UNIVERSITÉ DU QUÉBEC À MONTRÉAL

COHÉRENCE DES PRODUITS DE PRÉCIPITATION À HAUTE RÉSOLUTION POUR L'EST DE L'AMÉRIQUE DU NORD

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAITRISE EN SCIENCES DE L'ATMOSPHÈRE

PAR

TANGUI PICART

DÉCEMBRE 2022

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

REMERCIEMENTS

Petit, en Bretagne, mon père nous pointait souvent l'océan en nous disant, non sans une note d'ironie, que lui était suffisamment grand pour apercevoir les côtes de l'Amérique de l'autre côté de l'Atlantique. L'ironie a pris une forme toute particulière quand j'ai commencé ma maitrise 20 ans plus tard depuis les mêmes côtes, en regardant dans la même direction, attendant une accalmie dans la pandémie pour pouvoir finalement rejoindre les terres québécoises.

Plus que les simples remerciements d'usages naturellement mérités par mes superviseurs, je tiens à saluer particulièrement l'intarissable humanité du Professeur Alejandro Di Luca et le soutien du professeur René Laprise. Il me tient également à cœur de remercier Katja WINGER, Frédérick TOUPIN, la Professeur Julie Thériault et Leticia Hernandez Diaz et Éva Monteiro pour leurs contributions à cette petite aventure commencée à distance.

Béatrice, Charles-Elie, Charlotte, Karel, Delphine, Catherine et Nicolas ont également une juste place dans ces remerciements pour avoir mêlé leurs mots et idées à mes travaux. Ma dernière pensée sera pour Henry et le temps qu'il nous a consacré.

TABLE DES MATIÈRES

LISTE DES FIGURES

LISTE DES TABLEAUX

LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES

AO	Atlantic Ocean
CAGL	Southeast Canada and the Great Lakes
CMORPH	Climate Prediction Center morphing technique
CONUS	CONtiguous United States
CPC	Climate Prediction Center
ECMWF	European centre for Medium-range Weather Forecasts
EMC	Environmental Modeling Center
ERA	ECMWF Re-Analysis
GM	Gulf of Mexico
GPCC	Global Precipitation Climatology Centre
GPCP	Global Precipitation Climatology Project
GPM	Global Precipitation Measurement mission
GSMaP	Global Satellite Mapping of Precipitation
IMERG	Integrated Multi-satellitE Retrievals for Global precipitation measurement
MSWEP	Multi-Source Weighted-Ensemble Precipitation
NCEP	National Centers for Environmental Prediction
NEUS	Northeast United States
PERSIANN	The Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks
RFC	River Forecast Center
SEUS	Southeast United States
TRMM	Tropical Rainfall Measuring Mission multi-satellite precipitation analysis
US	United States
UTC	Coordinated Universal Time

LISTE DES SYMBOLES ET DES UNITÉS

$°$	Degré
$mm.(3h)^{.}$	Accumulation de précipitation à 3 h
$mm.d^{-1}$	Accumulation de précipitation journalière
P_M	Moyenne temporelle de la précipitation
N_T	Nombre total de pas de temps
ΔP_M	Différence absolue de précipitation moyenne, i.e., biais absolue
B_j	Intervalle d'intensité de précipitation
P_{B_j}	Accumulation totale des évènements de précipitation dont l'intensité est comprise dans B_j
ΔP_B	Différence absolue de P_{B_j}
F	Fréquence de précipitation, i.e. la fraction de périodes précipitantes
I	Intensité moyenne de précipitation, i.e. précipitation moyenne pour les périodes précipitantes
i	Intensité minimale entre deux produits
f	Fréquence minimale entre deux produits
ΔF	Différence de fréquence
ΔI	Différence d'intensité
ΔP_{IF}	Différence basée sur la décomposition de la précipitation moyenne en produit de l'intensité et la fréquence de précipitation
ΔP_A	Différence absolue moyenne, i.e., la moyenne temporelle de la différence absolue de la précipitation à 3 h
$\overline{(\Delta P_X)}^p$	Cohérence moyenne du produit p pour la distance ΔP_X
max_7	Incertitude de l'ensemble des produits
max_4	Incertitude des quatre produits les plus probables

RÉSUMÉ

Plusieurs estimations de la précipitation fondées sur des observations sont aujourd'hui disponibles pour de hautes résolutions temporelle (≤3 h) et spatiale (≤0,25°). La cohérence entre ces estimations dépend de plusieurs facteurs, comme la métrique avec laquelle les produits sont comparés, la région et la saison. Nous proposons ici une méthodologie permettant de quantifier l'incertitude d'un ensemble de produits et d'en identifier les moins probables en chaque point de grille. L'approche est appliquée sur une région zone couvrant l'est de l'Amérique du Nord entre 2015 et 2019 pour sept produits à haute résolution couramment utilisés. Quatre métriques sont utilisées pour quantifier la cohérence entre les produits selon différents aspects de la précipitation tels que l'accumulation totale, la distribution en intensité, l'intensité/fréquence et la chronologie des événements. D'importantes variations régionales et saisonnières de l'incertitude sont constatées. Les produits sont moins cohérents (et donc plus incertains) au Canada qu'aux États-Unis, ce qui reflète la moindre contribution des observations in situ en raison de la taille réduite du réseau d'observation. Aux latitudes moyennes septentrionales, la cohérence des produits est la plus élevée en été, tandis qu'aux latitudes moyennes méridionales, elle est la plus élevée en hiver. En été, les variations du cycle diurne des précipitations sont globalement plus cohérentes que l'amplitude du cycle ; la divergence de la chronologie des précipitations est cependant constante tout au long de la journée. L'absence de référence absolue pour comparer les produits et la grande variabilité de leur cohérence rendent difficile l'identification des produits les plus appropriés ; notre méthodologie permet cependant d'identifier les plus probables.

MOTS CLÉS

Précipitation, observations, Amérique du Nord, erreur

INTRODUCTION

La climatologie de précipitation est nécessaire pour de nombreuses applications : le calcul du bilan du cycle de l'eau sur Terre, l'estimation des tendances des extrêmes hydrométéorologiques, la vérification des prévisions météorologiques ou l'évaluation des simulations climatiques, pour n'en citer que quelques-unes (Hossain and Huffman 2008; Nissen and Ulbrich 2017; Trenberth 2011; Beck et al. 2020). Les mesures des jauges de précipitations fournissent généralement des estimations précises à l'échelle locale malgré certains problèmes instrumentaux (comme la sous-estimation des chutes de neige induite par le vent, Nitu et al. 2018). Cependant, pour certaines applications spécifiques, notamment l'évaluation de modèles, les observations de la précipitation souffrent de problèmes de représentativité spatiale (Prein and Gobiet 2017; Herold et al. 2016; Gervais et al. 2014; Chen and Knutson 2008). Pour cette raison, des estimations maillées de la précipitation ont été développées grâce à des approches mêlant interpolation, analyse objective ou assimilation de données, par exemple, l'ensemble de données de séries chronologiques maillées de la Climatic Research Unit (Harris et al. 2020) ; ou le Global Precipitation Climatology Centre (GPCC) (Meyer-Christoffer et al. 2011). Ces estimations des précipitations ne sont, cependant, fiables que dans les régions où la densité des stations est élevée, ce qui limite considérablement leur utilisation.

Au cours des dernières décennies, des estimations de la précipitation à hautes résolutions temporelle (échelle temporelle sub-journalière) et spatiale (généralement inférieure à 0,25°) sont apparues grâce à l'utilisation combinée de mesures de jauge de précipitations, satellitaires et radar (Beck et al. 2019a, 2020; Sun et al. 2018; Trenberth et al. 2017; Derin and Yilmaz 2014) ou par l'assimilation d'observations par des modèles pour produire des réanalyses (e.g. Kalnay et al. 1996). Leurs hautes résolutions temporelle et spatiale ainsi que leurs couvertures quasi globales en font des candidats privilégiés pour nombre d'applications, notamment l'évaluation des modèles météorologiques et climatiques (par exemple, (Huffman et al. 2015; Di Luca et al. 2021; Fosser et al. 2015; Prein and Gobiet 2017). Cependant, la qualité de ces produits dépend de plusieurs facteurs, comme la technologie de détection utilisée par le satellite (par exemple, infrarouge ou micro-ondes passives; Beck et al. 2020, 2019; Derin and Yilmaz 2014; Liu and Allan 2012; Sun et al. 2018), les caractéristiques de la surface (par exemple, les erreurs peuvent être plus importantes dans les zones enneigées ; Derin and Yilmaz 2014; Tian and Peters-Lidard 2010); les phénomènes météorologiques produisant les précipitations (par exemple, les produits satellitaires sont adaptés à la détection des précipitations issue de la convection mais ont tendance à en surestimer l'intensité ; (Gehne et al. 2016; Derin and Yilmaz 2014; Beck et al. 2020; Sun et al. 2018) ou la phase et le

type d'hydrométéores (par exemple, l'incertitude de mesure est plus grande pour la neige que pour la pluie ; (Prein and Gobiet 2017; Tian and Peters-Lidard 2010; Lockhoff et al. 2019). Ces estimations des précipitations peuvent être corrigées par comparaison avec des mesures de jauge et ce qui tend à réduire les disparités entre produits (Gehne et al. 2016; Prein and Gobiet 2017). Mais ces corrections dépendent fortement de la disponibilité et de la qualité des données utilisées (Sun et al. 2018; Prein and Gobiet 2017). La plupart des mesures de précipitation à partir de jauges ne sont disponibles qu'à une fréquence quotidienne, ce qui limite le développement de corrections à des échelles de temps sub-journalières (Zolina et al. 2014).

La moyenne temporelle des précipitations est régulièrement utilisée pour comparer les estimations de précipitation entre elles (par exemple, Gehne et al. 2016; Prein and Gobiet 2017; Sun et al. 2018). Sun et al. (2018) comparent la distribution spatiale mondiale de la précipitation moyenne entre 2003 et 2010 pour 30 produits utilisant des observations satellitaires, in situ ou des réanalyses. Les distributions sont généralement similaires, mais les disparités sont plus fortes pour certaines régions, dont le nord de l'Amérique du Nord. Sur la période étudiée, la précipitation annuelle de l'Amérique du Nord est représentée différemment par les produits, et son incertitude peut être aussi grande que la précipitation moyenne elle-même (c.à.d. l'erreur relative est proche de 100 %). La comparaison des estimations de la précipitation mensuelle moyenne entre 2001 et 2012 a été réalisée par Gehne et al. (2016) pour 13 produits de précipitations (sept produits basés sur l'observation et quatre réanalyses) pour l'ensemble du globe. Des différences importantes ont été constatées près de la zone de convergence intertropicale, pour les terres lors de la période estivale (notamment pour l'ouest de l'Amérique du Nord), et à proximité des courants océaniques (par exemple, le Gulf Stream). Cette même étude a également montré que la correction du biais mensuel augmente la cohérence de la précipitation journalière moyenne entre produits (Gehne et al. 2016), l'étude de la précipitation moyenne semble ainsi nécessaire à une étude des quantités plus complexes.

Une représentation similaire de la précipitation moyenne ne garantit pas une représentation similaire de variabilité temporelle ou de la chronologie de la précipitation (e.g., Sun et al. 2018; Gehne et al. 2016; Catto et al. 2015; Di Luca et al. 2021). Par exemple, dans Gehne et al. (2016), sur l'Amérique du Nord pour 2000-2012, des six produits fondés sur des observations satellitaires, les trois produits PERSIANN-CDR, GPCP et TRMM ont les moyennes annuelles et mensuelles les plus proches, mais des distributions intensité/fréquence différentes. La précipitation est estimée très différemment dans ces trois produits.

Des métriques ont été introduites pour identifier différentes représentations de la variabilité ou de la chronologie de la précipitation. Dans Aghakouchak et al. (2012), la variabilité spatiale et saisonnière de l'erreur systématique et aléatoire pour le domaine CONtiguous United States (CONUS) sur la période 2005-2007 est étudiée en comparant trois produits de précipitation par satellite (y compris CMORPH et TRMM) et le produit radar, STAGE IV. Pour les données à 3 h et journalières, le biais systématique (la différence dans les précipitations moyennes) ne représente qu'une petite partie de l'erreur totale par rapport à l'erreur aléatoire (causée par des facteurs tels que la variabilité ou la chronologie des précipitations). En hiver, la précipitation moyenne est la principale source de disparités entre produits dans le nord et l'ouest du domaine CONUS, alors qu'en été, elle ne joue un rôle notable que dans le nord-est et les Rocheuses.

La cohérence de la chronologie des précipitations extrêmes a été étudiée par Timmermans et al. (2019) pour le domaine CONUS et huit produits se fondant sur des observations, cinq in situ et trois satellitaires (comprenant TRMM et CMORPH). Les produits reposant sur des observations in situ partagent de fortes similitudes dans leur représentation des précipitations extrêmes alors que la cohérence des produits utilisant des observations satellitaires dépend fortement de la région : forte dans le sud-est, mais faible dans l'ouest et les régions montagneuses. Catto et al. (2013, 2015) ont introduit une métrique de distance qui a montré des résultats intéressants en distinguant et en quantifiant les différences d'intensité et de fréquence (au lieu de la seule différence de précipitation moyenne). Cette métrique a été utilisée principalement pour l'évolution des modèles en combinaison avec une décomposition environnementale (précipitations frontales ; Catto et al. 2013, 2015 ; ou régimes de précipitations dynamiques et thermodynamiques ; Di Luca et al. 2021) qui nécessitait des données complémentaires non fournies par les estimations de précipitations.

Lors de la comparaison de modèle à des observations, il est essentiel de s'assurer que la différence entre le modèle et les observations est plus faible que l'incertitude sur les observations. Si l'incertitude observationnelle est plus grande que l'erreur du modèle, la comparaison n'est pas conclusive. L'objectif de cette étude est de développer une méthodologie pour quantifier le niveau d'incertitude entre les estimations de précipitations et de l'appliquer à sept produits couramment utilisés sur l'est de l'Amérique du Nord. La cohérence entre produit est quantifiée à l'aide de quatre distances évaluant différents aspects des séries chronologiques de la précipitations à 3 h: de l'accumulation totale de la précipitation à des quantités plus complexes comme la variabilité des précipitations ou la chronologie des événements de précipitation. En outre, la quantification de l'incertitude repose sur une « approche démocratique » pour

identifier les produits les moins probables à chaque point de la grille de comparaison. L'étude comprend également une analyse de la sensibilité de nos résultats à la résolution temporelle (3 h, 6 h et quotidienne) et spatiale (0,25° et 0,75°) des données.

Le mémoire est organisé comme suit. La section suivante décrit les données utilisées, y compris les types de corrections utilisés par chaque produit. La méthodologie est présentée dans la section Méthodes. Dans la section Résultats, la métrique d'incertitude utilisée est introduite et ses variabilités régionale, saisonnière et diurne (en été) sont étudiées. Sa sensibilité au choix de la résolution de comparaison (temporelle et spatiale) est également testée. Une discussion sur la métrique d'incertitude et d'autres questions est présentée dans la section Discussion et un résumé et des conclusions sont donnés dans la section Conclusion.

CHAPITRE 1

CONSISTENCY OF HIGH-RESOLUTION PRECIPITATION PRODUCTS OVER EASTERN NORTH AMERICA

Ce chapitre est présenté sous forme d'un article scientifique rédigé en anglais. Il est basé sur mes analyses et mes graphiques, et a été rédigé sous la supervision de mes co-directeurs de recherche : Prof. Alejandro Di Luca et Prof. René Laprise. L'article est actuellement en rédaction et devrait être soumis au cours de la session d'automne 2022 au Journal of Hydrometeorology de l'American Meteorogical Society.

9 **CONSISTENCY OF HIGH-RESOLUTION PRECIPITATION PRODUCTS OVER EASTERN NORTH**

10 **AMERICA**

11 Tangui Picart,[a,b] Alejandro Di Luca[a,b] and René Laprise[a,b]

12 [a] Earth And Atmospheric Sciences Department, UQAM, Montreal, Canada

13 [b] Centre pour l'étude et la simulation du climat à l'échelle régionale (ESCER), UQAM, Montréal, Canada

14

15 *Corresponding author:* Tangui Picart, tangui.picart@laposte.net

16 ABSTRACT

17 Several observational precipitation products that provide high temporal ($\leq$ 3 h) and spatial ($\leq$ 0.25°)

18 resolution gridded estimations are available. Consistency across these estimates (i.e., their level of

19 agreement) depends on several factors, including the region, the season and the comparison metric.

20 Here, we propose a methodology to quantify the uncertainty of a set of products and to identify, at

21 individual grid points, the least likely ones. The methodology is applied over eastern North America

22 for the 2015–2019 period using seven commonly used high-resolution gridded precipitation products:

23 IMERG, GSMaP, STAGE IV, MSWEP, CMORPH, TRMM and ERA5. The consistency is quantified using

24 four metrics assessing different aspects of precipitation datasets such as the total accumulation, two

25 characteristics of the intensity - frequency distribution, and the chronology of precipitating events.

26 Large regional and seasonal variations in the observational uncertainty are found. Observational

27 uncertainty is found to be higher in Canada than in the United States, reflecting large differences in

28 the density of in weather stations measuring precipitations. In northern midlatitudes, the uncertainty

29 among products is lowest in summer, while in southern midlatitudes it is lowest in winter. In summer,

30 the timing of the diurnal cycle of the precipitation is generally consistent between all products, but

31 the amplitude is not; the absolute difference remains quite constant throughout the day. The lack of

32 an absolute reference to compare products and the high variability of their uncertainty makes it

33 difficult to identify the product nearest to the (un-observed) "truth", however, our methodology,

34 allows identifying the less likely ones to approach such absolute reference.

1.1 Introduction

Several applications require accurate historical precipitation values, there are, for example, the calculation of the Earth water cycle budget, estimation of trends in hydrometeorological extremes, verification of weather forecasts, and evaluation of climate simulations (Hossain and Huffman 2008; Nissen and Ulbrich 2017; Trenberth 2011; Beck et al. 2020). Despite some instrumental issues (e.g., wind-induced snowfall undercatch, Nitu et al. 2018), precipitation gauge measurements usually provide accurate precipitation estimates at the local scale except for solid precipitation (Rasmussen et al. 2012). However, for specific applications, including model evaluation, station data suffer from spatial representativity issues (Prein and Gobiet 2017; Herold et al. 2016; Gervais et al. 2014; Chen and Knutson 2008). For this reason, gridded precipitation datasets have been developed merging in situ and remote sensed observations with optimal interpolation or kriging, and -in the case of reanalyses- merging observations and background model with data assimilation techniques (e.g., Harris et al. 2020; Meyer-Christoffer et al. 2011a,b,c). However, gridded precipitation data can only provide accurate precipitation estimations in regions with high density of stations, which drastically limits their use.

In the past few decades, great progress has been made to improve estimates of precipitation at high temporal (sub-daily time scales) and spatial (usually smaller than 0.25°) resolutions by combining gauge measurements with remote sensing meteorological radar and satellite data (Beck et al. 2019a, 2020; Sun et al. 2018; Trenberth et al. 2017; Derin and Yilmaz 2014), or data assimilation of all available observations in reanalyses using models (e.g. Kalnay et al. 1996). Their refined temporal and spatial resolution and their nearly worldwide coverage make these products ideal in a number of applications, including the evaluation of precipitation simulated by weather and climate models that require high resolution areal-mean estimations (e.g., Huffman et al. 2015; Di Luca et al. 2021; Fosser et al. 2015; Prein and Gobiet 2017). The quality of these products depends on several factors, including the sensor (e.g., infrared or passive micro-wave) used by the satellite (Beck et al. 2020, 2019a; Derin and Yilmaz 2014; Liu and Allan 2012; Sun et al. 2018), the underlying coverage of the surface (e.g., errors can be larger in snow-covered areas; Derin and Yilmaz 2014; Tian and Peters-Lidard 2010), the meteorological phenomena producing precipitation (e.g., satellite-based products are suited to detect convection precipitations but tend to overestimate their intensity; (Gehne et al. 2016; Derin and Yilmaz 2014; Beck et al. 2020; Sun et al. 2018), or the phase and type of hydrometeors present in the atmosphere (e.g., larger uncertainty for snowfall than for rainfall; Prein and Gobiet 2017; Tian and Peters-Lidard 2010; Lockhoff et al. 2019). Corrections of these products based on precipitation gauge measurements tend to reduce the difference between them (Gehne et al. 2016; Prein and Gobiet 2017). But these

68 corrections depend strongly on the availability and quality of the data used to make the correction

69 (Sun et al. 2018). For example, in regions with high station density, gauge-corrected products are

70 usually unbiased regarding monthly or even daily time scales, depending on the type of correction

71 being applied. However, most in situ observations are only available at a daily frequency, limiting the

72 development of corrections at sub-daily time scales (Zolina et al. 2014).

73 Several studies have evaluated differences among precipitation products using temporal averaged

74 precipitation (e.g., Gehne et al. 2016; Prein and Gobiet 2017; Sun et al. 2018). Sun et al. (2018)

75 compared the spatial distribution of time-mean precipitation between 2003 and 2010 over the globe

76 across 30 products obtained using satellite, in situ and reanalysis data. Their findings show that most

77 products agree well on the overall spatial patterns, but they show large discrepancies over several

78 land regions including northern North America where mean precipitations vary from almost 900 $(y)^{-1}$

79 (for a reanalysis) to 600 mm (for CPC). A similar global analysis was carried out by Gehne et al. (2016)

80 for monthly mean precipitation between 2001 and 2012 using 13 precipitation products (seven

81 observation-based products and four reanalyses). They found large differences near the intertropical

82 convergence zone, over land areas in the summer hemisphere (notably over western North America),

83 and over the western boundary of the Gulf Stream. Gehne et al. (2016) also showed that monthly bias

84 correction improves daily mean consistency between the corrected products: the correlation between

85 the daily precipitation over land of GPCP and PERSIANN is null without correction and over 0.90 with

86 the monthly correction.

87 Agreement in the time-mean precipitation does not ensure, however, a consistent representation of

88 the daily or sub-daily temporal variability, nor of the chronology of precipitation (Catto et al. 2015;

89 Gehne et al. 2016; Sun et al. 2018; Di Luca et al. 2021). Over North America, Gehne et al. (2016) show

90 that PERSIANN, GPCP, and TRMM have similar annual and monthly mean values but different

91 intensity/frequency distribution, suggesting that the way precipitation occurs is different across the

92 three datasets. To identify differences in precipitation products arising from the variability or the

93 chronology of events, several metrics have been used. In Aghakouchak et al. (2012), disparities

94 between three satellite-based precipitation products (including CMORPH and TRMM) and STAGE IV

95 and are investigated over the CONtiguous United States (CONUS). For most of the domain, the bias

96 counts for a small portion of the difference between products for daily and 3-hourly data but different

97 representation of the precipitation variability and timing (captured in the study by the random error).

98 However, disagreements on the mean precipitation are the main error source for the North and the

99 West of CONUS in winter and in the Northeast and the Rockies in summer. Guilloteau et al. (2022)

100 also found that the representation of the fine-scale precipitation (spatial and temporal) is the main

source of disparities between IMERG, and another product based on precipitation gauge and radar measurements over southeastern US between January 2018 and April 2020. The consistency of the extreme precipitation chronology has been investigated by Timmermans et al. (2019) over CONUS for five precipitation gauge-based products and three satellite-based products (including TRMM and CMORPH). Products based on precipitation gauges share strong similarities in their representation of extreme precipitation, while the consistency between satellite-based products depends strongly on the region: good in the southeast and poor in the west and over mountainous regions. Catto et al. (2013, 2015) introduced a distance metric that showed interesting results in distinguishing intensity and frequency differences instead of only the mean precipitation difference. This metric has been used mainly for the evaluation of models in combination with a stratification based on the synoptic situation (frontal precipitation; Catto et al. 2013, 2015; or dynamic and thermodynamic precipitation regimes; Di Luca et al. 2021), which required complementary data not provided by precipitation estimates.

The aim of this study is to develop and apply a methodology to quantify the level of consistency (or conversely uncertainty) between 3-hourly gridded precipitation estimates provided by seven commonly used products over eastern North America. Consistency is quantified using four distance metrics that evaluate different aspects of the precipitation gridded products; from the relatively simple total precipitation amounts to the variability of precipitation by looking at intensity and frequency, and the timing of precipitation events. The quantification of the uncertainty associated with this ensemble of gridded observation datasets is based on the use of a "democratic approach" to identify the less likely products at individual grid points and seasons. The study also includes an analysis of the sensitivity of our results to the temporal (3-hourly, 6-hourly, and daily) and spatial (0.25° and 0.75°) resolution of the data.

The article is organised as follows. The next section (1.2) described the datasets used, including the types of corrections used by each product. The methodology is presented in Section 1.3. In Section 1.4, the uncertainty metric used in this article is induced and its regional, seasonal, and diurnal (in summer) variabilities are investigated. Its sensitivity to the choice of comparison resolution (temporal and spatial) is also tested. Some discussion about the uncertainty metric and other issues is presented in Section Discussion and a summary and conclusions are given in Section Conclusion.

130   1.2   Data

131   The precipitation products used in this study are based on satellite, radar, precipitation gauge, and

132   reanalysis data (Table 1). All products are corrected using precipitation gauge estimates, although the

133   applied correction depends on the product and is briefly discussed below.

134   CMORPH estimates are corrected to match the National Oceanic and Atmospheric Administration

135   Climate Prediction Center's (CPC) daily precipitation gauge analysis over land and Global Precipitation

136   Climatology Project (GPCP) monthly merged analysis over ocean (Xie et al. 2017). Over land, the

137   monthly mean precipitation of CMORPH and the monthly distribution of daily intensities are corrected

138   by matching probability density functions against CPC data. Over ocean, CMORPH daily data is

139   upscaled to GPCP grid (2.5° lat/lon grid). First, the mean precipitation computed for the point and the

140   nearest points (under three points) is daily unbiased and secondly, the point mean precipitation is

141   unbiased for a 19-days period centered on the corrected day. The multiplicative correction is applied

142   uniformly for all time steps and points.

143   TRMM and IMERG satellite estimates are corrected over land grid points using a version of the 1°

144   lat/lon monthly precipitation gauge analysis of the Global Precipitation Climatology Centre (GPCC)

145   (Huffman et al. 2018, 2007, 2010) taking into account the precipitation undercatch in gauge

146   measurements. The correction is applied so that the monthly mean values of TRMM and IMERG,

147   upscaled on the GPCC grid, match the GPCC values. IMERG estimates use an updated correction

148   method that, among other things, applies a smoother correction for grid points near coasts.

149   GSMaP satellite estimates are corrected over land grid points based on the CPC Unified Gauge-Based

150   Analysis of Global Daily Precipitation (GPM Global Rainfall Map Algorithm Development Team 2014).

151   The correction is carried on by linearly interpolating 0.5° CPC daily values into the 0.1° GSMaP grid.

152   Daily precipitations are debiased and sub-daily precipitation rates are corrected based on

153   considerations as the regime or the duration of the precipitation event (Mega et al. 2019). The

154   correction is only applied over grid boxes where clouds are detected by infrared measurements (Mega

155   et al. 2019).

156   STAGE IV estimates are based on precipitation estimates based on radar observation from National

157   Weather Service River Forecast Centers (RFC) over CONUS (Lin and Mitchell 2005; Nelson et al. 2016).

158   Each RFC uses the multi-sensor precipitation estimator and a field bias correction to generate its

159   precipitation estimates from radar and precipitation gauge measurements. The RFC precipitation

160   estimates are then merged and bias-adjusted in near real-time based on observations from

161   hydrometeorological automated data system gauges, automated surface observing system, and
162   automated airport weather station reports (Prat and Nelson 2015). Manual quality control is also
163   performed (Nelson et al. 2016).

164   MSWEP 3-hourly precipitation estimates are obtained by merging gauge observations with IMERG and
165   ERA5 precipitation estimates (Beck et al. 2017, 2019; GloH2O 2021). Before merging, the ERA5
166   precipitation frequency is corrected at each point to match the frequency estimated by nearby
167   precipitation gauges using a minimum daily precipitation threshold of 0.5 mm d^{-1}. The IMERG and
168   ERA5 precipitation estimates are then merged based on their correlation with local precipitation
169   gauge measurements. A cumulative distribution function is also used within the merging procedure
170   to correct the intensity distribution of the final estimate. Finally, the merged estimate is bias corrected
171   based on the five closest gauges for daily values, and then monthly debiased based on the GPCC Full
172   Data Reanalysis, version 7.

173   The precipitation estimate from the ERA5 reanalysis (Hersbach et al. 2020) was also included in our
174   study. ERA5 has a finer temporal and spatial resolution than the previous ERA-Interim reanalysis (Dee
175   et al. 2011; Hersbach et al. 2019). ERA5 has benefited from notable improvements, such as the
176   assimilation humidity-sensitive satellite channels using the new all-sky assimilation approach and the
177   use of new parameterizations of physical processes (large-scale clouds, microphysics). In addition,
178   ERA5 assimilates most measurements on which previously described precipitation estimates are
179   based. Precipitation rate from the STAGE IV product over the US are also assimilated within the
180   reanalysis.

181   Table 1. Abbreviation, name and version, reference, native spatial Δx and temporal Δt resolution, temporal and spatial coverage, and input data sources of the
182   precipitation datasets used in the study.

Abbreviation	Name/version	Reference	Δx (°)	Δt (h)	Available period	Spatial coverage	Data Source
CMORPH	Climate Prediction Center morphing technique (V1.0)	(Joyce et al. 2004)	0.25	1	1998/01 – present	Quasi-global (60°N – 60°S)	Satellite/gauge
GSMaP	Global Satellite Mapping of Precipitation (V7)	(Ushio et al. 2009)	0.1	1	2014/03 – present	Quasi-global (60°S to 60°N only)	Satellite/gauge
IMERG	Integrated Multi-satellitE Retrievals for Global precipitation measurement (precipitationCal), V6	(Huffman et al., 2014, 2018, 2020)	0.1	0.5	2000/06 – present	Global	Satellite/gauge
TRMM	Tropical Rainfall Measuring Mission multi-satellite precipitation analysis (3B42)	(Huffman et al. 2007)	0.25	3	1979/01 – present	Quasi-global (50°N – 50°S)	Satellite/gauge/ reanalyse
STAGE IV	NCEP/EMC US Gridded Radar-Estimated Precipitation with Bias Removal (stage IV)	(Lin and Mitchell 2005)	~0.05	1	2002/01 – present	United States	Radar/gauge
MSWEP	Multi-Source Weighted-Ensemble Precipitation (V2.8)	(Beck et al. 2019b; GloH2O 2021)	0.1	3	1998/01 – 2019/12	Global	IMERG/ERA5/gauge
ERA5	European Centre for Medium-range Weather Forecasts Reanalysis 5 High-Resolution	(Hersbach et al. 2020)	0.25	1	2008/01 – present	Global	Model with data assimilation (satellite/gauge/radar)

183

184   1.2.1   Region and period of study

185   The study focuses over eastern North America, bounded by latitudes 25°N and 50°N, and longitudes 100°W

186   and 60°W (Figure 1.1). Figure 1.1 shows the 5 sub-domains used in our analysis. The comparison between

187   precipitation products is carried over a 5-year period from 2015 to 2019, which is the longest period

188   common to all products (Table 1).

189
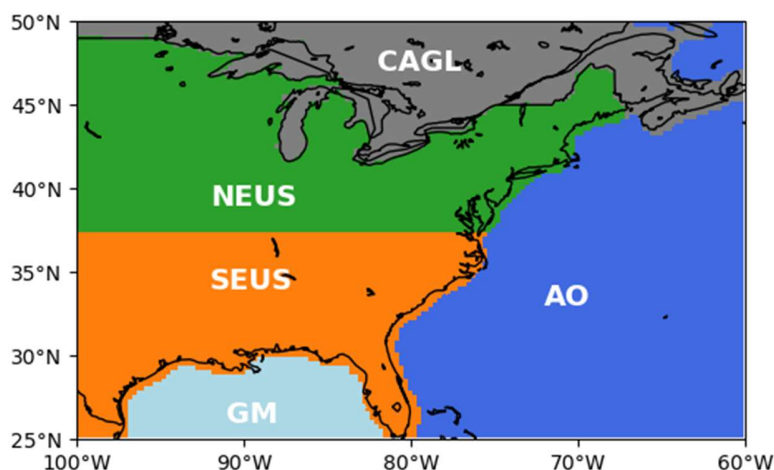
190   Figure 1.1: Region of the study including the five sub-domains: northeast US (NEUS, green), southeast US (SEUS,
191   orange), southeast Canada and the Great Lakes (CAGL, grey), Atlantic Ocean (AO, blue), and Gulf of Mexico (GM,
192   pale blue).

193   1.3   Methods

194   1.3.1   Precipitation Data Preprocessing

195   The seven datasets provide precipitation estimates at different temporal and horizontal resolutions. To

196   facilitate the comparison, all datasets were spatially interpolated to the coarser grid mesh (ERA5: 0.25°

197   lat/lon grid) and upscaled at the lowest temporal resolution (TRMM/MSWEP 3-hourly frequency). Spatial

198   interpolation was performed using the "conservative" algorithm from the xESMF python library

199   (https://doi.org/10.5281/zenodo.1134365), which ensures the mass conservation of the precipitation

200   over the coarser ERA5 grid boxes as similarly done in previous studies (e.g., Di Luca et al. 2021; Prein and

201   Gobiet 2017).

202   Specifically, the STAGE IV estimate was first interpolated using a nearest-neighbour approach from its

203   native 4-km polar stereographic projection to a 0.01° lat/lon grid before being conservatively interpolated

204   to the ERA5 grid. This product (based on radar observations) is only considered for the NEUS and SEUS

205   sub-domains.

13

206 MSWEP and GMSAP products (available at 0.1°) were first shifted by +0.025° in latitude and longitude to

207 ensure that the number of input cells (2.5) used to produce an output cell (0.25°) was constant over the

208 studied region. More details are given in the APPENDIX A.

209 Due to file compression issues, ERA5 precipitation estimates can contain some small negative values ($\geq$

210 -10^{-8} mm (3 h)$^{-1}$). These values were set to zero before interpolation.

1.3.2 Distance metrics

212 The distance metrics are used to characterize the difference between two products, all calculated at

213 individual grid points, by comparing the time series of 3 h precipitation accumulations of two products.

214 Note that given eight datasets, a total of 28 distinct pairs are possible, $8!/(2! \times 6!)$. The metrics are

215 distance metrics in the mathematical sense and satisfy the separation, positivity, symmetry, and the

216 triangular inequality.

217 The first metric denoted as ΔP_M is the absolute value of the difference between the time-mean

218 precipitation (i.e., the absolute bias) of two given datasets, noted by superscripts (1) and (2) hereinafter

$$\Delta P_M = \frac{1}{N_T} \left| \sum_t \left(P_t^{(1)} - P_t^{(2)} \right) \right| = \left| P_M^{(1)} - P_M^{(2)} \right| \tag{1}$$

220 where the index t stands for the 3-hourly time interval and N_T the total number of intervals. Figure E.1 in

221 the APPENDIX E shows the mean precipitation maps of each product. The comparison of two datasets

222 might be zero even if the products are not identical, due to compensating negative and positive errors

223 within the temporal average.

224 The second metric denoted as ΔP_A is the time-average of absolute differences between 3-hourly

225 precipitation values (i.e., the mean absolute difference):

$$\Delta P_A = \frac{1}{N_T} \sum_t \left| P_t^{(1)} - P_t^{(2)} \right| \tag{2}$$

227 This metric is null only if the two precipitation time series are identical, thus preventing any form of error

228 compensation. ΔP_A will be used as a reference in our analysis. In contrast to ΔP_M (and the other two

14

229 metrics described later), this metric can only be used when the timing of meteorological events coincides

230 between the two datasets (i.e when there is an exact 1:1 pairing of the two datasets to be compared)

231 To illustrate the difference between metrics ΔP_M and ΔP_A, Figure 1.2a shows TRMM and MSWEP

232 precipitation time series for the grid point closest to Montréal, Canada, during the month of June 2015. At

233 this grid point and for this period, ΔP_A has a value of 0.74 mm (3 h)$^{-1}$ while ΔP_M has a value of 0.017 mm

234 (3 h)$^{-1}$. This shows that ΔP_A is affected by timing and intensity disparities in the precipitation event,

235 whereas ΔP_M is not affected by time mismatches, and large part of the intensity disparities cancels out in

236 the time aggregation.
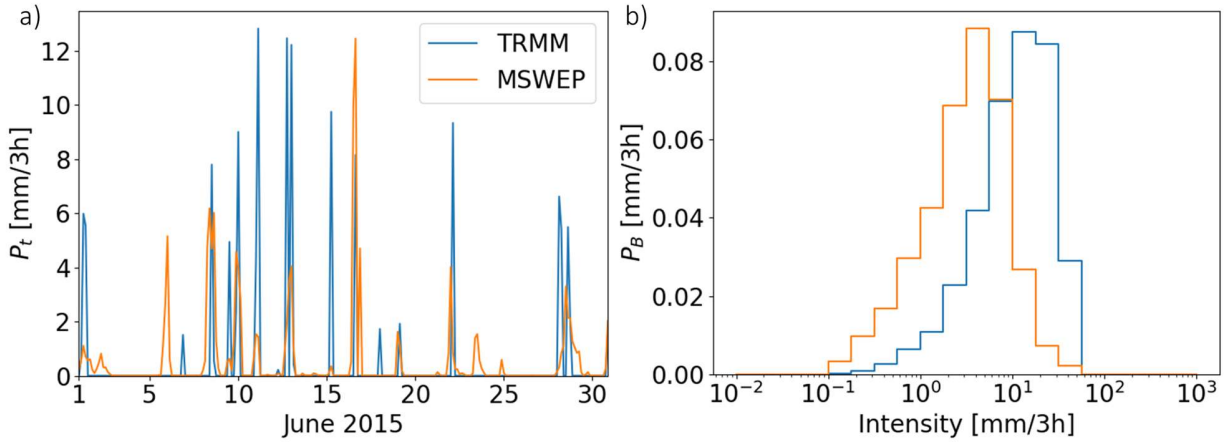
237

238 Figure 1.2: TRMM and IMERG (a) 3-hourly precipitation rates as a function of time during June 2015 and (b) total
239 precipitation as a function of the 3-hourly precipitation intensity for a grid point located near Montreal, Canada
240 (45.5°N, 73.5°W). The intensity distribution is represented for 20 logarithmic bins.

241 The third metric quantifies differences in the intensity distribution. For a given selection of intensity bins,

242 the precipitation within the bin $B_j(P_{B_j})$ is the sum of precipitation amounts for events with an intensity

243 in $B_j = (I_j, I_{j+1}]$:

$$P_{B_j} = \sum_{I_j < P_t \leq I_{j+1}} P_t \qquad (3)$$

244

245 The binning precipitation metric ΔP_B is defined as:

$$\Delta P_B = \frac{1}{N_T} \sum_{B_j} \left| P_{B_j}^{(1)} - P_{B_j}^{(2)} \right| \qquad (4)$$

246

247   The metric ΔP_B is useful to identify compensations within the intensity distribution and it provides an

248   intermediate estimation between ΔP_M (that allows for error compensations) and ΔP_A (that does not allow

249   for any compensation): $\Delta P_A \geq \Delta P_B \geq \Delta P_M$. A graphical interpretation of ΔP_B is given in the Figure 1.2b

250   for the two time series previously discussed in Figure 2; in this case the metric ΔP_B has a value of 0.53 mm

251   (3 h)$^{-1}$ when computed using 20 bins (of which only 11 are effective, the others having very small

252   occurence).

253   The choice of bins B_j must be made with care. It must span the full range of possible intensities. Also, the

254   bin number must be chosen to ensure a low noise (which decreases with the bin number) and a precise

255   representation of the intensity distributions. In our study, a logarithmic distribution with 130 bins was

256   used when analysing the entire period, with the lowest intensity being a small threshold near 0 and the

257   highest being 10^3 mm (3 h)$^{-1}$. Note that the maximum precipitation for our study is 350 mm (3 h)$^{-1}$ for

258   GSMaP. When calculations are made at the monthly basis (hence with fewer data), we used 20 bins.

259   Further details regarding the choice of bin number are given in the APPENDIX C.

260   A fourth metric is introduced based on the commonly used intensity-frequency precipitation

261   decomposition. At a grid point, the mean precipitation is given by:

$$262 \qquad P_M = \frac{\sum_t P_t}{N_T} = \frac{\sum_t P_t}{N_{P_t>k}} \cdot \frac{N_{P_t>k}}{N_T} = I.F \qquad (5)$$

263   With $N_{P_t>k}$ the number of 3-hourly periods with precipitation over the threshold k, I the mean intensity,

264   which is the precipitation averaged over all 3-hourly periods with precipitation over the threshold, and F

265   the precipitation frequency, which is the fraction of 3-hourly periods with precipitation. The mean

266   precipitation used here is the same as in equation 1. Figure E.2 in the APPENDIX E shows the frequency

267   maps of each product.

268   The intensity-frequency distance metric ΔP_{IF} is defined as:

$$269 \qquad \Delta P_{IF} = i.|\Delta F| + f.|\Delta I| + R \qquad (6)$$

270   with

$$i = \min\left(I^{(1)}, I^{(2)}\right), f = \min\left(F^{(1)}, F^{(2)}\right) \qquad (7)$$

$$\Delta F = F^{(1)} - F^{(2)}, \Delta I = I^{(1)} - I^{(2)}, \qquad (8)$$

$$R = \left|\left(I^{(1)} - i\right).\left(F^{(1)} - f\right) - \left(I^{(2)} - i\right).\left(F^{(2)} - f\right)\right| \qquad (9)$$

The terms $i.|\Delta F|$ and $f.|\Delta I|$ measure differences in mean precipitation amounts of two products due to the disparity in the frequency or the intensity of precipitation, respectively. The residual term R reflects whether there is compensation between the differences in the intensity and in the frequency of precipitation. In case of a compensation between the frequency and the intensity, the differences $(I_1 - I_2)$ and $(F_1 - F_2)$ have opposite signs (i.e., the product with the highest frequency also has the lowest intensity) and the residual term $R = 0$ with $\Delta P_{IF} = i.|\Delta F| + f.|\Delta I|$. Otherwise, if no compensation occurs, $R > 0$ and $\Delta P_{IF} = \Delta P_M$. The metric ΔP_{IF} is constructed to highlight intensity-frequency compensation within ΔP_M and by construction $\Delta P_{IF} \geq \Delta P_M$. However, because ΔP_{IF} only considers wet events during the period, its value should not be compared with ΔP_B and ΔP_A that pertain to all cases.

In our study, precipitation frequencies are estimated using a threshold of 1 mm day[-1] (i.e., 0.125 mm (3 h)[-1]) and only precipitation events with a higher intensity, are considered. For this threshold value, 38% of the precipitation events are neglected and these neglected events represent 1.2% of the total precipitation. More details can be found the Table B.1 of the supplementary material. This threshold is used to ensure the comparability of our results with previous studies using an intensity-frequency decomposition of the mean precipitation (Catto et al. 2015; Di Luca et al. 2021). The sensitivity of ΔP_{IF} to the choice of this threshold is assessed in APPENDIX B (Figure B.2). As long as the threshold remains small, the threshold choice has only a small effect on the results.

All distance metrics have units of precipitation rate, mm (3 h)[-1]. To better reflect local differences, relative differences are calculated by normalising all distance metrics (e.g., ΔP_M) by the sum of mean precipitation in both products:

$$\Delta P(\%) = 100 \times \frac{\Delta P(mm\ (3h)^{-1})}{P_M^{(1)} + P_M^{(2)}} \qquad (10)$$

Defined in this way, relative distance metrics vary between 0 and 100% (see APPENDICE D "Bounded distance metrics") and shows the fraction of the mean precipitation (i.e., the total accumulation), which is not consistent for a given metric. A value of 0% suggests that the compared aspects is similar for both the

298   precipitation time series, while a 100% indicates no similarity. For example, using the June 2015 3-hourly

299   precipitation time series shown in Figure 2a, the relative differences between MSWEP and TRMM are

300   ΔP_M=1.3%, ΔP_A=69%, ΔP_B=43% and ΔP_{IF}=71%.

301   As mentioned above, a total of 21 differences can be obtained, for each metric, based on the seven

302   products. Moreover, for any given distance metric ΔP_X, the mean difference for the product p, i.e., its

303   mean consistency, can be calculated by averaging differences between the product p and all other

304   products:

305
$$\overline{\Delta P_X}^p = \frac{1}{6} \sum_{j \neq p}^{6} \Delta P_X \left(P^p, P^j \right) \tag{11}$$

306

307 1.4   Results

308 1.4.1   Constructing an uncertainty metric

309 In this section, we evaluate step-by-step the score assessing the agreement of a set of products at

310 representing the precipitation at individual grid points. The process of deriving the uncertainty metric is

311 illustrated in Figure 1.3 using the distance metric ΔP_A and precipitation time series in winter for a single

312 grid point located at (46°N, 68°W).

313 Figure 1.3a shows all 21 individual differences obtained using the seven precipitation datasets and the

314 distance metric ΔP_A. Values vary from about 20% for the MSWEP-ERA5 pair to about 85% for the CMORPH-

315 TRMM pair. The largest ΔP_A value is denoted here as $max_7(\Delta P_A)$ and provides the largest dataset

316 difference at the grid point. As such, max_7 could be used to characterise the upper limit of the

317 observational uncertainty. However, max_7 can be affected by outliers or erroneous values and so its value

318 will tend to overestimate the uncertainty. For example, CMORPH shows little or no precipitation in winter

319 at high latitudes, a characteristic that appears to be related to a deficiency in the CMORPH algorithm (Xie

320 et al. 2017). Note that in Figure 1.3 all large ΔP_A value include CMORPH in the comparison. Consequently,

321 the uncertainty would be exaggerated if including the unrealistic CMORPH product.

322 Figure 1.3b shows the average distance for each individual product ($\overline{\Delta P^p}$). At this specific grid point,

323 CMORPH, TRMM, and IMERG have the largest mean $\overline{\Delta P_A^p}$ values (81%, 69% and 64% respectively). The 11

324 comparisons (out of 21) with the largest ΔP_A values are comparisons that include either TRMM or

325 CMORPH. Therefore, we assume that products less likely to be close to the truth are characterized by large

326 $\overline{\Delta P_A}^p$ values. To get a better estimation of the observational uncertainty, we assume that there may be

327 three products that provide erroneous estimations, and these three products are excluded from the

328 uncertainty calculation. That is, the observational uncertainty is estimated using the max_4 value obtained

329 by only considering the four most consistent products (the four products with the lowest $\overline{\Delta P_A}^p$). Here after,

330 the uncertainty refers to the measured max_4 values for a set of products while the consistency refers to

331 the $\overline{\Delta P^p}$ value a product. The four most likely products are shown under the dashed line in Figure 1.3b,

332 and comparisons between the four most likely products are displayed in a dashed rectangle in Figure 1.3a.

333 Small ΔP_A distances between two datasets suggest a high level of agreement and can be related with

334 products being dependent on each other. Figure 1.3b shows that ERA5 and MSWEP are close, which

335 probably reflects the fact that MSWEP is constructed using ERA5 precipitation estimates. Similarly, Figure

336 1.3b shows the strong agreement between ERA5 and STAGE IV, likely related to ERA5 assimilating the

337 same radar measurements as the STAGE IV product.
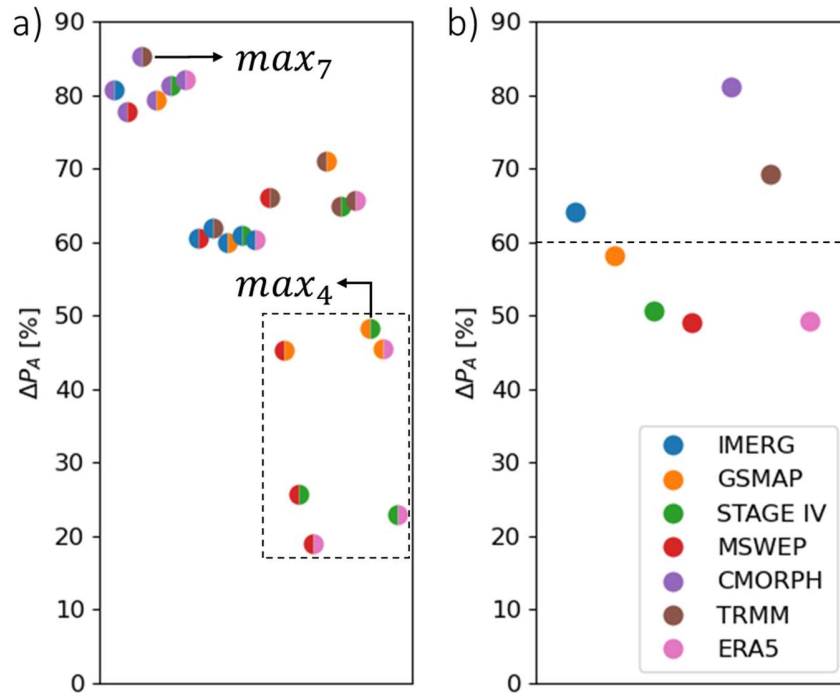
339 Figure 1.3: Relative distance metric ΔP_A calculated for (a) the 21 pairs of datasets, with the two colors in the dot
340 indicating the datasets being compared, and (b) each individual dataset $\overline{\Delta P_A}^p$ (see Eq. 11). Distance metrics were
341 calculated using 3-hourly precipitation time series from the grid point (46°N, 68°W) in winter. See section 1.4.1 for
342 the description of max_7 and max_4.

343 1.4.2   Selecting a distance metric

344 In this section, we will compare the four-distance metrics that were introduced in the Methods section.

345 The comparison will be illustrated for the annual cycle of the monthly average for each sub-domain. Figure

346 1.4 shows the annual cycle of the monthly-mean for the seven products ($\overline{\Delta P}^p$) and for each distance

347 metric. All values are for the NEUS sub-domain. For the metric ΔP_M, Figure 1.4a shows that the mean

348 distance of the CMORPH product ($\overline{\Delta P_M}^{CMORPH}$) is much higher than any other product from October to

349 May, reaching a value of nearly 60% in January. Products other than CMORPH exhibit a mean distance

350 $\overline{\Delta P_M}^p$ that varies from about 20% in winter to approximately 10% in summer. Interestingly, the max_4 is

351 larger in winter than in summer, which shows that the large uncertainty in mean precipitation in winter is

352 not only the result of including CMORPH in the averaging process.

353
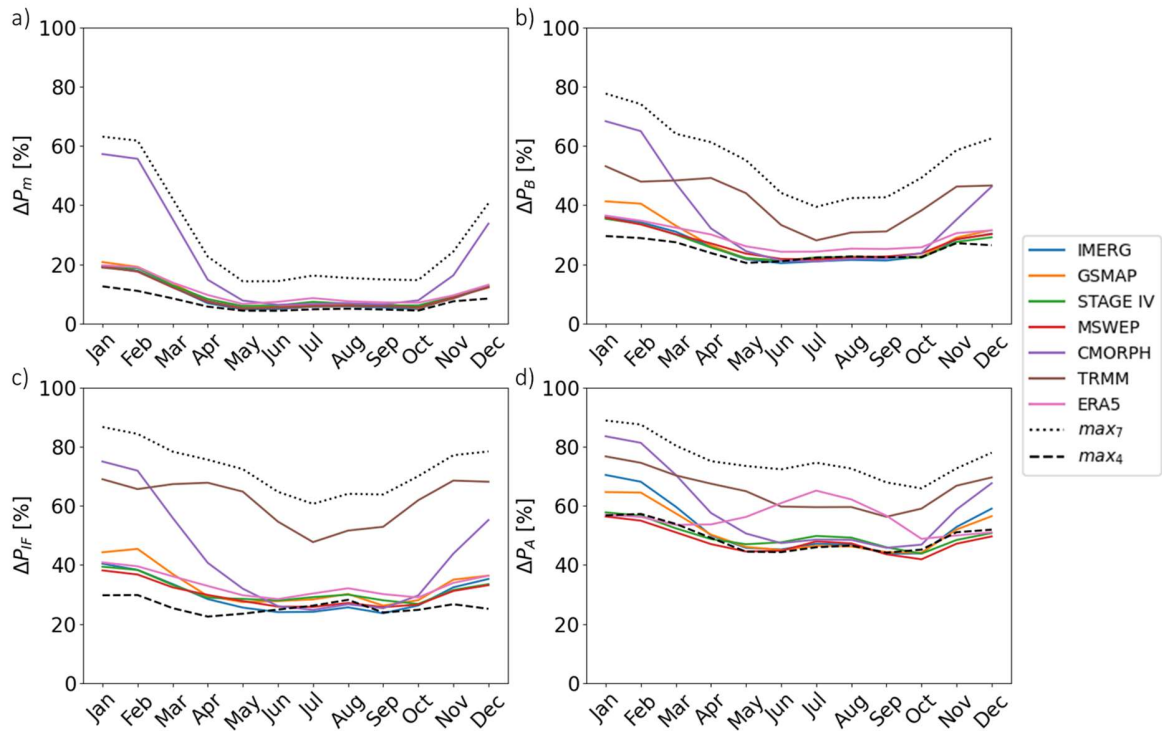
Figure 1.4: Monthly $\overline{\Delta P^p}$ for each product for each distance metric a) ΔP_M, b) ΔP_B, c) ΔP_{IF} and d) ΔP_A for grid
points in the NEUS sub-domain. Monthly mean values for the max7 and max4 metrics are also included.

The mean distances for the ΔP_B and the ΔP_{IF} metrics show qualitatively similar results to the ΔP_M metric although they are systematically higher. The large disparity between CMORPH and other products in winter remains but, in addition, ΔP_B and ΔP_{IF} also reveal that TRMM is substantially different compared to the other products all year round. While TRMM shows differences in mean precipitation of at most 20%, similar to other products, it shows differences of about 40% and 60% for $\overline{\Delta P_B}^{TRMM}$ and $\overline{\Delta P_{IF}}^{TRMM}$, respectively. This suggests that the intensity and frequency distribution of 3-hourly precipitation of TRMM is at odds with other products.

Figure 1.4d shows that the ΔP_A metric captures well all disparities between any two products and thus has greater values than ΔP_M and ΔP_B. For ΔP_A, IMERG and GSMaP appear to be, as CMORPH, less consistent with others in the winter with values attaining 70% in January. ERA5 is less consistent in summer when $\overline{\Delta P_A}^{ERA5}$ reaches a maximum value of 70% in July. TRMM has a poor consistency year-long with a mean value of around 75%. Interestingly, some of the disparities between IMERG/ERA5 and other products are only captured by the ΔP_A metric. According to the ΔP_A metric, MSWEP and STAGE IV are among the most consistent products for the whole year with mean values around 50%.

370   For the four metrics, Figure 1.4 shows that the upper bound of the uncertainty of the seven products

371   measured by max_7 (dotted line) and the uncertainty of the four most likely ones measured by max_4

372   (dashed line) vary differently throughout the year. For example, for the ΔP_A metric, max_4 varies from 57%

373   in February to 44% in October and max_7 varies from 89% in January to 66% in September. They vary also

374   differently and for all distance metrics (with maybe the exception of ΔP_{IF}), max_4 shows that the largest

375   uncertainty occurs during the winter months and the lowest during the summer. Systematically, max_7

376   shows a minimum in summer due to the major disparities in winter, while this is not the case for max4.
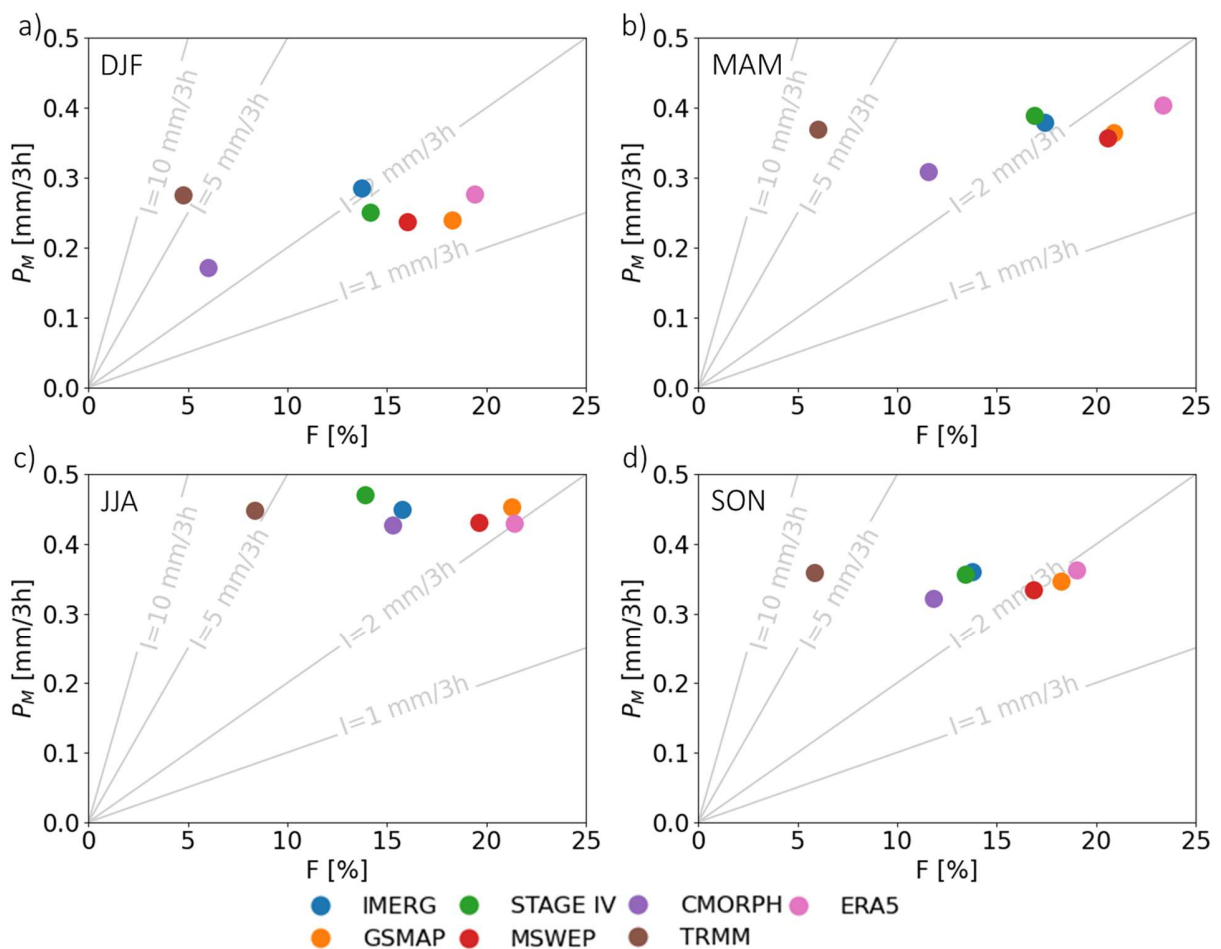
378   Figure 1.5: Mean precipitation, P_M, as a function of frequency of precipitation events, F, for each product for DJF
379   (a), MAM (b), JJA (c) and SON (d). Results are for a domain comprising the 2 sub-domains NEUS and SEUS. Constant
380   precipitation intensity lines, I, are shown in gray (see equation 5).

381   Figure 1.5 shows the mean precipitation and frequency of precipitating events for the 7 product and each

382   season. As presented in equation 5, The mean precipitation, P_M, can be expressed as the product of the

383   precipitation frequency, F, and mean intensity, I. Grey lines represent constant intensity. TRMM shows a

384   mean precipitation similar to most other products (hence $\overline{\Delta P_M}^{TRMM}$ is small) but tends to systematically

385   generate fewer precipitation events with a higher intensity than other products (hence $\overline{\Delta P_{IF}}^{TRMM}$ and

386     $\overline{\Delta P_B}^{TRMM}$ are large); its mean intensity is near 5 mm (3 h)$^{-1}$ while other products have intensities below 3

387   mm (3 h)$^{-1}$. CMORPH tends to underestimate the mean precipitation largely due to an underestimation of

388   the frequency of events in both winter and spring. In winter, CMORPH has a mean precipitation of 0.17

389   mm (3 h)$^{-1}$ and a frequency of 6%, while other products (except TRMM) have a mean precipitation and a

390   frequency of more than 0.24 mm (3 h)$^{-1}$ and 14%.

391   Figure 1.6 shows the fraction of grid points that each product is within the four most consistent products

392   (i.e., within max_4) over the NEUS sub-domain in winter, for ΔP_M and ΔP_A. The sum of all fractions equals

393   400% because four products are used for each grid point. For the ΔP_M metric, the monthly fraction is

394   between 40% and 80% for most products, except for CMORPH that shows values below 10% in winter

395   months. ERA5 exhibits low agreement, especially in the summer. For the ΔP_A metric, results are more

396   dependent on the season and on the product. The most likely products are MSWEP, STAGE IV, ERA5 and

397   GSMaP in winter, and IMERG, GSMaP, MSWEP and STAGE IV in summer. Regardless of the month, for the

398   ΔP_A metric, TRMM is nearly never considered among the four most consistent products over this sub-

399   domain (<0.3% of the grid points).
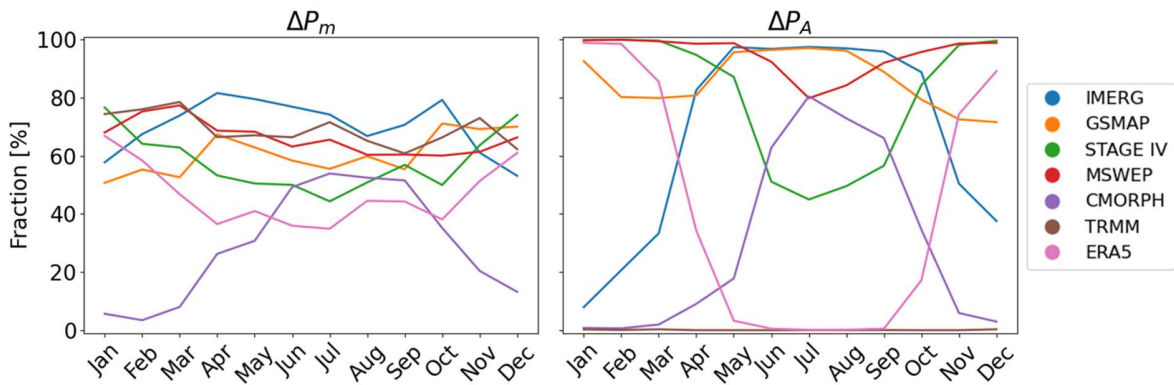
400

401 Figure 1.6: Fraction of grid points for which each product is selected within max_4 for ΔP_M (left) and ΔP_A (right), for
402 each month of the year over the NEUS sub-domain.

403   The nature of the differences between the products varies and only ΔP_A identifies disparities in the mean

404   precipitation (as ΔP_M), in its variability (as ΔP_B and ΔP_{IF}), and in the precipitation chronology. Thus, for

405   the rest of this article, results are shown mainly for ΔP_A, and other metrics are occasionally used to identify

406   the source of the disparities.

1.4.3 Regional variability of products uncertainty

The analysis is now carried for each sub-domain to identify regional differences in the uncertainty across products. Figure 1.7a shows the spatial distribution of $max_4(\Delta P_A)$ values over the full region. A strong spatial variability is visible across eastern North America, with high latitudes over the CAGL sub-domain showing the largest uncertainty values. Over the CAGL sub-domain, max_4 has a mean value of 63%, reaching 80% locally. The max_4 uncertainty over US sub-domains is lower than for the CAGL one, with sub-domain average values of 43% and 49% for SEUS and NEUS, respectively. The NEUS and CAGL subdomains are separated by a strong discontinuity in max_4 (an increase by about 10% over the CAGL subdomain) likely due to the decrease in station density or different treatment of the data between US and Canada, which strongly affects product corrections.
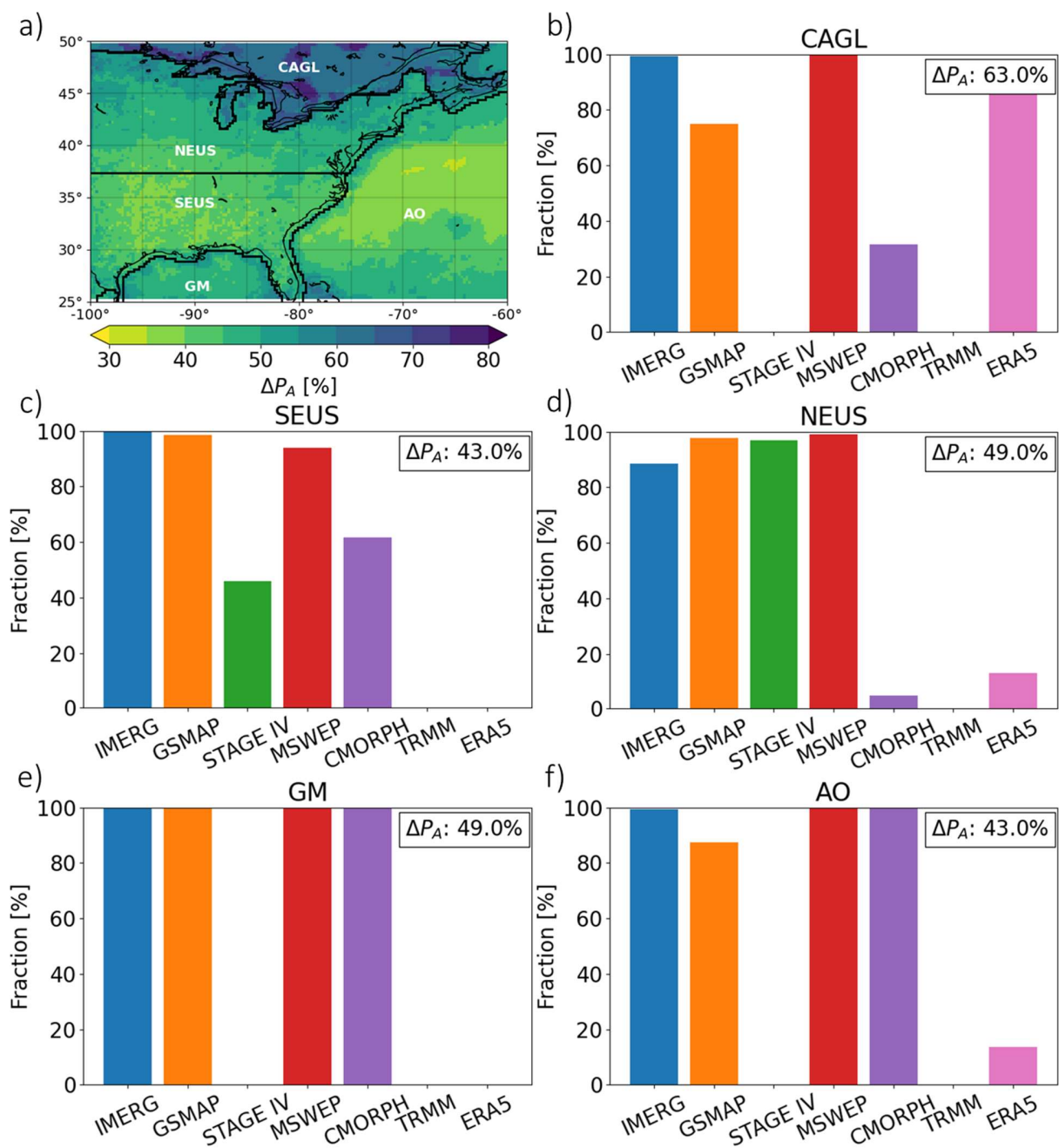
Figure 1.7: (a) Map of max_4, and (b-f) fraction of grid points for which each product is selected within max_4 for each sub-domain. The statistics are computed for ΔP_A.

AO and GM show sub-domain mean max_4 values of 43 and 49% respectively, thus showing similar values to those over land. Locally, large max_4 values are found in coastal and islands grid points, with max_4 values attaining 70%. Along the coasts, the discrepancies are mainly caused by different precipitation timing. The use of different land-sea masks within products and interpolation issues are also sources of

424   uncertainty in particular for comparisons with GSMaP. Discrepancies in points being either considered as

425   land or ocean are source of uncertainty. For example, the estimation of the precipitation of GSMaP near

426   the 33.25°N; 64.75°W point, close to the Bermuda archipelago, differs from other products as this point is

427   considered as a land point by GSMaP (i.e., the precipitation is corrected with in situ data) but as a water

428   point in other products. These differences lead to a large uncertainty in some grid points around island

429   (see also around The Bahamas). For other metrics (see Appendix F), the largest uncertainties tend to be

430   between the latitudes 25° and 35° and in the Gulf of St. Lawrence.

431   Figure 1.7b-f show the fraction of grid points for which each product is selected within max_4 for individual

432   sub-domains. For the CAGL sub-domain (Figure 1.7b), IMERG (99%), MSWEP (100%) and ERA5 (94%) are

433   among the four most likely products at most grid points, while GSMaP is selected in about 75% of grid

434   points. Grid points where GSMaP is not selected are also points where the uncertainty is particularly large,

435   over 65%. These points are gathered east of the Georgian Bay (45.3°N; 79.2°W), south of the Lake Nipissing

436   (47.9°N; 80.2°W), north of the Gouin Reservoir (49.6°N; 74.5°W), and over New Brunswick (46.9°N 65.9°W).

437   For these points, GSMaP has a poor consistency with other products, which is mainly caused by its strong

438   underestimation of the mean precipitation (ΔP_M values around 45%).

439   Figure 1.7c,d show that the most likely products differ between NEUS and SEUS sub-domains, with NEUS

440   using MSWEP, GSMaP, STAGE IV and IMERG for most grid points, while SEUS uses IMERG, GSMaP and

441   MSWEP. These panels show that the consistency between CMORPH and other products is much lower in

442   NEUS than in SEUS. On the contrary, ERA5 has a poor consistency with others in SEUS and is only selected

443   east of the NEUS sub-domain (< 75°W) where it becomes more consistent with others than IMERG. Figure

444   1.7e,f show that for the ocean sub-domains, IMERG, MSWEP and CMORPH are always used. GSMaP is used

445   for most of the ocean points, except over northern AO (15% of AO points) where ERA5 is usually considered.

446   1.4.4   Seasonality of products uncertainty

447   Figure 1.8a shows monthly mean values of max_4 averaged for each sub-domain. The uncertainties of

448   CAGL and the NEUS sub-domains are clearly larger in winter than in summer. The CAGL sub-domain has

449   an uncertainty of 73% in winter and 57% in summer. Other sub-domains generally show small differences

450   throughout the year, although SEUS, MG and AO sub-domains exhibit somewhat larger values in summer,

451   with values of 46 and 55% for the GM sub-domain. Further analysis (see Figure F.4 in Appendix F) shows

452 that the high uncertainty in summer is particularly high for lower latitude and near coasts (including

453 islands).
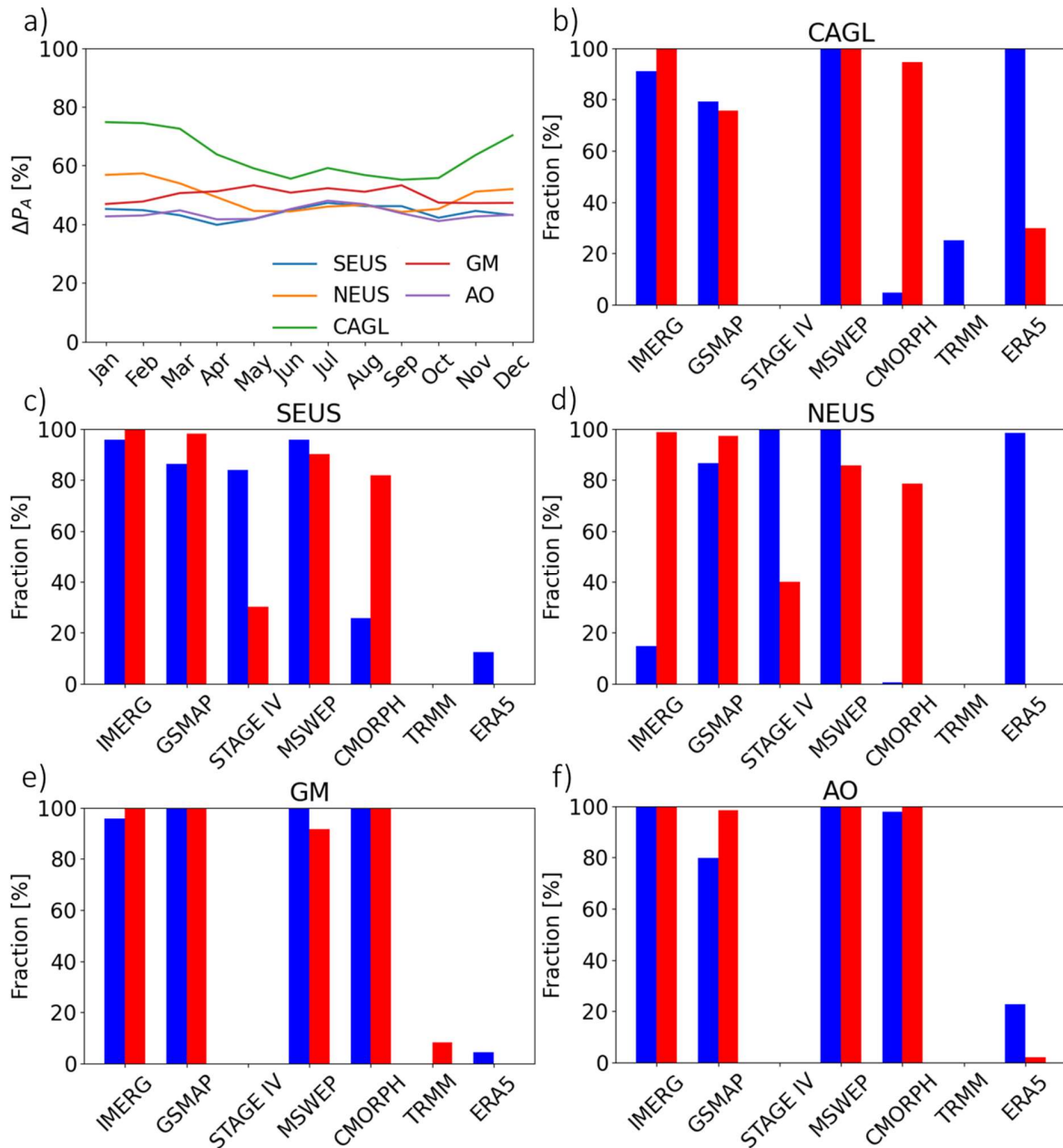
454

455 Figure 1.8: (a) Annual cycle of the max_4 uncertainty metric for each sub-domain and (b-f) fraction of grid points
456 that are selected within max_4 for each product in (red) summer and (blue) winter. The statistics are computed for
457 the metric ΔP_A.

458 Figure 1.8b-d show that over land, the fraction of CMORPH, ERA5, STAGE IV and IMERG have a strong

459 seasonal variability. CMORPH is less used in winter than in summer and for the NEUS sub-domain, its

460   fraction varies from near 0% in winter to 80% in summer. IMERG has a similar fraction for SEUS and Canada

461   sub-domains (with slightly lower values in winter) but for NEUS, it is clearly replaced by ERA5 in winter

462   For the NEUS sub-domain, STAGE IV fraction varies from 40% in the summer to 100% in the winter. The

463   ERA5 reanalysis is one of the least consistent products in the summer in the south, when the precipitation

464   mainly results from convective meteorological systems, which are poorly simulated by reanalyses as ERA5

465   (Beck et al. 2020). Figure 1.8e,f show that the fraction is less dependent on the season for ocean sub-

466   domains. The products IMERG, MSWEP, GSMaP and CMORPH are the most likely for most of the ocean

467   grid points in winter and in summer.

468   1.4.5   Diurnal cycle of products uncertainty

469   In this section, we investigate whether the uncertainty across products shows some dependence on the

470   time of the day by looking at the diurnal cycle during the summer season. Hours are expressed in UTC, and

471   it should be noted that there is a 2h40 difference in local time between the eastern and western part of

472   the studied region, which is 40° longitude wide. For example, the period 1500 and 1800 UTC is equivalent

473   in local time to the period 8:20 am. – 11:20 am. at the longitude 100°W and to 11:00 am. – 2:00 pm. at the

474   longitude 60°W.

475   Figure 1.9 shows the diurnal cycle of the mean summer precipitation for each product in the five sub-

476   domains. Most sub-domains display a clear diurnal cycle of P_M and the amplitude of this cycle tends to be

477   greater for southern domains (GM, SEUS, and AO) than for northern ones (NEUS and CAGL). Over GM,

478   most products show that mean precipitation values reach a maximum at 1500 UTC (values between 0.75

479   and 1 mm $(3\ h)^{-1}$) and a minimum at 0300 UTC (values between 0.25 and 0.35 mm $(3\ h)^{-1}$). The maximum

480   of P_M diurnal cycle occurs for the SEUS sub-domain 6h later for all products. For AO, the maximum is not

481   consistently represented across products, occurring between 0900 UTC for ERA5, MSWEP and CMOPRH

482   and 1500 UTC for TRMM. For CAGL, mean precipitation values vary between 0.2 and 0.45 mm $(3\ h)^{-1}$ and

483   the daily cycle do not seem to be consistent across the products. Figure A.4 in the APPENDIX E displays the

484   diurnal cycle of the mean precipitation in winter; it shows that 3-hourly precipitation estimates do not

485   seem to have a clear diurnal for ΔP_M for the most likely products, the mean precipitation has small

486   variation during the day.

P_M ΔP_A
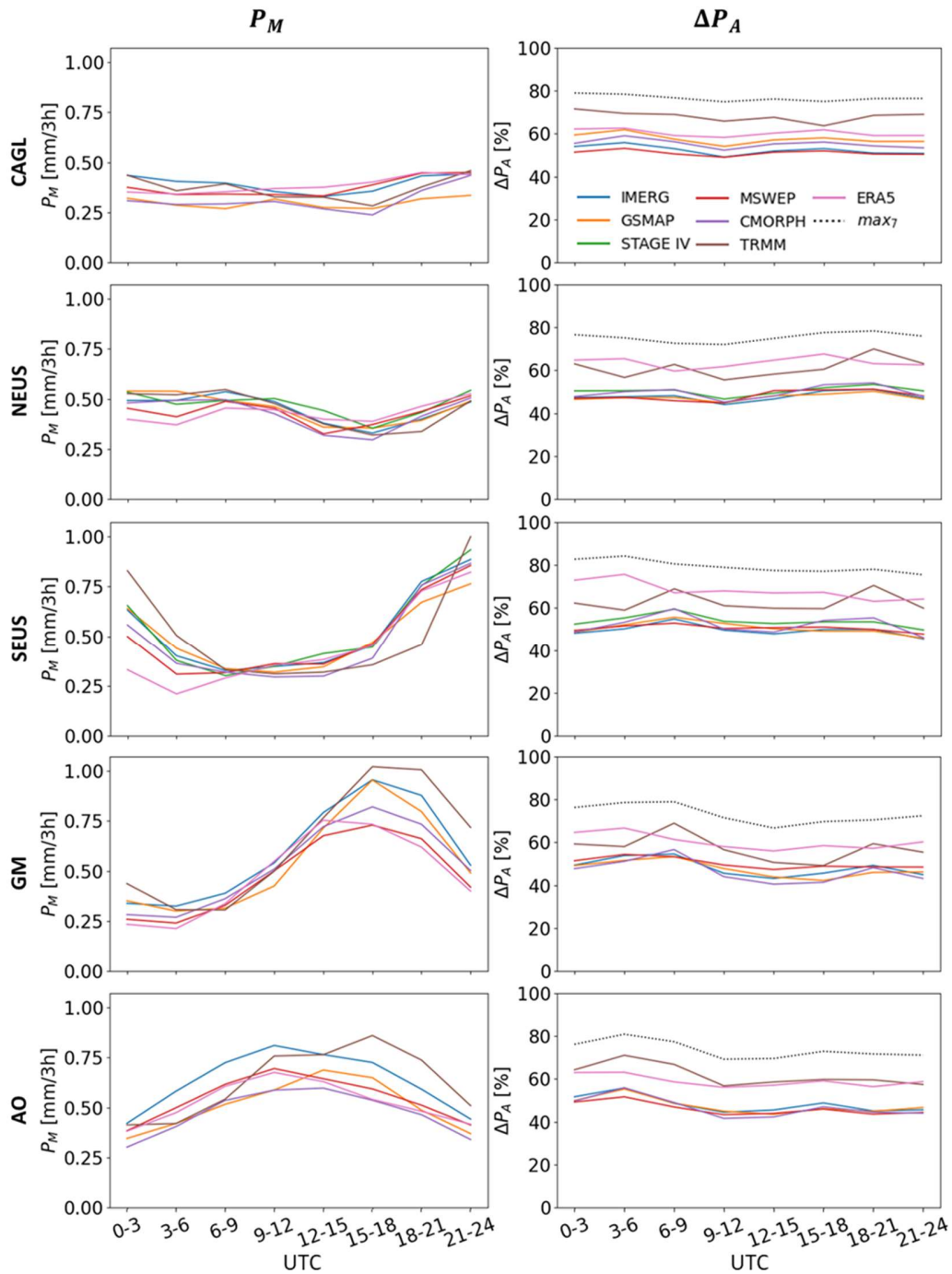
Figure 1.9: (left) Diurnal cycle of the mean precipitation and (right) 3-hourly values of $\overline{\Delta P_A}^p$ and max_7, for each sub-domain in summer.

Figure 1.9 shows that $\overline{\Delta P_A}^p$ exhibits small diurnal variation for most products. For NEUS, $\overline{\Delta P_A}^p$ varies between 45 and 55% for all products, except ERA5 and TRMM during the day. These two products, which are also the less consistent products for the five sub-domains, are the only products with a notable diurnal

493   variation of $\overline{\Delta P_A}^p$. TRMM is less consistent with other products at 0600 and 1800 UTC (1-2 a.m. and 13-14

494   p.m. local time) for SEUS, NEUS and GM sub-domains and at 0300 UTC for AO. For GM, $\overline{\Delta P_A}^{TRMM}$ reaches

495   70% and 60% at 0600 and 1800 UTC and 49% at 1500 UCT. ERA5 tends to be less consistent at 0300 UTC

496   for Southern sub-domains (GM, SEUS, and AO), and at 15 UCT for Northern sub-domains (CAGL and NEUS).

497   For NEUS, these two peaks are visible and $\overline{\Delta P_A}^{ERA5}$ reaches two maximums at 0300 and 1500 UTC with 65

498   and 69% for a minimum value of 58%.

499   1.4.6   Sensitivity of the uncertainty to the choice of temporal and spatial resolutions

500   Most products were interpolated from their native grid to the coarser resolution 0.25° ERA5 grid to be

501   able to compare them. The sensitivity of the uncertainty to a change in the resolution of the output grid

502   was carried by decreasing the spatial resolution from 0.25° to 0.75°. Similarly, the sensitivity of the

503   uncertainty to the temporal resolution of the data is assessed by comparing 3-hourly, 6-hourly, and daily

504   data.
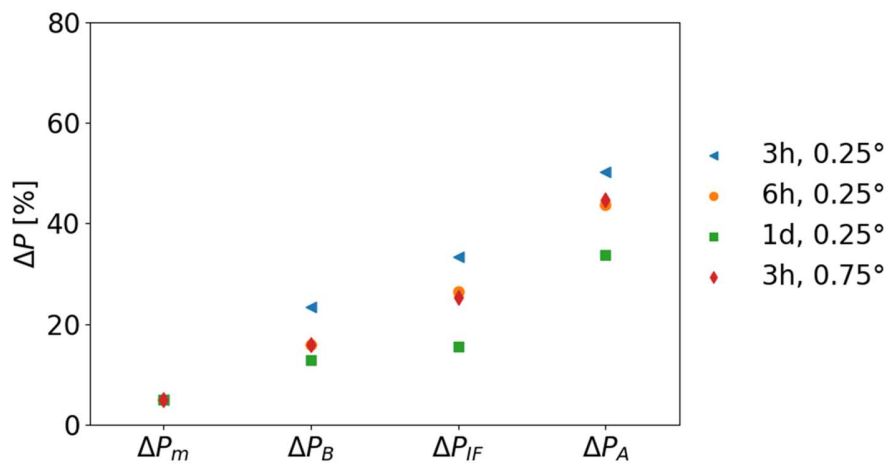
506   Figure 1.10. Impact of a change in resolution on the differences between products. Differences are averaged for US
507   sub-domains and for all comparisons.

508   Decreasing the spatial or temporal resolution leads to a decrease of the uncertainty (increase in the

509   consistency) for ΔP_B, ΔP_{IF} and ΔP_A but, as expected, it has no impact on ΔP_M. The ΔP_{IF} distance metric is

510   the most sensitive to a change of temporal resolution, a change of resolution from 3 h to 1 day leading to

511   a decrease from 38% to 20% of its values.

512   1.5   Discussion

513   The lack of a high-resolution precipitation dataset that can be assumed as truth implies that it is

514   challenging to characterise the error of individual gridded products. Even when assuming that gauge-based

515   estimates could provide the best precipitation estimate, the difference in the representativity of gauges

516   and satellite estimates limit their direct comparison. Considering in situ observations as a reference also

517   raises issues about the evaluation and use of satellite products for regions with little or no direct

518   observation (Timmermans et al. 2019; Prein and Gobiet 2017). In the absence of a reference dataset,

519   multiple products can be used to provide an estimation of the level of consistency or uncertainty. If the

520   observational uncertainty is low (i.e., differences are small), it can be assumed that the products are

521   accurate.

522   For a given distance metric, the uncertainty of a set of products can be defined and quantified following

523   different approaches, which are not all useful in our case. The average and median of the inter-product

524   comparisons gives little information on the local consistency of the precipitation estimates. The maximum

525   of the differences generally returns the distance between the worst products, which are potentially bad.

526   Therefore, in our study, we used the uncertainty of the four products with the lowest difference value

527   amongst seven, max_4, which returns a more relevant uncertainty for the evaluation of a set of products.

528   It also allows us to identify the least likely and potentially wrong products. The product likelihood is

529   investigated for four distance metrics that assess the discrepancies between the 3h-precipitation

530   estimates. The results are only shown for ΔP_A, which is the only metric capturing disparities in the

531   precipitation time series, not only associated with the total mean accumulation and the

532   intensity/frequency distribution, but also with the chronology. Using metrics of increasing complexity

533   helps to identify the sources of discrepancies.

534   The uncertainty exhibits strong regional and annual variabilities. Regional analysis showed that CAGL has

535   the highest mean uncertainty, and the Atlantic Ocean and southern CONUS have the lowest one. Moreover,

536   the uncertainty tends to be larger in winter for Northern domains (Canada, NEUS), while larger in summer

537   for Southern domains. These results area for the uncertainty of the four most likely product and represent

538   thus a general trend within the set of products. For North land, the low density of stations over Canada

539   that affect the precipitation correction (Prein and Gobiet, 2017) and the larger errors in satellite retrieval

540   algorithms when representing snowfall or precipitation in snow-covered areas (Henn et al. 2018;

541   Trenberth et al. 2017) may explain these results. In southern domains (SEUS, MG, and AO), the large

542   uncertainty in summer is probably caused by the predominance of convective precipitation that are

543   associated with high spatial and temporal variability (Prein and Gobiet 2017; Beck et al. 2020). Over Canada,

544   the four most probable products uncertainty reaches almost 3/4 of the total precipitation amounts. This

545   particularly large uncertainty shows one of the limits of high-resolution products available today. The

546   relationship between ERA5, MSWEP and STAGE IV is visible in our results. However, we have noted enough

547   disparities between these products to keep the three of them in our study.

548   Methodological choices affect the datasets and thus can modify their consistency. The distance between

549   products and the uncertainty of the set of products decreases with resolution for all metrics except ΔP_M.

550   Reducing of temporal resolution has a smoothing effect on the intensity of precipitation (Norris et al. 2019;

551   Trenberth and Zhang 2018; Chen and Dai 2018; Trenberth et al. 2017). Product uncertainty is smaller for

552   long term or spatial mean values than for patterns at specific times or when examining fine spatial detail.

553   However, lowering the resolution reduces the uncertainty (which can be seen as a way to reach more

554   practical values for a given application) but also reduces the information within datasets.

555   1.6   Conclusion

556   We developed a methodology to quantify the uncertainty among a set of seven high-resolution gridded

557   precipitation datasets and to identify, at individual grid points, the least/most likely products. At each grid

558   point, the uncertainty is estimated as the largest precipitation difference among the four most consistent

559   products of the seven, thus avoiding the overestimation of uncertainty arising from erroneous or

560   unrealistic products. We used four distance metrics to quantify several aspects of the precipitation time

561   series in a hierarchical way: from the simple total accumulation (denoted as ΔP_M), to the variability of

562   precipitation by looking at the intensity and frequency (ΔP_{IF}) or the intensity distribution (ΔP_B), to the

563   inclusion of the right chronology of events (ΔP_A). The absolute difference ΔP_A is interesting because it

564   captures discrepancies not only associated with the total mean accumulation and the intensity/frequency

565   distribution, but also with the chronology of precipitation events.

566   We found that the uncertainty shows strong spatial and seasonal variations. The uncertainty is greater in

567   the north of the domain, over southern Canada, where it is almost as large as the total precipitation.

568   Additionally, in the north of the domain, the uncertainty appears to be greater in winter than in summer.

569   These results are probably due to the low density of stations in these regions that affect the correction of

570   precipitation estimates from satellite (Prein and Gobiet 2017) and to the larger errors in satellite retrieval

571 algorithms when representing snowfall or precipitation over snow-covered areas (Henn et al. 2018;

572 Trenberth et al. 2017; Xie et al. 2017). In the south of the domain, the uncertainty is greater in summer,

573 probably due to the predominance of convective precipitation, which is associated with high spatial and

574 temporal variability (Prein and Gobiet 2017).

575 Our results show that no single product appears to be best over all regions and seasons, although we

576 clearly show that some individual products are very likely 'outliers' (in the sense that they behave

577 differently from the rest of the pack) for specific regions and seasons. For example, CMORPH produces

578 little or no precipitation in winter at high latitudes. Also, the precipitation chronology of ERA5 is not

579 consistent with other products in the summer for lower latitude of our domain. TRMM is inconsistent with

580 other products, and use of its successor, IMERG, is preferable. Recent precipitation estimates based on

581 satellite observations (IMERG, GSMaP, and MSWEP) are generally part of the four most likely ones,

582 especially over ocean. However, GSMaP and IMERG tend to be less consistent over Canada in winter,

583 where locally GSMaP produces little precipitation.

584 Further work is needed to understand how the uncertainty metric depends on an ensemble of products

585 that are not independent from one another. For example, the MSWEP product uses ERA5 and IMERG data.

586 In addition, it would be valuable to take advantage of the broader availability of sub-daily in situ

587 observations (e.g., Smith et al. 2011) to establish a reference product.

CONCLUSION

Nous avons développé une méthodologie pour quantifier l'incertitude parmi un ensemble de sept estimations maillées de la précipitation à haute résolution. Cette méthodologie permet par ailleurs d'identifier, aux points de grille individuels, les produits les moins/les plus probables. À chaque point de grille, l'incertitude est estimée comme étant la plus grande différence de précipitations parmi les quatre produits les plus probables, évitant ainsi la surestimation de l'incertitude découlant de produits erronés ou irréalistes. Quatre distances ont été utilisées pour quantifier les disparités dans l'estimation de la précipitation pour des quantités à la complexité croissante de l'estimation de la précipitation : de la simple accumulation totale (notée ΔP_M), à la variabilité des précipitations en examinant l'intensité et la fréquence (ΔP_{IF}) ou la distribution de l'intensité (ΔP_B), à l'inclusion de la bonne chronologie des événements (ΔP_A). Pour cet article, les résultats sont principalement présentés pour la différence absolue, ΔP_A, qui capture les disparités non seulement associées au cumul moyen total et à la distribution intensité/fréquence, mais aussi à la chronologie des événements de précipitations.

Nous avons constaté que l'incertitude présente de fortes variations spatiale et temporelle. L'incertitude est plus grande dans le nord du domaine, en particulier pour le sud du Canada, où elle est presque aussi grande que la précipitation totale. Dans le nord du domaine, l'incertitude est généralement plus grande en hiver qu'en été. Ces résultats sont probablement dus à la faible densité de stations dans ces régions qui affectent la correction des estimations de précipitations par satellite (Prein et Gobiet 2017) et aux erreurs plus importantes des algorithmes de récupération par satellite pour la neige ou les précipitations au-dessus des zones enneigées (Henn et al. 2018 ; Trenberth et al. 2017 ; Xie et al. 2017). A l'inverse, dans le sud du domaine, l'incertitude est plus importante en été, probablement en raison de la prédominance des précipitations convectives qui sont associées à une forte variabilité spatiale et temporelle (Prein et Gobiet 2017).

Nos résultats montrent qu'aucun produit n'est meilleur que les autres pour l'ensemble des régions et des saisons, bien que nous prouvions clairement que certains produits individuels sont très probablement faux pour des régions et des saisons spécifiques. Par exemple, CMORPH produit peu ou pas de précipitations en hiver aux hautes latitudes. De même, la chronologie des précipitations de ERA5 est incohérente avec les autres produits en été pour les plus basses latitudes de notre domaine. TRMM n'est pas cohérent avec les autres produits, et l'utilisation de son successeur, IMERG, est donc à privilégier. Il est cependant à noter

que les produits récents basés sur des observations satellitaires (IMERG, GSMaP et MSWEP) font généralement partie des quatre produits les plus probables, surtout sur les points de grille océaniques. GSMaP et IMERG ont tendance à être moins cohérents sur le Canada en hiver, où localement GSMaP produit peu de précipitations.

Des travaux supplémentaires sont nécessaires pour comprendre comment la métrique d'incertitude dépend d'un ensemble de produits qui ne sont pas indépendants les uns des autres. Par exemple, le produit MSWEP utilise des données ERA5 et IMERG. En outre, il serait utile de profiter de la plus grande disponibilité des observations in situ infra-quotidiennes (par exemple, Smith et al. 2011) pour établir un produit de référence.

APPENDIX A

Spatial interpolation and temporal resampling of precipitation datasets

The interpolation of 3-hourly precipitation high-resolution fields to a coarser grid should be made with caution as it can create strong discontinuities in the intensity and frequency of precipitation (due to file compression issues, ERA5). This effect is visible when the number of input cells (0.1°) used in the interpolation varies from one output cell (0.25°) to another. To correct this effect, the native grid of the MSWEP and GMSAP product were shifted in latitude and longitude by a factor +0.025°. This shift has a limited impact on the following comparisons for quantities based on the mean precipitation or the instantaneous precipitation and does not change the overall conclusion of this study.
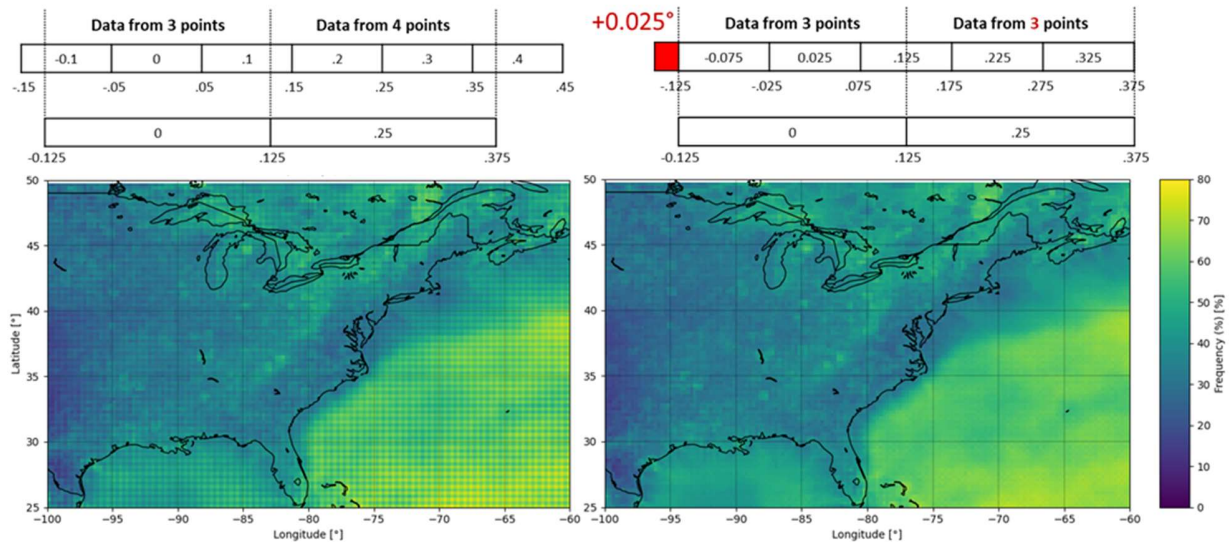

Figure A.1. (up) Graphical example of the grid shift used in the interpolation of MSWEP and GSMaP datasets (0.1°) to match the ERA5 grid (0.25°). (down) Graphics of the precipitation frequency for the 2015-2019 for the MSWEP product at 0.25° with (right) and without (left) shift in the native grid before interpolation.

APPENDIX B

Precipitating events

In our study, precipitating events were defined using a threshold of 1 mm day^{-1} (i.e., 0.125 mm (3 h)$^{-1}$). The sensitivity of the various distance metrics to the choice of the minimum precipitation value is assessed here.

Several factors including products interpolations or data compression can artificially generate new precipitation events within a dataset (see the section "Precipitation Data Preprocessing"). Due to their (usually) small intensity, generated precipitation events have a limited impact on precipitation accumulation but need to be removed to correctly estimate the precipitation frequency. A threshold is usually used to do so, only precipitations with higher intensities are considered. There is no agreement on the choice of threshold. It is usually chosen based on the precision of the observation used (e.g. Barbero et al. 2019; Trenberth et al. 2017) or to ensure comparability with earlier studies (e.g. Chen and Dai 2018). However, The number of precipitation occurrence depends on the k value and an increase in k value results to a lower frequency and higher intensity (Chen and Dai 2018; Zhou et al. 2008).
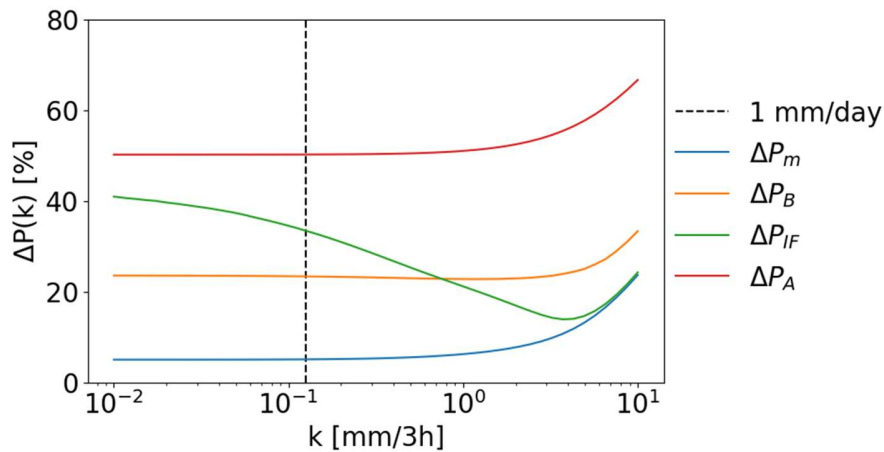

Figure A.2. Distance metrics values averaged over the NEUS and SEUS sub-domains and comparisons as a function of k. The 1 mm d^{-1} is shown by the dashed line.

Figure A.2 shows distances values averaged for all comparisons over the NEUS and SEUS sub-domains as a function of the threshold. The threshold choice has a limited impact on ΔP_M, ΔP_B and ΔP_A for values around 1 mm d^{-1}. ΔP_M and ΔP_A increase with the threshold value: the decrease of their absolute values is slower than the decrease of the mean precipitation. ΔP_B has a value of 23.3% for k =1 mm d^{-1} and reaches

a minimum of 22.7% for k =1.2 mm (3 h)$^{-1}$ (9.6 mm d^{-1}). ΔP_{IF} is more sensitive to the threshold value: it minimum value is near 15% for k =3.7 mm (3 h)$^{-1}$ (30 mm d^{-1}) when its value for k =1 mm d^{-1} is 33.4%. This gain in consistency is due to a decrease in the frequency uncertainty, the intensity uncertainty is little affected by the choice of threshold. These results are similar for other main domains of the regions (not shown).

Table B.1 shows the fraction of neglected precipitation events and amounts for a threshold of 1 mm d^{-1}. ERA5 has the highest fractions with 59% of the events and 2.7% of the total precipitation amount that are neglected while it only represents 2.6% of TRMM events and almost 0% of its precipitation accumulation.

Table B.1: Fraction of precipitation events with an intensity under 1 mm d^{-1} and the fraction of the total precipitation accumulation within these events over the studied domain and region.

Product	Neglected precipitation events	Neglected precipitation
CMORPH	32%	0.8%
IMERG	56%	1.3%
MSWEP	42%	2.3%
TRMM	2.6%	0.0037%
GSMAP	49%	1.8%
STAGE IV	31%	0.58%
ERA5	59%	2.7%
PERSIANN	29%	0.50%
Total	38%	1.2%

APPENDIX C

Sensitivity of ΔP_B to the number of bins

The ΔP_B value obtained when comparing two given datasets depends on the number and distribution of the intensity bins. The choice of bin number needs to consider the signal to noise ratio: A small bin number reduces noise within the distribution while a great number gives greater precision to the intensity distribution of the precipitation. Several methods exist to determine the number of bins for a distribution. However, in our case, the intensity distribution changes from one product to another. Using the combination of products distribution for these methods returns much larger bin numbers. For our analysis, the choices of 130 logarithmic bins for overall period and 20 bins for the monthly analysis were arbitrary and conservative in regard to the number of bins returned by other methods.
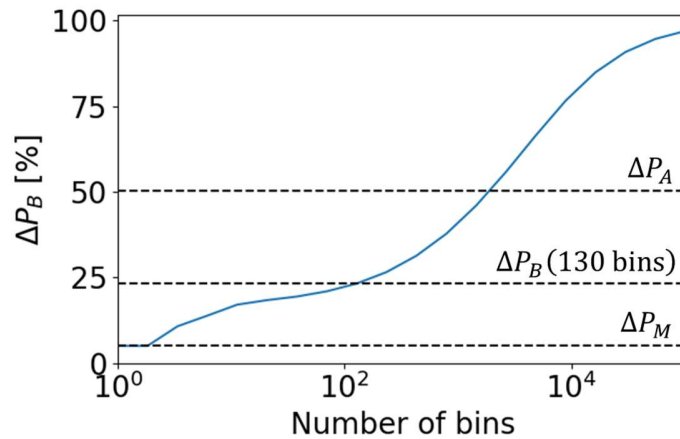

Figure A.3. Averaged values of ΔP_B over NEUS and SEUS domains as a function of the bin number. ΔP_A, ΔP_B for 30 bins and ΔP_M are shown by the dashed lines

Figure A.3 shows the ΔP_B value, averaged across all products, over the combined NEUS and SEUS domains as a function of the number of bins. The inequality $\Delta P_A \geq \Delta P_B \geq \Delta P_M$ is valid for bins numbers under 2 000 bins. For small variations of the bin number around 130, ΔP_B results changes in values but the analysis remains similar. For too high bin numbers, ΔP_B measures disparities the list of precipitation events and compares events to one another, and not the distribution as expected. In this case, the metric tends to 100%.

APPENDIX D

Bounded distance metrics

The equation 10 is verified for the four-distance metrics. For ΔP_M and ΔP_A, the demonstration is straight forward but is more complex for ΔP_B and ΔP_{IF}.

For the difference in the intensity distribution ΔP_B, we have:

$$\Delta P_B = \frac{1}{N_T} \sum_{B_j} \left| P_{B_j}^{(1)} - P_{B_j}^{(2)} \right| \le \frac{1}{N_T} \left(\sum_{B_j} P_{B_j}^{(1)} + \sum_{B_j} P_{B_j}^{(2)} \right) = P_M^{(1)} + P_M^{(2)}$$

For the difference based on the intensity-frequency precipitation decomposition ΔP_{IF}, the demonstration can be done for compensation between the intensity and frequency of two products and for no compensation. In the case of no compensation, the metric ΔP_{IF} equals ΔP_M for which the inequality is true.

For compensation, the product with the highest frequency also has the lowest intensity), we can assume that:

$$P_M^{(1)} + P_M^{(2)} = \min\left(I^{(1)}, I^{(2)}\right).\max\left(F^{(1)}, F^{(2)}\right) + \max\left(I^{(1)}, I^{(2)}\right).\min\left(F^{(1)}, F^{(2)}\right)$$

And we have:

$$\Delta P_{IF}\left(P^{(1)}, P^{(2)}\right) = \min\left(I^{(1)}, I^{(2)}\right).\left|F^{(1)} - F^{(2)}\right| + \min\left(F^{(1)}, F^{(2)}\right).\left|I^{(1)} - I^{(2)}\right|$$
$$\le \min\left(I^{(1)}, I^{(2)}\right).\max\left(F^{(1)}, F^{(2)}\right) + \min\left(F^{(1)}, F^{(2)}\right).\max\left(I^{(1)}, I^{(2)}\right) = P_M^{(1)} + P_M^{(2)}$$

APPENDIX E

Products mean precipitation and frequency
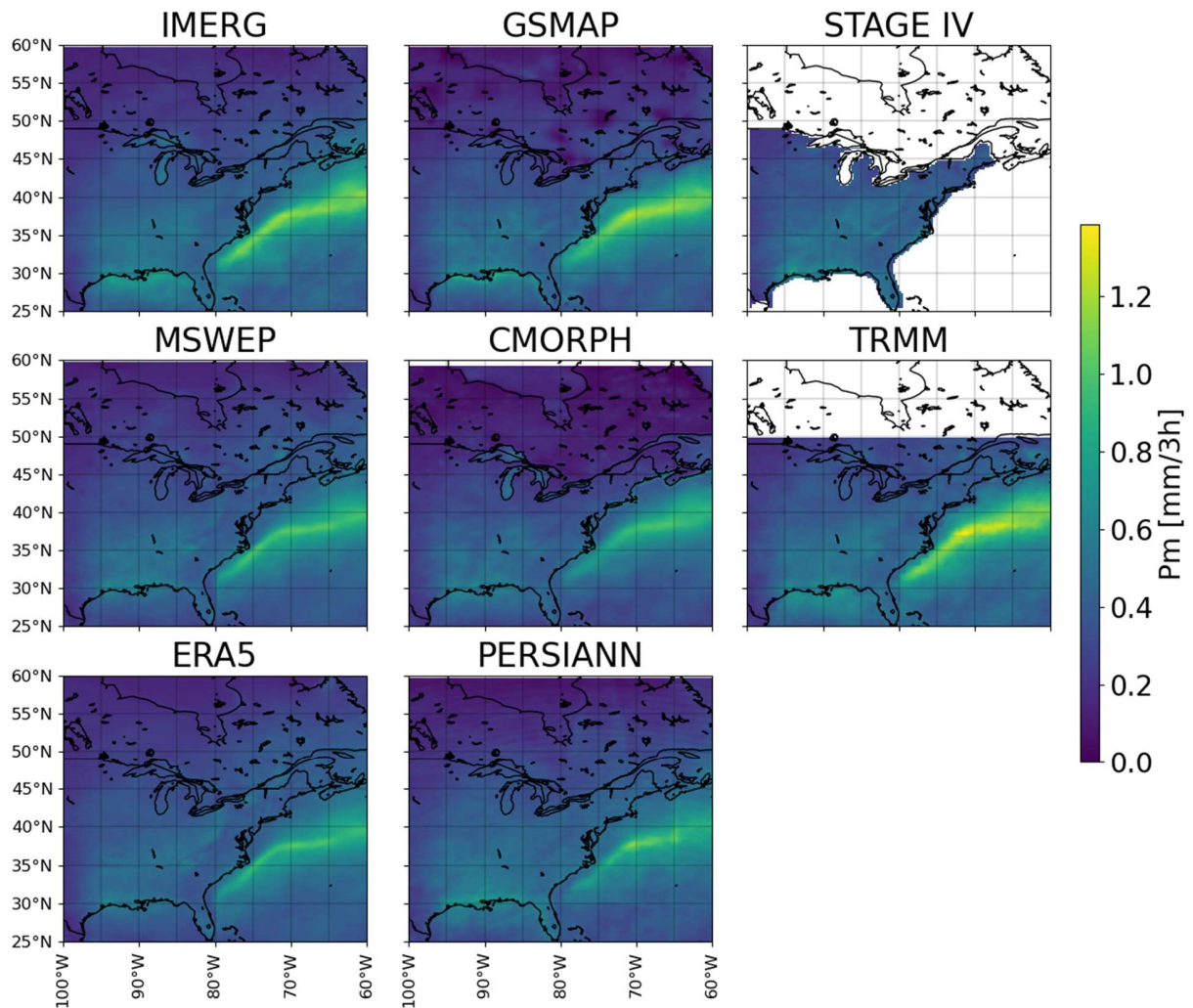

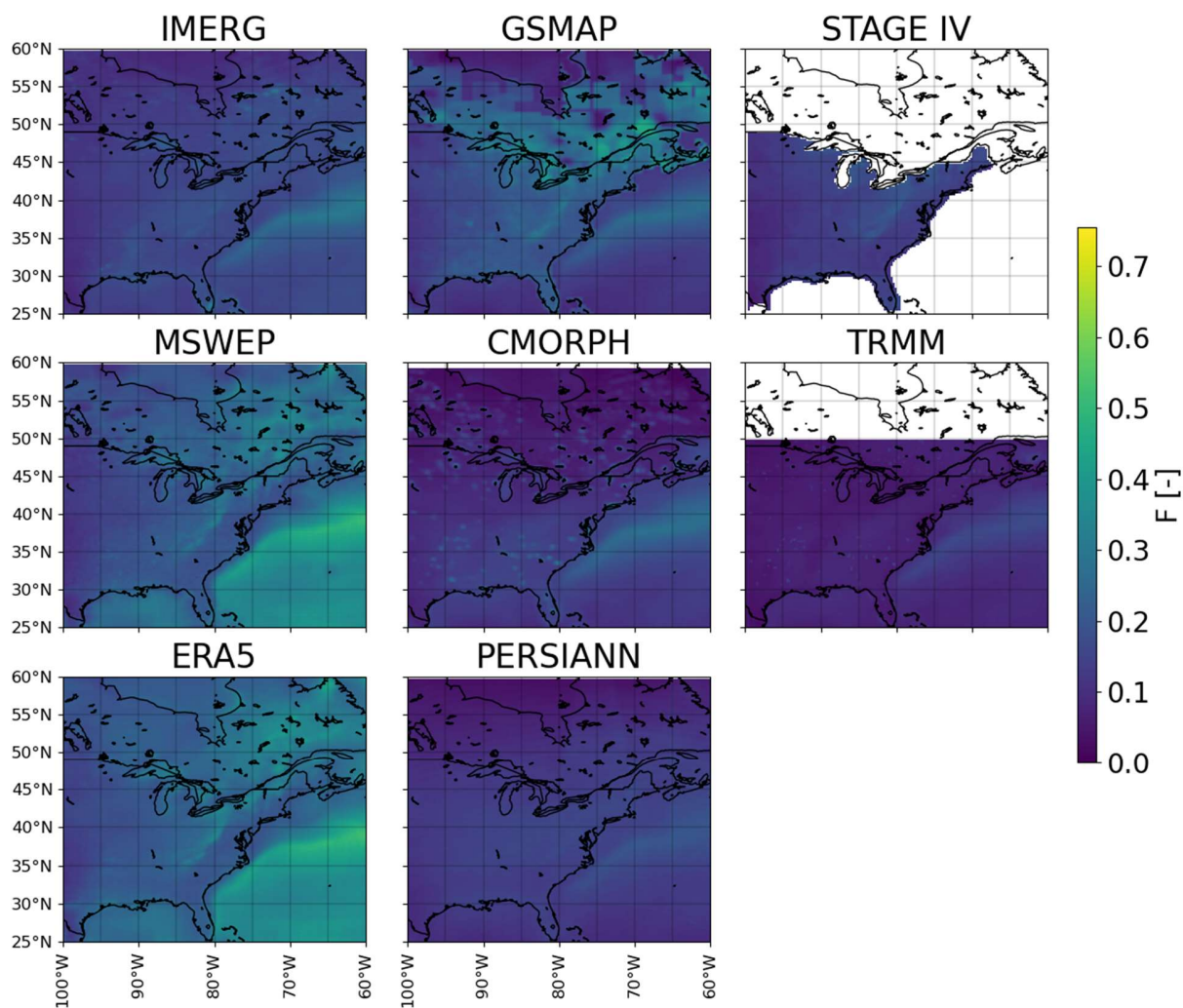Figure E.1: Mean precipitation for each product

Figure E.2: Precipitation frequency for each product
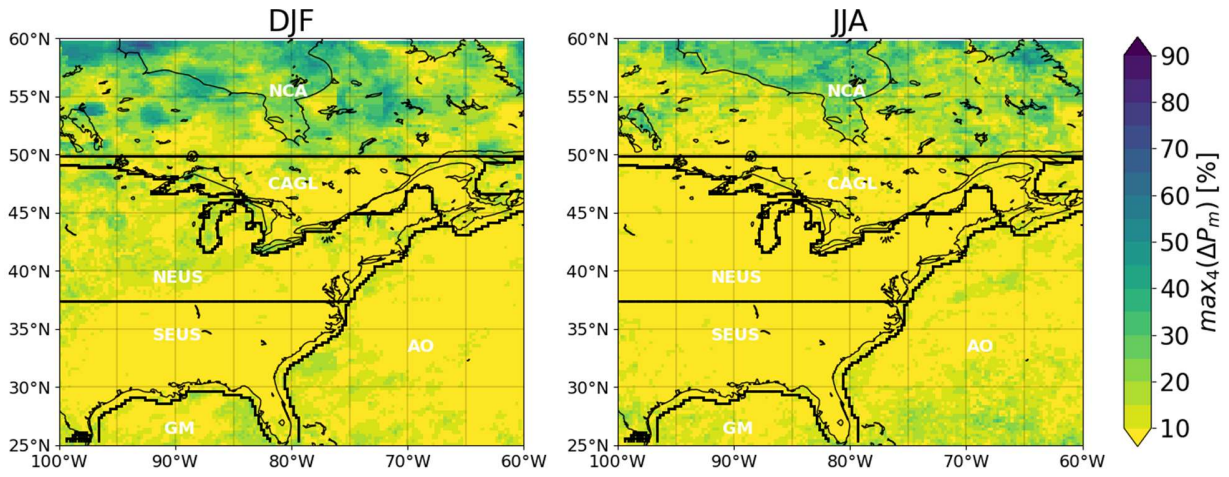
APPENDIX F

Summer and winter uncertainties

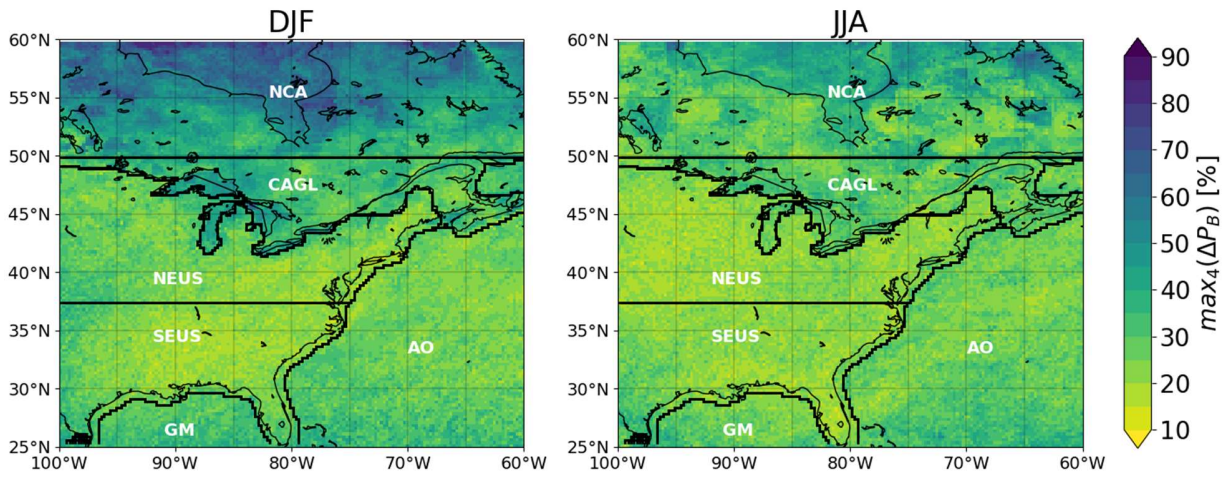Figure F.1: Winter (left) and summer (right) values of max_4 for ΔP_M
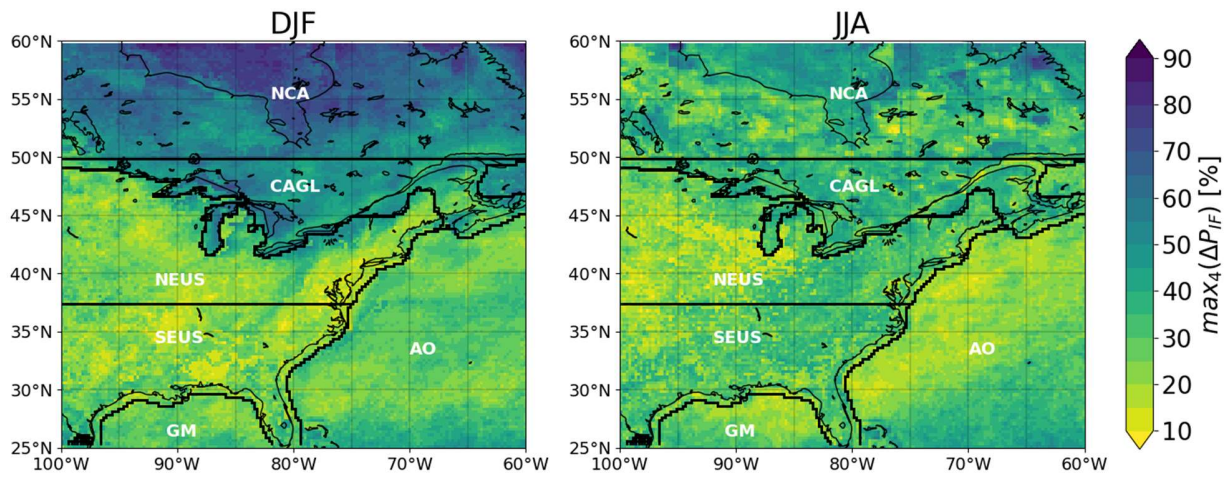

Figure F.2: Same as Figure F.1 for ΔP_B

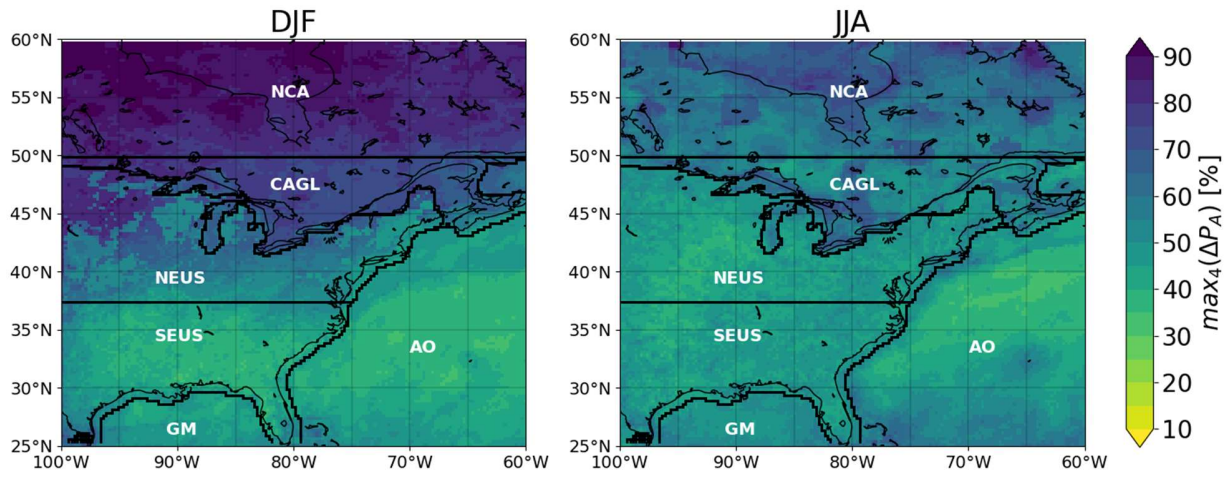Figure F.3: Same as Figure F.1 for ΔP_{IF}

Figure F.4: Same as Figure F.1 for ΔP_A

APPENDIX F

Diurnal cycle of the P_M in winter

Figure A.4 shows the mean precipitation in winter for each time step of the day for winter months. In winter, no diurnal cycle for the 3 h precipitation is visible for most of the domains. In contrast to summer, in winter, the amplitude and position of the maxima and minima are not consistent between the products for most domains. However, IMERG, MSWEP, ERA5 share a maximum of precipitation at 0900 UTC over AO which is also visible for summer precipitation. TRMM is the product with the most sub-daily variation notably over Canada where its mean precipitation varies from 0.23 mm (3 h)$^{-1}$ at 2100 UTC to 0.42 mm (3 h)$^{-1}$ at 0600 UTC.
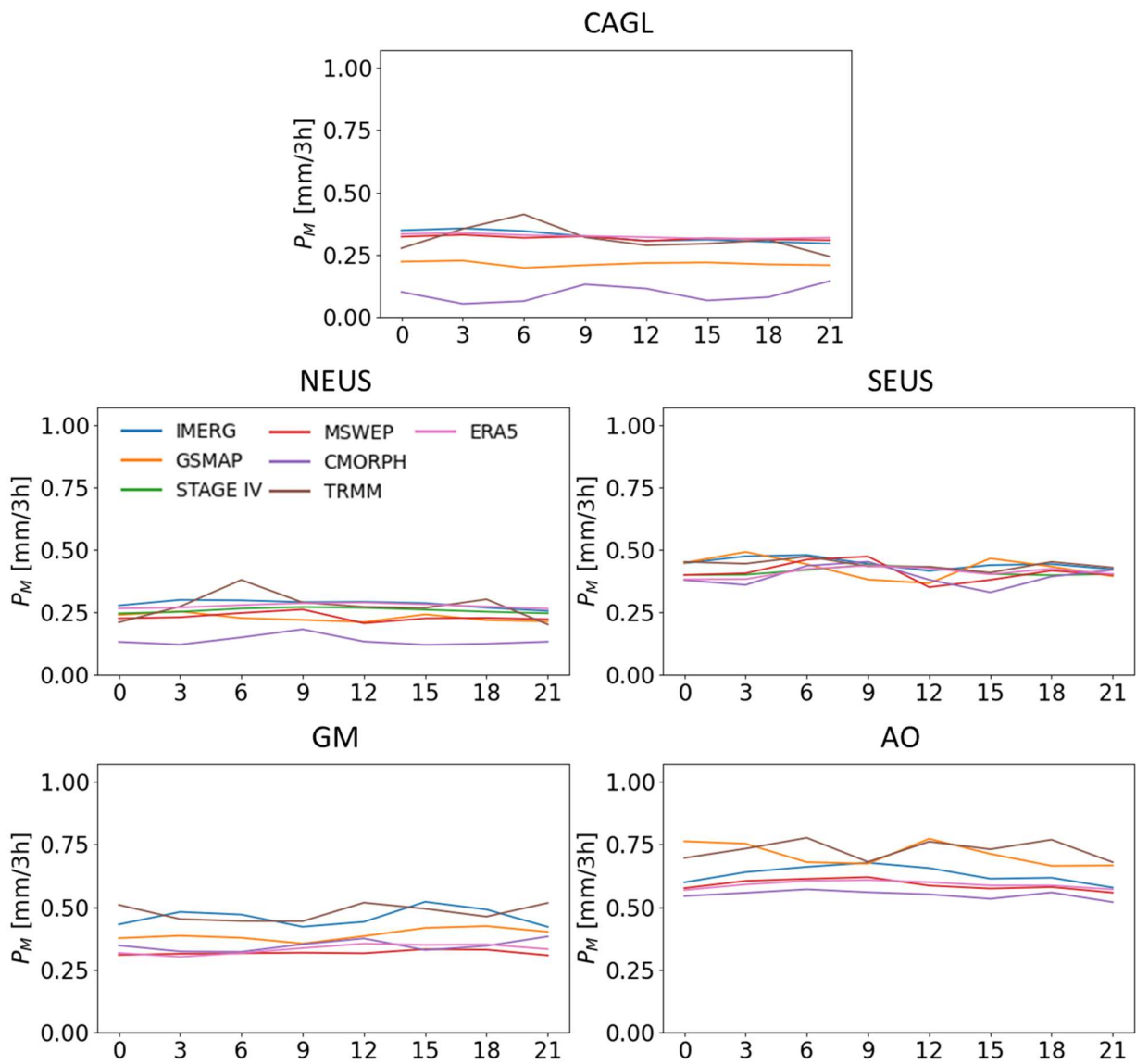
Figure A.4. diurnal cycle of the mean precipitation for each domain in winter.

RÉFÉRENCES

Aghakouchak, A., A. Mehran, H. Norouzi, and A. Behrangi, 2012: Systematic and random error components in satellite precipitation data sets. *Geophys. Res. Lett.*, **39**, 3–6, https://doi.org/10.1029/2012GL051592.

Barbero, R., and Coauthors, 2019: A synthesis of hourly and daily precipitation extremes in different climatic regions. *Weather Clim. Extrem.*, **26**, https://doi.org/10.1016/j.wace.2019.100219.

Beck, H. E., A. I. J. M. Van Dijk, V. Levizzani, J. Schellekens, D. G. Miralles, B. Martens, and A. De Roo, 2017: MSWEP: 3-hourly 0.25° global gridded precipitation (1979-2015) by merging gauge, satellite, and reanalysis data. *Hydrol. Earth Syst. Sci.*, **21**, 589–615, https://doi.org/10.5194/hess-21-589-2017.

——, and Coauthors, 2019a: Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. *Hydrol. Earth Syst. Sci.*, **23**, 207–224, https://doi.org/10.5194/hess-23-207-2019.

——, E. F. Wood, M. Pan, C. K. Fisher, D. G. Miralles, A. I. J. M. Van Dijk, T. R. McVicar, and R. F. Adler, 2019b: MSWep v2 Global 3-hourly 0.1° precipitation: Methodology and quantitative assessment. *Bull. Am. Meteorol. Soc.*, **100**, 473–500, https://doi.org/10.1175/BAMS-D-17-0138.1.

——, and Coauthors, 2020: Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Adv. Glob. Chang. Res.*, **69**, 625–653, https://doi.org/10.1007/978-3-030-35798-6_9.

Catto, J. L., C. Jakob, and N. Nicholls, 2013: A global evaluation of fronts and precipitation in the ACCESS model. *Aust. Meteorol. Oceanogr. J.*, **63**, 191–203, https://doi.org/10.22499/2.6301.012.

——, ——, and ——, 2015: Can the CMIP5 models represent winter frontal precipitation? *Geophys. Res. Lett.*, **42**, 8596–8604, https://doi.org/10.1002/2015GL066015.

Chen, C. T., and T. Knutson, 2008: On the verification and comparison of extreme rainfall indices from climate models. *J. Clim.*, **21**, 1605–1621, https://doi.org/10.1175/2007JCLI1494.1.

Chen, D., and A. Dai, 2018: Dependence of estimated precipitation frequency and intensity on data resolution. *Clim. Dyn.*, **50**, 3625–3647, https://doi.org/10.1007/s00382-017-3830-7.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.*, **137**, 553–597, https://doi.org/10.1002/qj.828.

Derin, Y., and K. K. Yilmaz, 2014: Evaluation of multiple satellite-based precipitation products over complex topography. *J. Hydrometeorol.*, **15**, 1498–1516, https://doi.org/10.1175/JHM-D-13-0191.1.

Fosser, G., S. Khodayar, and P. Berg, 2015: Benefit of convection permitting climate model simulations in the representation of convective precipitation. *Clim. Dyn.*, **44**, 45–60, https://doi.org/10.1007/s00382-014-2242-1.

Gehne, M., T. M. Hamill, G. N. Kiladis, and K. E. Trenberth, 2016: Comparison of global precipitation estimates across a range of temporal and spatial scales. *J. Clim.*, **29**, 7773–7795, https://doi.org/10.1175/JCLI-

D-15-0618.1.

Gervais, M., J. R. Gyakum, E. Atallah, L. B. Tremblay, and R. B. Neale, 2014: How well are the distribution and extreme values of daily precipitation over North America represented in the community climate system model? A comparison to reanalysis, satellite, and gridded station data. *J. Clim.*, **27**, 5219–5239, https://doi.org/10.1175/JCLI-D-13-00320.1.

GloH20, 2021: MSWEP V2.8 Technical Documentation. *MSWEP V2.8 Tech. Doc.*, 1–8.

GPM Global Rainfall Map Algorithm Development Team, 2014: Global Satellite Mapping of Precipitation (GSMaP) for GPM Algorithm Theoretical Basis Document (ATBD): Algorithm Theoretical Basis Document (ATBD): Algorithm Ver.6. 17.

Guilloteau, C., E. Foufoula-Georgiou, P. Kirstetter, J. Tan, and G. J. Huffman, 2022: How Well Do Multisatellite Products Capture the Space–Time Dynamics of Precipitation? Part II: Building an Error Model through Spectral System Identification. *J. Hydrometeorol.*, **23**, 1383–1399, https://doi.org/10.1175/JHM-D-22-0041.1.

Harris, I., T. J. Osborn, P. Jones, and D. Lister, 2020: Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Sci. Data*, **7**, 1–18, https://doi.org/10.1038/s41597-020-0453-3.

Henn, B., A. J. Newman, B. Livneh, C. Daly, and J. D. Lundquist, 2018: An assessment of differences in gridded precipitation datasets in complex terrain. *J. Hydrol.*, **556**, 1205–1219, https://doi.org/10.1016/j.jhydrol.2017.03.008.

Herold, N., L. V. Alexander, M. G. Donat, S. Contractor, and A. Becker, 2016: How much does it rain over land? *Geophys. Res. Lett.*, **43**, 341–348, https://doi.org/10.1002/2015GL066615.

Hersbach, H., and Coauthors, 2019: Global reanalysis: goodbye ERA-Interim, hello ERA5. *ECMWF Newsl.*, 17–24, https://doi.org/10.21957/vf291hehd7.

——, and Coauthors, 2020: The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Hossain, F., and G. J. Huffman, 2008: Investigating Error Metrics for Satellite Rainfall Data at Hydrologically Relevant Scales. *J. Hydrometeorol.*, **9**, 563–575, https://doi.org/10.1175/2007JHM925.1.

Huffman, G., D. Bolvin, D. Braithwaite, K. Hsu, and R. Joyce, 2018: Algorithm Theoretical Basis Document (ATBD) NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG). *Nasa*, 29.

Huffman, G. J., and Coauthors, 2007: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeorol.*, **8**, 38–55, https://doi.org/10.1175/JHM560.1.

Huffman, G. J., R. F. Adler, D. T. Bolvin, and E. J. Nelkin, 2010: The TRMM Multi-Satellite Precipitation Analysis (TMPA). *Satellite Rainfall Applications for Surface Hydrology*, M. Gebremichael and F. Hossain, Eds., Springer Netherlands, 3–22.

Huffman, G. J., D. T. Bolvin, D. Braithwaite, K. Hsu, R. Joyce, C. Kidd, E. J. Nelkin, and P. Xie, 2015: NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG). *Algorithm Theor. Basis Doc. Version 4.5*, 26.

Joyce, R. J., J. E. Janowiak, P. A. Arkin, and P. Xie, 2004: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeorol.*, **5**, 487–503, https://doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2.

Kalnay, E., W. Collins, D. Deaven, L. Gandin, M. Iredell, R. Jenne, and D. Joseph, 1996: The NCEP NCAR 40-year reanalysis project. 1996.pdf. Bull. Am. Meteorol. Soc., 77, 437–472.Lin, Y., and K. E. Mitchell, 2005: The NCEP stage II/IV hourly precipitation analyses: Development and applications. 85th AMS Annu. Meet. Am. Meteorol. Soc. - Comb. Prepr., 1649–1652.

Liu, C., and R. P. Allan, 2012: Multisatellite observed responses of precipitation and its extremes to interannual climate variability. *J. Geophys. Res. Atmos.*, **117**, 1–16, https://doi.org/10.1029/2011JD016568.

Lockhoff, M., O. Zolina, C. Simmer, and J. Schulz, 2019: Representation of Precipitation Characteristics and Extremes in Regional Reanalyses and Satellite- and Gauge-Based Estimates over Western and Central Europe. *J. Hydrometeorol.*, **20**, 1123–1145, https://doi.org/10.1175/JHM-D-18-0200.1.

Di Luca, A., D. Argüeso, S. Sherwood, and J. P. Evans, 2021: Evaluating Precipitation Errors Using the Environmentally Conditioned Intensity-Frequency Decomposition Method. *J. Adv. Model. Earth Syst.*, **13**, 1–45, https://doi.org/10.1029/2020MS002447.

Mega, T., T. Ushio, T. Matsuda, T. Kubota, M. Kachi, and R. Oki, 2019: Gauge-Adjusted Global Satellite Mapping of Precipitation. *IEEE Trans. Geosci. Remote Sens.*, **57**, 1928–1935, https://doi.org/10.1109/TGRS.2018.2870199.

Meyer-Christoffer, A., A. Becker, P. Finger, B. Rudolf, U. Schneider, and M. Ziese, 2011a: GPCC Climatology Version 2011 at 0.5°: monthly land-surface precipitation climatology for every month and the total year from rain-gauges built on GTS-based and historic data. https://doi.org/10.5676/DWD_GPCC/CLIM_M_V2011_050.

——, ——, ——, ——, ——, and ——, 2011b: GPCC Climatology Version 2011 at 1.0°: monthly land-surface precipitation climatology for every month and the total year from rain-gauges built on GTS-based and historic data. https://doi.org/10.5676/DWD_GPCC/CLIM_M_V2011_100.

——, ——, ——, ——, ——, and ——, 2011c: GPCC Climatology Version 2011 at 0.25°: monthly land-surface precipitation climatology for every month and the total year from rain-gauges built on GTS-based and historic data. https://doi.org/10.5676/DWD_GPCC/CLIM_M_V2011_025.

Nelson, B. R., O. P. Prat, D. J. Seo, and E. Habib, 2016: Assessment and implications of NCEP stage IV quantitative precipitation estimates for product intercomparisons. *Weather Forecast.*, **31**, 371–394, https://doi.org/10.1175/WAF-D-14-00112.1.

Nissen, K. M., and U. Ulbrich, 2017: Increasing frequencies and changing characteristics of heavy precipitation events threatening infrastructure in Europe under climate change. *Nat. Hazards Earth Syst. Sci.*, **17**,

1177–1190, https://doi.org/10.5194/nhess-17-1177-2017.

Nitu, R., and Coauthors, 2018: WMO Solid Precipitation Intercomparison Experiment (SPICE) (2012 – 2015). *WMO-IOM Rep. No. 131.*, https://doi.org/https://www.researchgate.net/publication/349114825_WMO_Solid_Precipitation_Intercomparison_Experiment_SPICE.

Norris, J., G. Chen, and J. David Neelin, 2019: Changes in frequency of large precipitation accumulations over land in a warming climate from the CESM large ensemble: The roles of moisture, circulation, and duration. *J. Clim.*, **32**, 5397–5416, https://doi.org/10.1175/JCLI-D-18-0600.1.

Prat, O. P., and B. R. Nelson, 2015: Evaluation of precipitation estimates over CONUS derived from satellite, radar, and rain gauge data sets at daily to annual scales (2002-2012). *Hydrol. Earth Syst. Sci.*, **19**, 2037–2056, https://doi.org/10.5194/hess-19-2037-2015.

Prein, A. F., and A. Gobiet, 2017: Impacts of uncertainties in European gridded precipitation observations on regional climate analysis. *Int. J. Climatol.*, **37**, 305–327, https://doi.org/10.1002/joc.4706.

Sun, Q., C. Miao, Q. Duan, H. Ashouri, S. Sorooshian, and K. L. Hsu, 2018: A Review of Global Precipitation Data Sets: Data Sources, Estimation, and Intercomparisons. *Rev. Geophys.*, **56**, 79–107, https://doi.org/10.1002/2017RG000574.

Tian, Y., and C. D. Peters-Lidard, 2010: A global map of uncertainties in satellite-based precipitation measurements. *Geophys. Res. Lett.*, **37**, 1–6, https://doi.org/10.1029/2010GL046008.

Timmermans, B., M. Wehner, D. Cooley, T. O'Brien, and H. Krishnan, 2019: An evaluation of the consistency of extremes in gridded precipitation data sets. *Clim. Dyn.*, **52**, 6651–6670, https://doi.org/10.1007/s00382-018-4537-0.

Trenberth, K. E., 2011: Changes in precipitation with climate change. *Clim. Res.*, **47**, 123–138, https://doi.org/10.3354/cr00953.

——, and Y. Zhang, 2018: How often does it really rain? *Bull. Am. Meteorol. Soc.*, **99**, 289–298, https://doi.org/10.1175/BAMS-D-17-0107.1.

——, ——, and M. Gehne, 2017: Intermittency in precipitation: Duration, frequency, intensity, and amounts using hourly data. *J. Hydrometeorol.*, **18**, 1393–1412, https://doi.org/10.1175/JHM-D-16-0263.1.

Ushio, T., and Coauthors, 2009: A kalman filter approach to the global satellite mapping of precipitation (GSMaP) from combined passive microwave and infrared radiometric data. *J. Meteorol. Soc. Japan*, **87 A**, 137–151, https://doi.org/10.2151/jmsj.87A.137.

Xie, P., R. Joyce, S. Wu, S. H. Yoo, Y. Yarosh, F. Sun, and R. Lin, 2017: Reprocessed, bias-corrected CMORPH global high-resolution precipitation estimates from 1998. *J. Hydrometeorol.*, **18**, 1617–1641, https://doi.org/10.1175/JHM-D-16-0168.1.

Zhou, T., R. Yu, H. Chen, A. Dai, and Y. Pan, 2008: Summer precipitation frequency, intensity, and diurnal cycle

over China: A comparison of satellite data with rain gauge observations. *J. Clim.*, **21**, 3997–4010, https://doi.org/10.1175/2008JCLI2028.1.

Zolina, O., C. Simmer, A. Kapala, P. Shabanov, P. Becker, H. MäcHel, S. Gulev, and P. Groisman, 2014: Precipitation variability and extremes in Central Europe: New View from STAMMEX Results. *Bull. Am. Meteorol. Soc.*, **95**, 995–1002, https://doi.org/10.1175/BAMS-D-12-00134.1.