG have obvious differences in their dominant patch fractions. When concentrated solely on the agricultural state of Nebraska, a more focused comparison is possible. Two main crop types by area in Nebraska are corn (maize), and soybeans. Corn is a C4 crop, while soybeans use the C3 pathway. With this in mind, and with already having compared corn yields in the state, much of the focus will continue to be on C4 crops and corn.

Figure A.5 compares the C4 patch fraction between ECOCLIMAP-II and ECOCLIMAP-SG at 0.25° resolution. These maps demonstrate large and widespread differences between the two ECOCLIMAP databases. While ECOCLIMAP-II has nearly the entire Eastern half of the domain at nearly 100% C4 fraction, ECOCLIMAP-SG shows a domain with C4 concentrated closer to the Platte River and Iowa, while also showing an overall reduction in C4 fraction where it is present.

While Figure A.5 does demonstrate some significant differences, when the model is run with both ECOCLIMAP databases, the results are very similar. Figure A.7 shows the same time series as Figure A.2 but with an identical plot using ECOCLIMAP-II below. The two experiments were run with the same forcing, observations, and domain, only differing in the ECOCLIMAP version. As it shows, the model and analysis (blue and red lines respectively) barely change at all.

A high resolution frequency map of corn land cover over the state of Nebraska is shown in Figure A.6. This map is created from the U.S. Department of Agriculture's (USDA) crop-specific data layer online tool (USDA, 2021), and depicts the frequency of corn cover from 2008-2020. Visually comparing this map to Figure A.5, it is clear that ECOCLIMAP-SG presents a better spatial representation of C4 crops over Nebraska, as the western third portion of the states does indeed have a significant presence of corn, which is not seen in ECOCLIMAP-II.

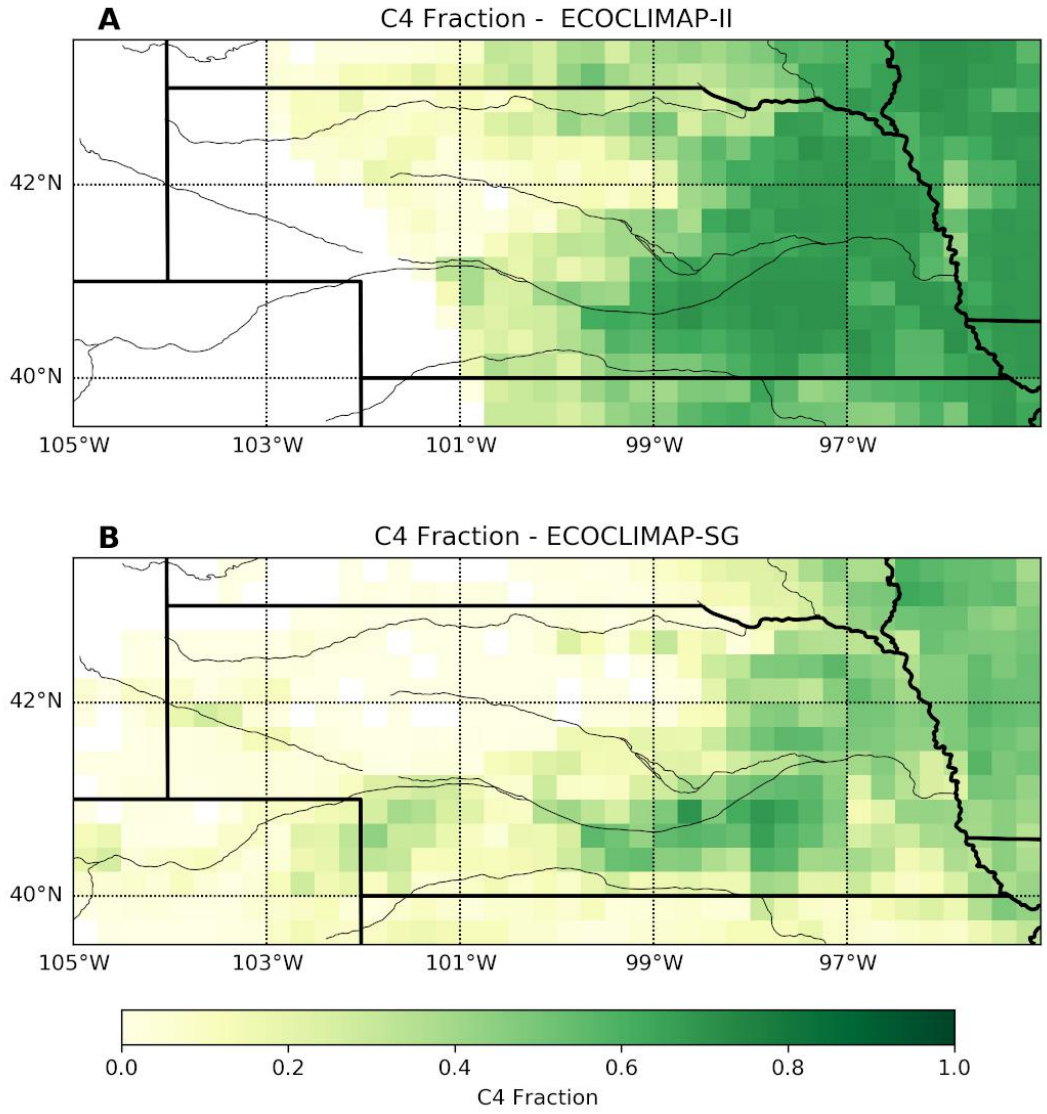

Figure A.5: Maps of C4 crop fraction over Nebraska between A) ECOCLIMAP-II and B) ECOCLIMAP-SG at 0.25° resolution.

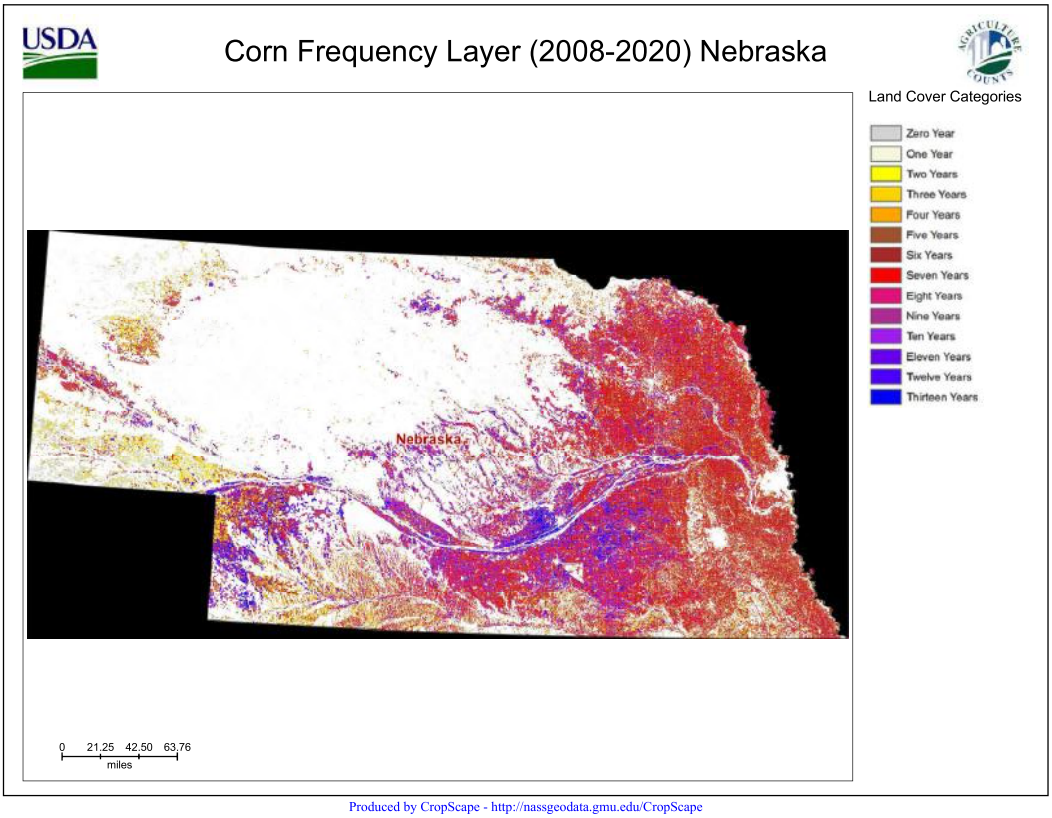

Figure A.6: A map of corn frequency over Nebraska between 2008 and 2020. The color categories indicate how many of the last thirteen years the land has had corn cover. Source: USDA, 2021

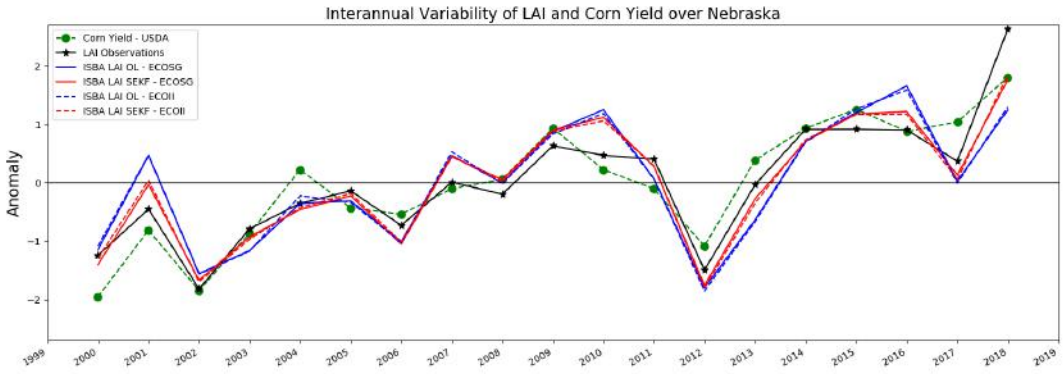

Figure A.7: A time series of annual LAI and corn yield anomalies as in Figure A.2 with solid lines representing the experiment run with ECOCLIMAP-SG, and dashed lines representing the experiment run with ECOCLIMAP-II.

In terms of changes to statistical scores, Table A.2 demonstrates that, rounded to two significant figures, the only change seen is a slight increase of corn yield correlation from 0.86 between the SEKF with ECOCLIMAP-II to

0.87 with ECOCLIMAP-SG. No significant differences are found in the scores when running the system with either ECOCLIMAP database. This indicates that while differences certainly exist, and are visually evident when focused on individual patch types, they play a smaller role in the outcome of model LAI. Additionally, the spatial and temporal averages of this comparison smooth out many differences, leading to very similar scores.

Table A.2: Correlation Coefficients between Model and Analysis LAI versus Corn Yield and LAI observations over Nebraska for two ECOCLIMAP versions. Values outside parentheses are correlations from runs with ECOCLIMAP-II and values within parentheses are from runs with ECOCLIMAP-SG.

	OL ECOII (OL ECOSG)	SEKF ECOII (SEKF ECOSG)
Corn Yield	0.79 (0.79)	0.86 (0.87)
LAI Observations	0.86 (0.86)	0.94 (0.94)

A.2 Summary of Appendix A

- Among these Nebraska experiments are those focusing on the testing of running LDAS-Monde at higher resolutions, and assessing the benefit's gleaned from the more detailed spatial view. The increased resolutions provide significantly more detail, and can identify topography, urban centers, and even large individual fields. This sort of capability may be used in future studies focused on precision agriculture modelling.

- Using the same domain, testing was performed on the LDAS-Monde system, complementing and expanding on information presented in the article, this assessment of the capability shows that not only does LAI work well as a indicator tracking annual crop yield anomalies, but the model and analysis LAI have similarly strong matches. Additionally, a small assessment of the ERA5 forcing was performed by comparing monthly precipitation at two locations in Nebraska over three separate years. Finally, differences between the ECOCLIMAP-II and ECOCLIMAP-SG are presented, highlighting the differences in spatial distribution of vegetation types, as well as comparing LAI anomaly outputs when the two versions are used in otherwise identical LDAS-Monde experiments.

Appendix B

Appendix B

This appendix provides the same figures for VODC as were shown in Chapter 4 for VODX. At this point in the study, VODCA VODX had been chosen as the band that was to be linearly re-scaled and assimilated in LDAS-Monde, however future studies may select different bands. Thus the same analysis was performed on both VODC and VODX. In this analysis, the differences were generally minor.

B.1 LAI versus VODC

Deciduous and coniferous vegetation shown potentially the biggest change, with LAI-VODC correlations for both ECOCLIMAP-II and ECOCLIMAP-SG significantly dropping compared to LAI-VODX. Additionally, the "noisiness" of VODC compared to VODX as seen in the domain comparisons, is still present when selecting vegetation types.

Table B.1: Seasonal and total correlations of VODC versus LAI for ECOCLIMAP-II patches

VODC ECOII					
Veg Type	All Seasons (575)	Winter (144)	Spring (143)	Summer (144)	Autumn (144)
All Vegetation	0.47	0.71	0.68	0.74	0.49
Deciduous	-0.06	-0.15	-0.26	-0.33	0.38
Coniferous	-0.38	0.23	-0.28	0.14	-0.24
C3 Crops	0.59	0.04	0.36	0.50	0.53
C4 Crops	0.63	0.06	-0.44	0.83	0.74
C3 Grasslands	0.49	0.54	0.57	0.63	0.21
Irrigated Crops	0.38	0.28	0.44	0.63	0.26

Table B.2: Seasonal and total correlations of VODC versus LAI for ECOCLIMAP-SG patches

VODC ECOSG					
Veg Type	All Seasons (575)	Winter (144)	Spring (143)	Summer (144)	Autumn (144)
All Vegetation	0.69	0.30	0.71	0.37	0.51
Deciduous	0.22	0.66	0.22	0.67	0.25
Coniferous	-0.61	0.12	-0.51	0.01	-0.38
C3 Crops	0.69	0.05	-0.04	0.76	0.54
C4 Crops	0.59	-0.23	-0.23	0.85	0.61
C3 Grasslands	0.44	0.52	0.45	0.55	0.33
Irrigated Crops	0.16	0.14	0.30	0.12	0.04

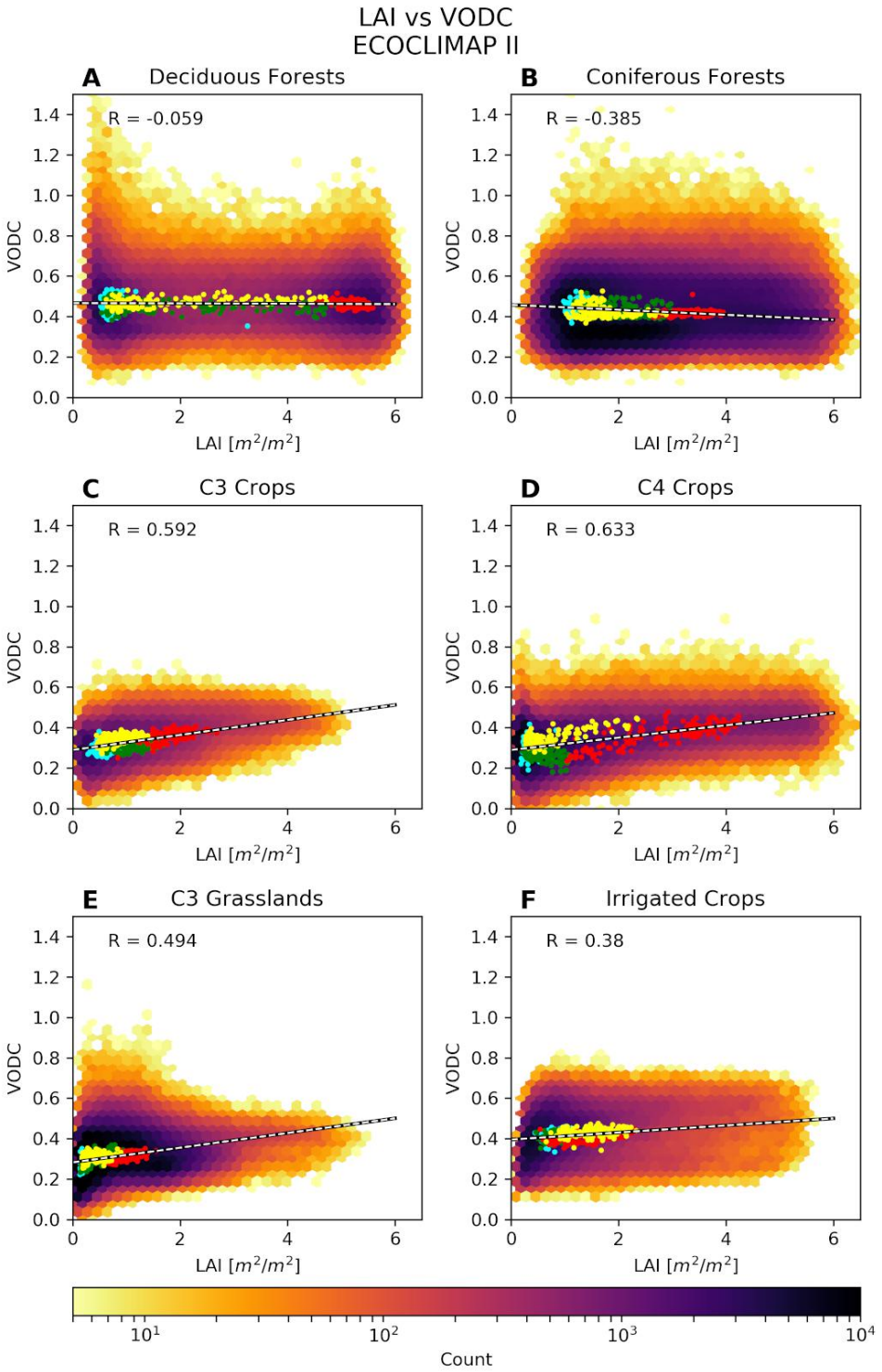

Figure B.1: A heatmap detailing the relationship between LAI and VODC from VODCA over six dominant vegetation types, Deciduous Forests, Coniferous Forests, C3 Crops, C4 Crops, C3 Grasslands, and Irrigated Crops. Dominant vegetation is defined as where 50% or more of the patch contains that single vegetation type. Black dots represent are where and when both observations of LAI and VODC are present. Colored dots represent the spatial average over the four seasons, where cyan is Winter, green is Spring, red is Summer, and yellow is Autumn. Purple dashed lines represent the linear regression of the seasonal dots.

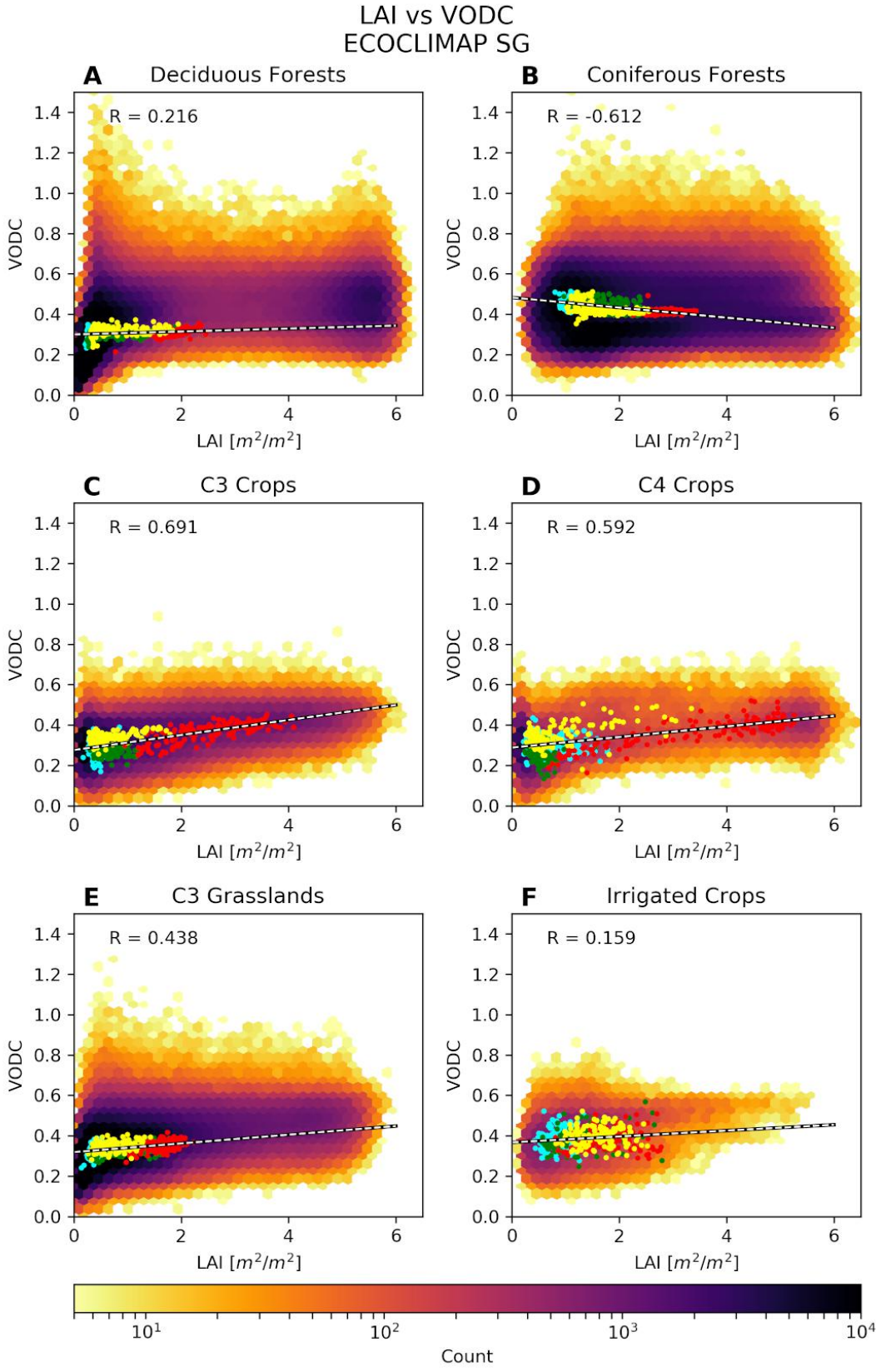

Figure B.2: Same as Figure B.1, but using ECOCLIMAP-SG instead of ECOCLIMAP-II

B.2 LAI versus Matched VODC

After linear re-scaling, the correlations and density scatter plots strongly resemble the LAI-matched VODX results in Chapter 4. The hysteresis pattern seen primarily in the C3 and C4 crops is still present when looking at matched VODC instead of matched VODX, indicating that this phenomenon is not VOD band dependent.

Table B.3: Seasonal and total correlations of Matched VODC versus LAI for ECOCLIMAP-II patches

	Matched VODC ECOII				
Veg Type	All Seasons (575)	Winter (144)	Spring (143)	Summer (144)	Autumn (144)
All Vegetation	0.92	0.79	0.89	0.73	0.90
Deciduous	0.97	0.15	0.93	0.25	0.96
Coniferous	0.96	0.18	0.73	0.67	0.90
C3 Crops	0.93	0.43	0.83	0.60	0.91
C4 Crops	0.94	0.09	0.89	0.89	0.93
C3 Grasslands	0.95	0.56	0.89	0.83	0.92
Irrigated Crops	0.93	0.66	0.60	0.89	0.87

Table B.4: Seasonal and total correlations of Matched VODC versus LAI for ECOCLIMAP-SG patches

	Matched VODC ECOSG				
Veg Type	All Seasons (575)	Winter (144)	Spring (143)	Summer (144)	Autumn (144)
All Vegetation	0.95	0.17	0.91	0.33	0.94
Deciduous	0.95	0.66	0.89	0.23	0.95
Coniferous	0.97	0.27	0.82	0.70	0.93
C3 Crops	0.95	0.25	0.74	0.83	0.95
C4 Crops	0.90	0.36	0.46	0.89	0.86
C3 Grasslands	0.96	0.36	0.91	0.70	0.93
Irrigated Crops	0.81	0.58	0.71	0.58	0.61

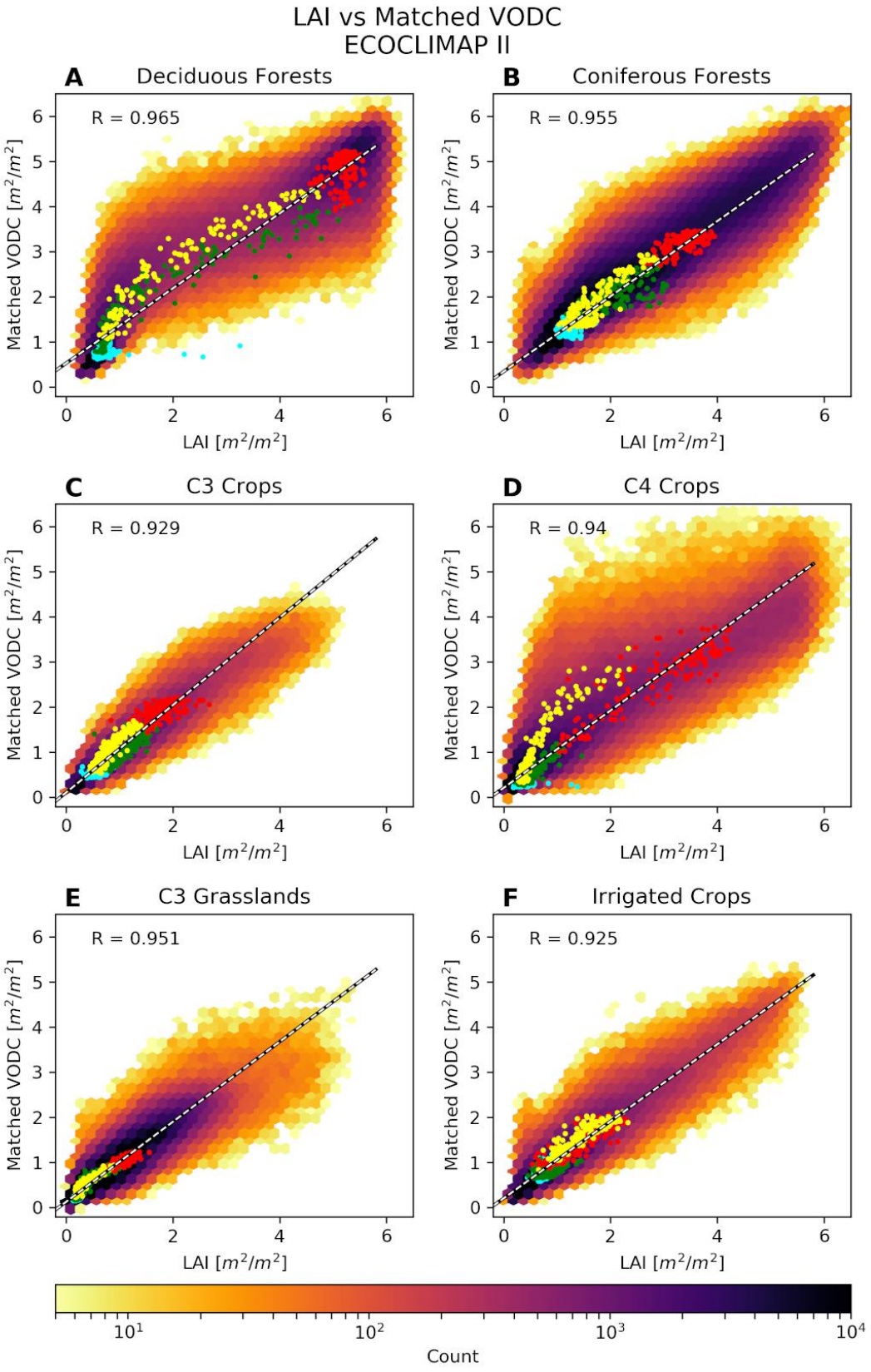

Figure B.3: Same as Figure B.1, but using Matched VOD instead of raw.

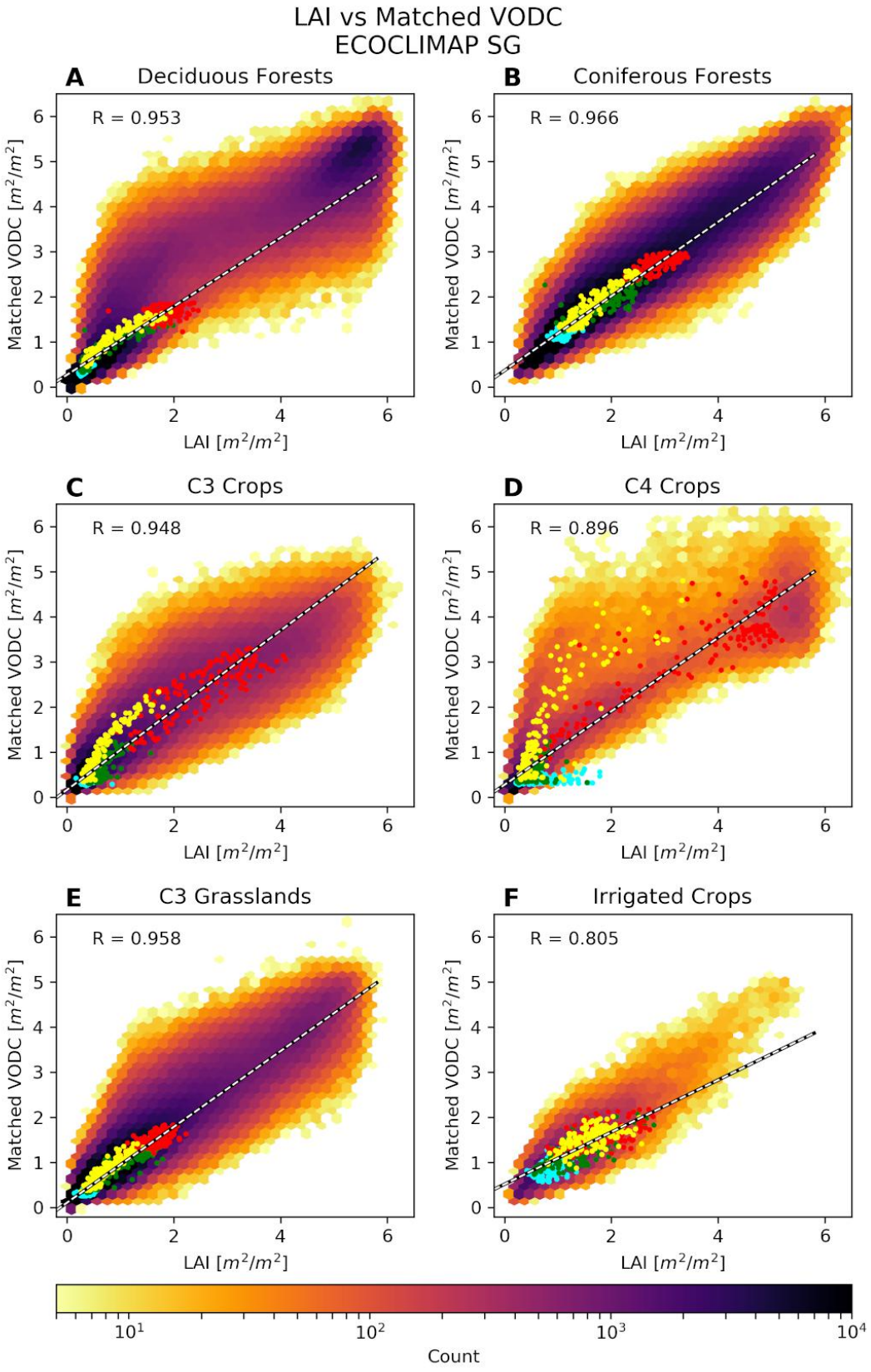

Figure B.4: Same as Figure B.2, but using Matched VOD instead of unmatched.

Appendix C

Appendix C

C.1 VOD vs LAI Comparison over Subdomains

This appendix plots the relationship between LAI and VOD over the domains shown in Chapter 2 and Figure 2.10. For each domain, there are two figures, each with two graphs. The first figure presents LAI versus VODC, while the second figure is LAI versus VODX. Panel A of each figure is LAI from observations in the CGLS dataset plotted against VOD from the VODCA dataset. Panel B of each figure is LAI from the ISBA LSM (with no data assimilation), using ECOCLIMAP-SG land surface parameterization plotted against VOD from VODCA.

California

Over California (or more specifically, the California domain from Figure 2.10), there is far weaker a relationship between VOD and LAI observations from CGLS in (Figures C.1 and C.2). For both VODC and VODX, the strength of the regression is low (correlations of 0.18 for both), primarily caused by a concentration of low value LAI observations with more variable VOD observations seen on the bottom left of Panel A for both Figures C.1 and C.2. While the rest of the observations at higher LAI and VOD values follow a good slope, they are heavily outweighed by the far higher concentrations at low values.

Compared to LAI from LDAS-Monde, however, a stronger relationship is shown, with lower variation at lower VOD and LAI values. Correlations are 0.54 for VODC-LAI ISBA and 0.56 for VODX-LAI ISBA. While the variation at lower LAI values is lower, there is increased VOD variation at LAI values higher than 1 m^2/m^2.

One potential reason for a strong disagreement between LAI observations and LAI from ISBA is that this California domain contains a large sampling of rocky and desert regions. These areas would have very low LAI values throughout the year, but the VOD values would be significantly affected by any precipitation.

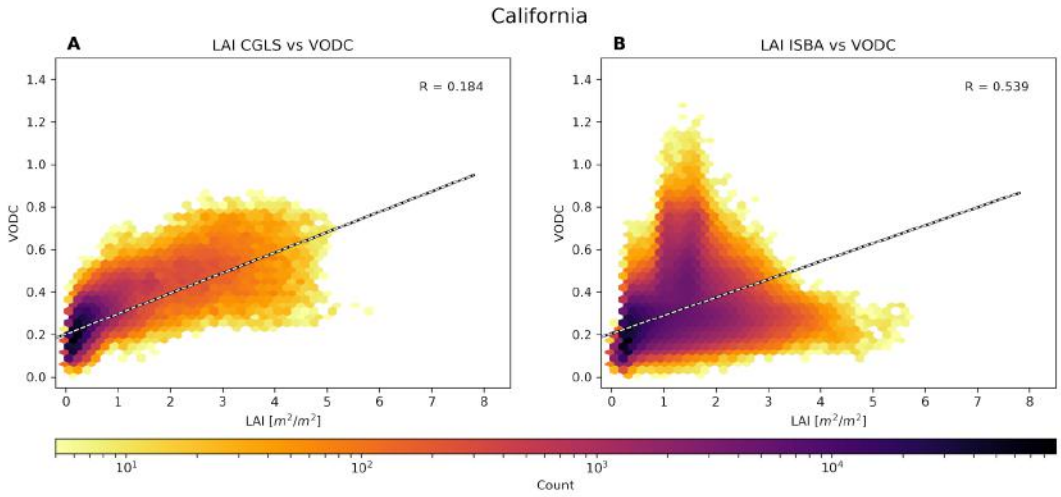

Figure C.1: Same as Figure 4.1, but for the California domain.

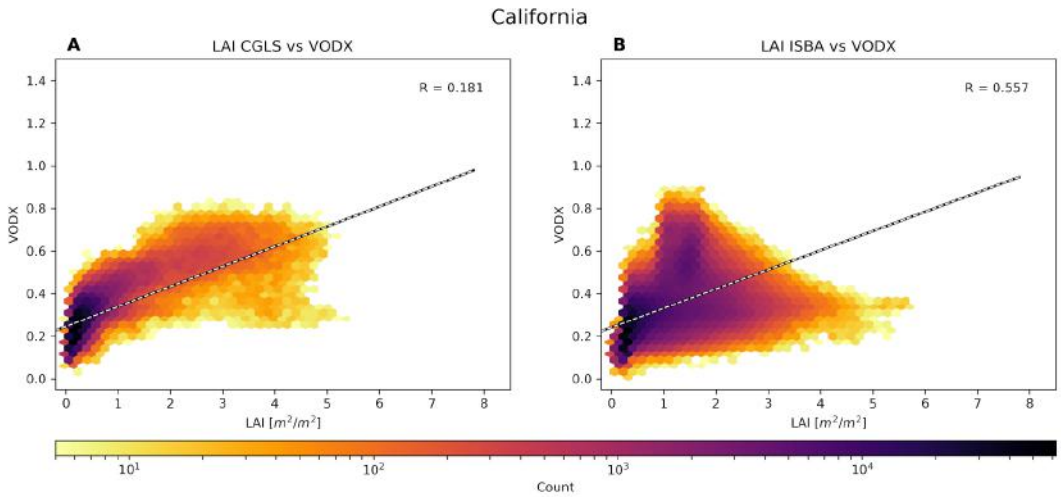

Figure C.2: Same as Figure 4.2, but for the California domain.

Midwest

The Midwest domain provides a relatively strong relationship between VOD and LAI from both observations and ISBA, as given in Figure C.3 and Figure C.4. This Midwest region is comprised of a large amount of agricultural land, mixed between C3 and C4 crops, as well as containing a fair amount of grasslands. The northern portions also contain a sizable number of deciduous and coniferous trees. Correlation scores between VOD and observed LAI are far stronger than that of California at 0.61 for VODC-LAI obs and 0.72 for VODX-LAI obs. Much like CONUS, as observed LAI increases over the Midwest, the VOD variation seems to flatten out. However, the density differences seen in the colors of the figures show that the VOD-LAI response is still in fact increasing on average.

The correlations between ISBA LAI and VODC and VODX are also quite strong (0.70 and 0.76 respectively). The difference between VODC and VODX is stark in this region, with the noise of higher variation very visible in VODC.

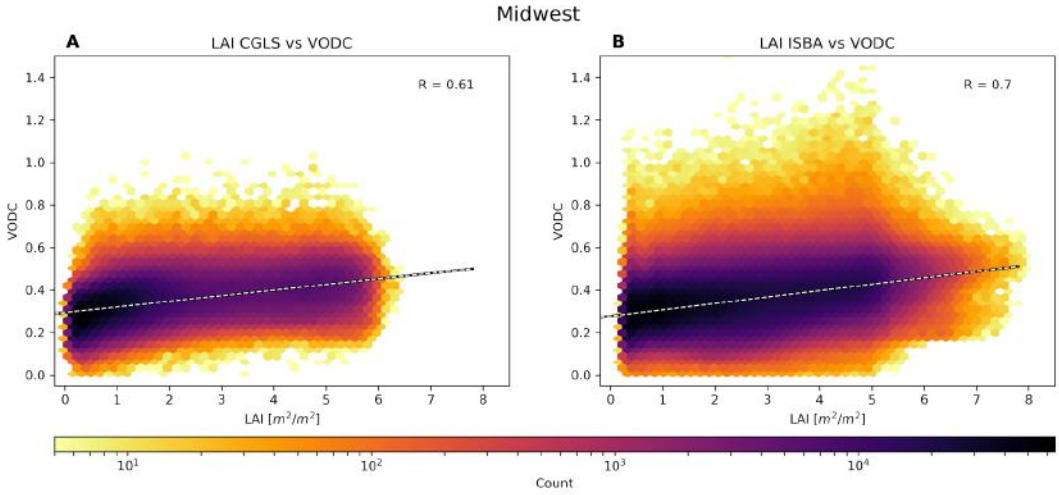

Figure C.3: Same as Figure 4.1, but for the Midwest domain.

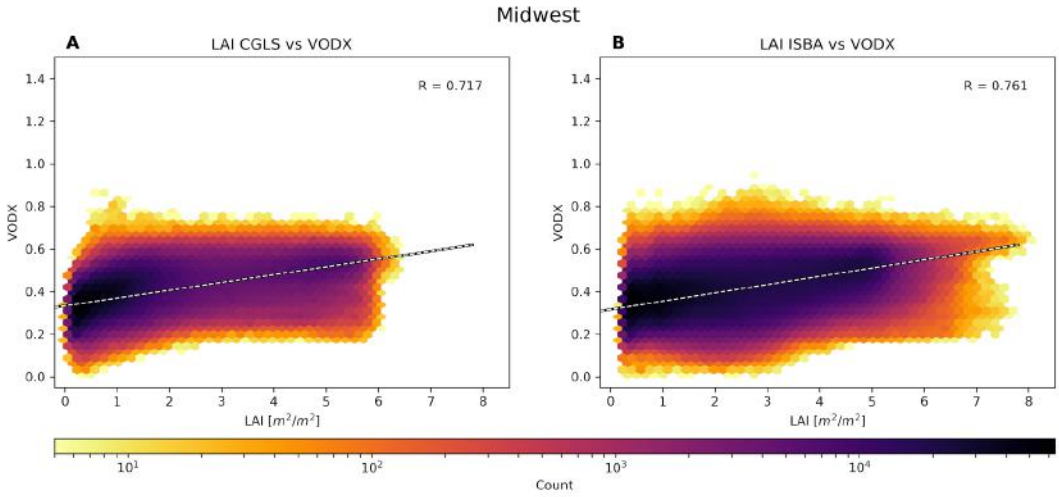

Figure C.4: Same as Figure 4.2, but for the Midwest domain.

Northeast

The Northeast region demonstrates a far weaker VOD-LAI relationship in Figure C.5 and Figure C.6, with very little change in VOD even as LAI values grow quite large. This response is in some respects the opposite seen in Figures C.1 and C.2 in California when compared to observed LAI from CGLS. In California, the LAI was concentrated at low values, while the VOD was variable. Here in the Northeast, the LAI is variable while the VOD concentration is more or less flat, with the highest concentration of VOD-LAI at high LAI values, in

contrast to all the other domains. This can be due to the Northeast region's high concentration of forests. Even with variable LAI values, the correlation between observed LAI and VODC is just barely positive at 0.01, and VODX strongly improves that score at 0.41 due to a more consistent distribution of high VODX values. The heavy forests also may be the cause of the poor correlations between observations. Saatchi et al. (2011) demonstrates that L-band satellite radar estimations of above ground biomass (AGB) are strongly impacted by forest structure, and Mialon et al. (2020) shows poor correlations between L-band VOD and estimated AGB over heavily forested areas of the Northern hemisphere. The higher energy wavelengths, such as the C and X-band used here, penetrate less deeply into the forest canopy, and that lack of penetration may reduce some variability, but also hides the information of the underlying vegetation. However, this relatively strong mismatch seems to fall in line with previous studies on the subject.

Correlations to ISBA LAI increase the correlations of VODC (0.33) and VODX (0.67). Again, the noise seen in the VODC figure is strongly reduced in the VODX figure, while VODX still has higher average values. The very visible artifact of the minimum 1 m^2/m^2 threshold for evergreen forests is seen in Panel B of both figures, indicating that evergreen forests are strongly represented in this region from ECOCLIMAP-SG. As the concentration of LAI values is far higher in the observations from CGLS, this suggests that ECOCLIMAP may be mischaracterizing the land surface in this region.

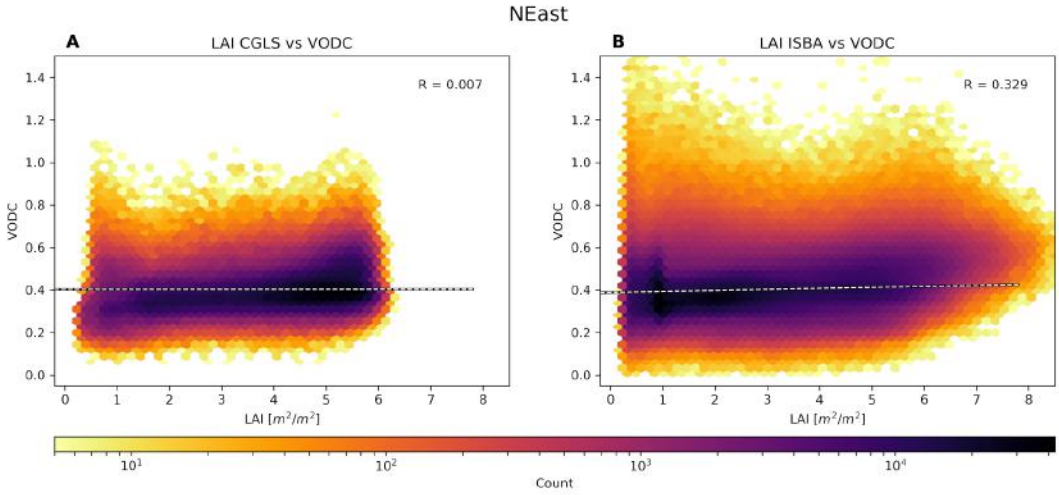

Figure C.5: Same as Figure 4.1, but for the Northeast domain.

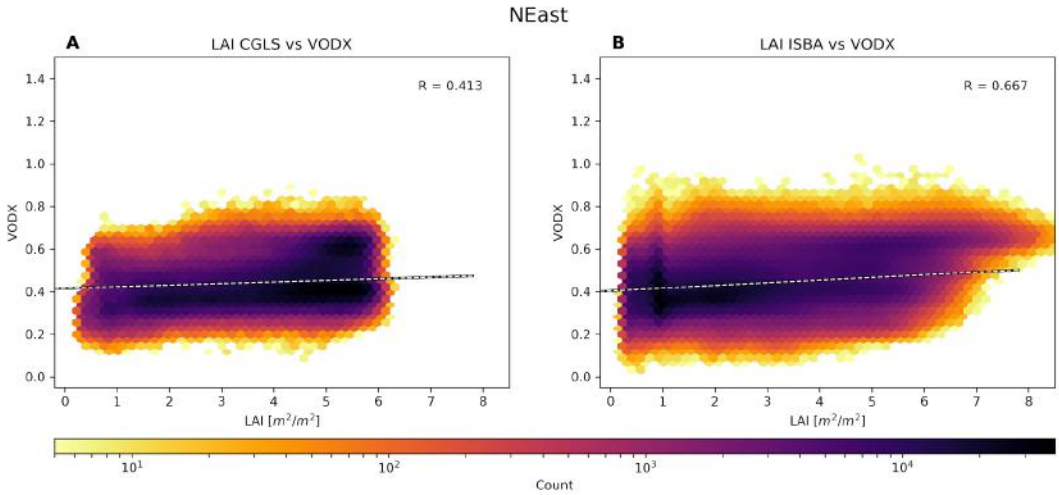

Figure C.6: Same as Figure 4.2, but for the Northeast domain.

Southern Plains

The Southern Plains region, with VOD-LAI responses given in Figure C.7 and Figure C.8, shows weak to moderate strength between observed LAI and VOD. This Southern Plains region contains a diverse range of vegetation, including a dry, rocky west and transitioning to a warm, subtropical zone in the east and on the Gulf of Mexico. The observed LAI-VODC correlation is 0.35, while the observed LAI-VODX correlation is 0.317. As with other regions, the high concentration of low LAI values is the driving force of the relationship.

Modeled LAI from ISBA is better correlated to VODC (0.60) and VODX (0.59). These graphs are characterized by significant densities of LAI values at the minimum 0.3 m^2/m^2 threshold, which falls in line with western portions of this domain marked by dry and low-vegetation conditions. The large range of

LAI also seems in accordance with what we know of the domain, as the eastern and coastal regions have far larger LAI values.

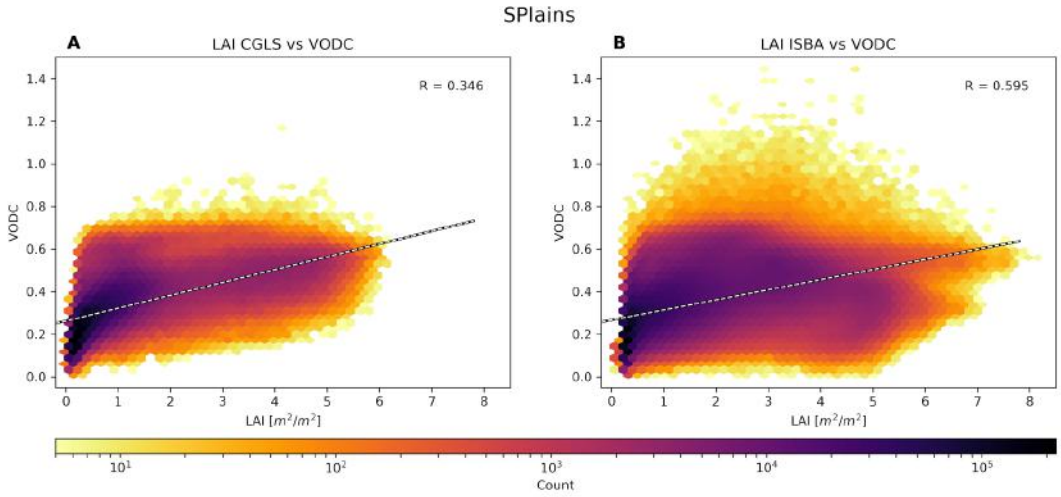

Figure C.7: Same as Figure 4.1, but for the Southern Plains domain.

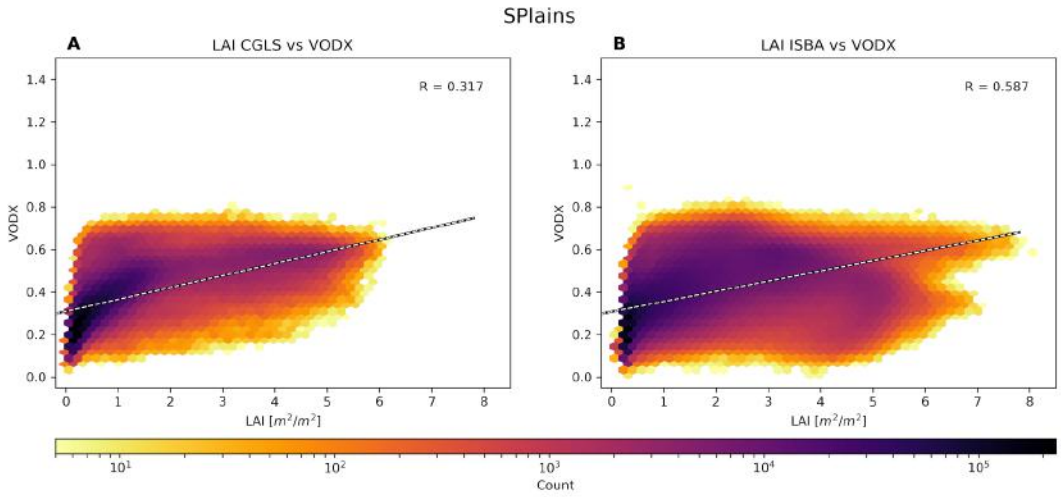

Figure C.8: Same as Figure 4.2, but for the Southern Plains domain.

Nebraska

The Nebraska domain shows a moderate strength correlation between LAI obs from CGLS and VODC (Figure C.9) at 0.52 and a higher correlation between LAI obs and VODX (Figure C.10) at 0.64. LAI values are generally low, less than 2 m^2/m^2, and VOD stays generally flat, or even slightly declines at higher LAI values. During this growing season, corn, soy, and grasslands are widespread, which generally agrees with the low LAI observation values.

When compared with LAI from ISBA, the regression scores remain close to those with the LAI observations. We do see, however, a larger range of VOD values at higher LAIs, and a generally higher amount of larger LAI values. This indicates that ISBA by itself is overestimating LAI with respect to the observations.

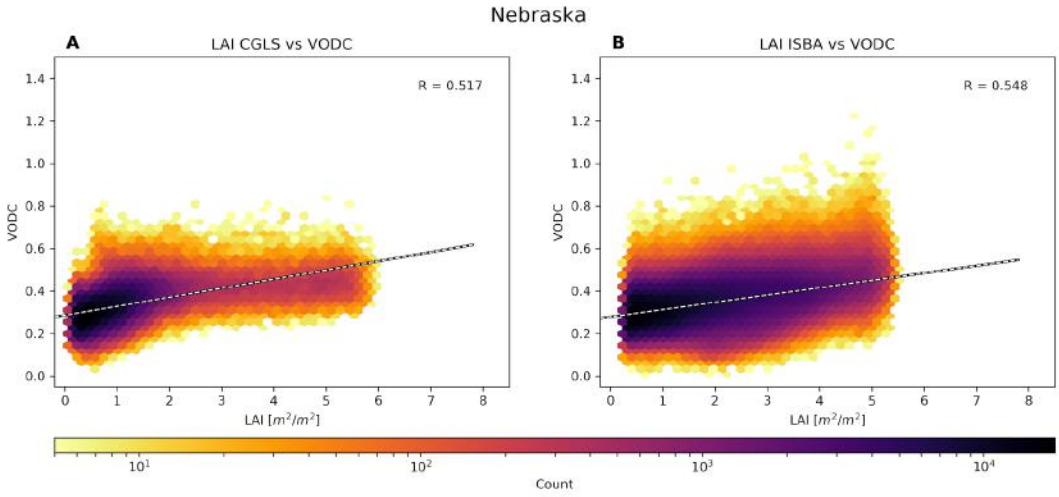

Figure C.9: Same as Figure 4.1, but for the Nebraska domain.

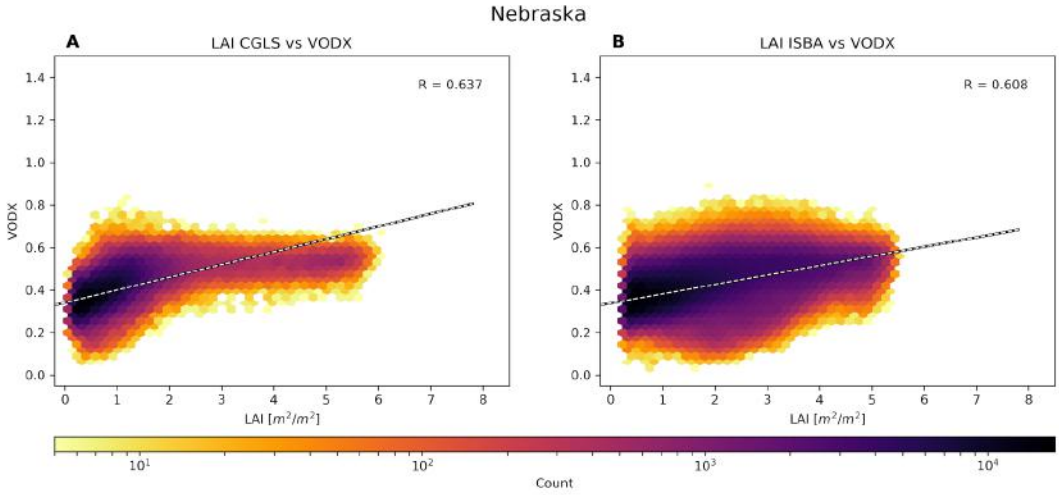

Figure C.10: Same as Figure 4.2, but for the Nebraska domain.

Appendix D

Appendix D

This appendix analyses the results of the experiments shown in Table 4.6 over the domains from Chapter 2 and Figure 2.10. The same analysis that is applied to CONUS is also applied to these subdomains.

D.1 Impact of Assimilating VOD as an LAI proxy - Subdomains

California

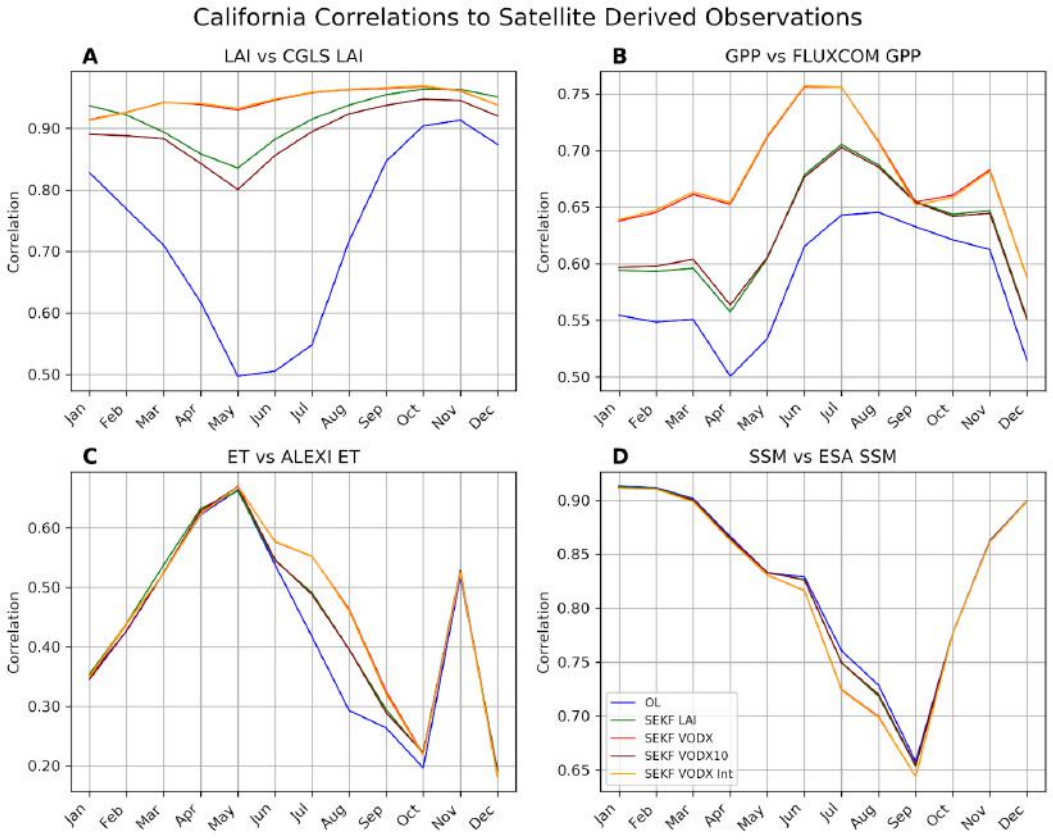

Figure D.1: Same as Figure 4.8, but for California

Figure D.1 shows the same monthly correlations for A) LAI, B) GPP, C) ET, and D) SSM, as previously seen for CONUS.

LAI: The month by month correlations of LAI over California show some similar trends as they were over the whole CONUS domain, but also some additional information can be gleaned. For example, during the months of April to August, the OL has far weaker scores than CONUS, and a similar, but far less pronounced impact is seen with SEKF LAI and SEKF VODX10 in May. This may have a similar explanation as seen with GPP over CONUS, where these months of rapid vegetation change are not well constrained by the model, and not optimally improved by data assimilation at a 10 day frequency. However, the California domain contains far less vegetation with seasonal extremes, and in fact contains a fair amount of desert. Still, for the months of March through September, SEKF VODX and SEKF VODX_Int perform better than SEKF LAI and SEKF VODX10, with correlations consistently above 0.9. As with CONUS, the SEKF VODX10 is consistently, slightly lower in score than SEKF LAI, meaning the added frequency of SEKF VODX is what is driving the improvement of these scores.

GPP: Scores compared to GPP estimations show the same pattern, that there is added value of SEKF VODX over SEKF LAI. For all months except September, the SEKF VODX (and SEKF VODX_Int) outperform all other experiments by a visible margin. A positive note for all experiments assimilating data is that they consistently outperform the OL against this independent dataset. Adding to the pile of evidence that show the more frequent assimilation of VODX better constrains these vegetation variables is that SEKF VODX10 and SEKF LAI are almost identical throughout the period, where the assimilation frequencies are exactly the same.

ET: Contrary to the entire CONUS domain, ET scores do show some more visible differences over California. From the months of June to September, SEKF VODX shows consistent improvement over SEKF LAI, which in turn is an improvement over the OL. Like the other variables, SEKF VODX10 being almost identical to SEKF LAI shows the frequency is the major factor improving the results. One oddity seen in this Panel C is that the month of November sees a strong increase in correlations compared to October and December for all experiments. While unlikely that this is simply a processing error, some more investigation is needed to explain this nearly sudden 0.3 jump not seen in other domains.

SSM: Similarly to CONUS, SSM correlations are only distinguishable from June to September, with SEKF VODX performing the worst, and the OL having the best scores. As with CONUS, the potential explanation is that all the control variables are adjusted by the assimilation, and assimilating values as LAI strongly affects deeper layers of soil moisture. These changes at deeper layers

come at the price of affecting shallow layers, which are not directly constrained by assimilation of SSM in these experiments.

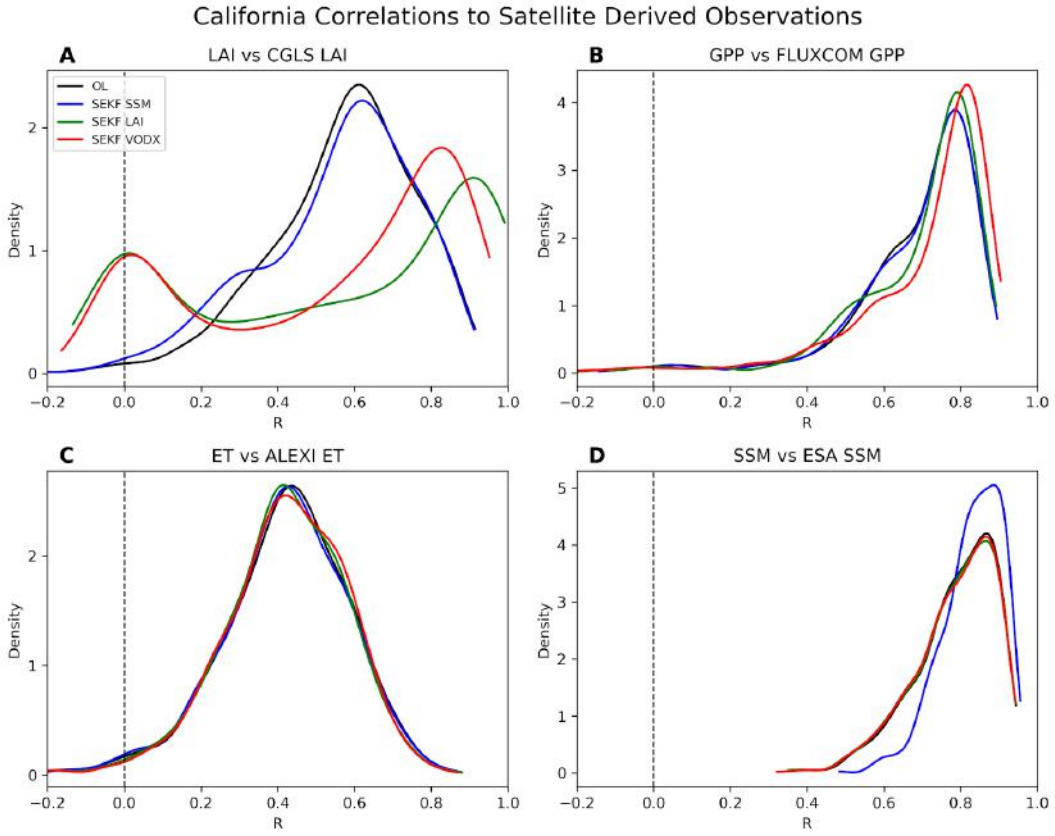

Figure D.2: Same as Figure 4.9, but for California

Figure D.2 shows the distributions of correlations over the California domain for A) LAI, B) GPP, C) ET, and D) SSM and analyzes the OL, SEKF SSM, SEKF LAI, and SEKF VODX experiments.

LAI: The small bump in concentrations of correlations near zero as seen at the CONUS scale has increased greatly when looking at California. Both SEKF LAI and SEKF VODX show a very dramatic increase at and around 0.0 that is not seen or even hinted at with the OL and SEKF SSM. Figure D.3, which shows the LAI correlation differences over the whole CONUS domain between OL and A) SEKF SSM, B) SEKF LAI, and C) SEKF VODX, clearly indicate that very dry zone in the Southwest US, comprised primarily of southeast California, south Nevada, and west Arizona has strongly diminished LAI correlations compared to the OL. In Panel A, the white gridcells are in fact exactly zero correlation difference. This is due to the filtering of points over 1500m on the ESA SSM. For those cells, there is no assimilation and thus no difference between it and the OL. This means that something about the assimilation of matched VODX and/or LAI observations is not meshing well within the model, and causes spurious

poor correlations specifically over this region characterized by habitually dry conditions. As these are compared against the very observations that were assimilated in SEKF LAI and that were used to match SEKF VODX, it indicates that the assimilation in the model does not correctly advance this LAI variable in these conditions, and is in fact pushing LAI values away from the values of the observations. More investigation is warranted to find the more precise cause of this mismatch.

While this region of strong degradation is evident in the PDFs by the large bump at and around zero, we also see more higher scores from SEKF LAI and SEKF VODX past 0.8. SEKF LAI does outperform SEKF VODX at these higher correlations. These very localized spikes of degradation (or improvement) are the reason why all the same analyses are done over all the selected subdomains.

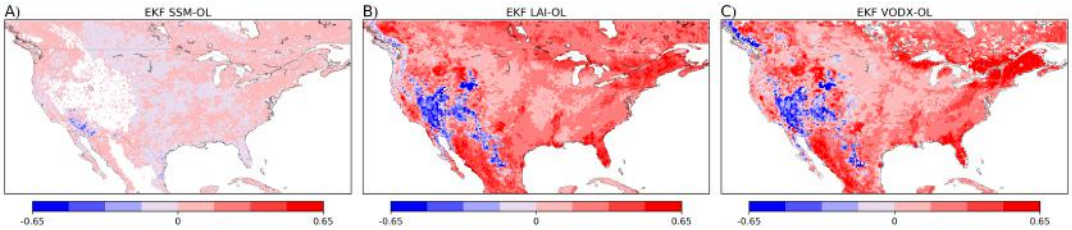

Figure D.3: Maps of differences in correlation between the OL and A) SEKF SSM, B) SEKF LAI, and C) SEKF VODX.

GPP: Moving on to Panel B, the distribution of scores does trend higher with the assimilation of SEKF VODX and SEKF LAI. In this case, starting at 0.4, SEKF VODX consistently outperforms SEKF LAI, as well as the OL and SEKF SSM. These distributions again agree with the hypothesis that assimilation of VODX improves the performance of model vegetation compared to independent datasets.

ET: The distributions of scores compared to ALEXI ET in Panel C are all very similar for California. At their peak densities, SEKF LAI and SEKF VODX are slightly lower in correlation than the OL and SEKF, but as the correlations trend higher, the reverse occurs. All of these differences are very small, and it is unlikely that these small changes show any significance.

SSM: Finally, coming to Panel D and the distribution of correlations compared to SSM observations. As expected, the assimilation of the very observations used in the correlations, yields the best scores, with SEKF SSM strongly outperforming all other experiments. The distributions of the OL, SEKF LAI, and SEKF VODX are all too close to glean any useful information from over this domain.

Midwest

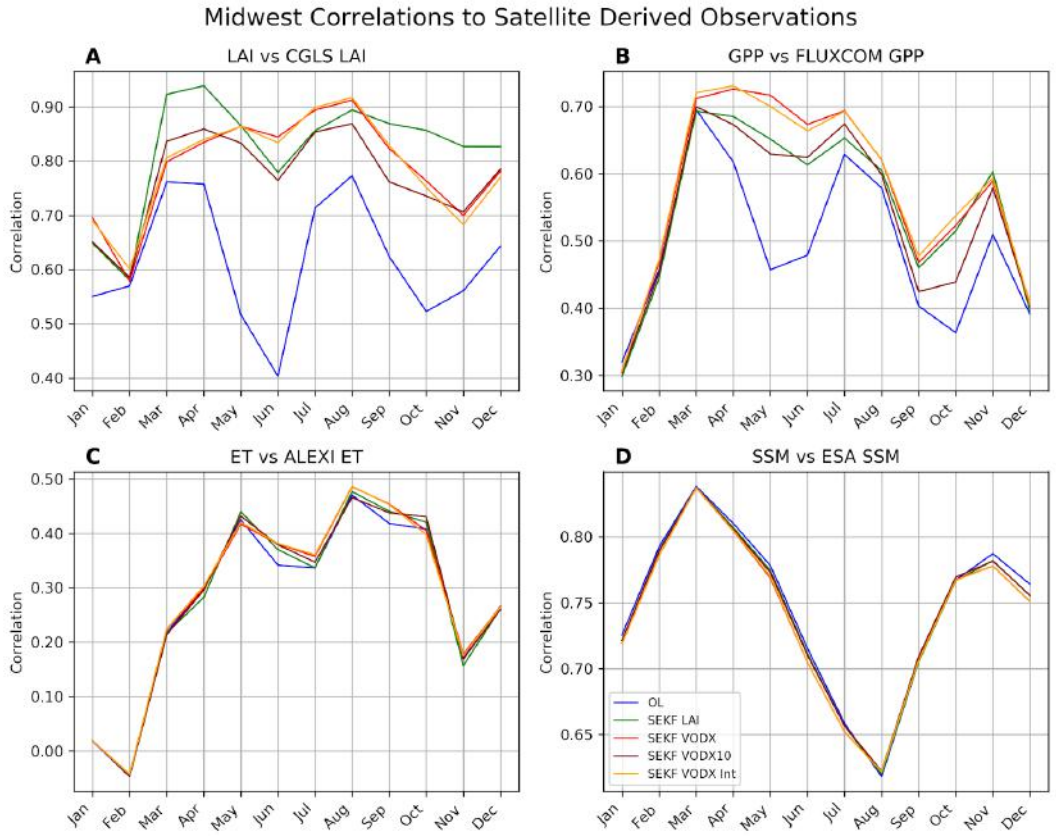

Figure D.4: Same as Figure 4.8, but for the Midwest

The analysis of monthly LAI, GPP, ET, and SSM correlations over the Midwest region of the United States is shown in Figure D.4.

LAI: While all assimilation scenarios provide better correlations than the OL at all months, the Midwest region provides the closest scores between SEKF LAI, SEKF VODX, SEKF VODX10, and SEKF VODX_Int. The months of March and April, as well as September to December, show SEKF LAI to be significantly higher than SEKF VODX, while June through August show the assimilation of VODX to provide moderately improved correlations to the LAI observations. Curiously, March and April are also the first indication of monthly scores where SEKF VODX10 is superior to SEKF VODX. While this improvement is only moderate, and coupled with the fact that SEKF LAI has the highest score in these months, it may indicate that the VOD that is assimilated was not sufficiently matched to the LAI. And the more frequent assimilation with SEKF VODX actually tended the model away from the observed LAI values. The Midwest region is heavily populated with C3 and C4 crops. It is potentially these C4 crops that could be a root cause, as looking at Figure 4.7 shows Winter months in blue falling heavily outside the overall regression. Spring months

in green area also weakly correlated. This matches with the March and April timing of this observed inversion caused by the worsening of correlations by the more frequent assimilation of VODX. Additionally, a dip in all scores is seen in the month of June, which was also seen over CONUS.

GPP: The monthly progression of scores against GPP estimates proceeds like in other domains. SEKF VODX and SEKF VODX_Int provide the highest correlation scores, improving the most in the Spring and early Summer months. All assimilation experiments are consistently outperforming the OL, with the largest difference seen in May and June when the OL scores dip. This may again be related to the fast evolution of vegetation during these months, which are not well captured in the model physics, but that the more frequent assimilation of VODX does improve.

ET: The evolution of monthly ET correlations are very similar, as seen over CONUS. During June and September, it is the most clear that the OL is the least correlated, but no clear and significant pattern of which assimilation experiment performs the best is seen. It is also noted that these scores are often very poor, with February correlations even dropping into the negatives for all experiments. However, Spring and Summer months improve these values to moderate correlations.

SSM: Likewise with ET, SSM correlations are nearly identical over the year. The months of November and December show that the OL slightly outperforms the other experiments.

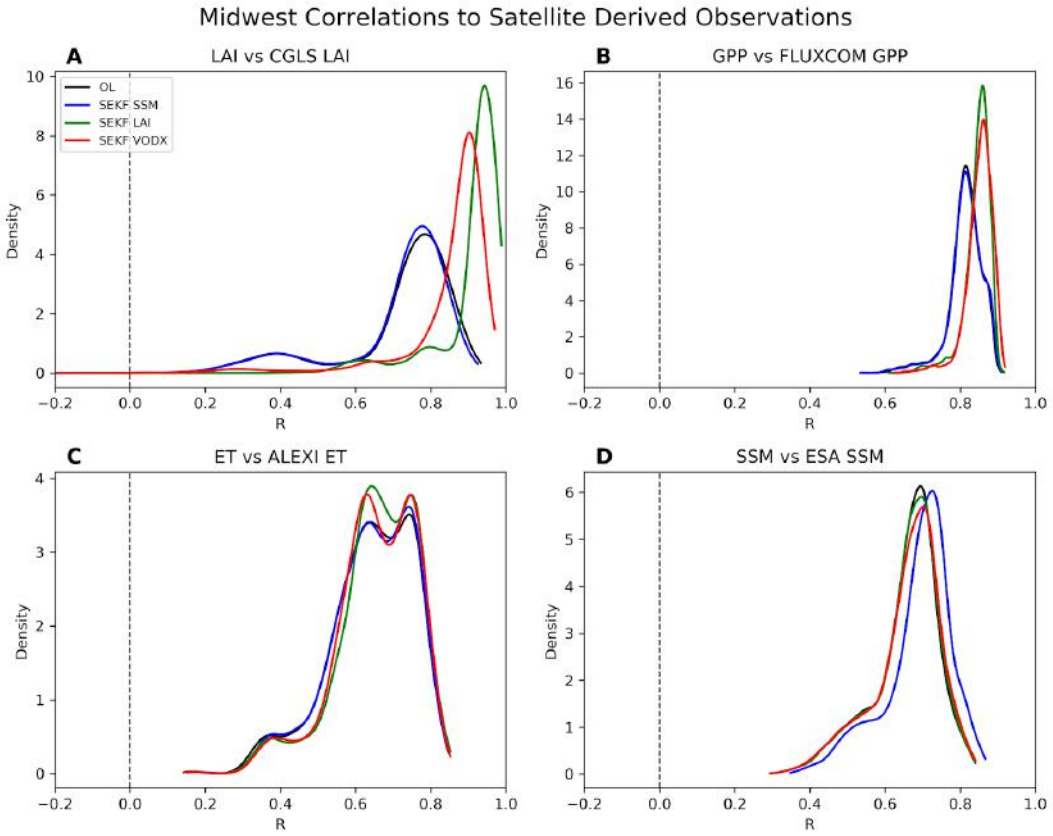

Figure D.5: Same as Figure 4.9, but for the Midwest

Figure D.5 shows, as in CONUS and California domains, the distribution of correlation scores compared to our satellite derived LSVs.

LAI: Starting with Panel A, LAI, the SEKF LAI and SEKF VODX experiments show a striking, consistent, and strong improvement over the domain. Unlike California, there is no bump at all around 0, which agrees with the theory that the model vegetation dynamics of specific dry southwest region of the California, Arizona, and Nevada border is insufficiently represented. The OL and SEKF SSM distributions show a semi bimodal pattern, with a surge around 0.4, and the main peak around 0.75. SEKF LAI and SEKF VODX show some small bumps in correlation density around 0.6, and for SEKF LAI, 0.8. The primary peak density of SEKF VODX is slightly below and at lower correlations than SEKF LAI, agreeing with previous domains that SEKF LAI compared to the LAI observations that it used in assimilation outperforms the additionally processed and assimilated VODX. Yet the SEKF VODX still shows massive improvement compared to no assimilation.

GPP: As with LAI, GPP also shows a distinct improvement from the assimilation of VODX or LAI, with a very concentrated number of gridcells having

high correlations. While the OL and SEKF SSM, which are virtually indistinguishable and also give a strong performance, the SEKF LAI and SEKF VODX exclusively shift the correlations to higher values. In this case, SEKF VODX slightly outperforms SEKF LAI, by having a small amount more pixels at the highest correlations.

ET: The distributions of ET correlations are also improved following VODX or LAI assimilation. The number of correlations in the range of 0.4 to 0.5 are lowered, but scores from 0.55 to 0.65 are generally increased. From 0.6 to 0.7, it is also seen that SEKF LAI has higher concentrations of scores compared to SEKF VODX, which are not made up at higher values. Therefore it seems that for the Midwest region, while both VODX and LAI assimilation improve the representation of ET, LAI is still a bit better.

SSM: Similarly to previous domains, the Midwest shows a strong, consistent shift to higher correlations following the assimilation of SEKF SSM. The dynamics of the distribution of SEKF VODX and SEKF LAI scores are far closer, but SEKF VODX shows an overall, albeit slight, increase in correlations over the domain.

Northeast

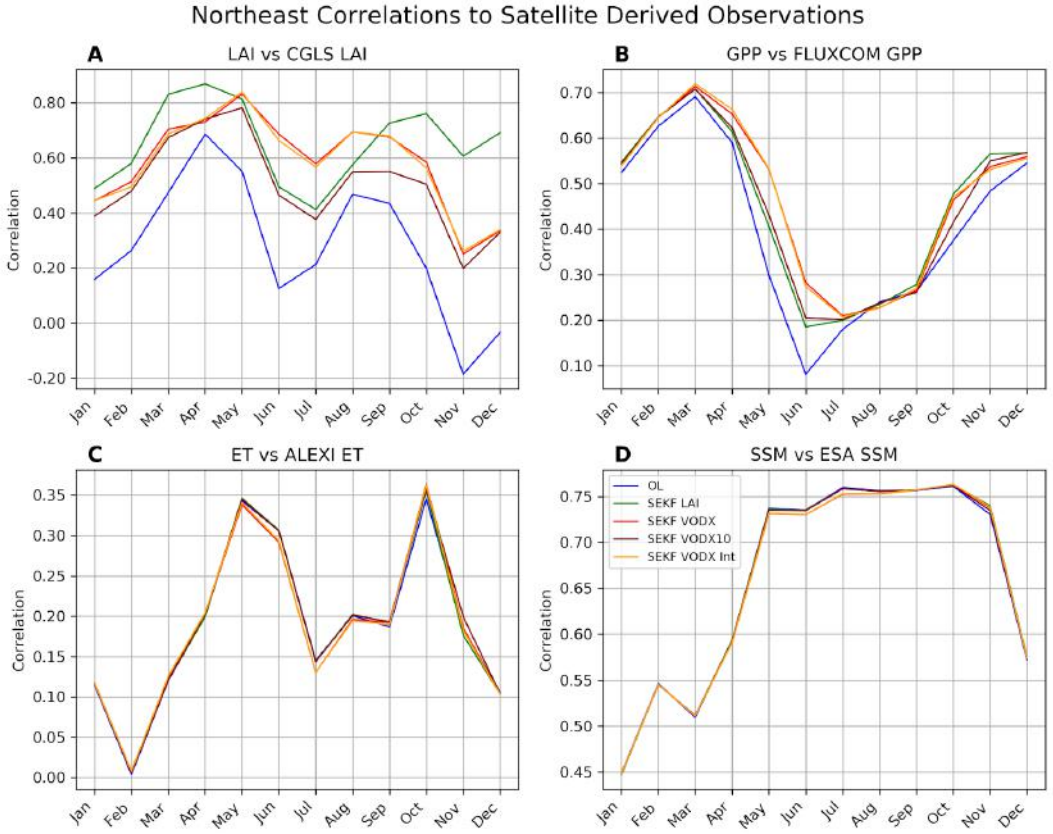

Figure D.6: Same as Figure 4.8, but for the Northeast

Figure D.6 illustrates the seasonal correlation over the Northeast domain as done for all the previous domains.

LAI: The trends seen in LAI correlations from month to month are similar to previous domains, with SEKF VODX outperforming SEKF LAI in the months of June through August, but with SEKF LAI having higher correlations the rest of the year. Again, similar to previous domains, any assimilation gives improvement compared to the OL. A different response seen in this Northeast region is the far larger variability of the monthly correlations. Here, correlations for the OL in November and December drop into the negatives, while most of the other experiments for the rest of the year range from 0.2 to 0.8. Other domains saw strong changes, but the minimum correlations were generally near 0.5 for LAI. This variation, along with the overall lower scores suggest that this domain is poorly represented in ECOCLIMAP or the model dynamics for these vegetation types poorly represented in ISBA. Or both. These scores just suggest a potential problem, but more in depth research is needed to fully assess the issue and explore possible solutions.

GPP: GPP over the Northeast, in Panel B, demonstrates the smallest overall change due to data assimilation that we have seen in any domain yet. Still, for the months of April to June, the pattern of SEKF VODX and SEKF VODX_Iint outperforming SEKF LAI and SEKF VODX10 is consistent, and the latter two equally outperform the OL. During the months of January to March and October to December, all the assimilation scenarios are relatively close in score, and all consistently outperform the OL. The Summer months where SEKF VODX was clearly the superior scenario, add even more evidence that VODX assimilation improves the representation of vegetation. As with LAI over the Northeast, GPP correlations are far more variable compared to other domains, with OL values dropping as low as 0.1 and assimilation experiments dropping to around 0.2 in June. This backs the general hypothesis made in the LAI section, that the vegetation in this region is systematically poorly represented. However, the seasonal timing of these poor correlations are not matched between the two variables, with LAI dropping to the lowest values in Winter, and GPP in Summer. Still, the far greater variability in scores of this region warrants more study.

ET: Correlation scores of ET in the Northeast do not indicate any consistent differences between experiments. While generally very low, and even dropping to near zero in February, all the scenarios are well within each other's potential error.

SSM: Similarly to ET, correlations to SSM observations over the Northeast show all the experiments being indistinguishable. However, the overall shape of the monthly correlations is nearly the inverse of that produced over CONUS, California, the Midwest, and the forthcoming correlations over the Southern Plains. While, in general, the correlations drop from a peak in February or March to a trough in August, the Northeast has steady and high correlations from May to October. This difference is likely caused by the dominant vegetation, which over the Northeast is forests. It is possible that the heavier forest foliage contributes to reduce SSM variability compared to more open and lower vegetation density regions such as agricultural fields during Summer. This reduction in variability may lead to ISBA being able to more accurately predict SSM during these months.

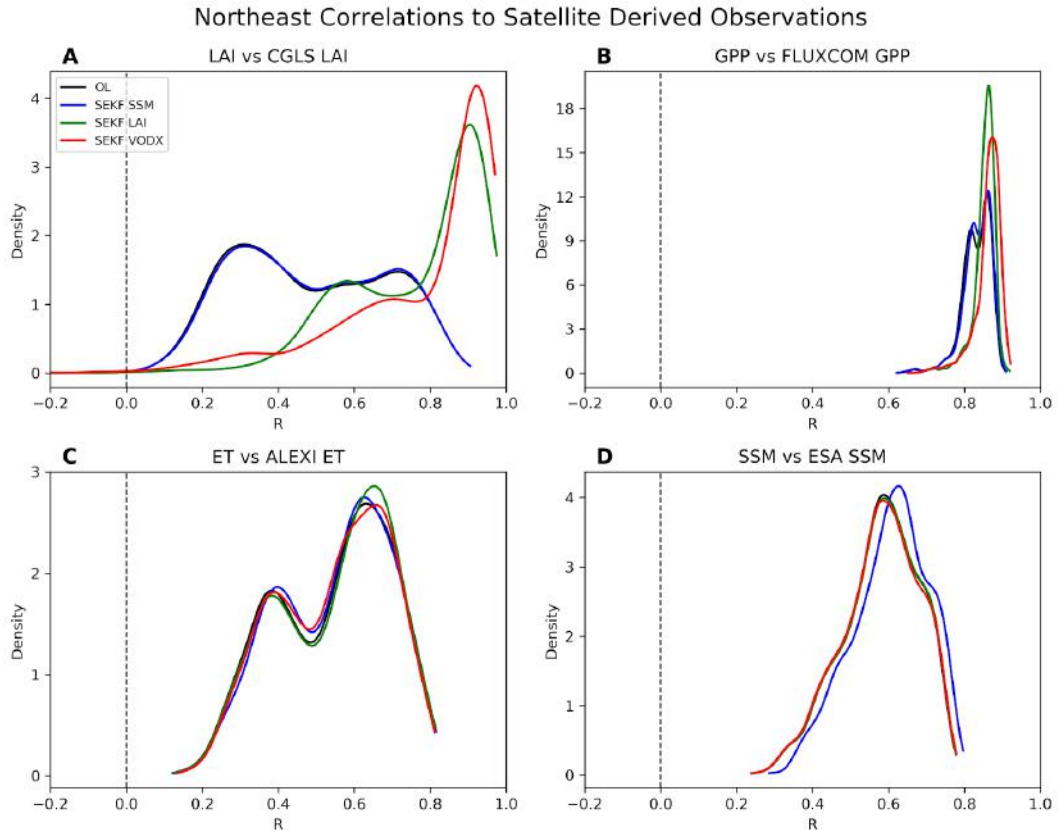

Figure D.7: Same as Figure 4.9, but for the Northeast

The distribution of individual gridcell correlations for the Northeast region are given in Figure D.7.

LAI: The OL and SEKF SSM experiments over the Northeast region has a visibly different distribution of LAI scores compared to CONUS or any subdomain. The distribution is close to bimodal, but the highest density of correlations is actually very low, near 0.3. While still overwhelmingly positive, this region is no doubt one of the weak links of the model representation of LAI, and as demonstrated by this figure, is greatly improved through data assimilation. SEKF LAI and SEKF VODX both strongly shift the distribution to higher correlations, but in this case the VODX assimilation actually slightly outperforms that of the LAI. Obviously it is still important to keep in mind that this comparison is more of a benchmark assessing the assimilation performance and not the improvement to vegetation monitoring.

GPP: And again, GPP observations being an independent measure, can be better used to gauge vegetation monitoring and any changes associated with data assimilation. In the case of the Northeast, the shift is strongly towards higher correlations by both SEKF LAI and SEKF VODX over that of the OL. Importantly, even though the assimilation does greatly improve the correlation

distribution to higher values, even the model by itself provides very consistent high correlations to GPP. This may seem in contrast to the monthly scores discussed above that demonstrated higher variation. But with this distribution we can surmise that it is the seasonality that is likely the cause. The scores in this distribution are calculated from the first day in the time frame to the last, whereas the monthly correlations do not capture the inter-seasonal and even inter-annual trends. These inter-annual trends are likely to increase the overall correlation score.

ET: The PDF of ET correlation scores in Panel C is clearly bimodal for all experiments. While there are some changes between the assimilation scenarios, it is hard to say which are definitively improving, if any. As the SEKF SSM and SEKF VODX show higher densities of correlations from 0.4 to 0.5, the SEKF LAI shows higher densities at 0.6. A more precise calculation would be required to demonstrate significant improvement or degradation.

SSM: As seen with every other domain, the distribution of SSM correlations is consistently improved by SEKF SSM. The OL, SEKF LAI, and SEKF VODX are all nearly indistinguishable from one another over this Northeast domain.

Southern Plains

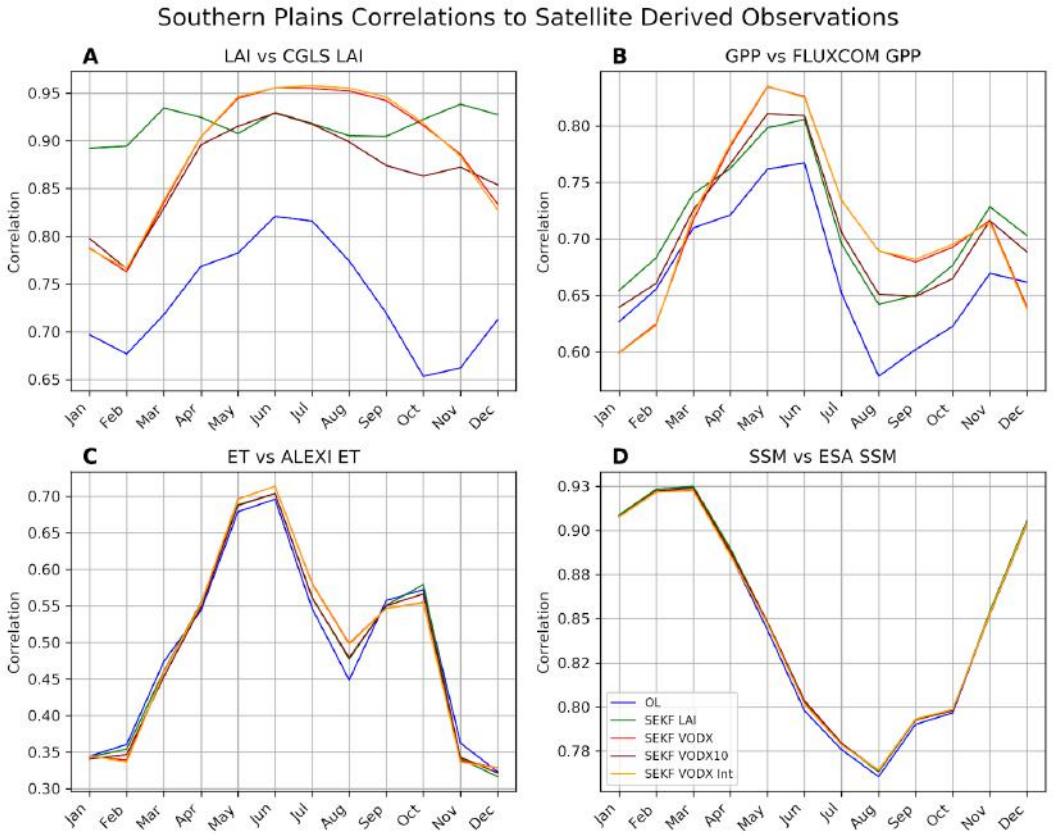

Figure D.8: Same as Figure 4.8, but for the Southern Plains

In Figure D.8, monthly correlations against the four LSVs of interest are displayed for the Southern Plains domain.

LAI: The Southern Plains regions shows much the same month to month LAI correlations as the overall CONUS domain. All assimilation scenarios strongly improve against the OL, and the months of May to September are characterized by the SEKF VODX and SEKF VODX_Int establishing the highest scores. As mentioned previously, this improvement in the Summer months may be directly linked to the faster evolution of vegetation being better captured with the more frequent VODX assimilation. In contrast to other domains, the months of January through March have a stronger difference between SEKF LAI and SEKF VODX.

GPP: Regarding GPP scores, the Southern Plains seem to be the only region that has several continuous months where the SEKF VODX performs the worst of all the experiments, including the OL. December through February all show lower correlations, with SEKF LAI having the highest during the same time. The differences are still relatively minor, around 0.05. This same time also sees

the SEKF VOD10 performing better than the more frequent assimilation of the same base data. This indicates that something about the matched VOD moving the analysis in the wrong direction, and it must be linked to the differences between LAI and matched VOD. This Southern Plains domain contains dry conditions in the western and southern extents and the evolution of vegetation over these regions in Winter may give clues as to why the matched VODX proves less adequate. More investigation into this specific dynamic would be required.

ET: With ET correlations in Panel C, there are some noticeable differences between experiments. Most notably, the SEKF VODX and SEKF VODX_Int show small, yet consistent improvements compared to SEKF LAI, SEKF VODX10, and the OL during the months of May to August. SEKF LAI and SEKF VODX10 also demonstrate slightly improved scores compared to the OL during this period, but still come short of the SEKF VODX.

SSM: And for the last variable, monthly SSM correlations are mostly intertwined, but it can be seen that the OL is very slightly outperformed by all the assimilation experiments in the months of May to October.

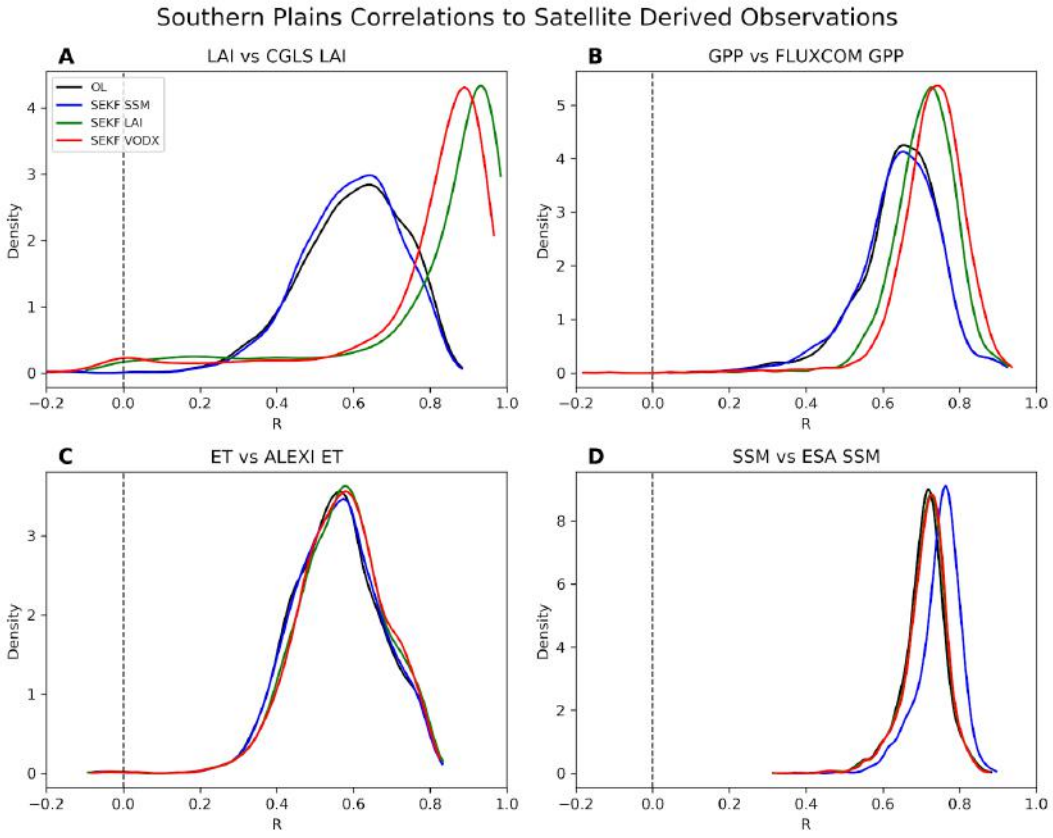

Figure D.9: Same as Figure 4.9, but for the Southern Plains

Lastly for this section, Figure D.9 provides the distribution of correlations for the LSVs over the Southern Plains.

LAI: The distribution of LAI correlations for the SEKF LAI and SEKF VODX experiments show a small similarity to that of California, seen in Figure D.2, with a noticeable increase of scores around zero. Looking again at Figure D.3, this discrepancy can be explained in the same way. LAI correlations in northern New Mexico as well as along the Texas-Mexico border, show strong degradation between the experiments assimilating LAI or VODX and the OL. This same degradation is not seen between SEKF SSM and the OL. The rest of this Southern Plains region sees a strongly improved correlation distribution after LAI or VODX assimilation, and in this case, the assimilation of LAI seems to slightly outperform the assimilation of VODX.

GPP: Panel B, regarding GPP scores, gives potentially the most straightforward example of simply improved scores from the assimilation of SEKF VODX. While the OL and SEKF SSM distribution peaks at around 0.65, SEKF LAI and SEKF VODX are shifted to about 0.75, but SEKF VODX demonstrates a remarkably consistent displacement towards higher correlations even over SEKF LAI. Along with all the other evidence, this panel successfully encapsulates the advantage to vegetation monitoring by assimilating VOD as an LAI proxy.

ET: While the distributions of ET scores are all very similar, one can clearly say that SEKF LAI and SEKF VODX do outperform that of the OL and SEKF SSM. The differences are relatively small, yet consistent from the buildup to higher density, as well as the decline, as those two experiments steadily shift the distributions of correlations slightly higher.

SSM: Much the same as the other domains, SSM correlation distributions are heavily and positively shifted via the assimilation of SSM. There is also a weak, but again persistent shift towards higher correlations from the SEKF LAI and SEKF VODX, although it is too close to determine which of the two performs better.

D.2 Impact of Jointly Assimilating Vegetation Variables and SSM - Subdomains

The same graphs as analyzed for the individual assimilation of observations are presented for the joint assimilation, and for all the subdomains, shown in Figure D.10, Figure D.11, Figure D.12, and Figure D.13. For all the subdomains, the changes in LAI correlations are small or non-existent between the single and joint assimilation experiments. And equally, for all the subdomains, the correlations to SSM observations greatly increase when jointly assimilating vegetation and SSM.

The responses of GPP and ET, however, are variable. Over California, there is virtually no difference in GPP correlations, while ET scores are actually degraded over the months of June and July caused by assimilating SSM with vegetation. Over the Midwest, we see the strongest responses of GPP and ET scores to the joint assimilation, improving both over July and August, albeit very slightly. The Northeast region displays no GPP score changes, and a small degradation of scores in July and September. Finally, the Southern Plains shows virtually no changes to either ET or GPP from jointly assimilating LAI or VOD plus SSM.

PDFs visualizing the distributions of the scores also showed no meaningful change over vegetation LSVs due to joint assimilation. These results clearly indicate that the assimilation of SSM, alone or jointly, is far less powerful in changing vegetation related LSVs. Over the entire CONUS domain, there is some small direct evidence indicating added value for GPP and ET variables from joint vegetation and soil moisture assimilation. Localized responses are far more variable. In conjunction with the results from the USCRN analysis, joint assimilation does overall show to provide more value than single vegetation assimilation, and very few drawbacks.

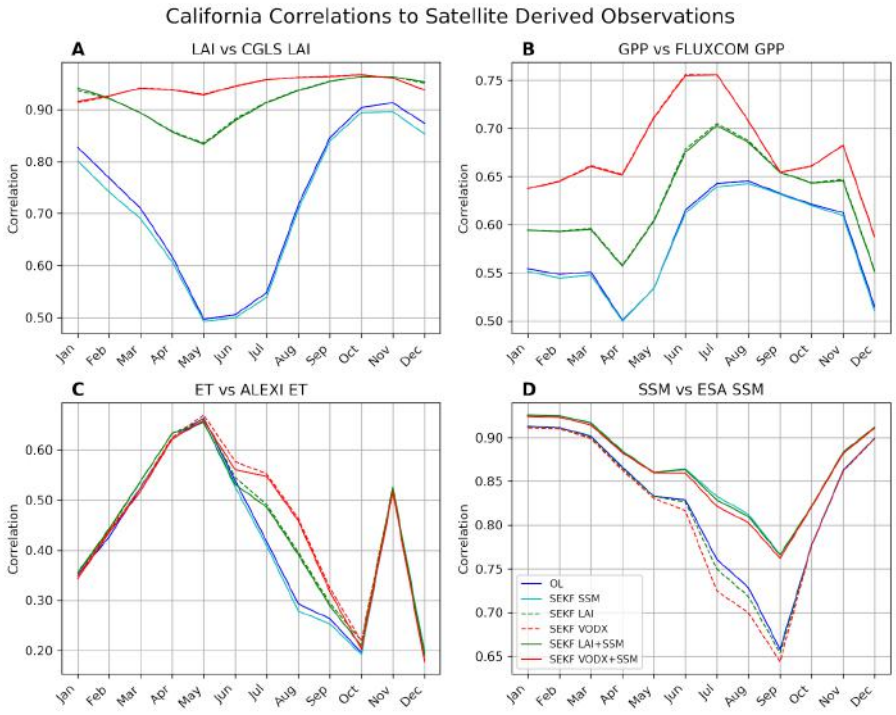

Figure D.10: Same as Figure 4.15, but over the California domain.

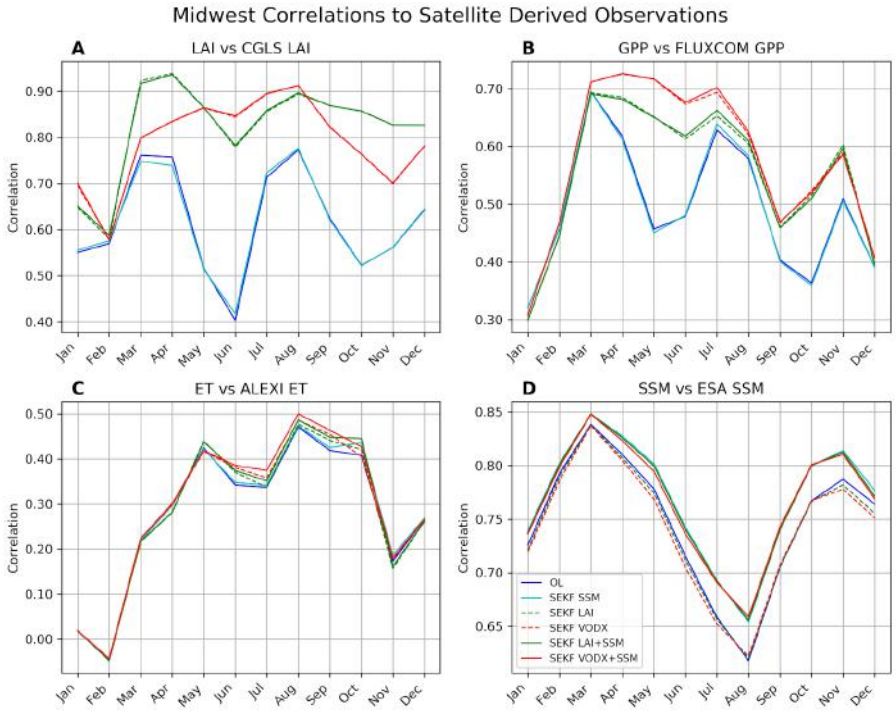

Figure D.11: Same as Figure 4.15, but over the Midwest domain.

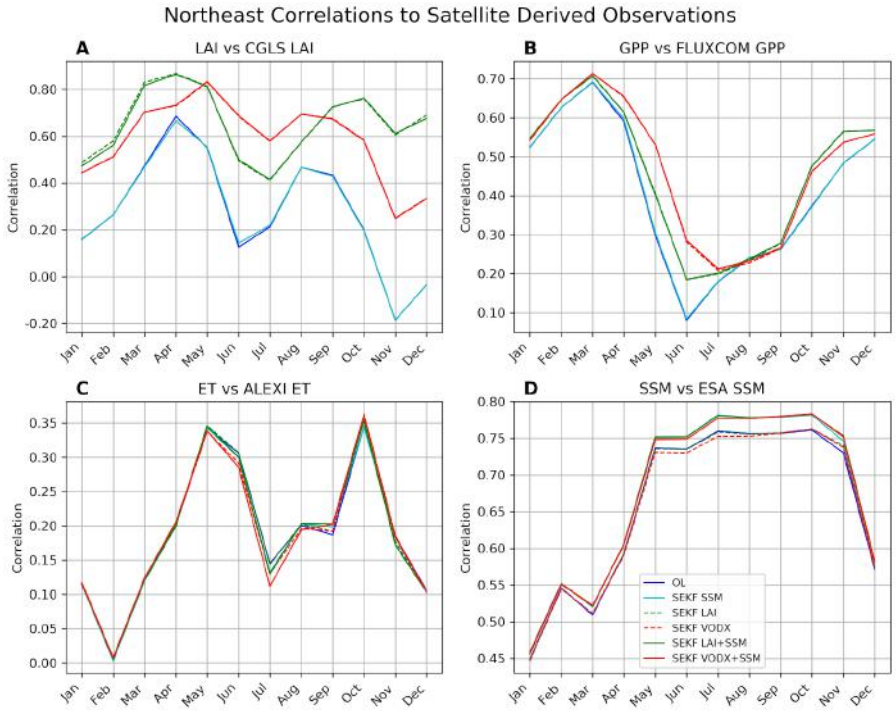

Figure D.12: Same as Figure 4.15, but over the Northeast domain.

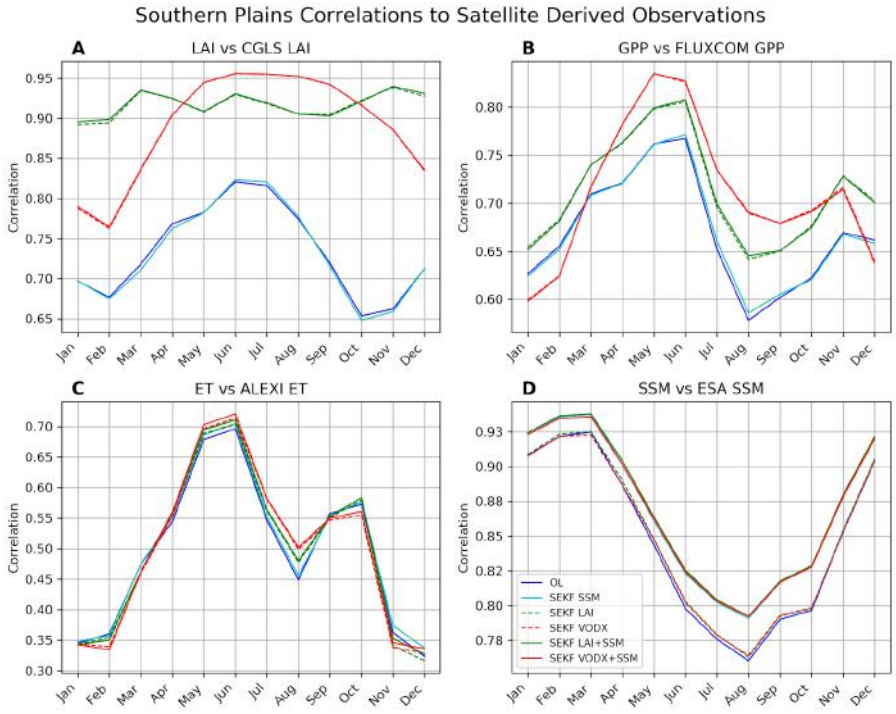

Figure D.13: Same as Figure 4.15, but over the Southern Plains domain.

Bibliography

Albergel, C. et al. (2008). "From near-surface to root-zone soil moisture using an exponential filter: an assessment of the method based on in-situ observations and model simulations". In: *Hydrology and Earth System Sciences* 12.6, pp. 1323–1337. ISSN: 1607-7938. DOI: 10.5194/hess-12-1323-2008. URL: https://hess.copernicus.org/articles/12/1323/2008/.

Albergel, C. et al. (2009). "An evaluation of ASCAT surface soil moisture products with in-situ observations in Southwestern France". In: *Hydrology and Earth System Sciences* 13.2, pp. 115–124. ISSN: 1607-7938. DOI: 10.5194/hess-13-115-2009. URL: https://www.hydrol-earth-syst-sci.net/13/115/2009/.

Albergel, C. et al. (2010). "Monitoring of water and carbon fluxes using a land data assimilation system: a case study for southwestern France". In: *Hydrology and Earth System Sciences* 14.6, pp. 1109–1124. ISSN: 1607-7938. DOI: 10.5194/hess-14-1109-2010. URL: https://hess.copernicus.org/articles/14/1109/2010/.

Albergel, Clément et al. (2017). "Sequential assimilation of satellite-derived vegetation and soil moisture products using SURFEX_v8.0: LDAS-Monde assessment over the Euro-Mediterranean area". In: *Geoscientific Model Development* 10.10, pp. 3889–3912. ISSN: 1991-9603. DOI: 10.5194/gmd-10-3889-2017. URL: https://www.geosci-model-dev.net/10/3889/2017/.

Albergel, Clement et al. (2018a). "ERA-5 and ERA-Interim driven ISBA land surface model simulations: which one performs better?" In: *Hydrology and Earth System Sciences* 22.6, pp. 3515–3532. ISSN: 1607-7938. DOI: 10.5194/hess-22-3515-2018. URL: https://www.hydrol-earth-syst-sci.net/22/3515/2018/.

Albergel, Clement et al. (2018b). "LDAS-Monde Sequential Assimilation of Satellite Derived Observations Applied to the Contiguous US: An ERA-5 Driven Reanalysis of the Land Surface Variables". In: *Remote Sensing* 10.10, p. 1627. ISSN: 2072-4292. DOI: 10.3390/rs10101627. URL: http://www.mdpi.com/2072-4292/10/10/1627.

Albergel, Clément et al. (2019). "Monitoring and Forecasting the Impact of the 2018 Summer Heatwave on Vegetation". In: *Remote Sensing* 11.5, p. 520. ISSN: 2072-4292. DOI: 10.3390/rs11050520. URL: https://www.mdpi.com/2072-4292/11/5/520.

Albergel, Clément et al. (2020). "Data assimilation for continuous global assessment of severe conditions over terrestrial surfaces". In: *Hydrology and Earth System Sciences* 24.9, pp. 4291–4316. ISSN: 1607-7938. DOI: 10.5194/hess-24-4291-2020. URL: https://hess.copernicus.org/articles/24/4291/2020/.

American Society of Civil Engineers (1998). *Sustainability Criteria for Water Resource Systems*. ISBN: 978-0-7844-0331-0.

Anav, Alessandro et al. (2015). "Spatiotemporal patterns of terrestrial gross primary production: A review". In: *Reviews of Geophysics* 53.3, pp. 785–818. ISSN: 87551209. DOI: 10.1002/2015RG000483. URL: http://doi.wiley.com/10.1002/2015RG000483.

Anderson, M et al. (1997). "A Two-Source Time-Integrated Model for Estimating Surface Fluxes Using Thermal Infrared Remote Sensing". In: *Remote Sensing of Environment* 60.2, pp. 195–216. ISSN: 00344257. DOI: 10.1016/S0034-4257(96)00215-5. arXiv: S0034-4"257(96)00215-5. URL: https://linkinghub.elsevier.com/retrieve/pii/S0034425796002155.

Anderson, Martha C. et al. (2007a). "A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing: 1. Model formulation". In: *Journal of Geophysical Research: Atmospheres* 112.D10. ISSN: 01480227. DOI: 10.1029/2006JD007506. URL: http://doi.wiley.com/10.1029/2006JD007506.

— (2007b). "A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing: 2. Surface moisture climatology". In: *Journal of Geophysical Research* 112.D11, p. D11112. ISSN: 0148-0227. DOI: 10.1029/2006JD007507. URL: http://doi.wiley.com/10.1029/2006JD007507.

Anderson, Martha C. et al. (2011). "Evaluation of Drought Indices Based on Thermal Remote Sensing of Evapotranspiration over the Continental United States". In: *Journal of Climate* 24.8, pp. 2025–2044. ISSN: 0894-8755. DOI: 10.1175/2010JCLI3812.1. URL: http://journals.ametsoc.org/doi/10.1175/2010JCLI3812.1.

Anderson, Martha C. et al. (2013). "An Intercomparison of Drought Indicators Based on Thermal Remote Sensing and NLDAS-2 Simulations with U.S. Drought Monitor Classifications". In: *Journal of Hydrometeorology* 14.4, pp. 1035–1056. ISSN: 1525-755X. DOI: 10.1175/JHM-D-12-0140.1. URL: http://journals.ametsoc.org/doi/10.1175/JHM-D-12-0140.1.

Anten, N. P. R. et al. (1995). "Optimal leaf area indices in C3 and C4 mono- and dicotyledonous species at low and high nitrogen availability". In: *Physiologia Plantarum* 95.4, pp. 541–550. ISSN: 0031-9317. DOI: 10.1111/j.1399-3054.1995.tb05520.x. URL: http://doi.wiley.com/10.1111/j.1399-3054.1995.tb05520.x.

Asseng, S. et al. (2014). "Simulation Modeling: Applications in Cropping Systems". In: *Encyclopedia of Agriculture and Food Systems*. Elsevier, pp. 102–112. DOI: 10.1016/B978-0-444-52512-3.00233-3. URL: https://linkinghub.elsevier.com/retrieve/pii/B9780444525123002333.

Baldocchi, Dennis (2008). "'Breathing' of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems". In: *Australian Journal of Botany* 56.1, p. 1. ISSN: 0067-1924. DOI: 10.1071/BT07151. URL: http://www.publish.csiro.au/?paper=BT07151.

Balsamo, G. et al. (2015). "ERA-Interim/Land: a global land surface reanalysis data set". In: *Hydrology and Earth System Sciences* 19.1, pp. 389–407. ISSN:

1607-7938. DOI: 10.5194/hess-19-389-2015. URL: https://hess.copernicus.org/articles/19/389/2015/.

Bamzai, A. S. and J. Shukla (1999). "Relation between Eurasian Snow Cover, Snow Depth, and the Indian Summer Monsoon: An Observational Study". In: *Journal of Climate* 12.10, pp. 3117–3132. ISSN: 0894-8755. DOI: 10.1175/1520-0442(1999)012<3117:RBESCS>2.0.CO;2. URL: http://journals.ametsoc.org/doi/10.1175/1520-0442(1999)012%3C3117:RBESCS%3E2.0.CO;2.

Barbu, A. L. et al. (2011). "Assimilation of Soil Wetness Index and Leaf Area Index into the ISBA-A-gs land surface model: grassland case study". In: *Biogeosciences Discussions* 8.1, pp. 1831–1877. ISSN: 1810-6285. DOI: 10.5194/bgd-8-1831-2011. URL: https://www.biogeosciences-discuss.net/8/1831/2011/.

Barbu, A. L. et al. (2014). "Integrating ASCAT surface soil moisture and GEOV1 leaf area index into the SURFEX modelling platform: a land data assimilation application over France". In: *Hydrology and Earth System Sciences* 18.1, pp. 173–192. ISSN: 1607-7938. DOI: 10.5194/hess-18-173-2014. URL: https://www.hydrol-earth-syst-sci.net/18/173/2014/.

Baret, Frédéric et al. (2007). "LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION". In: *Remote Sensing of Environment* 110.3, pp. 275–286. ISSN: 00344257. DOI: 10.1016/j.rse.2007.02.018. URL: https://linkinghub.elsevier.com/retrieve/pii/S0034425707000909.

Bartalis, Zoltan et al. (2007). "Initial soil moisture retrievals from the METOP-A Advanced Scatterometer (ASCAT)". In: *Geophysical Research Letters* 34.20, p. L20401. ISSN: 0094-8276. DOI: 10.1029/2007GL031088. URL: http://doi.wiley.com/10.1029/2007GL031088.

Bauer, Peter, Alan Thorpe, and Gilbert Brunet (2015). "The quiet revolution of numerical weather prediction". In: *Nature* 525.7567, pp. 47–55. ISSN: 0028-0836. DOI: 10.1038/nature14956. URL: http://www.nature.com/articles/nature14956.

Bauer, Peter et al. (2011). "Satellite cloud and precipitation assimilation at operational NWP centres". In: *Quarterly Journal of the Royal Meteorological Society* 137.661, pp. 1934–1951. ISSN: 00359009. DOI: 10.1002/qj.905. URL: http://doi.wiley.com/10.1002/qj.905.

Beer, Christian et al. (2010). "Terrestrial gross carbon dioxide uptake: Global distribution and covariation with climate". In: *Science* 329.5993, pp. 834–838. ISSN: 00368075. DOI: 10.1126/science.1184984.

Bell, Jesse E. et al. (2013). "U.S. Climate Reference Network Soil Moisture and Temperature Observations". In: *Journal of Hydrometeorology* 14.3, pp. 977–988. ISSN: 1525-755X. DOI: 10.1175/JHM-D-12-0146.1. URL: http://journals.ametsoc.org/doi/10.1175/JHM-D-12-0146.1.

Betts, Alan K. et al. (1996). "The land surface-atmosphere interaction: A review based on observational and global modeling perspectives". In: *Journal of Geophysical Research Atmospheres* 101.D3, pp. 7209–7225. ISSN: 01480227.

DOI: 10.1029/95JD02135. URL: http://doi.wiley.com/10.1029/95JD02135.

Bierkens, M. F. P. and L. P. H. van Beek (2009). "Seasonal Predictability of European Discharge: NAO and Hydrological Response Time". In: *Journal of Hydrometeorology* 10.4, pp. 953–968. ISSN: 1525-7541. DOI: 10.1175/2009JHM1034.1. URL: http://journals.ametsoc.org/doi/10.1175/2009JHM1034.1.

Black, C C (1973). "Photosynthetic Carbon Fixation in Relation to Net CO 2 Uptake". In: *Annual Review of Plant Physiology* 24.1, pp. 253–286. ISSN: 0066-4294. DOI: 10.1146/annurev.pp.24.060173.001345. URL: http://www.annualreviews.org/doi/10.1146/annurev.pp.24.060173.001345.

Blyth, Eleanor M. et al. (2021). "Advances in Land Surface Modelling". In: *Current Climate Change Reports* 7.2, pp. 45–71. ISSN: 2198-6061. DOI: 10.1007/s40641-021-00171-5. URL: https://link.springer.com/10.1007/s40641-021-00171-5.

Boken, Vijendra K, Arthur P Cracknell, and Ronald L Heathcote (2005). *Monitoring and predicting agricultural drought: a global study*. Oxford University Press.

Bonan, Bertrand et al. (2020). "An ensemble square root filter for the joint assimilation of surface soil moisture and leaf area index within the Land Data Assimilation System LDAS-Monde: application over the Euro-Mediterranean region". In: *Hydrology and Earth System Sciences* 24.1, pp. 325–347. ISSN: 1607-7938. DOI: 10.5194/hess-24-325-2020. URL: https://hess.copernicus.org/articles/24/325/2020/.

Bonan, Gordon B. (1995). "Land-Atmosphere interactions for climate system Models: coupling biophysical, biogeochemical, and ecosystem dynamical processes". In: *Remote Sensing of Environment* 51.1, pp. 57–73. ISSN: 00344257. DOI: 10.1016/0034-4257(94)00065-U. URL: https://linkinghub.elsevier.com/retrieve/pii/003442579400065U.

— (2016). *Ecolological Climatology*. Boulder, CO: Cambridge University Press. ISBN: 978-1-107-61905-0.

Boone, A. et al. (2000). "The Influence of the Inclusion of Soil Freezing on Simulations by a Soil–Vegetation–Atmosphere Transfer Scheme". In: *Journal of Applied Meteorology* 39.9, pp. 1544–1569. ISSN: 0894-8763. DOI: 10.1175/1520-0450(2000)039<1544:TIOTIO>2.0.CO;2. URL: http://journals.ametsoc.org/doi/abs/10.1175/1520-0450{\%}282000{\%}29039{\%}3C1544{\%}3ATIOTIO{\%}3E2.0.CO{\%}3B2.

Boone, Aaron, Jean-Christophe Calvet, and Joël Noilhan (1999). "Inclusion of a Third Soil Layer in a Land Surface Scheme Using the Force–Restore Method". In: *Journal of Applied Meteorology* 38.11, pp. 1611–1630. ISSN: 0894-8763. DOI: 10.1175/1520-0450(1999)038<1611:IOATSL>2.0.CO;2. URL: http://journals.ametsoc.org/doi/10.1175/1520-0450(1999)038{\%}3C1611:IOATSL{\%}3E2.0.CO;2.

Boone, Aaron and Pierre Etchevers (2001). "An Intercomparison of Three Snow Schemes of Varying Complexity Coupled to the Same Land Surface Model:

Local-Scale Evaluation at an Alpine Site". In: *Journal of Hydrometeorology* 2.4, pp. 374–394. ISSN: 1525-755X. DOI: 10.1175/1525-7541(2001)002<0374:AIOTSS>2.0.CO;2. URL: http://journals.ametsoc.org/doi/10.1175/1525-7541(2001)002{\%}3C0374:AIOTSS{\%}3E2.0.CO;2.

Borovikov, Anna et al. (2019). "GEOS-5 seasonal forecast system". In: *Climate Dynamics* 53.12, pp. 7335–7361. ISSN: 0930-7575. DOI: 10.1007/s00382-017-3835-2. URL: http://link.springer.com/10.1007/s00382-017-3835-2.

Bosilovich, M. et al. (May 2006). "NASA's Modern Era Retrospective-analysis for Research and Applications (MERRA)". In: *AGU Spring Meeting Abstracts*. Vol. 2007, A43D–01.

Breda, N. J. J. (2003). "Ground-based measurements of leaf area index: a review of methods, instruments and current controversies". In: *Journal of Experimental Botany* 54.392, pp. 2403–2417. ISSN: 1460-2431. DOI: 10.1093/jxb/erg263. URL: https://doi.org/10.1093/jxb/erg263.

Brier, Glenn W. (1950). "Verification of Forecasts Expressed in Terms of Probability". In: *Monthly Weather Review* 78.1, pp. 1–3. ISSN: 0027-0644. DOI: 10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2. URL: http://journals.ametsoc.org/doi/10.1175/1520-0493(1950)078{\%}3C0001:VOFEIT{\%}3E2.0.CO;2.

Bruce, James P. (1994). "Natural Disaster Reduction and Global Change". In: *Bulletin of the American Meteorological Society* 75.10, pp. 1831–1835. ISSN: 0003-0007. DOI: 10.1175/1520-0477(1994)075<1831:NDRAGC>2.0.CO;2. URL: http://dx.doi.org/10.1175/1520-0477(1994)075%3C1831:NDRAGC%3E2.0.CO;2.

Buchanan Smith, M. and S. Davies (1995). *Famine early warning and response: the missing link*. London, UK: Intermediate Technology Publications.

Burgan, Robert E. and Roberta A. Hartford (1993). *Monitoring Vegetation Greenness with Satellite Data*. Tech. rep. United States Department of Agriculture: Intermountain Research Station.

Calvet, Jean-Christophe (2000). "Investigating soil and atmospheric plant water stress using physiological and micrometeorological data". In: *Agricultural and Forest Meteorology* 103.3, pp. 229–247. ISSN: 01681923. DOI: 10.1016/S0168-1923(00)00130-1. URL: https://linkinghub.elsevier.com/retrieve/pii/S0168192300001301.

Calvet, Jean-Christophe and Jean-Louis Champeaux (2020). "L'apport de la télédétection spatiale à la modélisation des surfaces continentales". In: *La Météorologie* 108, p. 052. ISSN: 0026-1181. DOI: 10.37053/lameteorologie-2020-0016. URL: https://lameteorologie.fr/issues/2020/108/meteo_2020_108_52.

Calvet, Jean-Christophe and Jean-François Soussana (2001). "Modelling CO2-enrichment effects using an interactive vegetation SVAT scheme". In: *Agricultural and Forest Meteorology* 108.2, pp. 129–152. ISSN: 01681923. DOI: 10.1016/S0168-1923(01)00235-0. URL: https://linkinghub.elsevier.com/retrieve/pii/S0168192301002350.

Calvet, Jean-Christophe et al. (1998). "An interactive vegetation SVAT model tested against data from six contrasting sites". In: *Agricultural and Forest Meteorology* 92.2, pp. 73–95. ISSN: 01681923. DOI: 10.1016/S0168-1923(98)00091-4. URL: https://linkinghub.elsevier.com/retrieve/pii/S0168192398000914.

Calvet, Jean-Christophe et al. (2004). "Modelling forest transpiration and CO2 fluxes—response to soil moisture stress". In: *Agricultural and Forest Meteorology* 124.3-4, pp. 143–156. ISSN: 01681923. DOI: 10.1016/j.agrformet.2004.01.007. URL: https://linkinghub.elsevier.com/retrieve/pii/S0168192304000140.

Carrera, Marco L. et al. (2019). "Assimilation of Passive L-band Microwave Brightness Temperatures in the Canadian Land Data Assimilation System: Impacts on Short-Range Warm Season Numerical Weather Prediction". In: *Journal of Hydrometeorology* 20.6, pp. 1053–1079. ISSN: 1525-755X. DOI: 10.1175/JHM-D-18-0133.1. URL: http://journals.ametsoc.org/doi/10.1175/JHM-D-18-0133.1.

Centre National de Recherches Météorologiques - UMR 3589 (2018). *ECOCLIMAP-SG*. URL: https://opensource.umr-cnrm.fr/projects/ecoclimap-sg/wiki.

— (2020). *SURFEX*. URL: http://www.umr-cnrm.fr/spip.php?article145{\&}lang=fr.

Champeaux, Jean Louis, V. Masson, and F. Chauvin (2005). "ECOCLIMAP: A global database of land surface parameters at 1 km resolution". In: *Meteorological Applications* 12.1, pp. 29–32. ISSN: 13504827. DOI: 10.1017/S1350482705001519.

Chen, He et al. (2013). "Hydrological data assimilation with the Ensemble Square-Root-Filter: Use of streamflow observations to update model states for real-time flash flood forecasting". In: *Advances in Water Resources* 59, pp. 209–220. ISSN: 03091708. DOI: 10.1016/j.advwatres.2013.06.010. URL: https://linkinghub.elsevier.com/retrieve/pii/S0309170813001115.

Clark, Martyn P. et al. (2006). "Assimilation of snow covered area information into hydrologic and land-surface models". In: *Advances in Water Resources* 29.8, pp. 1209–1221. ISSN: 03091708. DOI: 10.1016/j.advwatres.2005.10.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S030917080500240X.

Collatz, G.James et al. (1991). "Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: a model that includes a laminar boundary layer". In: *Agricultural and Forest Meteorology* 54.2-4, pp. 107–136. ISSN: 01681923. DOI: 10.1016/0168-1923(91)90002-8. URL: https://linkinghub.elsevier.com/retrieve/pii/0168192391900028.

Cook, Edward R. et al. (2007). "North American drought: Reconstructions, causes, and consequences". In: *Earth-Science Reviews* 81.1-2, pp. 93–134. ISSN: 00128252. DOI: 10.1016/j.earscirev.2006.12.002. URL: https://linkinghub.elsevier.com/retrieve/pii/S0012825206001784.

Copernicus (2021). *Copernicus Global Land Service*. URL: https://land.copernicus.eu/global/.

Copernicus Global Land Operations (2019). *Copernicus Global Land Operations "Vegetation and Energy" Product User Manual*. Tech. rep. URL: https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS1_PUM_LAI1km-V2_I1.33.pdf.

— (2020). *Surface Soil Moisture*. URL: https://land.copernicus.eu/global/products/ssm.

Curlander, J.C. and R.N. McDonough (1991). *Synthetic Aperture Radar*. New York, NY: Wiley Interscience.

Dalu, G. (1986). "Satellite remote sensing of atmospheric water vapour". In: *International Journal of Remote Sensing* 7.9, pp. 1089–1097. ISSN: 0143-1161. DOI: 10.1080/01431168608948911. URL: https://www.tandfonline.com/doi/full/10.1080/01431168608948911.

De Lannoy, Gabriëlle J. M. et al. (2012). "Multiscale assimilation of Advanced Microwave Scanning Radiometer-EOS snow water equivalent and Moderate Resolution Imaging Spectroradiometer snow cover fraction observations in northern Colorado". In: *Water Resources Research* 48.1. ISSN: 00431397. DOI: 10.1029/2011WR010588. URL: http://doi.wiley.com/10.1029/2011WR010588.

Deardorff, J. W. (1978). "Efficient prediction of ground surface temperature and moisture, with inclusion of a layer of vegetation". In: *Journal of Geophysical Research* 83.C4, p. 1889. ISSN: 0148-0227. DOI: 10.1029/jc083ic04p01889.

Decharme, B. and H. Douville (2007). "Global validation of the ISBA sub-grid hydrology". In: *Climate Dynamics* 29.1, pp. 21–37. ISSN: 0930-7575. DOI: 10.1007/s00382-006-0216-7. URL: http://link.springer.com/10.1007/s00382-006-0216-7.

Decharme, Bertrand et al. (2016). "Impacts of snow and organic soils parameterization on northern Eurasian soil temperature profiles simulated by the ISBA land surface model". In: *The Cryosphere* 10.2, pp. 853–877. ISSN: 1994-0424. DOI: 10.5194/tc-10-853-2016. URL: https://tc.copernicus.org/articles/10/853/2016/.

Delire, Christine et al. (2020). "The Global Land Carbon Cycle Simulated With ISBA-CTRIP: Improvements Over the Last Decade". In: *Journal of Advances in Modeling Earth Systems* 12.9. ISSN: 1942-2466. DOI: 10.1029/2019MS001886. URL: https://onlinelibrary.wiley.com/doi/10.1029/2019MS001886.

Demarty, J. et al. (2007). "Assimilation of global MODIS leaf area index retrievals within a terrestrial biosphere model". In: *Geophysical Research Letters* 34.15. ISSN: 00948276. DOI: 10.1029/2007GL030014. URL: http://doi.wiley.com/10.1029/2007GL030014.

Devereux, Stephen (2007). "The impact of droughts and floods on food security and policy options to alleviate negative effects". In: *Agricultural Economics* 37, pp. 47–58. ISSN: 01695150. DOI: 10.1111/j.1574-0862.2007.00234.x. URL: http://doi.wiley.com/10.1111/j.1574-0862.2007.00234.x.

Diamond, Howard J. et al. (2013). "U.S. Climate Reference Network after One Decade of Operations: Status and Assessment". In: *Bulletin of the American Meteorological Society* 94.4, pp. 485–498. ISSN: 1520-0477. DOI: 10.1175/BAMS-D-12-00170.1. URL: https://journals.ametsoc.org/doi/10.1175/BAMS-D-12-00170.1.

Dickinson, R.E (1986). *Biosphere/atmosphere transfer scheme (BATS) for the NCAR community climate model*. Tech. rep. NCAR.

Dinar, Ariel and Robert O Mendelsohn (2011). *Handbook on climate change and agriculture*. Edward Elgar Publishing.

Dirmeyer, Paul A., Zhichang Guo, and Xiang Gao (2004). "Comparison, Validation, and Transferability of Eight Multiyear Global Soil Wetness Products". In: *Journal of Hydrometeorology* 5.6, pp. 1011–1033. ISSN: 1525-7541. DOI: 10.1175/JHM-388.1. URL: http://journals.ametsoc.org/doi/10.1175/JHM-388.1.

Dobson, M. and Fawwaz Ulaby (1986). "Active Microwave Soil Moisture Research". In: *IEEE Transactions on Geoscience and Remote Sensing* GE-24.1, pp. 23–36. ISSN: 0196-2892. DOI: 10.1109/TGRS.1986.289585. URL: http://ieeexplore.ieee.org/document/4072416/.

Dorigo, W.A. et al. (2015). "Evaluation of the ESA CCI soil moisture product using ground-based observations". In: *Remote Sensing of Environment* 162, pp. 380–395. ISSN: 00344257. DOI: 10.1016/j.rse.2014.07.023. URL: https://linkinghub.elsevier.com/retrieve/pii/S0034425714002727.

Dorigo, Wouter et al. (2017). "ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions". In: *Remote Sensing of Environment* 203, pp. 185–215. ISSN: 00344257. DOI: 10.1016/j.rse.2017.07.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S0034425717303061.

Draper, C. et al. (2011). "Assimilation of ASCAT near-surface soil moisture into the SIM hydrological model over France". In: *Hydrology and Earth System Sciences* 15.12, pp. 3829–3841. ISSN: 1607-7938. DOI: 10.5194/hess-15-3829-2011. URL: https://www.hydrol-earth-syst-sci.net/15/3829/2011/.

Draper, C. S., J.-F. Mahfouf, and J. P. Walker (2009). "An EKF assimilation of AMSR-E soil moisture into the ISBA land surface scheme". In: *Journal of Geophysical Research* 114.D20, p. D20104. ISSN: 0148-0227. DOI: 10.1029/2008JD011650. URL: http://doi.wiley.com/10.1029/2008JD011650.

Drought.gov (2021). *Drought by Sector: Agriculture*. URL: https://www.drought.gov/sectors/agriculture#:$\sim$:text=Thecostofdroughtevents,hazardwithsubstantialsocioeconomicconsequences..

Druel, Arsene et al. (2021). "Implementation and validation of a new irrigation scheme in the ISBA land surface model". In: *Geosci. Model Dev.* In Review. DOI: https://doi.org/10.5194/gmd-2021-332. URL: https://gmd.copernicus.org/preprints/gmd-2021-332/.

Drusch, M., E. F. Wood, and H. Gao (2005). "Observation operators for the direct assimilation of TRMM microwave imager retrieved soil moisture". In: *Geophysical Research Letters* 32.15, p. L15403. ISSN: 0094-8276. DOI:

10.1029/2005GL023623. URL: http://doi.wiley.com/10.1029/2005GL023623.

Drusch, Matthias and Pedro Viterbo (2007). "Assimilation of Screen-Level Variables in ECMWF's Integrated Forecast System: A Study on the Impact on the Forecast Quality and Analyzed Soil Moisture". In: *Monthly Weather Review* 135.2, pp. 300–314. ISSN: 1520-0493. DOI: 10.1175/MWR3309.1. URL: http://journals.ametsoc.org/doi/10.1175/MWR3309.1.

Entin, Jared K. et al. (1999). "Evaluation of Global Soil Wetness Project Soil Moisture Simulations". In: *Journal of the Meteorological Society of Japan. Ser. II* 77.1B, pp. 183–198. ISSN: 0026-1165. DOI: 10.2151/jmsj1965.77.1B_183. URL: https://www.jstage.jst.go.jp/article/jmsj1965/77/1B/77_1B_183/_article.

European Commission (2007a). "Communication from the commission to the council and the European parliament". In: 59.11 SUPPL. ISSN: 00167460. URL: https://ec.europa.eu/environment/water/quantity/pdf/comm{_}droughts/impact{_}assessment.pdf.

— (2007b). "Water Scarcity & Droughts. In-Depth Assessment. Second Interim Report – June 2007". In: 17.1, pp. 5–18. ISSN: 13679430. URL: http://ec.europa.eu/environment/water/quantity/pdf/comm{_}droughts/2nd{_}int{_}report.pdf.

European Commission and Global Vegetation Monitoring Unit (2000). *Global Land Cover 2000*. URL: https://forobs.jrc.ec.europa.eu/products/glc2000/glc2000.php.

European Environmental Agency and Copernicus (2000). *Corine Land Cover 2000*. URL: https://land.copernicus.eu/pan-european/corine-land-cover/clc-2000?tab=metadata.

Fairbairn, David et al. (2017). "The effect of satellite-derived surface soil moisture and leaf area index land data assimilation on streamflow simulations over France". In: *Hydrology and Earth System Sciences* 21.4, pp. 2015–2033. ISSN: 1607-7938. DOI: 10.5194/hess-21-2015-2017. URL: https://www.hydrol-earth-syst-sci.net/21/2015/2017/.

Faroux, S. et al. (2013). "ECOCLIMAP-II/Europe: a twofold database of ecosystems and surface parameters at 1 km resolution based on satellite information for use in land surface, meteorological and climate models". In: *Geoscientific Model Development* 6.2, pp. 563–582. ISSN: 1991-9603. DOI: 10.5194/gmd-6-563-2013. URL: https://www.geosci-model-dev.net/6/563/2013/.

Federal Emergency Management Agency (1995). *National mitigation strategy: Partnerships for building safer communities*. URL: http://hazardmitigation.calema.ca.gov/docs/1995{_}National{_}Mitigation{_}Strategy{_}web.pdf.

Fernandez-Moran, R. et al. (2015). "Roughness and vegetation parameterizations at L-band for soil moisture retrievals over a vineyard field". In: *Remote Sens. Environ.* 170, pp. 269–279. ISSN: 00344257. DOI: 10.1016/j.rse.2015.09.006. URL: https://linkinghub.elsevier.com/retrieve/pii/S0034425715301292.

Fisher, Rosie A. and Charles D. Koven (2020). "Perspectives on the Future of Land Surface Models and the Challenges of Representing Complex Terrestrial Systems". In: *Journal of Advances in Modeling Earth Systems* 12.4. ISSN: 1942-2466. DOI: 10.1029/2018MS001453. URL: https://onlinelibrary.wiley.com/doi/10.1029/2018MS001453.

Folving, Sten, Pamela Kennedy, and Jacques Megier (1995). *The Application of the FIRS Project's Foundation Action 1. The Regionalisation and Stratification of European Forest Ecosystems for Providing a European NOAA-AVHRR-Based Forest Map*. URL: https://publications.jrc.ec.europa.eu/repository/handle/JRC12308.

Food and Agriculture Organization (2003). *The State of Food Insecuirty in the World*. Tech. rep. Rome, Italy: Food and Agriculture Organization of the United Nations. URL: http://www.fao.org/3/j0083e/j0083e00.htm.

Ford, Trent W. and Christopher F. Labosier (2017). "Meteorological conditions associated with the onset of flash drought in the Eastern United States". In: *Agricultural and Forest Meteorology* 247, pp. 414–423. ISSN: 01681923. DOI: 10.1016/j.agrformet.2017.08.031. URL: https://linkinghub.elsevier.com/retrieve/pii/S0168192317302885.

Ford, Trent W. et al. (2015). "On the utility of in situ soil moisture observations for flash drought early warning in Oklahoma, USA". In: *Geophysical Research Letters* 42.22, pp. 9790–9798. ISSN: 00948276. DOI: 10.1002/2015GL066600. URL: http://doi.wiley.com/10.1002/2015GL066600.

Forseth, Irwin N. (2010). "The Ecology of Photosynthetic Pathways". In: *Nature Educational Knowledge* 3.10. URL: https://www.nature.com/scitable/knowledge/library/the-ecology-of-photosynthetic-pathways-15785165/#:$\sim$:text=Plantshaveevolvedthreephotosynthetic,patternsofgrowthanddistribution.&text=Plantfitnessrevolvesaroundthe,carbonforgrowthandreproduction..

Frappart, Frédéric et al. (2020). "Global Monitoring of the Vegetation Dynamics from the Vegetation Optical Depth (VOD): A Review". In: *Remote Sensing* 12.18, p. 2915. ISSN: 2072-4292. DOI: 10.3390/rs12182915. URL: https://www.mdpi.com/2072-4292/12/18/2915.

Friedl, M. A. et al. (1994). "Estimating grassland biomass and leaf area index using ground and satellite data". In: *International Journal of Remote Sensing* 15.7, pp. 1401–1420. ISSN: 0143-1161. DOI: 10.1080/01431169408954174. URL: https://www.tandfonline.com/doi/full/10.1080/01431169408954174.

Friedlingstein, Pierre et al. (2020). "Global Carbon Budget 2020". In: *Earth System Science Data* 12.4, pp. 3269–3340. ISSN: 1866-3516. DOI: 10.5194/essd-12-3269-2020. URL: https://essd.copernicus.org/articles/12/3269/2020/.

Fritz, S. and H. Wexler (1960). "Cloud Pictures from Satellite TIROS I*". In: *Monthly Weather Review* 88.3, pp. 79–87. ISSN: 0027-0644. DOI: 10.1175/1520-0493(1960)088<0079:CPFSTI>2.0.CO;2. URL: http://journals.ametsoc.org/doi/10.1175/1520-0493(1960)088{\%}3C0079:CPFSTI{\%}3E2.0.CO;2.

Garratt, J. R. (1992). "The atmospheric boundary layer". In: *The atmospheric boundary layer* 37, pp. 89–134. ISSN: 0066-4189. DOI: 10.1146/annurev.fl.02.010170.001301.

Gelaro, Ronald et al. (2017). "The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2)". In: *Journal of Climate* 30.14, pp. 5419–5454. ISSN: 0894-8755. DOI: 10.1175/JCLI-D-16-0758.1. URL: https://journals.ametsoc.org/doi/10.1175/JCLI-D-16-0758.1.

Getirana, Augusto et al. (2020). "GRACE Improves Seasonal Groundwater Forecast Initialization over the United States". In: *Journal of Hydrometeorology* 21.1, pp. 59–71. ISSN: 1525-755X. DOI: 10.1175/JHM-D-19-0096.1. URL: https://journals.ametsoc.org/view/journals/hydr/21/1/jhm-d-19-0096.1.xml.

Gibelin, Anne-Laure et al. (2006). "Ability of the land surface model ISBA-A-gs to simulate leaf area index at the global scale: Comparison with satellites products". In: *Journal of Geophysical Research* 111.D18, p. D18102. ISSN: 0148-0227. DOI: 10.1029/2005JD006691. URL: http://doi.wiley.com/10.1029/2005JD006691.

GIEC (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)] Intergovernmental Panel on Climate Change, p. 151. ISBN: 9789291691432.

Gillette, HP (1950). "A creeping drought under way". In: *Water and sewage works* 104.5.

Gitelson, Anatoly A. et al. (2003). "Remote estimation of leaf area index and green leaf biomass in maize canopies". In: *Geophysical Research Letters* 30.5, n/a–n/a. ISSN: 00948276. DOI: 10.1029/2002GL016450. URL: http://doi.wiley.com/10.1029/2002GL016450.

Gruber, Alexander et al. (2017). "Triple Collocation-Based Merging of Satellite Soil Moisture Retrievals". In: *IEEE Transactions on Geoscience and Remote Sensing* 55.12, pp. 6780–6792. ISSN: 0196-2892. DOI: 10.1109/TGRS.2017.2734070. URL: http://ieeexplore.ieee.org/document/8046118/.

Gruber, Alexander et al. (2019). "Evolution of the ESA CCI Soil Moisture climate data records and their underlying merging methodology". In: *Earth System Science Data* 11.2, pp. 717–739. ISSN: 1866-3516. DOI: 10.5194/essd-11-717-2019. URL: https://essd.copernicus.org/articles/11/717/2019/.

Gutman, G. Garik (1990). "Towards Monitoring Droughts from Space". In: *Journal of Climate* 3.2, pp. 282–295. ISSN: 0894-8755. DOI: 10.1175/1520-0442(1990)003<0282:TMDFS>2.0.CO;2. URL: http://journals.ametsoc.org/doi/10.1175/1520-0442(1990)003{\%}3C0282:TMDFS{\%}3E2.0.CO;2.

Hagman, Gunnar et al. (1984). *Prevention better than cure. Report on human and environmental disasters in the Third World*.

Haile, Menghestab (2005). "Weather patterns, food security and humanitarian response in sub-Saharan Africa". In: *Philosophical Transactions of the Royal*

Society B: Biological Sciences 360.1463, pp. 2169–2182. ISSN: 0962-8436. DOI: 10.1098/rstb.2005.1746. URL: https://royalsocietypublishing.org/doi/10.1098/rstb.2005.1746.

Han, Xujun et al. (2014). "Soil moisture and soil properties estimation in the Community Land Model with synthetic brightness temperature observations". In: *Water Resources Research* 50.7, pp. 6081–6105. ISSN: 00431397. DOI: 10.1002/2013WR014586. URL: http://doi.wiley.com/10.1002/2013WR014586.

Hansen, M. C. et al. (2000). "Global land cover classification at 1 km spatial resolution using a classification tree approach". In: *International Journal of Remote Sensing* 21.6-7, pp. 1331–1364. ISSN: 0143-1161. DOI: 10.1080/014311600210209. URL: https://www.tandfonline.com/doi/full/10.1080/014311600210209.

Hayes, Michael et al. (2011). "The lincoln declaration on drought indices: Universal meteorological drought index recommended". In: *Bulletin of the American Meteorological Society* 92.4, pp. 485–488. ISSN: 00030007. DOI: 10.1175/2010BAMS3103.1.

Hayes, Michael J. et al. (2012). "Drought monitoring: Historical and current perspectives". In: *Remote Sensing of Drought: Innovative Monitoring Approaches*, pp. 1–19. DOI: 10.1201/b11863.

Hayes, Mike et al. (2005). "Drought Monitoring: New Tools for the 21st Century". In: *Drought and Water Crises: Science, Technology, and Management Issues*. Boca Raton, FL: CRC Press, pp. 53–69.

Heidke, P. (1926). "Berechnung Des Erfolges Und Der Güte Der Windstärkevorhersagen Im Sturmwarnungsdienst". In: *Geografiska Annaler* 8.4, pp. 301–349. ISSN: 2001-4422. DOI: 10.1080/20014422.1926.11881138. URL: https://www.tandfonline.com/doi/full/10.1080/20014422.1926.11881138.

Heim, Richard R. (2002). "A Review of Twentieth-Century Drought Indices Used in the United States". In: *Bulletin of the American Meteorological Society* 83.8, pp. 1149–1166. ISSN: 0003-0007. DOI: 10.1175/1520-0477-83.8.1149. URL: https://journals.ametsoc.org/doi/10.1175/1520-0477-83.8.1149.

Heim, Richard R. and Michael J. Brewer (2012). "The Global Drought Monitor Portal: The Foundation for a Global Drought Information System". In: *Earth Interactions* 16.15, pp. 1–28. ISSN: 1087-3562. DOI: 10.1175/2012EI000446.1. URL: https://journals.ametsoc.org/doi/10.1175/2012EI000446.1.

Hersbach, H. et al. (2018). *ERA5 hourly data on single levels from 1979 to present*. DOI: 10.24381/cds.adbb2d47.

Hersbach, Hans (2000). "Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems". In: *Weather and Forecasting* 15.5, pp. 559–570. ISSN: 0882-8156. DOI: 10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2. URL: http://journals.ametsoc.org/doi/10.1175/1520-0434(2000)015{\%}3C0559:DOTCRP{\%}3E2.0.CO;2.

Hornbuckle, Brian K. et al. (2016). "SMOS optical thickness changes in response to the growth and development of crops, crop management, and weather". In:

Remote Sens. Environ. 180, pp. 320–333. ISSN: 00344257. DOI: 10.1016/j.rse.2016.02.043. URL: https://linkinghub.elsevier.com/retrieve/pii/S0034425716300712.

Houborg, Rasmus et al. (2012). "Drought indicators based on model-assimilated Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage observations". In: *Water Resources Research* 48.7. ISSN: 00431397. DOI: 10.1029/2011WR011291. URL: http://doi.wiley.com/10.1029/2011WR011291.

Hunt, Eric D. et al. (2009). "The development and evaluation of a soil moisture index". In: *International Journal of Climatology* 29.5, pp. 747–759. ISSN: 08998418. DOI: 10.1002/joc.1749. URL: http://doi.wiley.com/10.1002/joc.1749.

Hunt, Eric D. et al. (2014). "Monitoring the effects of rapid onset of drought on non-irrigated maize with agronomic data and climate-based drought indices". In: *Agricultural and Forest Meteorology* 191, pp. 1–11. ISSN: 01681923. DOI: 10.1016/j.agrformet.2014.02.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S0168192314000409.

Hutchinson, C. F. (1991). "Uses of satellite data for famine early warning in sub-Saharan Africa". In: *International Journal of Remote Sensing* 12.6, pp. 1405–1421. ISSN: 0143-1161. DOI: 10.1080/01431169108929733. URL: https://www.tandfonline.com/doi/full/10.1080/01431169108929733.

Inoue, Toshiro (1985). "On the Temperature and Effective Emissivity Determination of Semi-Transparent Cirrus Clouds by Bi-Spectral Measurements in the 10μm Window Region". In: *Journal of the Meteorological Society of Japan. Ser. II* 63.1, pp. 88–99. ISSN: 0026-1165. DOI: 10.2151/jmsj1965.63.1_88. URL: https://www.jstage.jst.go.jp/article/jmsj1965/63/1/63{_}1{_}88/{_}article.

Intergovernmental Panel on Climate Change (2013). "The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change". In: *IPCC Climate Change*, pp. 465–570. ISSN: 1098-6596. DOI: 10.1017/CBO9781107415324.015. arXiv: arXiv:1011.1669v3. URL: http://www.ipcc.ch/report/ar5/wg1/docs/review/WG1AR5_SOD_Ch06_All_Final.pdf%5Cnhttp://ebooks.cambridge.org/ref/id/CBO9781107415324A023.

Intergovernmental Panel on Climate Change (2018). "Impacts of 1.5ºC Global Warming on Natural and Human Systems". In: *Global Warming of 1.5°C. An IPCC Special Report*, pp. 175–311. URL: https://www.ipcc.ch/sr15.

Intergovernmental Panel on Climate Change (2018). "Impacts of 1.5ºC Global Warming on Natural and Human Systems". In: *Global Warming of 1.5°C. An IPCC Special Report*, pp. 175–311. URL: https://www.ipcc.ch/sr15.

International Water Management Institute (2015). *Development of South Asia Drought Monitoring System*. Tech. rep. June, pp. 1–22.

Ionita, Monica et al. (2017). "The European 2015 drought from a climatological perspective". In: *Hydrology and Earth System Sciences* 21.3, pp. 1397–1419. ISSN: 1607-7938. DOI: 10.5194/hess-21-1397-2017. URL: https://www.hydrol-earth-syst-sci.net/21/1397/2017/.

IPCC (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)] Intergovernmental Panel on Climate Change, p. 151. ISBN: 9789291691432.

Jackson, T. J. and T. J. Schmugge (1991). "Vegetation effects on the microwave emission of soils". In: *Remote Sensing of Environment* 36.3, pp. 203–212. ISSN: 00344257. DOI: 10.1016/0034-4257(91)90057-D.

Jackson, Thomas J., Thomas J. Schmugge, and James R. Wang (1982). "Passive microwave sensing of soil moisture under vegetation canopies". In: *Water Resources Research* 18.4, pp. 1137–1142. ISSN: 00431397. DOI: 10.1029/WR018i004p01137. URL: http://doi.wiley.com/10.1029/WR018i004p01137.

Japan Aerospace Exploration Agency (2016). *Mechanisms of Remote Sensing*. URL: https://www.eorc.jaxa.jp/en/hatoyama/experience/rm{_}kiso/mecha{_}howto{_}e.html (visited on 03/23/2021).

Jiao, Xianfeng et al. (2010). "The sensitivity of multi-frequency (X, C and L-band) radar backscatter signatures to bio-physical variables (LAI) over corn and soybean fields". In: *ISPRS TC VII Symposium—100 Years ISPRS*, pp. 317–325.

Jung, M., M. Reichstein, and A. Bondeau (2009). "Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model". In: *Biogeosciences* 6.10, pp. 2001–2013. ISSN: 1726-4189. DOI: 10.5194/bg-6-2001-2009. URL: https://bg.copernicus.org/articles/6/2001/2009/.

Jung, Martin et al. (2019). "The FLUXCOM ensemble of global land-atmosphere energy fluxes". In: *Scientific Data* 6.1, p. 74. ISSN: 2052-4463. DOI: 10.1038/s41597-019-0076-8. URL: http://www.nature.com/articles/s41597-019-0076-8.

Kiehl, Jeffrey T. and Kevin E. Trenberth (1997). "Earth ' s Annual Global Energy Budget". In: *Bulletin of the American Meteorological Society* 78.2, pp. 197–208.

Koeppe, Clarence E. and George De Long (1958). *Weather and Climate*. London: MacGraw-Hill, p. 341.

Kogan, F. N. (1990). "Remote sensing of weather impacts on vegetation in non-homogeneous areas". In: *International Journal of Remote Sensing* 11.8, pp. 1405–1419. ISSN: 0143-1161. DOI: 10.1080/01431169008955102. URL: https://www.tandfonline.com/doi/full/10.1080/01431169008955102.

Kogan, F.N. (1995). "Application of vegetation index and brightness temperature for drought detection". In: *Advances in Space Research* 15.11, pp. 91–100. ISSN: 02731177. DOI: 10.1016/0273-1177(95)00079-T. URL: https://linkinghub.elsevier.com/retrieve/pii/027311779500079T.

Konings, Alexandra G. et al. (2016). "Vegetation optical depth and scattering albedo retrieval using time series of dual-polarized L-band radiometer observations". In: *Remote Sensing of Environment* 172, pp. 178–189. ISSN:

00344257. DOI: 10.1016/j.rse.2015.11.009. URL: https://linkinghub.elsevier.com/retrieve/pii/S003442571530198X.

Koster, Randal D. and P. C. D. Milly (1997). "The Interplay between Transpiration and Runoff Formulations in Land Surface Schemes Used with Atmospheric Models". In: *Journal of Climate* 10.7, pp. 1578–1591. ISSN: 0894-8755. DOI: 10.1175/1520-0442(1997)010<1578:TIBTAR>2.0.CO;2. URL: http://journals.ametsoc.org/doi/10.1175/1520-0442(1997)010{\%}3C1578:TIBTAR{\%}3E2.0.CO;2.

Koster, Randal D. et al. (2010). "Skill in streamflow forecasts derived from large-scale estimates of soil moisture and snow". In: *Nature Geoscience* 3.9, pp. 613–616. ISSN: 1752-0894. DOI: 10.1038/ngeo944. URL: http://www.nature.com/articles/ngeo944.

Kumar, Sujay et al. (2020). "Assimilation of vegetation optical depth retrievals from passive microwave radiometry". In: *Hydrology and Earth System Sciences Discussions* 2015, pp. 1–39. ISSN: 1027-5606. DOI: 10.5194/hess-2020-36.

Kumar, Sujay V. et al. (2009). "Role of Subsurface Physics in the Assimilation of Surface Soil Moisture Observations". In: *Journal of Hydrometeorology* 10.6, pp. 1534–1547. ISSN: 1525-755X. DOI: 10.1175/2009JHM1134.1. URL: http://journals.ametsoc.org/doi/10.1175/2009JHM1134.1.

Kumar, Sujay V. et al. (2019). "Assimilation of Remotely Sensed Leaf Area Index into the Noah-MP Land Surface Model: Impacts on Water and Carbon Fluxes and States over the Continental United States". In: *Journal of Hydrometeorology* 20.7, pp. 1359–1377. ISSN: 1525-755X. DOI: 10.1175/JHM-D-18-0237.1. URL: http://journals.ametsoc.org/doi/10.1175/JHM-D-18-0237.1.

Kussul, Natalia et al. (2008). "Data Assimilation Technique For Flood Monitoring and Prediction". In: *Institute of Mathematics and Informatics, Bulgarian Academy of Sciences* 15.1. URL: http://sci-gems.math.bas.bg/jspui/handle/10525/304.

Lacaze, Roselyne et al. (2010). "GEOLAND2 - TOWARDS AN OPERATIONAL GMES LAND MONITORING CORE SERVICE; FIRST RESULTS OF THE BIOGEOPHYSICAL PARAMETER CORE MAPPING SERVICE". In: Vienna, Austria, pp. 354–359. URL: https://publik.tuwien.ac.at/files/PubDat_192445.pdf.

Lahoz, William A. and Gabriëlle J. M. De Lannoy (2014). "Closing the Gaps in Our Knowledge of the Hydrological Cycle over Land: Conceptual Problems". In: *Surveys in Geophysics* 35.3, pp. 623–660. ISSN: 0169-3298. DOI: 10.1007/s10712-013-9221-7. URL: http://link.springer.com/10.1007/s10712-013-9221-7.

Lambin, Eric F. and Alan H. Strahlers (1994). "Change-vector analysis in multitemporal space: A tool to detect and categorize land-cover change processes using high temporal-resolution satellite data". In: *Remote Sensing of Environment* 48.2, pp. 231–244. ISSN: 00344257. DOI: 10.1016/0034-4257(94)90144-9. URL: https://linkinghub.elsevier.com/retrieve/pii/0034425794901449.

Lande, R (2009). "Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation". In: *Journal of Evolutionary Biology* 22.7, pp. 1435–1446. ISSN: 1010061X. DOI: 10.1111/j.1420-9101.2009.01754.x. URL: http://doi.wiley.com/10.1111/j.1420-9101.2009.01754.x.

Langley, Susan Kathleen, Heather M. Cheshire, and Karen S. Humes (2001). "A comparison of single date and multitemporal satellite image classifications in a semi-arid grassland". In: *Journal of Arid Environments* 49.2, pp. 401–411. ISSN: 01401963. DOI: 10.1006/jare.2000.0771. URL: https://linkinghub.elsevier.com/retrieve/pii/S0140196300907717.

Lasslop, GITTA et al. (2010). "Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation". In: *Global Change Biology* 16.1, pp. 187–208. ISSN: 13541013. DOI: 10.1111/j.1365-2486.2009.02041.x. URL: http://doi.wiley.com/10.1111/j.1365-2486.2009.02041.x.

Law, B.E et al. (2002). "Environmental controls over carbon dioxide and water vapor exchange of terrestrial vegetation". In: *Agricultural and Forest Meteorology* 113.1-4, pp. 97–120. ISSN: 01681923. DOI: 10.1016/S0168-1923(02)00104-1. URL: https://linkinghub.elsevier.com/retrieve/pii/S0168192302001041.

Lawrimore, Jay et al. (2002). "Beginning A New Era Of Drought Monitoring Across North America". In: *Bulletin of the American Meteorological Society* 83.8, pp. 1191–1192. ISSN: 0003-0007. DOI: 10.1175/1520-0477-83.8.1191. URL: https://journals.ametsoc.org/doi/10.1175/1520-0477-83.8.1191.

Le Moigne, Patrick et al. (2009). *SURFEX Scientific Documentation*. Tech. rep. URL: http://www.umr-cnrm.fr/surfex/IMG/pdf/surfex{_}scidoc.pdf.

Le Moigne, Patrick et al. (2012). *SURFEX Scientific Documentation*. Tech. rep. URL: http://www.umr-cnrm.fr/surfex/IMG/pdf/surfex{_}scidoc{_}v2-2.pdf.

Le Moigne, Patrick et al. (2018). *SURFEX Scientific Documentation*. Tech. rep. URL: http://www.umr-cnrm.fr/surfex/IMG/pdf/surfex{_}scidoc{_}v8.1.pdf.

Li, Xiaojun et al. (2021). "Global-scale assessment and inter-comparison of recently developed/reprocessed microwave satellite vegetation optical depth products". In: *Remote Sensing of Environment* 253, p. 112208. ISSN: 00344257. DOI: 10.1016/j.rse.2020.112208. URL: https://linkinghub.elsevier.com/retrieve/pii/S0034425720305812.

Lievens, H. et al. (2015). "Optimization of a Radiative Transfer Forward Operator for Simulating SMOS Brightness Temperatures over the Upper Mississippi Basin". In: *Journal of Hydrometeorology* 16.3, pp. 1109–1134. ISSN: 1525-755X. DOI: 10.1175/JHM-D-14-0052.1. URL: http://journals.ametsoc.org/doi/10.1175/JHM-D-14-0052.1.

Lievens, H. et al. (2017). "Assimilation of global radar backscatter and radiometer brightness temperature observations to improve soil moisture and land

evaporation estimates". In: *Remote Sensing of Environment* 189, pp. 194–210. ISSN: 00344257. DOI: 10.1016/j.rse.2016.11.022. URL: https://linkinghub.elsevier.com/retrieve/pii/S0034425716304709.

Liu, W. T. and F. N. Kogan (1996). "Monitoring regional drought using the Vegetation Condition Index". In: *International Journal of Remote Sensing* 17.14, pp. 2761–2782. ISSN: 0143-1161. DOI: 10.1080/01431169608949106. URL: https://www.tandfonline.com/doi/full/10.1080/01431169608949106.

Liu, Yuqiong et al. (2013). "Assimilating satellite-based snow depth and snow cover products for improving snow predictions in Alaska". In: *Advances in Water Resources* 54, pp. 208–227. ISSN: 03091708. DOI: 10.1016/j.advwatres.2013.02.005. URL: https://linkinghub.elsevier.com/retrieve/pii/S0309170813000304.

Lo, C.P. (1986). "Applied remote sensing". In: *Geocarto International* 1.4, pp. 60–60. ISSN: 1010-6049. DOI: 10.1080/10106048609354071. URL: http://www.tandfonline.com/doi/abs/10.1080/10106048609354071.

Mahfouf, J-F. et al. (1995). "The Land Surface Scheme ISBA within the Météo-France Climate Model ARPEGE. Part I. Implementation and Preliminary Results". In: *Journal of Climate* 8.8, pp. 2039–2057. ISSN: 0894-8755. DOI: 10.1175/1520-0442(1995)008<2039:TLSSIW>2.0.CO;2. URL: http://journals.ametsoc.org/doi/10.1175/1520-0442(1995)008{\%}3C2039:TLSSIW{\%}3E2.0.CO;2.

Mahfouf, Jean-François (2010). "Assimilation of satellite-derived soil moisture from ASCAT in a limited-area NWP model". In: *Quarterly Journal of the Royal Meteorological Society*, n/a–n/a. ISSN: 00359009. DOI: 10.1002/qj.602. URL: http://doi.wiley.com/10.1002/qj.602.

Mahfouf, J.-F. et al. (2009). "A comparison of two off-line soil analysis schemes for assimilation of screen level observations". In: *Journal of Geophysical Research* 114.D8, p. D08105. ISSN: 0148-0227. DOI: 10.1029/2008JD011077. URL: http://doi.wiley.com/10.1029/2008JD011077.

Maisongrande, P., B. Duchemin, and G. Dedieu (2004). "VEGETATION/SPOT: an operational mission for the Earth monitoring; presentation of new standard products". In: *International Journal of Remote Sensing* 25.1, pp. 9–14. ISSN: 0143-1161. DOI: 10.1080/0143116031000115265. URL: https://www.tandfonline.com/doi/full/10.1080/0143116031000115265.

Manabe, Syukuro (1969). "The Atmospheric Circulation And The Hydrology Of The Earth's Surface". In: *Monthly Weather Review* 97.11, pp. 739–774. ISSN: 0027-0644. DOI: 10.1175/1520-0493(1969)097<0739:CATOC>2.3.CO;2. URL: http://journals.ametsoc.org/doi/10.1175/1520-0493(1969)097%3C0739:CATOC%3E2.3.CO;2.

Marschallinger, Bernhard, Christoph Paulik, and Tim Jacobs (2019). "Soil Water Index. Collection 1Km". In: p. 31. URL: https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS1_PUM_SWI1km-V1_I1.10.pdf.

Massari, Christian et al. (2018). "Exploiting Satellite-Based Surface Soil Moisture for Flood Forecasting in the Mediterranean Area: State Update Versus Rainfall Correction". In: *Remote Sensing* 10.2, p. 292. ISSN: 2072-4292. DOI: 10.3390/rs10020292. URL: http://www.mdpi.com/2072-4292/10/2/292.

Masson, Valéry (2000). "A Physically-Based Scheme For The Urban Energy Budget In Atmospheric Models". In: *Boundary-Layer Meteorology* 94.3, pp. 357–397. ISSN: 0006-8314. DOI: 10.1023/A:1002463829265. URL: http://link.springer.com/10.1023/A:1002463829265.

Mätzler, C. and A. Standley (2000). "Technical note: Relief effects for passive microwave remote sensing". In: *Int. J. Remote Sens.* 21.12, pp. 2403–2412. ISSN: 0143-1161. DOI: 10.1080/01431160050030538. URL: https://www.tandfonline.com/doi/full/10.1080/01431160050030538.

McGuffie, K. and A. Henderson-Sellers (2001). "Forty years of numerical climate modelling". In: *International Journal of Climatology* 21.9. ISSN: 0899-8418. DOI: 10.1002/joc.632.

McKee, Thomas B (1995). "Drought monitoring with multiple time scales". In: *Proceedings of 9th Conference on Applied Climatology, Boston, 1995*.

McNally, A. P. (2002). "A note on the occurrence of cloud in meteorologically sensitive areas and the implications for advanced infrared sounders". In: *Quarterly Journal of the Royal Meteorological Society* 128.585, pp. 2551–2556. ISSN: 1477870X. DOI: 10.1256/qj.01.206. URL: http://doi.wiley.com/10.1256/qj.01.206.

McNally, Amy et al. (2017). "A land data assimilation system for sub-Saharan Africa food and water security applications". In: *Scientific Data* 4.1, p. 170012. ISSN: 2052-4463. DOI: 10.1038/sdata.2017.12. URL: http://www.nature.com/articles/sdata201712.

Meesters, A.G.C.A., R.A.M. DeJeu, and M. Owe (2005). "Analytical Derivation of the Vegetation Optical Depth From the Microwave Polarization Difference Index". In: *IEEE Geoscience and Remote Sensing Letters* 2.2, pp. 121–123. ISSN: 1545-598X. DOI: 10.1109/LGRS.2005.843983. URL: http://ieeexplore.ieee.org/document/1420287/.

Meroni, M. et al. (2009). "Remote sensing of solar-induced chlorophyll fluorescence: Review of methods and applications". In: *Remote Sensing of Environment* 113.10, pp. 2037–2051. ISSN: 00344257. DOI: 10.1016/j.rse.2009.05.003. URL: https://linkinghub.elsevier.com/retrieve/pii/S003442570900162X.

Mesinger, Fedor et al. (2006). "North American Regional Reanalysis". In: *Bulletin of the American Meteorological Society* 87.3, pp. 343–360. ISSN: 0003-0007. DOI: 10.1175/BAMS-87-3-343. URL: https://journals.ametsoc.org/doi/10.1175/BAMS-87-3-343.

Mialon, Arnaud et al. (2020). "Evaluation of the Sensitivity of SMOS L-VOD to Forest Above-Ground Biomass at Global Scale". In: *Remote Sensing* 12.9, p. 1450. ISSN: 2072-4292. DOI: 10.3390/rs12091450. URL: https://www.mdpi.com/2072-4292/12/9/1450.

Mironov, Dmitrii (2008). *Parameterization of Lakes in Numerical Weather Prediction: Description of a Lake Model*. Tech. rep. German Weather Service,

Offenbach am Main, Germany. URL: http://nwpi.krc.karelia.ru/flake/papers/tr{_}11{_}flake.pdf.

Mo, T. et al. (1982). "A model for microwave emission from vegetation-covered fields". In: *Journal of Geophysical Research* 87.C13, p. 11229. ISSN: 0148-0227. DOI: 10.1029/JC087iC13p11229. URL: http://doi.wiley.com/10.1029/JC087iC13p11229.

Mocko, David M. et al. (2021). "Assimilation of Vegetation Conditions Improves the Representation of Drought over Agricultural Areas". In: *Journal of Hydrometeorology* 22.5, pp. 1085–1098. ISSN: 1525-755X. DOI: 10.1175/JHM-D-20-0065.1. URL: https://journals.ametsoc.org/view/journals/hydr/22/5/JHM-D-20-0065.1.xml.

Moesinger, Leander et al. (2020). "The global long-term microwave Vegetation Optical Depth Climate Archive (VODCA)". In: *Earth System Science Data* 12.1, pp. 177–196. ISSN: 1866-3516. DOI: 10.5194/essd-12-177-2020. URL: https://essd.copernicus.org/articles/12/177/2020/.

Momen, Mostafa et al. (2017). "Interacting Effects of Leaf Water Potential and Biomass on Vegetation Optical Depth". In: *Journal of Geophysical Research: Biogeosciences* 122.11, pp. 3031–3046. ISSN: 21698953. DOI: 10.1002/2017JG004145. URL: http://doi.wiley.com/10.1002/2017JG004145.

Mozny, Martin et al. (2012). "Use of a soil moisture network for drought monitoring in the Czech Republic". In: *Theoretical and Applied Climatology* 107.1-2, pp. 99–111. ISSN: 0177-798X. DOI: 10.1007/s00704-011-0460-6. URL: http://link.springer.com/10.1007/s00704-011-0460-6.

Mucia, Anthony et al. (2020). "From Monitoring to Forecasting Land Surface Conditions Using a Land Data Assimilation System: Application over the Contiguous United States". In: *Remote Sensing* 12.12, p. 2020. ISSN: 2072-4292. DOI: 10.3390/rs12122020. URL: https://www.mdpi.com/2072-4292/12/12/2020.

Nassiri Mahallati, Mehdi (2020). "Advances in modeling saffron growth and development at different scales". In: *Saffron*. Elsevier, pp. 139–167. DOI: 10.1016/B978-0-12-818638-1.00009-5. URL: https://linkinghub.elsevier.com/retrieve/pii/B9780128186381000095.

National Aeronautics and Space Administration (2020). *Landsat Science: Practical Uses*. URL: https://landsat.gsfc.nasa.gov/about/practical-uses (visited on 03/22/2021).

Nicotra, A.B. et al. (2010). "Plant phenotypic plasticity in a changing climate". In: *Trends in Plant Science* 15.12, pp. 684–692. ISSN: 13601385. DOI: 10.1016/j.tplants.2010.09.008. URL: https://linkinghub.elsevier.com/retrieve/pii/S1360138510001986.

Njoku, Eni G. and Dara Entekhabi (1996). "Passive microwave remote sensing of soil moisture". In: *Journal of Hydrology* 184.1-2, pp. 101–129. ISSN: 00221694. DOI: 10.1016/0022-1694(95)02970-2. URL: https://linkinghub.elsevier.com/retrieve/pii/0022169495029702.

Noilhan, J. and J.-F. Mahfouf (1996). "The ISBA land surface parameterisation scheme". In: *Global and Planetary Change* 13.1-4, pp. 145–

159. ISSN: 09218181. DOI: 10.1016/0921-8181(95)00043-7. URL: https://linkinghub.elsevier.com/retrieve/pii/0921818195000437.

Noilhan, J. and S. Planton (1989). "A Simple Parameterization of Land Surface Processes for Meteorological Models". In: *Monthly Weather Review* 117.3, pp. 536–549. ISSN: 0027-0644. DOI: 10.1175/1520-0493(1989)117<0536:ASPOLS>2.0.CO;2. URL: http://journals.ametsoc.org/doi/10.1175/1520-0493(1989)117%3C0536:ASPOLS%3E2.0.CO;2.

Noy-Meir, I (1973). "Desert Ecosystems: Environment and Producers". In: *Annual Review of Ecology and Systematics* 4.1, pp. 25–51. ISSN: 0066-4162. DOI: 10.1146/annurev.es.04.110173.000325. URL: http://www.annualreviews.org/doi/10.1146/annurev.es.04.110173.000325.

Obasi, G.O.P. (1994). "WMO's role in the international decade for natural disaster reduction". In: *Bulletin of the American Meteorological Society* 75.9, pp. 1655–1662.

Oki, Taikan and Y. C. Sud (1998). "Design of Total Runoff Integrating Pathways (TRIP)—A Global River Channel Network". In: *Earth Interactions* 2.1, pp. 1–37. ISSN: 1087-3562. DOI: 10.1175/1087-3562(1998)002<0001:DOTRIP>2.3.CO;2. URL: http://journals.ametsoc.org/doi/10.1175/1087-3562(1998)002{\%}3C0001:DOTRIP{\%}3E2.3.CO;2.

Onoda, Masami (2017). "Conclusion". In: *Satellite Earth Observations and Their Impact on Society and Policy*. Singapore: Springer Singapore, pp. 205–213. DOI: 10.1007/978-981-10-3713-9_17. URL: http://link.springer.com/10.1007/978-981-10-3713-9{_}17.

Otkin, Jason A. et al. (2013). "Examining Rapid Onset Drought Development Using the Thermal Infrared–Based Evaporative Stress Index". In: *Journal of Hydrometeorology* 14.4, pp. 1057–1074. ISSN: 1525-755X. DOI: 10.1175/JHM-D-12-0144.1. URL: http://journals.ametsoc.org/doi/10.1175/JHM-D-12-0144.1.

Otkin, Jason A. et al. (2018). "Flash Droughts: A Review and Assessment of the Challenges Imposed by Rapid-Onset Droughts in the United States". In: *Bulletin of the American Meteorological Society* 99.5, pp. 911–919. ISSN: 0003-0007. DOI: 10.1175/BAMS-D-17-0149.1. URL: https://journals.ametsoc.org/view/journals/bams/99/5/bams-d-17-0149.1.xml.

Owe, M., R. De Jeu, and J. Walker (2001). "A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization difference index". In: *IEEE Transactions on Geoscience and Remote Sensing* 39.8, pp. 1643–1654. ISSN: 01962892. DOI: 10.1109/36.942542.

Owe, Manfred, Richard de Jeu, and Thomas Holmes (2008). "Multisensor historical climatology of satellite-derived global land surface moisture". In: *Journal of Geophysical Research* 113.F1, F01002. ISSN: 0148-0227. DOI: 10.1029/2007JF000769. URL: http://doi.wiley.com/10.1029/2007JF000769.

Owens, R G and Tim Hewson (2018). "ECMWF Forecast User Guide". In: DOI: 10.21957/m1cs7h. URL: https://www.ecmwf.int/node/16559.

Palmer, Wayne C. (1965). *Meteorological Drought*. Tech. rep., p. 58. URL: https://www.ncdc.noaa.gov/temp-and-precip/drought/docs/palmer.pdf.

Pastorello, Gilberto et al. (2020). "The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data". In: *Scientific Data* 7.1, p. 225. ISSN: 2052-4463. DOI: 10.1038/s41597-020-0534-3. URL: http://www.nature.com/articles/s41597-020-0534-3.

Pitman, A. J. (2003). "The evolution of, and revolution in, land surface schemes designed for climate models". In: *International Journal of Climatology* 23.5, pp. 479–510. ISSN: 0899-8418. DOI: 10.1002/joc.893. URL: http://doi.wiley.com/10.1002/joc.893.

Poorter, Hendrik, Carlo Remkes, and Hans Lambers (1990). "Carbon and nitrogen economy of 24 wild species differing in relative growth rate". In: *Plant Physiology* 94.2, pp. 621–627. ISSN: 00320889. DOI: 10.1104/pp.94.2.621.

Price, J.C. (1993). "Estimating leaf area index from satellite data". In: *IEEE Transactions on Geoscience and Remote Sensing* 31.3, pp. 727–734. ISSN: 01962892. DOI: 10.1109/36.225538. URL: http://ieeexplore.ieee.org/document/225538/.

Reichle, R. H. and R. D. Koster (2004). "Bias reduction in short records of satellite soil moisture". In: *Geophysical Research Letters* 31.19, p. L19501. ISSN: 0094-8276. DOI: 10.1029/2004GL020938. URL: http://doi.wiley.com/10.1029/2004GL020938.

Reichle, Rolf H. (2005). "Global assimilation of satellite surface soil moisture retrievals into the NASA Catchment land surface model". In: *Geophysical Research Letters* 32.2, p. L02404. ISSN: 0094-8276. DOI: 10.1029/2004GL021700. URL: http://doi.wiley.com/10.1029/2004GL021700.

Reichle, Rolf H. et al. (2007). "Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR)". In: *Journal of Geophysical Research* 112.D9, p. D09108. ISSN: 0148-0227. DOI: 10.1029/2006JD008033. URL: http://doi.wiley.com/10.1029/2006JD008033.

Reichle, Rolf H. et al. (2010). "Assimilation of Satellite-Derived Skin Temperature Observations into Land Surface Models". In: *Journal of Hydrometeorology* 11.5, pp. 1103–1122. ISSN: 1525-7541. DOI: 10.1175/2010JHM1262.1. URL: http://journals.ametsoc.org/doi/10.1175/2010JHM1262.1.

Reichstein, Markus et al. (2005). "On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm". In: *Global Change Biology* 11.9, pp. 1424–1439. ISSN: 1354-1013. DOI: 10.1111/j.1365-2486.2005.001002.x. URL: http://doi.wiley.com/10.1111/j.1365-2486.2005.001002.x.

Rodell, M. et al. (2004). "The Global Land Data Assimilation System". In: *Bulletin of the American Meteorological Society* 85.3, pp. 381–394. ISSN: 0003-0007. DOI: 10.1175/BAMS-85-3-381. URL: http://journals.ametsoc.org/doi/10.1175/BAMS-85-3-381.

Rodell, M. et al. (2005). "Evaluation of 10 Methods for Initializing a Land Surface Model". In: *Journal of Hydrometeorology* 6.2, pp. 146–155. ISSN: 1525-7541. DOI: 10.1175/JHM414.1. URL: http://journals.ametsoc.org/doi/10.1175/JHM414.1.

Rodríguez-Fernández, Nemesio J. et al. (2018). "An evaluation of SMOS L-band vegetation optical depth (L-VOD) data sets: high sensitivity of L-VOD to above-ground biomass in Africa". In: *Biogeosciences* 15.14, pp. 4627–4645. ISSN: 1726-4189. DOI: 10.5194/bg-15-4627-2018. URL: https://bg.copernicus.org/articles/15/4627/2018/.

Rouse, JW et al. (1974). "Monitoring vegetation systems in the Great Plains with ERTS". In: *NASA special publication* 351.1974, p. 309.

Saatchi, Sassan et al. (2011). "Impact of spatial variability of tropical forest structure on radar estimation of aboveground biomass". In: *Remote Sensing of Environment* 115.11, pp. 2836–2849. ISSN: 00344257. DOI: 10.1016/j.rse.2010.07.015. URL: https://linkinghub.elsevier.com/retrieve/pii/S0034425711001313.

Saleh, Kauzar et al. (2006). "Impact of rain interception by vegetation and mulch on the L-band emission of natural grass". In: *Remote Sensing of Environment* 101.1, pp. 127–139. ISSN: 00344257. DOI: 10.1016/j.rse.2005.12.004. URL: https://linkinghub.elsevier.com/retrieve/pii/S0034425705004219.

Sawada, Yohei and Toshio Koike (2016). "Towards ecohydrological drought monitoring and prediction using a land data assimilation system: A case study on the Horn of Africa drought (2010-2011)". In: *Journal of Geophysical Research: Atmospheres* 121.14, pp. 8229–8242. ISSN: 2169897X. DOI: 10.1002/2015JD024705. URL: http://doi.wiley.com/10.1002/2015JD024705.

Sawada, Yohei et al. (2020). "Monitoring and Predicting Agricultural Droughts for a Water-Limited Subcontinental Region by Integrating a Land Surface Model and Microwave Remote Sensing". In: *IEEE Transactions on Geoscience and Remote Sensing* 58.1, pp. 14–33. ISSN: 0196-2892. DOI: 10.1109/TGRS.2019.2927342. URL: https://ieeexplore.ieee.org/document/8847390/.

Schalie, R. van der et al. (2017). "The merging of radiative transfer based surface soil moisture data from SMOS and AMSR-E". In: *Remote Sensing of Environment* 189, pp. 180–193. ISSN: 00344257. DOI: 10.1016/j.rse.2016.11.026. URL: https://linkinghub.elsevier.com/retrieve/pii/S0034425716304734.

Schellekens, Jaap et al. (2017). "A global water resources ensemble of hydrological models: the eartH2Observe Tier-1 dataset". In: *Earth System Science Data* 9.2, pp. 389–413. ISSN: 1866-3516. DOI: 10.5194/essd-9-389-2017. URL: https://essd.copernicus.org/articles/9/389/2017/.

Schulze, E. D. (1970). "The CO2-gas exchange of Fagus sylvatica in relation to climatic factors in the field." In: *Flora, Jena* 159.1/2, pp. 177–232.

Scipal, K., M. Drusch, and W. Wagner (2008). "Assimilation of a ERS scatterometer derived soil moisture index in the ECMWF numerical weather prediction system". In: *Advances in Water Resources* 31.8, pp. 1101–1112. ISSN: 03091708. DOI: 10.1016/j.advwatres.2008.04.013. URL: https://linkinghub.elsevier.com/retrieve/pii/S0309170808000730.

Scott, David W. (1992). *Multivariate Density Estimation*. Wiley Series in Probability and Statistics. Wiley. ISBN: 9780471547709. DOI: 10.1002/9780470316849. URL: https://onlinelibrary.wiley.com/doi/book/10.1002/9780470316849.

Sellers, P. J. et al. (1986). "A Simple Biosphere Model (SIB) for Use within General Circulation Models". In: *Journal of the Atmospheric Sciences* 43.6, pp. 505–531. ISSN: 0022-4928. DOI: 10.1175/1520-0469(1986)043<0505:ASBMFU>2.0.CO;2. URL: http://journals.ametsoc.org/doi/10.1175/1520-0469(1986)043%3C0505:ASBMFU%3E2.0.CO;2.

Sellers, P.J. et al. (1992). "Canopy reflectance, photosynthesis, and transpiration. III. A reanalysis using improved leaf models and a new canopy integration scheme." In: *Remote Sensing of Environment* 42.3, pp. 187–216. ISSN: 00344257. DOI: 10.1016/0034-4257(92)90102-P. URL: https://linkinghub.elsevier.com/retrieve/pii/003442579290102P.

Seneviratne, Sonia I. et al. (2006). "Land–atmosphere coupling and climate change in Europe". In: *Nature* 443.7108, pp. 205–209. ISSN: 0028-0836. DOI: 10.1038/nature05095. URL: http://www.nature.com/articles/nature05095.

Shamambo, Daniel et al. (2019). "Interpretation of ASCAT Radar Scatterometer Observations Over Land: A Case Study Over Southwestern France". In: *Remote Sensing* 11.23, p. 2842. ISSN: 2072-4292. DOI: 10.3390/rs11232842. URL: https://www.mdpi.com/2072-4292/11/23/2842.

Shamambo, Daniel Chiyeka (2020). "Assimilation de données satellitaires pour le suivi des ressources en eau dans la zone Euro-Méditerranée". PhD thesis. URL: http://www.theses.fr/2020TOU30143/document.

Sheffield, Justin et al. (2014). "A Drought Monitoring and Forecasting System for Sub-Sahara African Water Resources and Food Security". In: *Bulletin of the American Meteorological Society* 95.6, pp. 861–882. ISSN: 0003-0007. DOI: 10.1175/BAMS-D-12-00124.1. URL: http://journals.ametsoc.org/doi/10.1175/BAMS-D-12-00124.1.

Shiklomanov, Igor A (1993). "World Fresh Water Resources". In: *Water in Crisis: A guide to the World's Fresh Water Resources*. Ed. by Peter H Gleick. New York, NY: Oxford University Press. Chap. 2. ISBN: 0-19-507627-3.

Skamarock, William et al. (2005). *A Description of the Advanced Research WRF Version 2*. Tech. rep. Boulder, CO: National Center for Atmospheric Research,Mesoscale and Microscale Meteorology Division.

Sternberg, Troy (2011). "Regional drought has a global impact". In: *Nature* 472.7342, pp. 169–169.

SURFEX, C Szczypta, and P. Le Moigne. *SURFEX*. URL: http://www.umr-cnrm.fr/surfex/spip.php?article428 (visited on 04/09/2021).

Svoboda, M et al. (2002). "The Drought Monitor". In: *Bulletin of the American Meteorological Society* 83.8, pp. 1181–1190. ISSN: 0003-0007. DOI: 10.1175/1520-0477(2002)083<1181:TDM>2.3.CO;2. URL: http://journals.ametsoc.org/doi/abs/10.1175/1520-0477(2002)083{\%}3C1181:TDM{\%}3E2.3.CO;2.

Tall, Moustapha et al. (2019). "Towards a Long-Term Reanalysis of Land Surface Variables over Western Africa: LDAS-Monde Applied over Burkina Faso from 2001 to 2018". In: *Remote Sensing* 11.6, p. 735. ISSN: 2072-4292. DOI: 10.3390/rs11060735. URL: https://www.mdpi.com/2072-4292/11/6/735.

Teubner, Irene E. et al. (2018). "Assessing the relationship between microwave vegetation optical depth and gross primary production". In: *International Journal of Applied Earth Observation and Geoinformation* 65, pp. 79–91. ISSN: 03032434. DOI: 10.1016/j.jag.2017.10.006. URL: https://linkinghub.elsevier.com/retrieve/pii/S0303243417302258.

Teubner, Irene E. et al. (2021). "Impact of temperature and water availability on microwave-derived gross primary production". In: *Biogeosciences* 18.11, pp. 3285–3308. ISSN: 1726-4189. DOI: 10.5194/bg-18-3285-2021. URL: https://bg.copernicus.org/articles/18/3285/2021/.

Tramontana, Gianluca et al. (2016). "Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms". In: *Biogeosciences* 13.14, pp. 4291–4313. ISSN: 1726-4189. DOI: 10.5194/bg-13-4291-2016. URL: https://bg.copernicus.org/articles/13/4291/2016/.

Trenberth, Kevin E., John T. Fasullo, and Jeffrey Kiehl (2009). "Earth's global energy budget". In: *Bulletin of the American Meteorological Society* 90.3, pp. 311–323. ISSN: 00030007. DOI: 10.1175/2008BAMS2634.1.

Tucker, C. J. and P. J. Sellers (1986). "Satellite remote sensing of primary production". In: *International Journal of Remote Sensing* 7.11, pp. 1395–1416. ISSN: 0143-1161. DOI: 10.1080/01431168608948944. URL: https://www.tandfonline.com/doi/full/10.1080/01431168608948944.

Tucker, C. J. et al. (1991). "Mean and inter-year variation of growing-season normalized difference vegetation index for the Sahel 1981-1989". In: *International Journal of Remote Sensing* 12.6, pp. 1133–1135. ISSN: 0143-1161. DOI: 10.1080/01431169108929717. URL: https://www.tandfonline.com/doi/full/10.1080/01431169108929717.

Tucker, Compton J (1979). "Red and photographic infrared linear combinations for monitoring vegetation". In: *Remote sensing of Environment* 8.2, pp. 127–150.

Ulaby, F.T. et al. (1984). "Relating the microwave backscattering coefficient to leaf area index". In: *Remote Sens. Environ.* 14.1-3, pp. 113–133. ISSN: 00344257. DOI: 10.1016/0034-4257(84)90010-5. URL: https://linkinghub.elsevier.com/retrieve/pii/0034425784900105.

Unganai, Leonard S and Felix N Kogan (1998). "Drought Monitoring and Corn Yield Estimation in Southern Africa from AVHRR Data". In: *Remote Sensing of Environment* 63.3, pp. 219–232. ISSN: 00344257. DOI: 10.1016/S0034-4257(97)00132-6. URL: https://linkinghub.elsevier.com/retrieve/pii/S0034425797001326.

United States Geological Survey (2019). *The Natural Water Cycle*. URL: https://www.usgs.gov/media/images/natural-water-cycle-0.

USCRN. *USCRN Soil Climate Observations Documentation*. URL: https://www.ncei.noaa.gov/pub/data/uscrn/documentation/site/sensors/soil/Descriptions/Soil{_}StevensHydra-SDI-12.pdf.

USDA National Agricultural Statistics Service Cropland Data Layer (2021). URL: https://nassgeodata.gmu.edu/CropScape/.

Vasiliades, L., A. Loukas, and G. Patsonas (2009). "Evaluation of a statistical downscaling procedure for the estimation of climate change impacts on droughts". In: *Natural Hazards and Earth System Sciences* 9.3, pp. 879–894. ISSN: 1684-9981. DOI: 10.5194/nhess-9-879-2009. URL: https://nhess.copernicus.org/articles/9/879/2009/.

Vasiliades, Lampros and Athanasios Loukas (2009). "Hydrological response to meteorological drought using the Palmer drought indices in Thessaly, Greece". In: *Desalination* 237.1-3, pp. 3–21. ISSN: 00119164. DOI: 10.1016/j.desal.2007.12.019. URL: https://linkinghub.elsevier.com/retrieve/pii/S001191640800667X.

Verger, Aleixandre, Frederic Baret, and Marie Weiss (2014). "Near Real-Time Vegetation Monitoring at Global Scale". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7.8, pp. 3473–3481. ISSN: 1939-1404. DOI: 10.1109/JSTARS.2014.2328632. URL: http://ieeexplore.ieee.org/document/6842660/.

Vitart, F. et al. (2017). "The Subseasonal to Seasonal (S2S) Prediction Project Database". In: *Bulletin of the American Meteorological Society* 98.1, pp. 163–173. ISSN: 0003-0007. DOI: 10.1175/BAMS-D-16-0017.1. URL: https://journals.ametsoc.org/doi/10.1175/BAMS-D-16-0017.1.

Viterbo, Pedro et al. (1999). "The representation of soil moisture freezing and its impact on the stable boundary layer". In: *Quarterly Journal of the Royal Meteorological Society* 125.559, pp. 2401–2426. ISSN: 00359009. DOI: 10.1002/qj.49712555904. URL: http://doi.wiley.com/10.1002/qj.49712555904.

Vogt, J. et al. (Dec. 2011). "Developing a European Drought Observatory for Monitoring, Assessing and Forecasting Droughts across the European Continent". In: *AGU Fall Meeting Abstracts*. Vol. 2011, NH24A–07.

Wagner, Wolfgang, Guido Lemoine, and Helmut Rott (1999). "A Method for Estimating Soil Moisture from ERS Scatterometer and Soil Data". In: *Remote Sensing of Environment* 70.2, pp. 191–207. ISSN: 00344257. DOI: 10.1016/S0034-4257(99)00036-X. URL: https://linkinghub.elsevier.com/retrieve/pii/S003442579900036X.

Watson, D.J. (1947). "Comparative Physiological Studies on the Growth of Field Crops: I. Variation in Net Assimilation Rate and Leaf Area between Species and Varieties, and within and between Years". In: *Annals of Botany* 11.41, pp. 41–76. URL: https://www.jstor.org/stable/42907002?seq=1.

Western Governors Policy Office (1977). *Directory of Federal Drought Assistance*. Washington, DC: Institute for Policy Research for the Western Region Drought Action Task Force, the US Department of Agriculture, pp. 5–11. ISBN: 0148-5091.

Wigneron, Jean-Pierre (2002). "Modeling approaches to assimilating L band passive microwave observations over land surfaces". In: *Journal of Geophysical Research* 107.D14, p. 4219. ISSN: 0148-0227. DOI: 10.1029/2001JD000958. URL: http://doi.wiley.com/10.1029/2001JD000958.

Wigneron, Jean-Pierre et al. (1993). "Inversion of surface parameters from passive microwave measurements over a soybean field". In: *Remote Sensing of Environment* 46.1, pp. 61–72. ISSN: 00344257. DOI: 10.1016/0034-4257(93)90032-S. URL: https://linkinghub.elsevier.com/retrieve/pii/003442579390032S.

Wilhite, D. A., M. J. Hayes, and M. D. Svoboda (2000). "Drought Monitoring and Assessment: Status and Trends in the United States". In: pp. 149–160. DOI: 10.1007/978-94-015-9472-1_11.

Wilhite, Donald A (2000). "Drought as a Natural Hazard: Conceptsand Definitions". In: URL: https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1068{\&}context=droughtfacpub.

Wilhite, Donald A. and Michael H. Glantz (1985). "Understanding the drought phenomenon: The role of definitions". In: *Water International*, pp. 11–27. DOI: 10.4324/9780429301735-2.

Wilhite, Donald A., M.V.K. Sivakumar, and Deborah A. Wood (2000). "Early Warning Systems for Drought Preparedness and Drought Management". In: *Proceedings of an Expert Group Meeting held 5-7 September, 2000, in Lisbon, Portugal*. 1037, pp. 57–73. URL: http://www.droughtmanagement.info/literature/WMO_early_warning_systems_drought_prepardness_2000.pdf#page=105.

Yang, Wenze et al. (2006). "MODIS leaf area index products: from validation to algorithm improvement". In: *IEEE Transactions on Geoscience and Remote Sensing* 44.7, pp. 1885–1898. ISSN: 0196-2892. DOI: 10.1109/TGRS.2006.871215. URL: http://ieeexplore.ieee.org/document/1645288/.

Yevjevich, Vujica (1969). "An objective approach to definitions and investigations of continental hydrologic droughts". In: *Journal of Hydrology* 7.3, p. 353. ISSN: 00221694. DOI: 10.1016/0022-1694(69)90110-3.

Yucel, I. et al. (2015). "Calibration and evaluation of a flood forecasting system: Utility of numerical weather prediction model, data assimilation and satellite-based rainfall". In: *Journal of Hydrology* 523, pp. 49–66. ISSN: 00221694. DOI: 10.1016/j.jhydrol.2015.01.042. URL: https://linkinghub.elsevier.com/retrieve/pii/S0022169415000591.

Zaitchik, Benjamin F., Matthew Rodell, and Rolf H. Reichle (2008). "Assimilation of GRACE Terrestrial Water Storage Data into a Land Surface Model: Results for the Mississippi River Basin". In: *Journal of Hydrometeorology* 9.3, pp. 535–548. ISSN: 1525-7541. DOI: 10.1175/2007JHM951.1. URL: http://journals.ametsoc.org/doi/10.1175/2007JHM951.1.

Zakharova, E. et al. (2012). "Spatial and temporal variability of biophysical variables in southwestern France from airborne L-band radiometry". In: *Hydrology and Earth System Sciences* 16.6, pp. 1725–1743. ISSN: 1607-7938. DOI: 10.5194/hess-16-1725-2012. URL: https://hess.copernicus.org/articles/16/1725/2012/.

Zhang, Yong-Fei et al. (2014). "Assimilation of MODIS snow cover through the Data Assimilation Research Testbed and the Community Land Model version 4". In: *Journal of Geophysical Research: Atmospheres* 119.12, pp. 7091–7103. ISSN: 2169897X. DOI: 10.1002/2013JD021329. URL: http://doi.wiley.com/10.1002/2013JD021329.

Zink, Matthias et al. (2016). "The German drought monitor". In: *Environmental Research Letters* 11.7, p. 074002. ISSN: 1748-9326. DOI: 10.1088/1748-9326/11/7/074002. URL: https://iopscience.iop.org/article/10.1088/1748-9326/11/7/074002.