UNIVERSITY OF LJUBLJANA
FACULTY OF MATHEMATICS AND PHYSICS
DEPARTMENT OF PHYSICS

Gregor Skok

OBJECT-BASED ANALYSIS AND VERIFICATION OF PRECIPITATION OVER LOW- AND MID-LATITUDES

Doctoral thesis

ADVISER: prof. dr. Jože Rakovec
CO-ADVISOR: dr. Joe Tribbia

Ljubljana, 2009

UNIVERZA V LJUBLJANI
FAKULTETA ZA MATEMATIKO IN FIZIKO
ODDELEK ZA FIZIKO

Gregor Skok

OBJEKTNA ANALIZA IN VERIFIKACIJA PADAVIN NAD NIZKIMI IN SREDNJIMI GEOGRAFSKIMI ŠIRINAMI

Doktorska disertacija

MENTOR: prof. dr. Jože Rakovec
SOMENTOR: dr. Joe Tribbia

Ljubljana, 2009

Abstract

The Method for Object-based Diagnostic Evaluation (MODE) developed by Davis et al (2006a,b) is implemented and extended to characterize the temporal behavior of objects and to perform a diagnostic analysis on the spatial distribution and properties of precipitation systems over the equatorial Pacific. The analysis is performed on two satellite derived datasets (TRMM and PERSIANN) and WRF model Tropical Channel Simulation dataset (performed by NCAR). A sensitivity analysis showed that the temporal convolution produces an unwanted "spillover" effect and that a large spatial convolution radius produces too much smoothing which results in unrealistically large objects. The analysis of satellite datasets showed that the largest and most long-lived precipitation systems in the tropical Pacific are typically located in the western part. A good ability to track precipitation systems in tropical Pacific was demonstrated: movement of precipitation systems in the satellite datasets in the ITCZ is both westward and eastward although westward movement is more frequent and in the eastern part of the Pacific ITCZ the westward movement is dominant. Movement of systems in the mid latitudes was predominantly eastward. These findings were common to both satellite products. A comparison of TRMM dataset to the WRF simulation was also performed. The comparison of three years of data showed a general agreement but also some differences between the datasets. Notably, there is more precipitation in the model, the trajectories in the ITCZ are straighter in the WRF simulation and in the central part of ICTZ the movement in the TRMM dataset is both eastward and westward, while in the WRF simulation the westward movement is dominant.

PACS(2005)
92.60.Bh, 92.60.Jq, 92.70.Gt, 92.70.Ly

Key words
precipitation; PERSIANN; TRMM; WRF model; objects; tropics; ITCZ

Kratek Povzetek

V okviru doktorske naloge smo nadgradili osnovno metodo MODE za analizo padavin z uporabo objektov, ki jo je razvil Davis (2006a,b). Nadgrajena metoda MODE omogoča analizo časovnega razvoja objektov. Metodo smo uporabili pri analizi in primerjavi padavin treh padavinskih produktov – TRMM in PERSIANN, ki padavine merita predvsem preko inštrumentov na satelitih, ter simulacije tropskega kanala z modelom WRF. Za analizo smo si izbrali področje tropskega Pacifika. Analiza občutljivosti je pokazala, da konvolucija v časovni dimenziji ni primerna in da prevelike vrednosti prostorskega radija konvolucije niso primerne, ker padavinsko polje postane preveč gladko. Analiza dveh satelitskih padavinskih produktov je pokazala, da se padavinski sistemi, ki so zelo veliki ali imajo zelo dolgo življenjsko dobo, pojavljajo predvsem v zahodnem delu Pacifika ali pa na manjšem območju blizu obale Mehike. Metoda je pokazala dobro sposobnost sledenja padavinskim sistemom. Pokazalo se je, da gibanje padavinskih sistemov v ICTZ lahko poteka v obe smeri (proti vzhodu in proti zahodu), čeprav je gibanje v vzhodnem delu običajno bolj pogosto proti zahodu, medtem ko je v vzhodnem delu ICTZ prevladuje gibanje sistemov proti zahodu. Gibanje sistemov v zmernih širinah je bilo predvsem v smeri proti vzhodu. Te ugotovitve so bile skupne obema satelitskima produktoma. Izvedli smo tudi primerjavo med satelitskim produktom TRMM in modelsko simulacijo z modelom WRF. Primerjava podatkov triletnega časovnega obdobjea je pokazala veliko razlik vendar tudi nekaj ujemanj. Npr. količina padavin je precej večja v simulaciji z WRF, v simulaciji z WRF so na območju ICTZ trajektorije veliko bolj ravne kot v satelitskem produktu TRMM, gibanje v ICTZ v TRMM poteka v obeh smereh (proti vzhodu in proti zahodu), medtem ko v simulaciji z WRF prevladuje gibanje v smeri proti zahodu.

PACS (2005)
92.60.Bh, 92.60.Jq, 92.70.Gt, 92.70.Ly

Ključne besede
padavine; PERSIANN; TRMM; model WRF; objekti; tropi; ITCZ

Contents

1 INTRODUCTION

The tropics as a whole are play an immensely important role in the global hydrological cycle - it is estimated that two-thirds of global precipitation falls in the Tropics (Kummerow et al. 1998). The region of tropical Pacific is especially interesting to study since it hosts processes such as Madden-Julian oscillation (Madden and Julian 1971), the Intertropical Convergence Zone and the El Niño-Southern Oscillation.

The capability to develop climate statistics of precipitation in the Pacific in terms of objects has only now been realized since in recent years a number of new satellite rainfall products with high spatial-temporal resolution (0.25 degrees and one to three hours) over low- and mid-latitudes (40S to 40N) have become available. These include the Tropical Rainfall Measuring Mission 3B42 (TRMM, Huffman et al. 2007), Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN, Hsu et al. 1997) and Climate Prediction Center Morphed precipitation (CMORPH, Joyce et al. 2004). Although some analysis of the datasets has already been done [e.g. Dai et al. (2007)], where spatial patterns, seasonal mean precipitation amount, daily precipitation frequency and intensity were studied, the use of the object-based analysis provides a whole new set of precipitation properties, namely size and number of precipitation areas, their location, orientation, spatial distribution, frequency, direction of movement and lifespan.

Diagnostic analysis and verification as such are critical components for the development of both observing and forecasting systems, and play a key role in the evaluation of the accuracy and credibility of climate simulations. Ideally, the analysis of observed data should be used as a tool to gain additional useful knowledge on the overall behavior and properties of a few selected aspects of a modeling system since the underlying physical processes are usually too complex to be examined in their entirety. On the other hand verification should play a role in monitoring the quality of forecasts, provide feedback to developers and forecasters to help improve forecasts, and provide meaningful information to forecast users to apply in their decision making processes. As noted by Mahoney et al. (2002) forecast verification can also help to identify differences among forecasts made by different models or forecasters. Because forecast quality is intimately related to forecast value, albeit through relationships that are sometimes quite complex, verification has an important role to play in assessments of the value of particular types of forecasts (Murphy 1993).

Unfortunately, traditional approaches for the verification of spatial forecasts, including quantitative precipitation forecasts (QPFs) and convection forecasts, are inadequate to meet current needs. Typically, verification techniques have been based on simple grid overlays in which the forecast grid is matched to an observation grid or set of observation points. From these overlays, counts of forecast–observation (yes–no) pairs are computed to complete the standard 2×2 contingency table (Wilks 1995). The counts in this table can be used to compute a variety of verification measures and skill scores, such as the probability of detection (POD), false alarm ratio (FAR), critical success index (CSI), and equitable threat score (e.g., Doswell et al. 1990; Wilks 1995). These statistics, which represent a "measures-oriented approach" to verification, provide numbers that can be used to monitor system performance over time; however, they do not provide specific information regarding the way(s) in which a forecast went wrong or did well. That is, it is difficult to use these statistics to diagnose particular errors in the forecasts to provide meaningful information that can be used to improve the forecasts or provide guidance to forecast users; essentially, all incorrect forecasts are treated in the same way, with little information provided regarding the cause of a poor score. Thus, a potential user of the forecast does not know which aspects of the forecast provide valid information, and a forecast developer must use trial and error or some other approach to determine how to improve the forecast. In addition to these issues, the results of a measure oriented verification analysis are often not consistent with what a forecaster or analyst might infer by more subjective visual evaluation of a forecast. This inconsistency and other disadvantages of measures-oriented methods have led to the idea that an objective approach that would more closely mimic the subjective approach could provide more useful, diagnostic information about the quality of spatial forecasts.

Despite the fact that a traditional synoptic analysis decomposes atmospheric states into well-defined structural entities like fronts and rainbands, the quantitative evaluation of model performance using metrics tied to such structural elements has proven to be challenging. Recently, however object based analysis methods have been developed which can provide both qualitative and quantitative measures related to the spatial field structure and features. One example of such a method, the object-based analysis first presented by Davis et al. (2006a,b), also commonly know as MODE (Method for Object-Based Diagnostic Evaluation; Davis et al. 2009), defines objects as self-enclosed areas of a variable, for example precipitation, above a certain threshold. Prior to thresholding, the object borders are also smoothed using convolution in order to better reflect the object borders which would be drawn by a human. This method thus tries to reflect what a forecaster or analyst might infer by a more subjective visual evaluation of a field.

Quite a few studies have dealt with satellite-observed cloud "objects". For example, Welch et al. (1988) studied structural characteristics of stratocumulus cloud fields off the coast of southern California using the LANDSAT Multispectral Scanner and found that the cloud size distributions obey a power law. Machado et al. (1998) studied the evolution of morphological and radiative properties of clouds over life cycles of deep convective systems over the Americas using

GOES-7 ISCPP-B3 satellite data from the period 1987-88. Mapes and Houze (1993) objectively processed infrared satellite images of the oceanic warm-pool region (8O°E-160°W) to reveal tropical “cloud clusters” with temperatures colder than a given threshold. Wilcox (2003) used an automated cloud detection and tracking algorithm to observe and track overcast decks of clouds in a consecutive set of hourly Meteosat-5 images and NCAR CCM3 model output for 49 days during the 1999 winter monsoon period over the Indian Ocean. Past studies of satellite-observed cloud "objects" have mainly used data on brightness temperatures detected by a single satellite while the current satellite derived datasets, such as TRMM 3B42 and PERSIANN, combine data from many satellites and their instruments and obtain precipitation data in a high spatial-temporal resolution which was not available before. This data is also readily available in a large domain (whole tropics) for a period of a few years.

In the scope of this thesis the object-oriented analysis was firstly done by using the basic MODE object-based method developed by Davis et al. (2006a,b). However, since precipitation is a time-evolving process with inherent properties such as lifespan and direction of movement, we extended the basic MODE method to characterize the temporal behavior of objects. The analysis was first performed on TRMM and PERSIANN satellite derived precipitation datasets with high spatial-temporal resolution, for a six year period 2000-2005. Secondly, the object-based method was also used for verifying the WRF model Tropical Channel Simulation precipitation[1] against the TRMM satellite precipitation. The WRF model Tropical Channel Simulation is a 5-year, high resolution, climate simulation of the global tropics and surrounding regions. The verification was done for a three year period 1998-2000, when the WRF simulation and the TRMM dataset precipitation were simultaneously available.

Chapter 2 presents the basic MODE object-based method and the extended method which was developed in the scope of this thesis. Chapter 3 presents the analysis and comparison of TRMM and PERSIANN satellite derived precipitation datasets. It consist of description of the TRMM and PERSIANN datasets, comparison of precipitation accumulations, analysis of missing data, a sensitivity analysis and final results. Chapter 4 presents the analysis and verification of precipitation data provided by WRF model Tropical Channel Simulation. It consist of description of the WRF model and simulation, comparison of precipitation accumulations, analysis of missing data, a sensitivity analysis and final results.

[1]performed and provided by NCAR

2 METHODOLOGY

2.1 The basic MODE object-based analysis

An object-based analysis first presented by Davis et al. (2006a,b), also commonly know as MODE (Method for Object-Based Diagnostic Evaluation; Davis et al. 2009), defines objects as self-enclosed areas of a variable, for example precipitation, above a certain threshold. The process of spatial object identification consists of three steps: i) convolution, ii) thresholding and iii) object identification. Objects are considered as if there is no relation between spatial objects in different time intervals. Each time interval is treated as a separate case and no effort is made to correlate or link objects from consecutive time intervals.

Convolution is performed on the original precipitation field as a spatial moving average with a horizontal convolution radius R_{s}. All grid point precipitation amounts $Z_{\mathrm{x},y}$ inside the circle with radius R_{s} are summed up in a calculation of the average:

$$\widetilde{Z} = \frac{1}{N_{\mathrm{ell}}} \sum_{-R_{\mathrm{s}}}^{R_{\mathrm{s}}} \sum_{-R_{\mathrm{s}}}^{R_{\mathrm{s}}} Z_{x,y} \times \begin{cases} 1; & \text{if inside circle} \\ 0; & \text{if outside} \end{cases} , \tag{2.1}$$

where N_{ell} is the number of grid points inside a circle (can be pre-calculated in advance), and $Z_{x,y}$ is rainfall at location x,y. The convolution has the effect of smoothing the original field. As a result the spatial objects have smoother outlines which better reflect outlines which would be made by a human. The value of R_{s} greatly affects the object distribution. Essentially, R_{s} defines the "scale" of precipitation to be considered.

Thresholding is performed in a simple manner by disregarding any of the smoothed precipitation values that fall below a certain threshold value TH. If the value of TH is large, the spatial objects represent regions with heavy precipitation while lower values of TH include regions with light precipitation. The value of TH also greatly affects the object distribution.

The last step is the identification of spatial objects. This is done by the algorithm identifying all self-enclosed regions in the domain. The algorithm checks all grid points and checks whether their neighboring grid points are either part of an already identified object, part of a so far unidentified object, or not part of an object at all. Finally spatial object properties can be calculated, such as object center and object area.

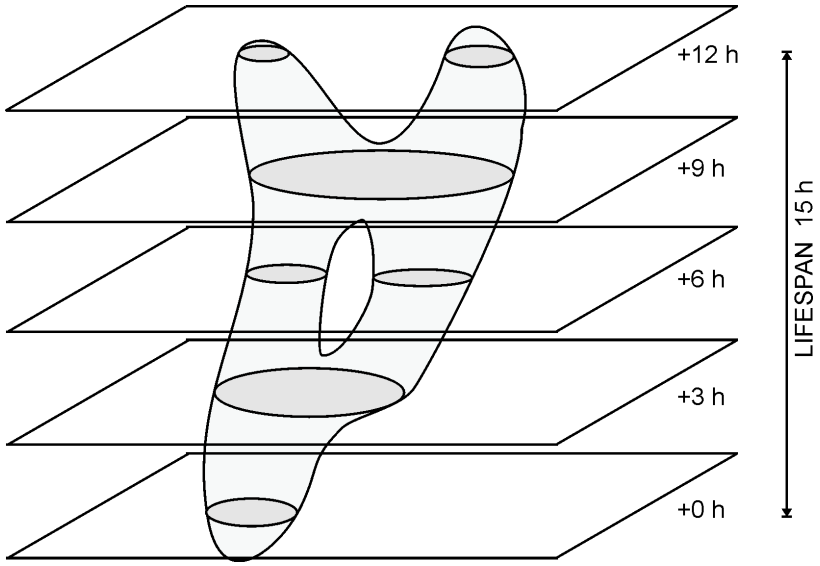


Figure 2.1: An example of temporal object with lifespan 15 hours.

2.2 Extension of the MODE method to include temporal behaviour of objects

The basic MODE method enables the study of spatial objects that were unrelated in time. While spatial objects have a few practical properties such as size and location, precipitation is a time-evolving process with inherent properties such as lifespan and direction of movement. The shortened term "temporal object" (TO) corresponds to a temporal sequence of spatial objects which are deemed to be part of the same transient precipitation feature (e.g., a storm, convective system, frontal system, ...). Temporarily disregarding the difference between temporal and spatial dimensions, a temporal object can be though of as a self enclosed object in three dimensions.

The shape of a TO can be nontrivial – for example, the object can split, merge or have holes (Figure 2.1). The lifespan of an object is defined as the difference between the time of the last appearance and the time of first appearance. If an object is present in a single time interval its lifespan is defined as the duration of the time interval, if in two its lifespan is double the duration of the time interval, etc. The movement of the object is defined as the movement of its centre in consecutive time intervals.

In order to implement the use of TOs the basic MODE algorithm for spatial objects had to be enhanced. It is important to mention that Davis et al. (2006b) also considers time-evolution of objects where temporal objects are identified by matching spatial objects, separated temporally by one time interval, if their centers are separated by less than a threshold distance. This process is repeated for the each next time interval thereby obtaining objects with longer lifespan. However, we chose a different approach, namely, the convolution is made in three space-time dimensions (previously only two spatial), objects are defined as self-

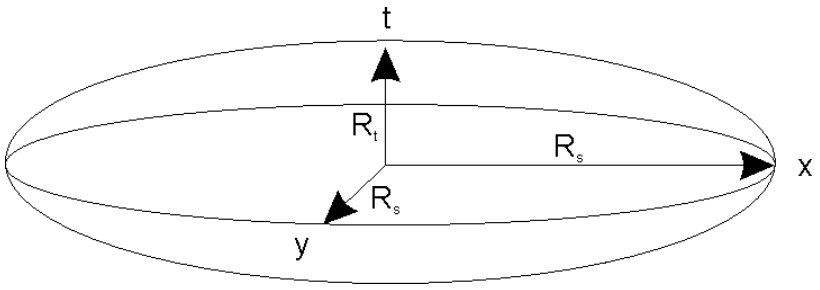


Figure 2.2: Convolution ellipsoid

enclosed objects in three dimensions (previously only two spatial). As a result a number of previous time intervals (or all) have to be kept in memory. The minimum number of time intervals that have to be kept in memory is defined by the longest life an object has for the given dataset, and depends also on TH and R_s.

2.2.1 Convolution and thresholding

Three-dimensional convolution is in principle the same as in two dimensions, except that it can be a computationally very demanding process. The amount of computational expense (CE) depends on the size of the convolution radius R and can be estimated as CE $\propto R^3$ since this expression represents the "volume" of a sphere – and therefore the rough number of surrounding grid points, that have to be summed up at each location. In general, the CE of convolution can be improved by using fast Fourier transformation. However this was not done since the speed of calculation was deemed sufficient for the purposes of the present study.

Two convolution radii were used: one in spatial (R_s) and one in temporal (R_t) dimension. The two values define an ellipsoid (Figure 2.2):

$$\frac{x^2}{R_\mathrm{s}^2} + \frac{y^2}{R_\mathrm{s}^2} + \frac{t^2}{R_\mathrm{t}^2} = 1, \tag{2.2}$$

where x, y are the spatial and t is the temporal dimension. All grid points inside the ellipsoid are again summed up:

$$\widetilde{Z} = \frac{1}{N_\mathrm{ell}} \sum_{-R_\mathrm{t}}^{R_\mathrm{t}} \sum_{-R_\mathrm{s}}^{R_\mathrm{s}} \sum_{-R_\mathrm{s}}^{R_\mathrm{s}} Z_{x,y,t} \times \begin{cases} 1; & \text{if inside circle} \\ 0; & \text{if outside} \end{cases}, \tag{2.3}$$

where N_ell is the number of grid points inside ellipsoid (can be pre-calculated in advance), and $Z_{x,y,t}$ is the rainfall at location x, y and time t. The CE can be improved by a factor of $6/\pi$ (the volume of a cube compared to an ellipsoid) if the summations are done only inside the ellipsoid:

$$\widetilde{Z} = \frac{1}{N_{\text{ell}}} \sum_{-R_{\text{t}}}^{R_{\text{t}}} \sum_{-R_{\text{A}}}^{R_{\text{A}}} \sum_{-R_{\text{B}}}^{R_{\text{B}}} Z_{x,y,t} \times \begin{cases} 1; & \text{if inside circle} \\ 0; & \text{if outside} \end{cases}, \tag{2.4}$$

where the lower and upper limits are $R_{\text{A}}(t) = R_{\text{s}}\sqrt{1 - \frac{t^2}{R_{\text{t}}^2}}$ and $R_{\text{B}}(x,t) = \sqrt{R_{\text{A}}^2 - x^2}$. Functions R_{A} and R_{B} can be pre-calculated in advance.

After the convolution, the thresholding is done in the same way as in the spatial object algorithm. It consists of looping over all grid points in the current time interval and checking whether the precipitation exceeds the threshold value or not. If the value is smaller than the threshold, the grid point is defined as a non-object.

2.2.2 Object identification

There are substantial differences between the extended algorithm for temporal objects and the original MODE algorithm for spatial objects. First, all data concerning TOs in the previous few time intervals have to be kept in memory due to the possibility of TOs merging at a later time interval. The minimum number of time intervals that have to be kept in memory is defined by the longest life an object has for the given dataset, and according to its TH, R_{s} and R_{t}.

Temporal objects are identified by merging the grid-points of the convoluted field which exceed the threshold TH and which are neighbors, either in time or in space (or both). At every grid-point in current time interval the procedure shown in flowchart in Figure 2.3 is performed.

The merging of temporal objects A and B is performed by reversing through time and setting all grid points with value A to value B. The computational expense of the TO merging can be notably improved by going back through time only as long as the old TO still exists. For example, if an TO exists at time interval –6 h but not at –9 h there is no need to check –12 h (or further back into the past) since the TO cannot be self enclosed if it does not exist in all intermediary time intervals.

2.2.3 Counting and re-indexing

After the object identification procedure has finished the number of the TOs in the current time interval can be determined. This can be achieved in a number of ways. The most straightforward approach is to sequentially check all the grid-points and at each location check for a TO and whether it is a new or an old one (i.e., one already found in previously checked grid-points). However the CE of this method can be large if the number of objects in the current time interval is large (for example more than 1000 objects) since at each grid-point where a TO is found a list of previously found TOs has to be checked. The CE can be greatly reduced by remembering the index of the object from the previously checked grid-point.

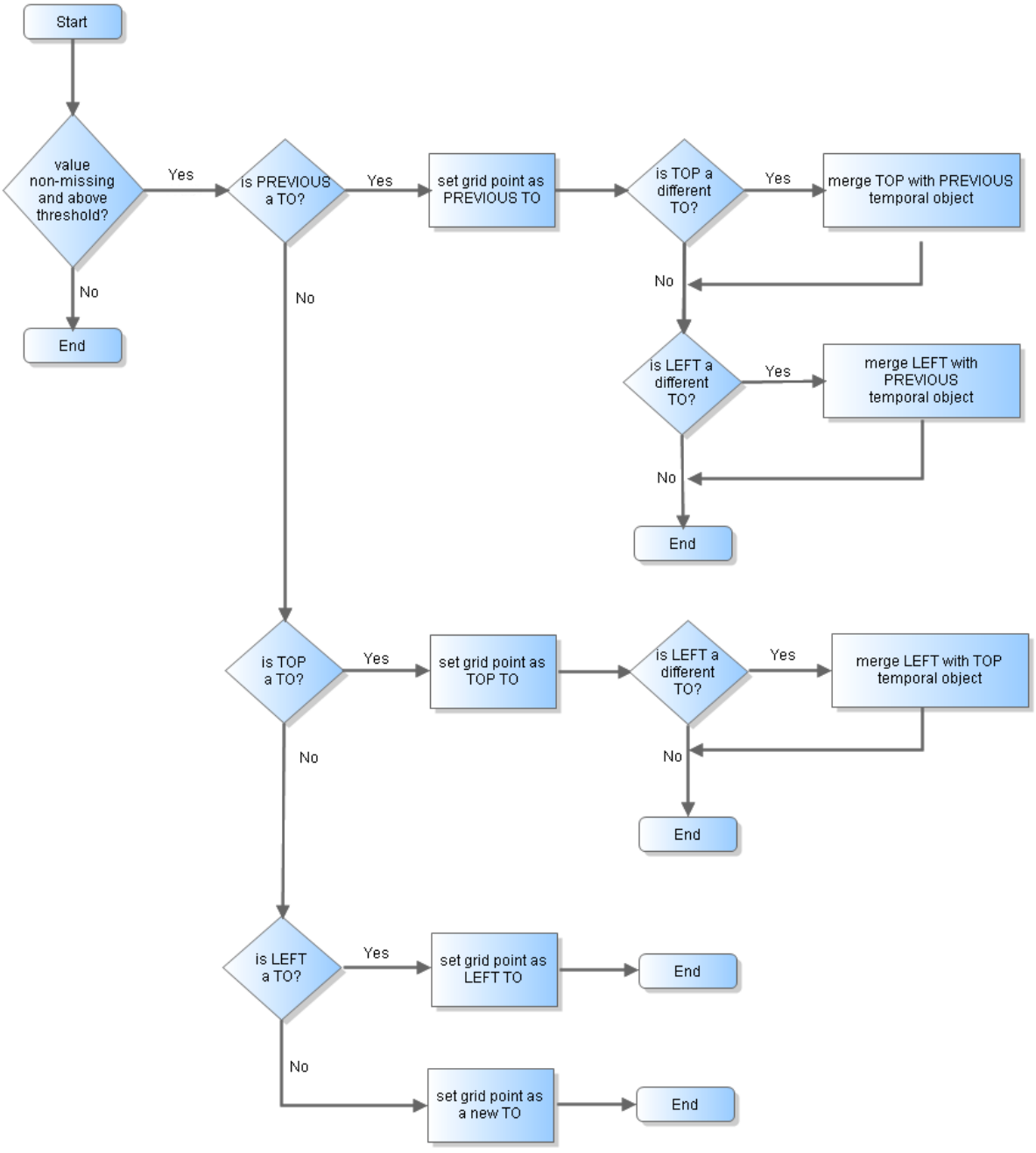


Figure 2.3: The flowchart for object identification algorithm which has to be performed on all grid points. The term PREVIOUS coresponds to the grid point at the same location but in a previous time interval. The term TOP and LEFT corespond to north and west grid points in the same time interval.

In this way the list of previously found objects does not have to be checked for grid-points located inside a TO (inside its boundary), since the object index is known from a previously checked grid-point.

The indices of objects should also be renumbered. The non-renumbered indices can easily take ten times more index space as renumbered indices – this is due to the manner in which the algorithm performs merging of objects. Merging produces many indices which do not belong to any object (the objects with these indices were merged with other objects and thereby acquired new indices). The renumbering is not necessary if the number of TOs in all time intervals is considerably smaller (by at least an order of magnitude) than the scope of the indexing variable (usually float or long). In our case where we analyzed the satellite data there were 17056 three-hourly intervals, each of which could have more than 1000 objects potentially resulting in more than 17×10^6 objects. Float variable type on our system could hold only 34×10^6 non-rounded values. With no renumbering the object index could easily exceed 170×10^6 which would result in an error.

2.2.4 Properties of objects

The properties of TOs can only be estimated at the last timestep kept in memory – before it is erased from memory. This is the only way to ensure that the object will not be merged at a later timestep.

TO lifespan (L) is defined as:

$$L = t_{\text{last}} - t_{\text{first}} + \Delta t, \tag{2.5}$$

where t_{last} is time of last appearance and t_{first} time of first appearance and Δt is the duration of time interval. The shortest lifespan can be the duration of one time interval, the longest is defined by $(TT \times \Delta t)$ where TT is the number of previus time intervals kept in memory. If $(TT \times \Delta t)$ is smaller than the lifespan the TO could either be unaffected or split in two (or more) objects with variable (usually short) lifetimes. This would produce an artificial increase in frequency of the smaller objects and produce an error in the TO distribution.

The two-dimensional area of a TO at a certain timestep is estimated as:

$$S = \sum_{0}^{\dim x} \sum_{0}^{\dim y} \Delta x \Delta y \cos\varphi(y) \cdot O(x, y), \tag{2.6}$$

where φ is latitude and $O(x, y)$ is 1 inside object and 0 otherwise.

The center of the object is calculated using:

$$x_{\text{c}} = \frac{1}{S_{\text{n}}} \sum_{0}^{\dim x} \sum_{0}^{\dim y} x \Delta x \Delta y \cdot O(x, y), \tag{2.7}$$

$$y_{\text{c}} = \frac{1}{S_{\text{n}}} \sum_{0}^{\dim x} \sum_{0}^{\dim y} y \Delta x \Delta y \cdot O(x, y). \tag{2.8}$$

S_n is the latitude non-adjusted object area (similar to S but without the $\cos\varphi$). The object displacement and speed of movement is estimated as difference in x_c and y_c between consecutive timesteps.

3 ANALYISIS AND COMPARISON OF TWO SATELLITE DATASETS

In this Chapter, the TRMM and PERSIANN satellite precipitation datasets are analyzed and compared in the tropical Pacific using the object-based methods. Firstly the datasets are presented and studied for missing data which is followed by an analysis of annual and seasonal precipitation accumulations. Secondly, the two datasets are analyzed using the basic MODE method (presented in Section 2.1) and the extended method (developed in this thesis and presented in Section 2.2). The study by both methods starts with the sensitivity analysis performed on a shorter time period, which provides the optimal values for parameters R_{s}, $R_{\mathrm{t}}$and TH, and is followed by a full study of the datasets in the Pacific domain for the entire 6-year period.

3.1 Dataset description

For our object-based analysis and comparison of the results we selected two of the datasets mentioned in the Introduction, namely TRMM and PERSIANN, because of their availability and completeness. Some selected properties of these two datasets can be seen in Table 3.1 while more detalied overview of the datasets are available in Subsections 3.1.1 and 3.1.2.

Since the precipitation accumulation times differ between the datasets (three hours compared to one hour) three PERSIANN one hour measurements were summed to correspond to the three hour accumulation time of TRMM. Also the start time of the accumulation differs by 30 minutes between datasets. For example, the TRMM three-hour accumulation starts at 22:30 UTC whereas the PERSIAN one-hour accumulation starts at 23:00 UTC (Figure 3.1). Therefore the three-hourly accumulation intervals were offset for half an hour with PERSIAN starting 30 minutes later than TRMM.

3.1.1 Overview ob the TRMM 3B42 dataset

The TRMM 3B42 (Huffman et al. 2007) is a fine scale ($0.25^\circ \times 0.25^\circ$ and 3 hourly) satellite derived precipitation dataset available in the latitude band $50^\circ\mathrm{S} - 50^\circ\mathrm{N}$. Most of the coverage in the TRMM 3B42 depends on input from two different sets

Table 3.1: Selected properties of analyzed Satellite derived precipitation datasets.

	TRMM	**PERSIANN**
Name	TRMM Multi-satellite precipitation analysis (TMPA, a.k.a. 3B42)	Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN)
Provider	GSFC (G. Huffman)	UC Irvine (K.-L. Hsu)
Data	Geo-IR, microwave from SSM/I, TRMM, AMSU, AMSR	Geo-IR, TRMM microwave
Method	Merged microwave and microwave-calibrated infrared (IR)	Adaptive neural network calibration of geo-IR to TRMM TMI
Domain	50° S – 50° N	60° S – 60° N
Start date	1. January 1998	1. March 2000
End date	31. December 2005	31. December 2005
Spatial resolution	0.25 deg	0.25 deg
Temporal resolution	3 h	1 h

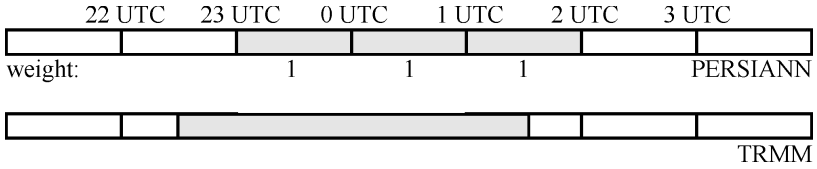


Figure 3.1: Three hour TRMM accumulations of and corresponding to one hour PERSIANN accumulations.

of sensors. First, precipitationrelated passive microwave data are collected by a variety of low earth orbit (LEO) satellites, including the Microwave Imager (TMI) on TRMM, Special Sensor Microwave Imager (SSM/I) on Defense Meteorological Satellite Program (DMSP) satellites, Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E) on Aqua, and the Advanced Microwave Sounding Unit-B (AMSU-B) on the National Oceanic and Atmospheric Administration (NOAA) satellite series. These data have a strong physical relationship to the hydrometeors that result in surface precipitation, but each individual satellite provides a very sparse sampling of the time–space occurrence of precipitation. Even taken together, there are significant gaps in the current 3-hourly coverage by the passive microwave estimates.

The second major data source for the TMPA is the window-channel ($\sim 10.7\ \mu$m) infrared (IR) data that are being collected by the international constellation of geosynchronous earth orbit (GEO) satellites. In contrast to the sparse temporal sampling of the passive microwave data, the GEO–IR data provide excellent time–space coverage. The CPC of the NOAA/National Weather Service (NWS) merges the international complement of GEO–IR data into half-hourly 4 km×4 km equivalent latitude–longitude grids (hereafter the (Janowiak et al. 2001). The IR brightness temperatures (T_b) are corrected for zenith angle viewing effects and

intersatellite calibration differences.

For estimates generated prior to the start of the CPC merged IR dataset in early 2000 a GPCP dataset is used (also produced at CPC) that contains 24-class histograms of GEO–IR T_b data on a 3-hourly, $1° \times 1°$ latitude–longitude grid covering the latitude band $40°\text{S} - 40°\text{N}$ (Huffman et al. 2001). This dataset also includes grid-box-average Geostationary Operational Environmental Satellite (GOES) Precipitation Index (GPI; Arkin and Meisner 1987) estimates computed from LEO–IR data recorded by the NOAA satellite series. These LEO–GPI data are used in the TMPA to fill gaps in the GEO–IR coverage, most notably in the Indian Ocean sector, where there was no GEO–IR coverage before Meteorological Satellite 5 (Meteosat-5) began providing observations there in June 1998.

Finally, the product also makes use of three additional data sources: the TRMM Combined Instrument (TCI) estimate, which employs data from both TMI and the TRMM precipitation radar (PR) as a source of calibration (TRMM product 2B31; Haddad et al. 1997a,b); the GPCP monthly rain gauge analysis developed by the Global Precipitation Climatological Center (GPCC; Rudolf 1993); and the Climate Assessment and Monitoring System (CAMS) monthly rain gauge analysis developed by CPC (Xie and Arkin 1996).

The precipitation estimates are produced in four stages: 1) the microwave precipitation estimates are calibrated and combined, 2) infrared precipitation estimates are created using the calibrated microwave precipitation, 3) the microwave and IR estimates are combined, and 4) rain gauge data are incorporated. Each three-hourly precipitation field is best interpreted as the precipitation rate effective at the nominal observation time.

The last step is the rescaling of precipitation to fit monthly rain gauge data. It is highly advantageous to include rain gauge data in combination datasets (Huffman et al. 1997, among others). However, experience shows that on any time scale shorter than a month the gauge data are neither reported with sufficient density nor reported with consistent observational intervals to warrant direct inclusion in a global algorithm. This issue was solved in the GPCP One-Degree Daily combination dataset by scaling the short-period estimates to sum to a monthly estimate that includes monthly gauge data (Huffman et al. 2001). In the TRMM dataset, a similar approach is taken: all available 3-hourly merged estimates are summed over a calendar month to create a monthly multisatellite (MS) product. The MS and gauge are combined as in Huffman et al. (1997) to create a postreal- time monthly SG combination, which is a TRMM research-grade product in its own right (3B43). Then the field of SG/MS ratios is computed on the $0.25° \times 0.25°$ grid (with controls) and applied to scale each 3-hourly field in the month, producing the version-6 3B42 product. The result is to provide the high resolution typical of satellite data and the typically small bias of gauge analyses over land.

3.1.2 Overview of the PERSIANN dataset

The PERSIANN (Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks, Hsu et al. 1997) is a fine scale ($0.25° \times 0.25°$ and 1 hourly) satellite derived precipitation dataset available in the latitude band $60°\text{S} - 60°\text{N}$.

The PERSIANN system for rainfall estimation under development at The University of Arizona undergoes periodic enhancement. The fundamental algorithm is based on a neural network and can therefore be easily adapted to incorporate relevant information as it becomes available. An adaptive training feature facilitates rapid updating of the network parameters whenever independent estimates of rainfall are available. The original system (Hsu et al. 1997) was based on geostationary infrared imagery and later extended (Hsu et al. 1999) to include the use of both infrared and daytime visible imagery. The product used in this study is based on the geostationary infrared imagery approach, but uses the TMI 2A12 product provided by the low-inclination orbiting TRMM satellite for regular updating of the network parameters.

The type of artificial neteork used by PERSIANN is the Modified Counter Propagation Network (Hsu, 1996). As is usually the case with artificial neural networks the network has first to be "trained" on some part of the dataset to be able to estimate the precipitation. The network structure usually consists of six normalized input variables in the hidden layer and 225 nodes each arranged in a 15×15 matrix in the hidden and output layers.

3.2 Domain and time period

A global study of rainfall data would doubtless be of interest, however due to the data availability and the overwhelming size of the task we decided to restrict our domain to the tropical Pacific. The tropical Pacific, which is associated with the ICTZ and El-Niño/La-Niña induced precipitation anomalies, is a very interesting region to study. The domain used here encompasses the following area: 120° E – 285° E and 39° S – 39° N (Figure 3.2). The period includes times when both datasets are available: 1st March 2000 – 31th December 2005. The domain consists of 622 x 313 grid points in 17,048 three-hourly time intervals.

According to the Oceanic Nino Index available on NOAA web-page[1], some El-Niño/La-Niña precipitation anomalies appeared during the 2000-2005 period, namely two El-Niño events (one weak in 2004 and one moderate event in 2002) and one weak La-Niña event in 2000.

[1] http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml

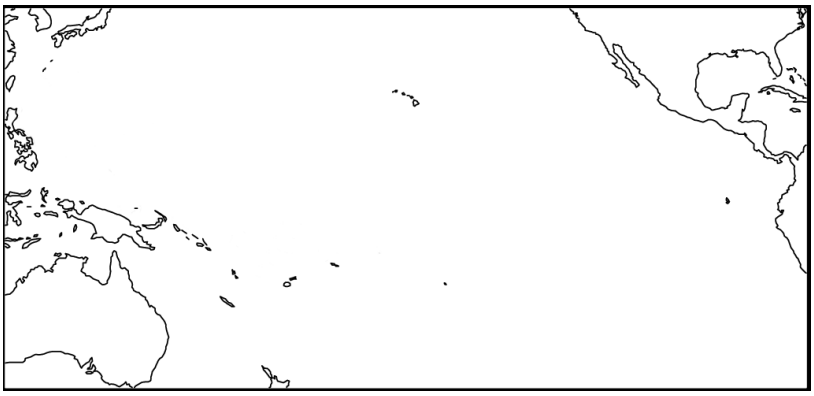


Figure 3.2: Pacific domain considered for the analysis.

3.3 Missing data

Missing data were present in both datasets, although much more frequently in PERSIANN (Figure 3.3). For some areas in the PERSIANN dataset (middle of Pacific, south Pacific) the data availability was as low as 35% and for some areas the data availability was considerably diminished (75% - 90% for southeast Pacific and west Australia). For the TRMM dataset the data availability was generally much better, never falling below 95%.

The severely diminished data availability in the PERSIANN dataset would present a problem if not properly treated. A correction was done in the following way: if at least one of the three one-hourly accumulations was available, its value was assumed for the whole three hour period and its value multiplied by three. In this way, only those three hourly accumulations for which all three one hourly periods were unavailable remained defined as missing. Additionally, to synchronize the missing data in the TRMM and PERSIANN datasets, a location was declared missing in both datasets if it was not available in either of the two. For example, if the data were available in the TRMM dataset but missing in the PERSIANN dataset, the data were also declared missing in the TRMM dataset. After applying this technique, the data availability everywhere exceeded 90% (Figure 3.3 bottom). On the other hand, the data availability was nowhere higher than 96%. This result is linked to the fact that the PERSIANN dataset is missing about 5% of the one-hourly accumulations (data are missing globally). This 5% of missing data would not considerably influence the availability of three-hourly accumulations if they were randomly distributed in time, but they are usually clumped in a long uninterrupted sequence.

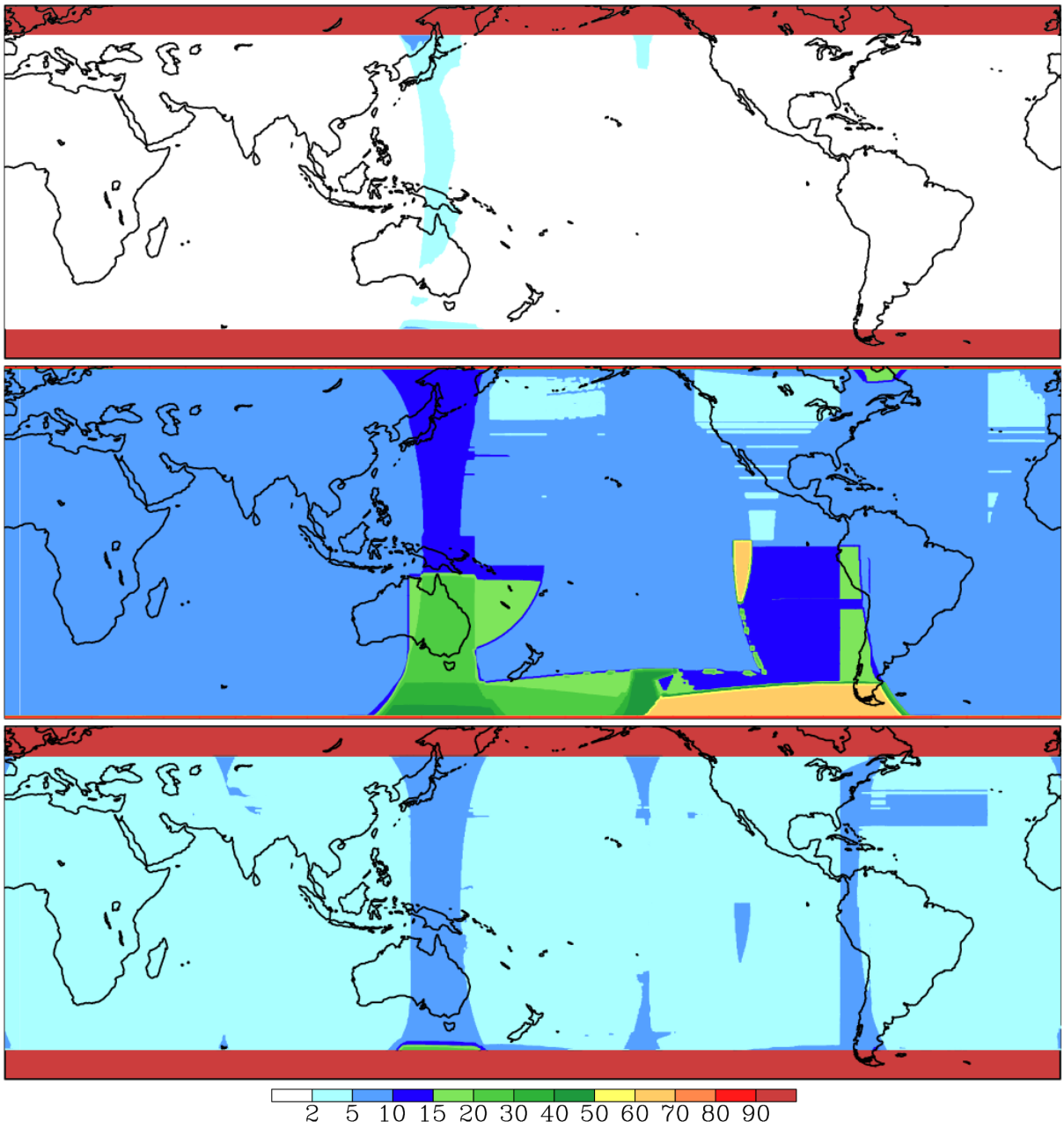


Figure 3.3: Missing data [%] for TRMM (top), PERSIANN (middle) and after synchronization (bottom) over the whole time period from March 2000 to December 2005.

3.4 The measured precipitation – annual and seasonal accumulations

With the aim of better understanding the spatial distribution of precipitation as detected by sensors on the satellites we first analyze the average annual and seasonal rainfall. These summary values were calculated for both datasets for the period from 1 March 2000 to 31 December 2005 and are shown in Figure 3.4.

If some data was missing the average precipitation accumulation was multiplied by

$$\frac{N_{\text{all}}}{N_{\text{non-missing}}}, \tag{3.1}$$

where N_{all} is number of all 3-hourly intervals in a period and $N_{\text{non-missing}}$ is number of non-missing 3-hourly intervals, thereby obtaining the missing data corrected average precipitation accumulation.

The overall spatial distribution of annual mean precipitation (i.e., the clear shape of the ITCZ with two maxima, one located in the east and one in the west) is similar for both datasets, but the amount of precipitation in regions with high precipitation is considerably different. In these regions the amount of precipitation in the PERSIANN dataset seems to be higher by approximately 20-30%. It is not clear why there is such a large difference between the datasets. Perhaps it due to somewhat different input datasets that are used by the two products (see Subsections 3.1.1 and 3.1.2), perhaps it is due to fact that TRMM precipitation is adjusted with the rain gauge data while PERSIANN precipitation is not, or perhaps due to the differences in algorithms used by the products, or the fact that the products use different ground validation data. It is not valid to say that one product performs better than the other. All one might say that the uncertancy in estimating precipitation by satellites is of similar order.

The datasets are also compared to the precipitation provided by Global Precipitation Climatology Project (GPCP Version 2.1, Adler et al., 2003). GPCP provides precipitation data[2] for daily (from 1997 to present) and monthly (from 1979 to present) accumulations in a lower $2.5° \times 2.5°$ resolution. The GPCP uses a multitude of satellite measurements as input as well as some rain gauge data. The microwave estimates are based on Special Sensor Microwave/Imager (SSM/I) data from the Defense Meteorological Satellite Program (DMSP, United States) satellites that fly in sun-synchronous low-earth orbits. The infrared (IR) precipitation estimates are computed primarily from geostationary satellites (United States, Europe, Japan), and secondarily from polar-orbiting satellites (United States). Additional low-Earth orbit estimates include the Atmospheric Infrared Sounder (AIRS data from the NASA Aqua, and Television Infrared Observation Satellite Program (TIROS) Operational Vertical Sounder (TOVS) and Outgoing Longwave Radiation Precipitation Index (OPI) data from the NOAA series satellites. The gauge data are assembled and analyzed by the Global Precipitation Climatology Centre (GPCC) of the Deutscher Wetterdienst and by the Climate Prediction Center of NOAA. Figure 3.4 bottom shows the GPCP average annual accumulations. Although the resolution is lower than with TRMM and PERSIANN datasets, the precipitation does not look too different. This may not be surprising since all three datasets use similar satellite data as input. In terms of precipitation amounts the GPCP seems to be somewhere in between the TRMM and PERSIANN precipitation amounts (TRMM dataset has less precipitation

[2] http://precip.gsfc.nasa.gov/

while PERSIANN dataset has more precipitation than the GPCP dataset).

Similar comments can be made for seasonal accumulations (Figure 3.5). The overall shape of precipitation distribution is similar for both datasets, but the amounts in areas with high precipitation differ – again the PERSIANN gives higher accumulations. The spatial distribution for the MAM and SON periods are rather similar to the annual precipitation. For the DJF period the maximum is clearly on the west side of domain, with very little precipitation in the east. For the JJA period the situation is reversed: the maximum is on the east side of the domain, but there is also some precipitation on the west side - although not as pronounced as in DJF period.

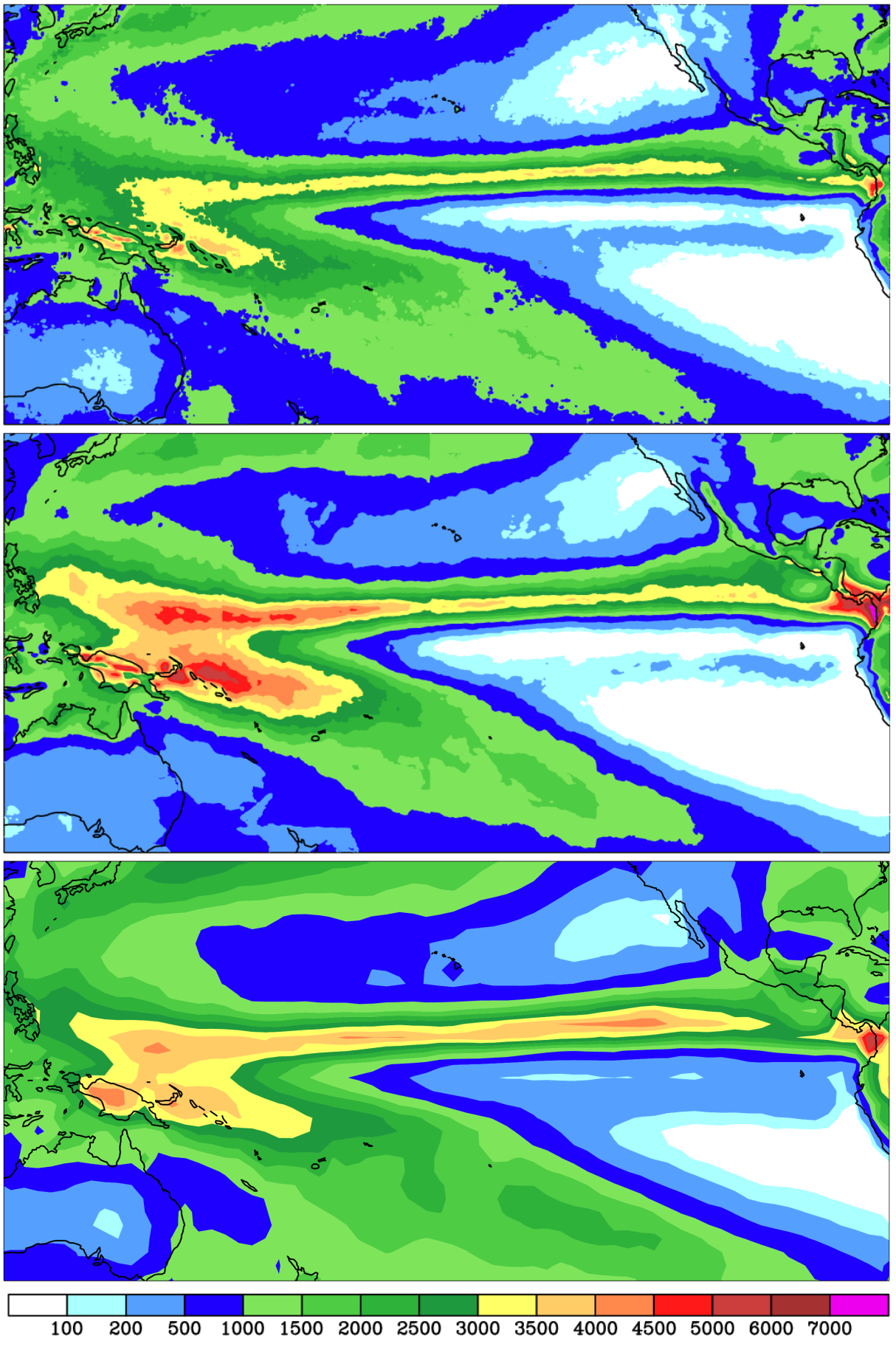


Figure 3.4: Average annual precipitation (mm) for TRMM (top), PERSIANN (middle) and GPCP (bottom).

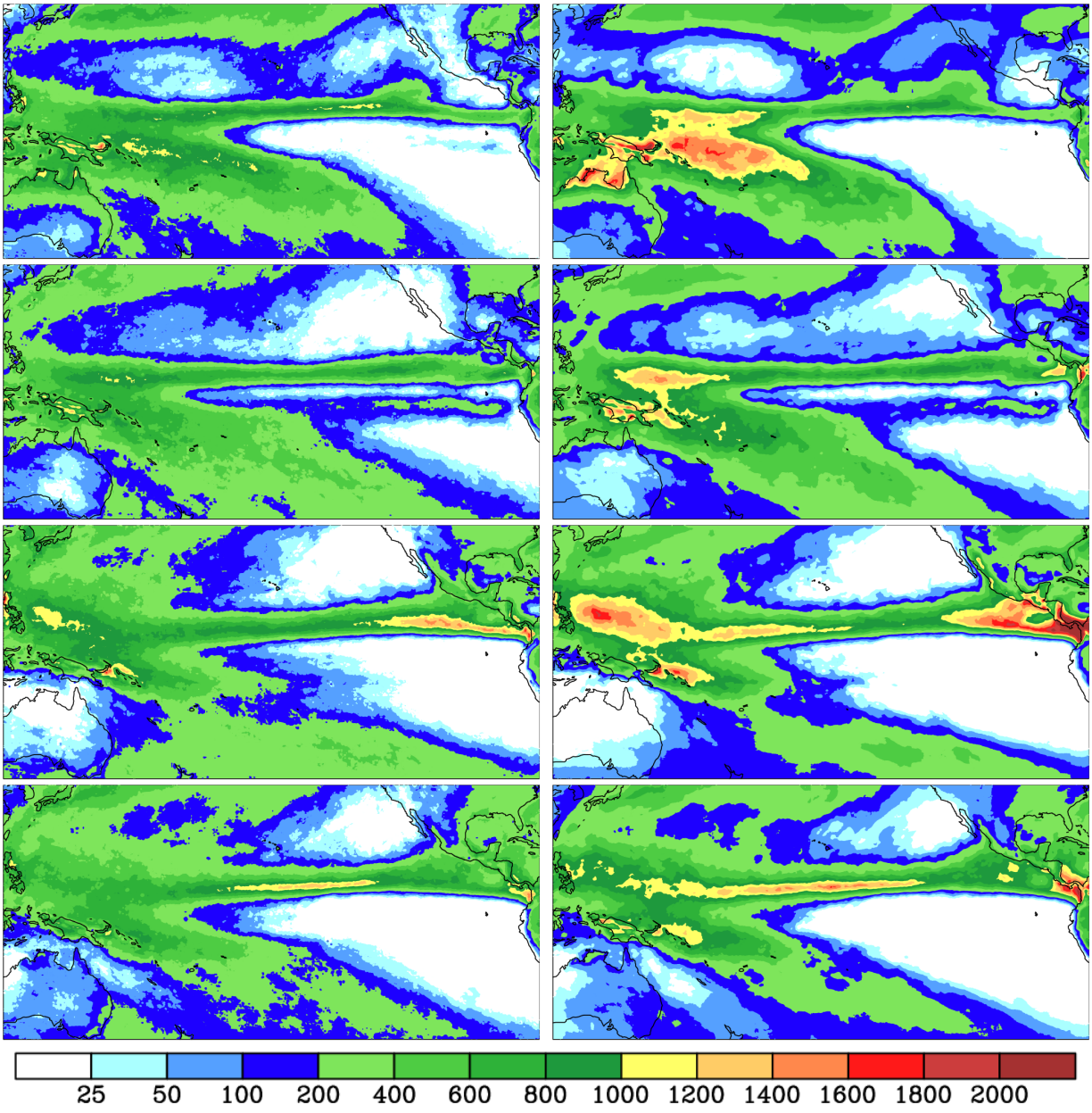


Figure 3.5: Average seasonal (three-month) precipitation (mm) for TRMM (left) and PERSIANN (right). From top to bottom: DJF, MAM, JJA and SON.

3.5 Analysis of spatial objets

It this section the TRMM and PERSIANN satellite precipitation datasets are analyzed using the basic MODE method (presented on page 15 in Section 2.1) - therefore only spatial objects (objects unrelated in time) are identified. A thorough sensitivity analysis is performed which provided optimal values for parameters R_s and TH. Using these values the datasets are analyzed in the Pacific domain for the 6-year period.

3.5.1 Sensitivity analysis

The values of R_s and TH are the most important parameters influencing the distribution of spatial objects. They heavily influence important properties such as the total number of identified objects, spatial distribution of objects and average size of objects. Therefore, a thorough sensitivity analysis is needed.

The analysis was done as follows: i) using a random number generator 100 three hourly time intervals were randomly chosen, ii) for each time interval objects were identified using a wide range of values of R_s and TH.

The convolution operates as a smoothing function and when applied to precipitation, the precipitation features are usually slightly enlarged while at the same time the precipitation maximum diminishes (and the larger the R_s is, the more they get enlarged and the maximum gets smaller). At the same time, the total amount of precipitating water in the domain is conserved (with the exception of regions near the domain borders), since the type of convolution we use is a moving average, so no precipitation is artificially lost or produced through this process.

Figure 3.6 top shows how the portion of the domain covered by objects for the TRMM and PERSIANN datasets depends on the threshold and convolution radius. Both datasets have similarly shaped graphs. There is a value of TH, invariant according to the R_s (approximately 2-3 mm of rain in 3 hours) where the choice of R_s does not influence the portion of domain covered by precipitation. The invariant value separates the graph into two distinctly different sections. If TH is smaller than the invariant threshold then the objects would always cover more area if R_s is increased. The opposite is true if TH is larger than the invariant. The reason for existence of the invariant value is explained in more detail on page 36 in Subsection 3.5.2.

If $TH = 0$ mm/3h (no thresholding) the raw objects (i.e., for $R_s = 0$), on average, cover approximately 10% of the domain for both datasets. That means that, on average, any precipitation that occurs does not normally cover more than 10% of the domain. It is therefore reasonable to not allow the processes of convolution and thresholding to lead to sets of objects that exceed that value. This means that any combination of R_s and TH that exceeds 10% coverage does not appear in nature. The precipitation features in these cases are too artificially

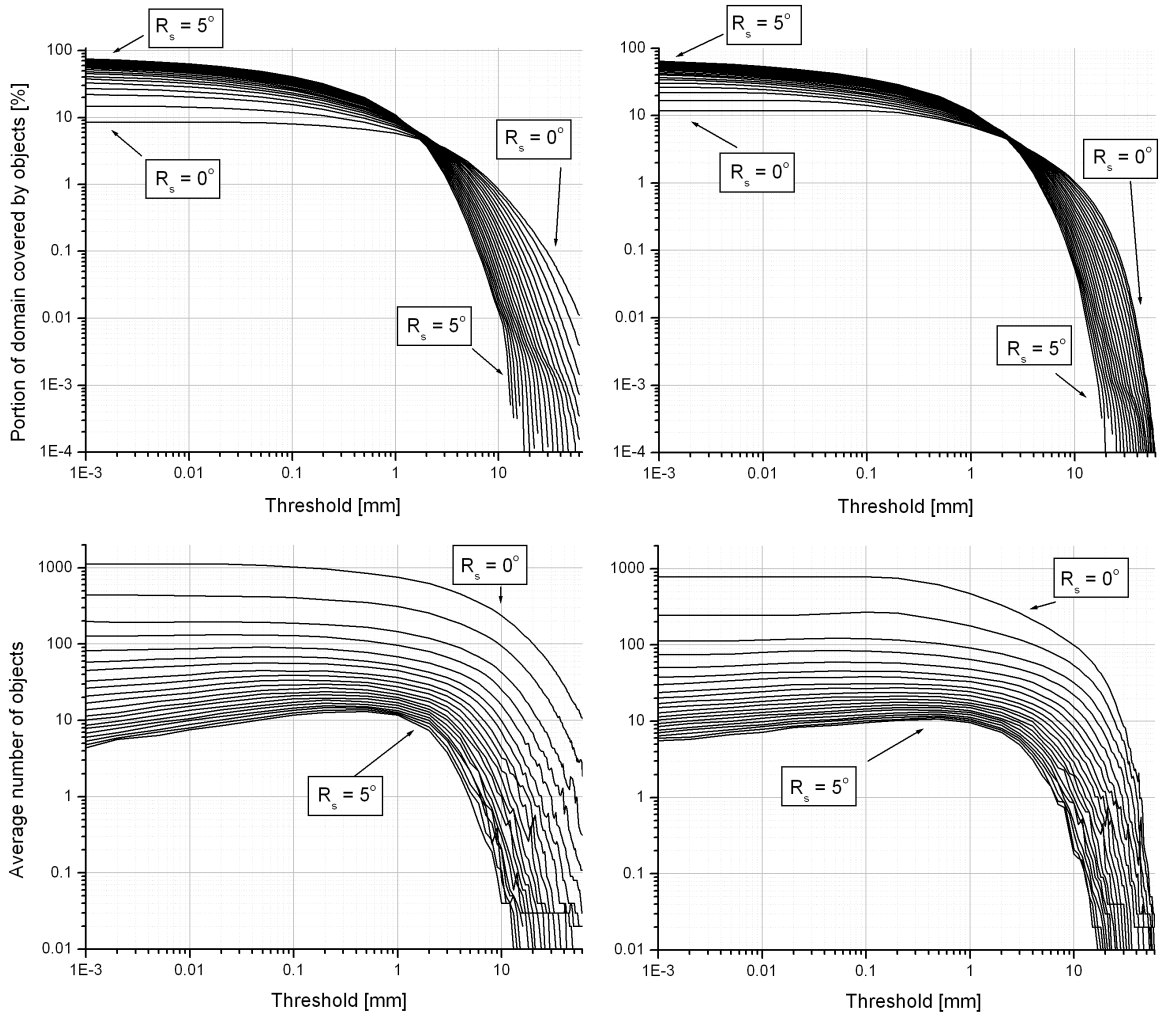


Figure 3.6: The dependence of the portion of domain covered by objects [%] (top) and of the average number of objects (bottom) on the threshold: for TRMM (left) and PERSIANN (right). Spatial convolution radii R_s span from 0 deg to 5 deg in increments of 0.25 deg (one grid point).

enlarged due to the strong smoothing effect of convolution. Thresholds spanning from 0 to 0.25 mm/3h are therefore not considered in this study, and larger thresholds (from 0.25 to the invariant) are considered only when paired with an appropriate R_s, not to exceed 10% of precipitation coverage over the whole domain.

For the non-convolved case at small values of TH (< 1 mm/3h) the objects in PERSIANN dataset cover a larger area than in TRMM – implying that the area with lower precipitation is on average bigger in PERSIANN dataset (12% compared to 9%). For TH values from 1 to 30 mm/3h, the datasets perform almost identically. For TH values from 30 to 60 mm/3h TRMM has a much larger value compared to PERSIANN. This indicates that the area with heavy

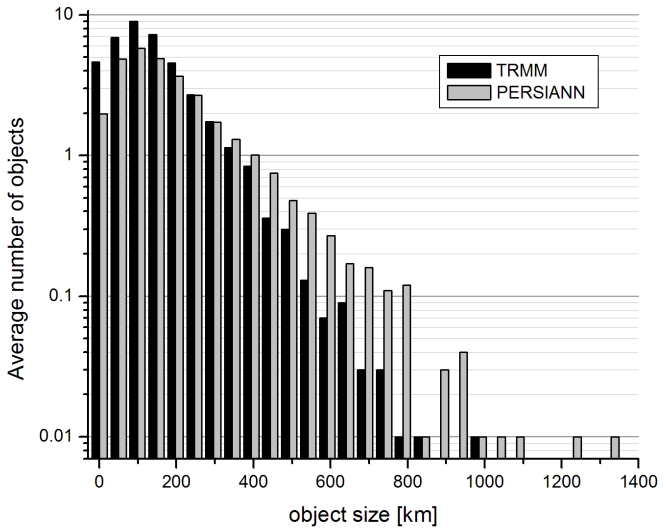


Figure 3.7: Average number of objects by size (a square root of object area) for $R_s = 0.75$ deg and $TH = 7$ mm/3h. For each case objects were separated due to their size and then the results averaged over all 100 cases.

precipitation is usually much larger in the TRMM dataset.

The average maximum number of objects when no convolution and thresholding is used is close to 1000 objects in both datasets (Figure 3.6 bottom). In general the TRMM dataset produces more objects than PERSIANN (independent of the specific R_s and TH values). This result indicates that precipitation in the PERSIANN dataset is smoother than in the TRMM dataset since thresholding produces more objects in the TRMM dataset if no convolution is applied. If the number of objects is small (< 3), the lines become somewhat non-smooth, which is due to the fact that the sample was small (100 cases).

Another conclusion that can be based on Figure 3.6 is that larger convolution radii always produce fewer objects. This result is related to the smoothing effect of the convolution process, since larger Rs values produce larger objects which are more likely to merge, which results in smaller numbers of objects. From the subjective point of view, a human would normally never identify more than 100 objects. This characteristic could lead to the conclusion that those values of convolution radii and thresholds that produce more than 100 objects are not meaningful.

The distributions of objects by size typically indicate that small objects are more numerous than the large ones. This attribute is illustrated in Figure 3.7 for $R_s = 0.75$ deg and $TH = 7$ mm/3h. This property is similar for most other combinations R_s and TH. As can be observed, for the two above specified values of R_s and TH, the PERSIANN dataset has fewer small objects (smaller than 200 km $\times$ 200 km) but more large objects compared to TRMM.

3.5.2 Study of Invariant value in the Portion of domain covered by objects

The existence of an invariant value in the of portion of domain covered by objects, shown in Figure 3.6, can be explained by studying a simplified model of convolution. Namely:

- assume that the precipitation object has the shape of a two dimensional Gaussian function:

$$f_{\mathrm{o}}(x,y) = \frac{1}{2\pi R_{\mathrm{o}}^2} \mathrm{e}^{-\frac{x^2+y^2}{2R_{\mathrm{o}}^2}}, \tag{3.2}$$

  the parameter $2R_{\mathrm{o}}$ can be understood as the size of the object.

- assume that the averaging function has the shape of a two dimensional Gaussian function:

$$f_{\mathrm{a}}(x,y) = \frac{1}{2\pi R_{\mathrm{a}}^2} \mathrm{e}^{-\frac{x^2+y^2}{2R_{\mathrm{a}}^2}}, \tag{3.3}$$

  the parameter R_{a} represents the R_{s}, although this is not the same parameter as in a non-simplified case. In fact, a moving average uses a step function (constant value inside R_{s} and zero elsewhere) for convolution.

The convolution of two Gaussian functions is a Gaussian function:

$$F(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\mathrm{o}}(h,u) f_{\mathrm{a}}(x-h, y-u)\mathrm{d}u\mathrm{d}h = \frac{1}{2\pi(R_{\mathrm{o}}^2 + R_{\mathrm{a}}^2)} \mathrm{e}^{-\frac{x^2+y^2}{2(R_{\mathrm{o}}^2+R_{\mathrm{a}}^2)}}. \tag{3.4}$$

Figure 3.8 a) shows function $F(x,y)$ using some selected values of R_{a}. The cross-section of an convolved object at a certain precipitation threshold TH, c, can be obtained from the above equation by setting $F(x,y) = TH$, $x = c/2$, $y = 0$ and then expressing c. An example of cross-section is shown in Figure 3.8. The cross-section can be expressed as:

$$c(TH) = 2\sqrt{2\left(R_{\mathrm{o}}^2 + R_{\mathrm{a}}^2\right) \cdot \ln\left[\frac{1}{TH \cdot 2\pi\left(R_{\mathrm{o}}^2 + R_{\mathrm{a}}^2\right)}\right]}. \tag{3.5}$$

Figure 3.8 b), representing the cross-section, is similar to the figures from the previous sections. The invariant threshold value is very well visible, but in Fig. 3.8 it exists only if the convolution radius (R_{a}) is smaller or similar than the half

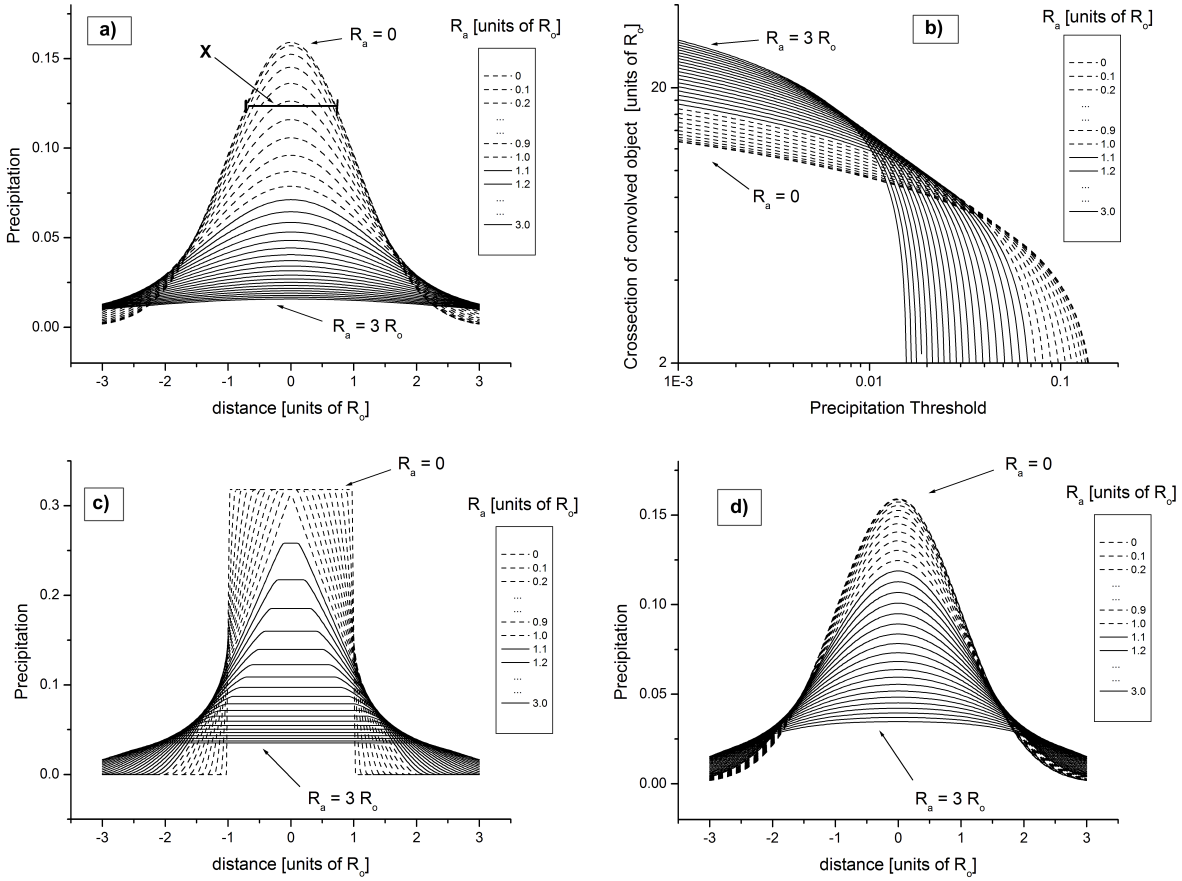


Figure 3.8: a) Gaussian precipitation object convolved with a Gaussian function. As an example the cross-section at threshold 0.125 and R_{a}=0 is shown and marked with X. b) Cross-section of a Gaussian object convolved with a Gaussian function at a certain precipitation threshold in log-log graph. c) Step shaped precipitation object convolved using a moving average. d) Gaussian precipitation object convolved using a moving average. In all four figures the lines representing convolution radius smaller than the half size of the object (R_{o}) are dotted.

size of the object (R_{o}). The cross-section at the invariant value is a bit larger than the object size (approximately $2.5R_{\mathrm{o}}$ compared to object size $2R_{\mathrm{o}}$).

Since our model does not use Gaussian function for convolution, but instead uses a moving average, the effect of a moving average (with a spatial convolution radius R_{a}) on two types of objects was tested, namely an object in the shape of a step function (constant value inside radius R_{o} and zero elsewhere) and an object in the shape of a Gaussian function. For these two cases it is not possible to obtain a simple expression for c, therefore the results were obtained numerically. The results are shown in Figure 3.8 c) and d). For both cases the invariant value is still clearly visible if R_{a} is smaller or similar than R_{o}. The existence of the invariant value does not seem to depend on the choice of the convolution function.

In general it can be said that if $2R_{\mathrm{s}}$ is considerably larger than the object

size, the cross-section at the invariant value is smaller or zero, and the invariant threshold does not appear. However, in the data, the smaller objects only cover a small part of total object area and the precipitation inside is usually low-intensity. The majority of the total object area and larger precipitation values are associated with bigger objects with sizes larger or similar to $2R_s$. This factor explains why the invariant value is visible in our dataset despite the fact that many smaller objects are present.

3.5.3 Selection of convolution radius and threshold

As shown in the previous sub-section, the proper choice of R_s and TH is of the utmost importance for object-based analysis of precipitation datasets. The selection procedure can consist of the following three steps.

1. Define goal of the analysis. In this case we are interested in convective precipitation, especially in organized convection in the tropics, which is oftentimes accompanied by heavy precipitation. We chose to define heavy precipitation as precipitation > 8 mm/3h

2. Determine the average portion of domain, covered by precipitation, for a non-convolved case with threshold 8 mm/3h. In this case the values can be obtained from Figure 3.6 and are 1.12% (TRMM) and 1.41% (PERSIANN).

3. Find a combination of R_s and TH that reflect the average portion of domain, covered by precipitation, while at the same time not breaking any of the rules defined previously. In the present case a combination of $R_s = 0.75$ deg and $TH = 7$ mm/3h produces coverage values of 1.22% (TRMM) and 1.44% (PERSIANN), while at the same time does not produce more than 100 objects (approximately 40 objects for TRMM and 31 for PERSIANN).

3.5.4 Spatial distribution of objects in Pacific

Results of an analysis of the data from March 2000 to December 2005 using $R_s = 0.75$ deg and $TH = 7$ mm/3h are graphically presented in Figure 3.9. For each grid point, object frequencies were calculated. Object frequency is the number of time intervals when an object appeared on a grid point divided by the total number of time intervals. The overall impression provided by the two graphs at the top of Fig. 3.9 is that the object frequencies are very similar to the average annual accumulations. Despite the fact that the values in the PERSIANN

dataset seem to be consistently larger than in the TRMM dataset, the general shape of the spatial distribution is somewhat similar. The higher frequency in the PERSIANN dataset indicates that heavy precipitation (> 8 mm/3h) is more frequent than in the TRMM dataset, which might be a result of higher precipitation accumulations in the PERSIANN dataset (Figure 3.4) which causes more objects being identified. Two maxima can be identified, a local maximum in the West and the domain maximum in the East. The domain maxima are approximately 12% for PERSIANN and 7.5% for TRMM. Another difference between the distributions for the two datasets is apparent in the area just east of Japan where objects are more frequent in the TRMM dataset.

The spatial distribution of the smallest objects (second row of Fig. 3.9) is similar to the distribution of all-sized objects. The highest frequency of middle-sized objects (third row of Fig. 3.9) seems to be in the East, although this is more apparent in PERSIANN dataset. The overall frequency of the larger objects (bottom row of Fig. 3.9) seems to differ considerably between datasets (maxima of 0.5% for TRMM and 3% for PERSIANN). On the other hand for both datasets the larger objects appear only in the western part of domain (and east of Japan for TRMM dataset).

An analysis of seasonal accumulations (available on page 89 in APPENDIX I) indicates that the seasonal object frequency distributions are consistent with the distributions of seasonal accumulations. The only observable difference seems to be for large objects, which consistently seem to be more frequent on the western side of the domain.

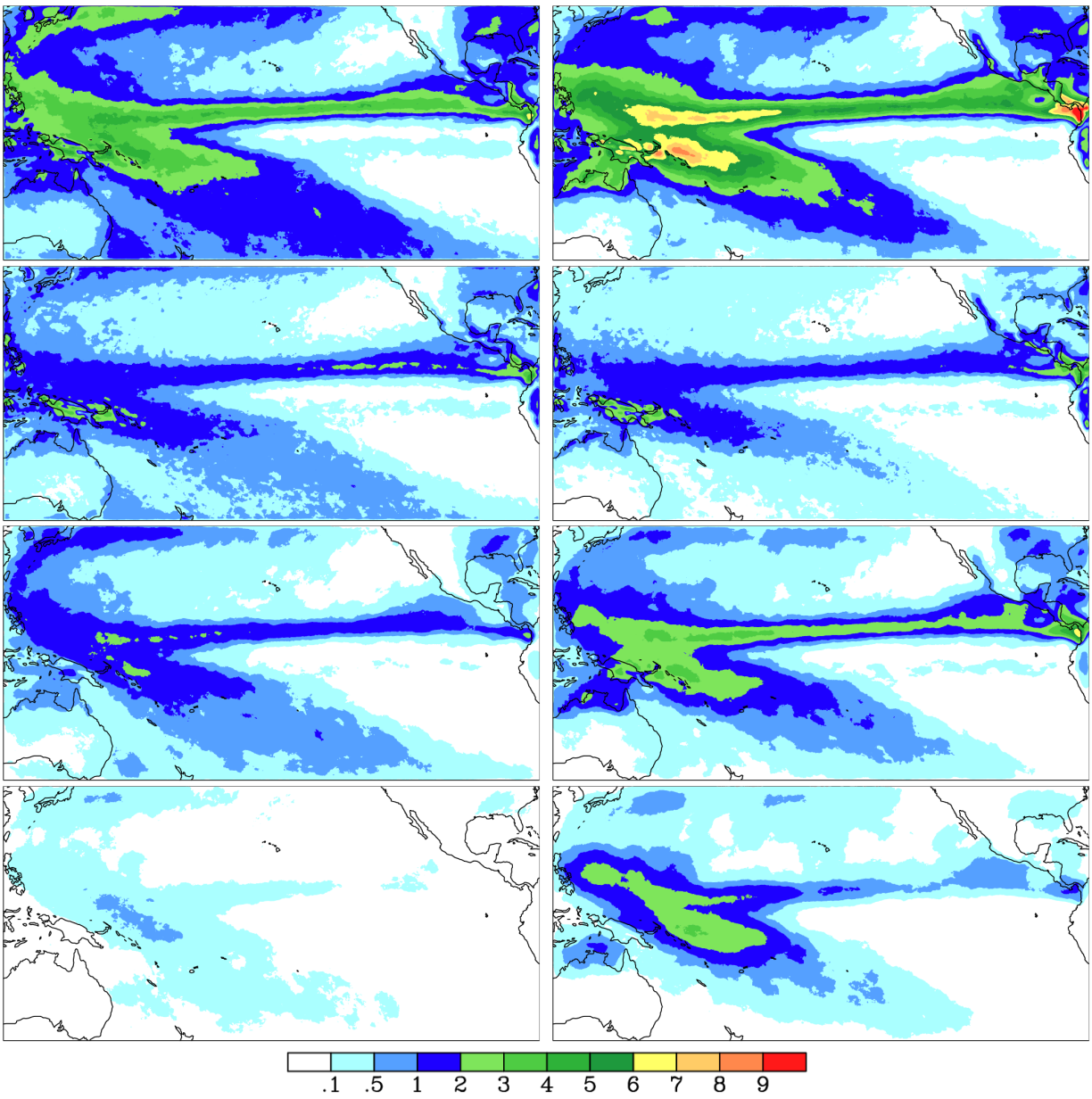


Figure 3.9: Annual object frequency [%] for TRMM (left) and PERSIANN (right): top - all objects regardless of size, 2nd row - objects with area 0 – 87,198 km^2, 3rd row - 87,198 – 348,792 km^2 and bottom - larger than 348,792 km^2.

3.6 Temporal behavior of objects

It this section the TRMM and PERSIANN satellite precipitation datasets are analyzed using the extended method, which was developed in this thesis and presented on page 16 in Section 2.2. This method enables the study of temporal evolution of objects - for example their lifespan and movement. A thorough sensitivity analysis is performed which provides optimal values for parameters R_s, $R_\mathrm{t}$and TH. Using these values the datasets is analyzed in the Pacific domain for the 6-year period.

3.6.1 Preliminary sensitivity analysis

The analysis was performed in two steps. The first step used a wide range of parameter values, but due to the computational limitations only a few of the simpler properties were studied. The second step used a smaller range of parameter values but the analysis included all properties, including lifespan.

The preliminary analysis was performed in the following manner: 100 three-hourly time intervals were randomly chosen. For each case, objects were identified by convolving with a wide range of R_t and R_s values and thresholding with different TH values, as presented in Table 3.2.

Due to the computational limitations the process of object identification was only done for randomly chosen time intervals (no preceding or subsequent time intervals were analyzed). Since only one time interval was processed, only limited properties could thus be estimated. For example, object lifespan could not be determined with such a simplified approach. For this part of the analysis only the TRMM dataset was studied.

Figure 3.10 shows that area covered by objects is very similar to the analysis for only spatial objects. The invariant threshold value (approximately 2 – 3 mm/3h) is again present. At thresholds smaller than the invariant value the portion of domain is larger than in a non-convolved case making this an invalid choice according to our criteria. At larger thresholds the area is smaller than in a non-convolved case. The temporal convolution behaves similarly to the spatial convolution in that it leads to coverage of larger portions of the domain if the threshold is smaller than the invariant value and to less coverage otherwise.

The effects of the temporal convolution on the number of objects (Figure 3.11) differs from the effects of the spatial convolution. In the analysis of spatial objects,

Table 3.2: Used values of convolution radii and threshold. All possible combinations of parameter values were analyzed.

Parameter	Values
R_t [h]	0, 3, 6, 9, 12 , 15
R_s [deg]	0, 0.25, 0.5, 1.25, 2.5, 5
TH [mm/3h]	0, 0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1, 2, 5, 10, 20, 50, 100

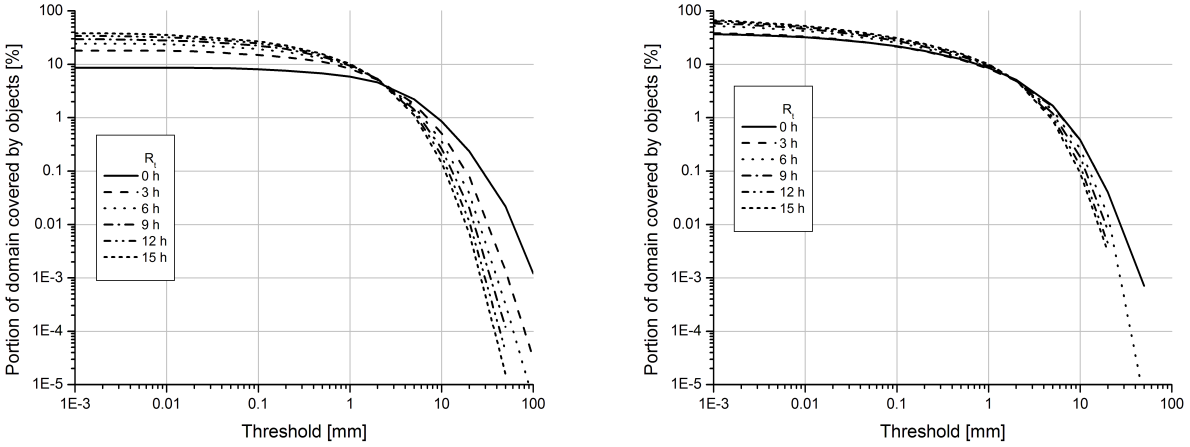


Figure 3.10: Figure 7: Portion of domain covered by objects [%] dependence on the threshold for increasingly larger R_t while R_s is constant. R_s is either 0 (left) or 1.25 deg (right).

larger R_s always produced fewer objects. This is true for R_t only as long as the threshold is larger than the invariant value. If the threshold is smaller, temporal convolution always produces more objects compared to cases where only a spatial convolution is applied. The increase in the number of objects due to temporal convolution demonstrates the "spillover" effect, whereby objects from nearby time intervals "spill over" to the current time interval due to temporal convolution. For cases with $R_s = 0$ deg, an increase in R_t always produces more objects, while for cases with $R_s = 1.25$ deg, the maximum number of objects is for $R_t = 3$ h. In contrase, if R_t is larger, the number of objects decreases (but is never smaller than the case with $R_t = 0$ h). The maximum at 3 hours, and the gradual decrease for larger R_t, can be attributed to an increased amount of spatial averaging when comparing $R_s = 0$ deg and 1.25 deg. The number of grid points inside the averaging ellipsoid is many times greater for $R_s = 1.25$ deg, where averaging prevails over the "spill over" effect at low thresholds.

The "spill over" effect can happen even if no object or precipitation is currently present at a location. This effect is clearly undesirable since it artificially increases the object lifespan. The spillover can happen regardless of the threshold value but seems to be much stronger if the threshold is smaller than the invariant value. Since the size of the R_t directly impacts the temporal scope of the spillover effect (the larger the R_t, the larger the temporal scope), we concluded that any value of R_t greater than one time interval (three hours) is not appropriate, because the objects can "spillover" to more than one time interval. Thus, only R_t values of zero and three hours were deemed appropriate for further analysis.

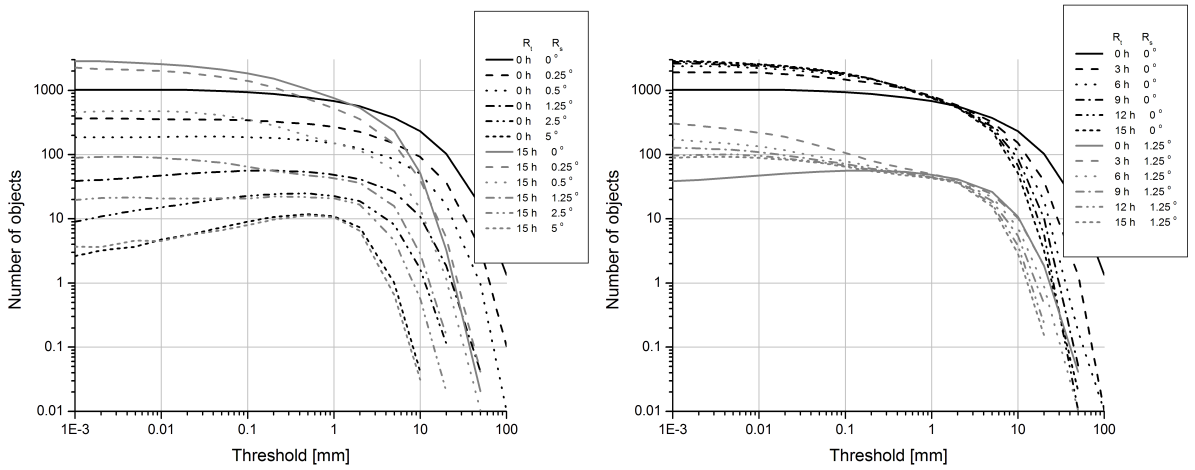


Figure 3.11: Number of objects dependence on the threshold for different combinations of temporal and spatial convolution radii. Left: lines with different spatial conv. radii are marked with different line styles. Two different temporal conv. radii are marked with black and grey colors. Right: lines with two different spatial conv. radii are marked with black and grey colors. Different temporal conv. radii are marked with different line styles.

Table 3.3: Used values of convolution radii and threshold. All possible combinations of parameter values were analyzed.

Parameter	Values
R_t [h]	0, 3
R_s [deg]	0, 0.25, 0.5, 1.25, 2.5, 5
TH [mm/3h]	1, 2, 5, 10, 20, 50, 100

3.6.2 Lifespan sensitivity analysis

The previous analysis included processing of only two consecutive time intervals for each of the 100 randomly selected cases. This enabled the selection of a wide range of convolution radii and threshold parameters. It was shown that the use of threshold values smaller than the invariant value (approximately 2 – 3 mm/3h) is not advisable. Additionally, the use of a temporal convolution produces an undesirable "spill-over" effect. The remaining range of the parameters is therefore considerably smaller (Table 3.3). This enables a more robust analysis where properties such as lifespan could be studied.

For each randomly chosen case the next 89 time intervals were processed. For this part of the analysis only the TRMM dataset was studied.

An initial hint on the lifespan of the objects can be based on the results of the Machado et al. (1998) sttudy, in which an algorithm was developed and applied to study the evolution of morphological and radiative properties of clouds over life cycles of deep convective systems over the Americas. In particular, Machado et al. found a function similar to a power law was appropriate to describe the relationship between the number of convective systems vs. the life cycle duration

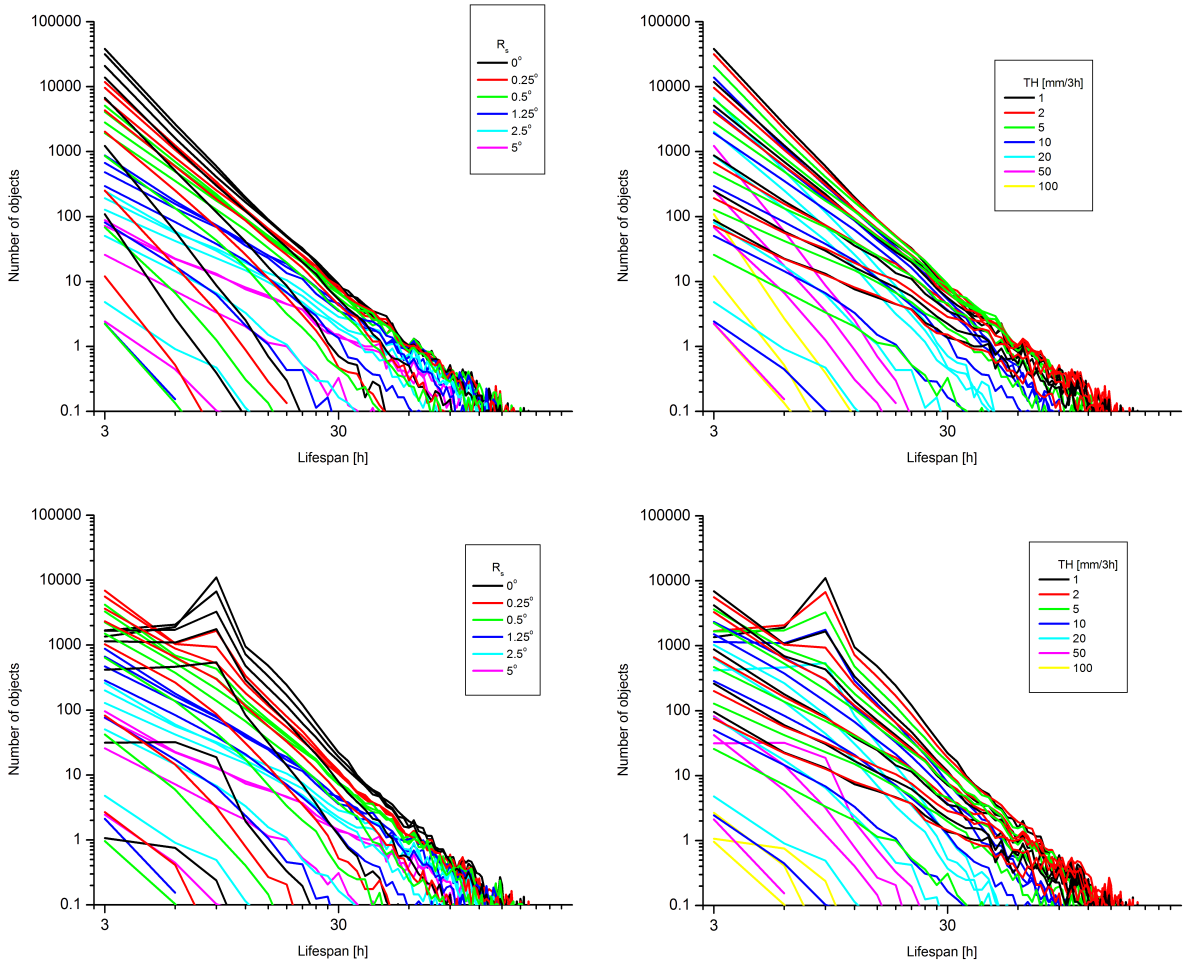


Figure 3.12: Average number of objects per case with respect to Lifespan on log-log scale for $R_t = 0$h (top) and $R_t = 3$ h (bottom). On the left side the lines are colored with respect to R_s and on the right with respect to threshold.

of the systems (see Fig 2 of Machado et al., 1998); their study was based on an analysis of GOES-7 ISCPP-B3 satellite data from the period 1987-88. In a related study, Wilcox (2003) used an automated cloud detection and tracking algorithm to observe and track overcast decks of clouds in a consecutive set of hourly Meteosat-5 images and NCAR CCM3 model output for 49 days in the 1999 winter monsoon period over the Indian Ocean. Wilcox also found a function similar to power law was appropriate to model the relationship between the number of clouds vs. cloud lifetime (see Fig 3a of Wilcox 2003). Following the approach taken in these studies we have examined whether the number of TOs vs. lifespan follows the power law in our case as well. A power law would appear as a straight line in a log-log graph so a log-log scale is presented in our Figure 3.12. A power law can be expressed as

$$N = N_0 L^{-a}, \tag{3.6}$$

where N is the number of objects, L is the lifespan and N_0, a are the parameters. N_0 describes the number of objects at the shortest lifespan, $L = 1$, while a represents how fast the number of objects decreases with lifespan. The different curves in Figure 3.12 do not collapse due, first, to different N_0 values associated with different samples, and second, to differences in the parameter a. Nevertheless, if no temporal convolution is applied, all of the curves appear as straight lines if the average number of objects is greater than one. However if a temporal convolution of 3 h is applied, some lines do not appear straight. This nonlinearity only is apparent in lines with small spatial convolution radii (0 and 0.25 deg) and the lines get straighter when threshold is increased. Also, the maximum number of objects shifts from shortest lifespan of three hours to nine hours. This characteristic leads to the conclusion that this is a spillover effect that occurs when relatively small objects with the shortest lifespan of three hours spill into two nearby time intervals and artificially increase the objects' lifespan to nine hours.

Each of the lines could be approximated with parameters N_0 and a, by fitting a linear function to a log-log graph. This was done by applying a logarithmic function to the above equation:

$$\ln N = \ln N_0 - a \ln L, \tag{3.7}$$

and estimating parameters a and $\ln(N_0)$ with a least squares fit. Only cases with at least five objects were used to estimate the fitted parameters. The line was fitted if more than two points existed with the number of objects greater than five.

In Figure 3.13 we show the results of estimating a and N_0 according to the choice of the spatial convolution radius R_s. As shown,in Fig. 3.13, larger R_s values produce smaller a and N_0 values. In fact, a convolution with a larger R_s value smooths the precipitation fields more than a convolution with a small R_s value; therefore, after thresholding fewer (and larger) objects are obtained. The smaller a associated with larger R_s values is due to the longer lifespan associated with the large-scale objects that are obtained. Additional temporal convolution only influences the parameters a and N_0 if R_s is small (smaller than 0.5 deg). This result is probably due to the fact that for smaller R_s values the number of grid-points in the convolution ellipsoid that are contributed by the current time interval is comparable to the number of grid-points in the two neighboring time intervals. On the other hand if, R_s is larger than R_t (comparing number of grid point distances and number of time intervals), the contribution from the current time interval is much larger. But even if R_s is small, one would expect that a would decrease with introduction of temporal convolution. Temporal convolution should produce more overlapping and consequently fewer objects (smaller N_0)

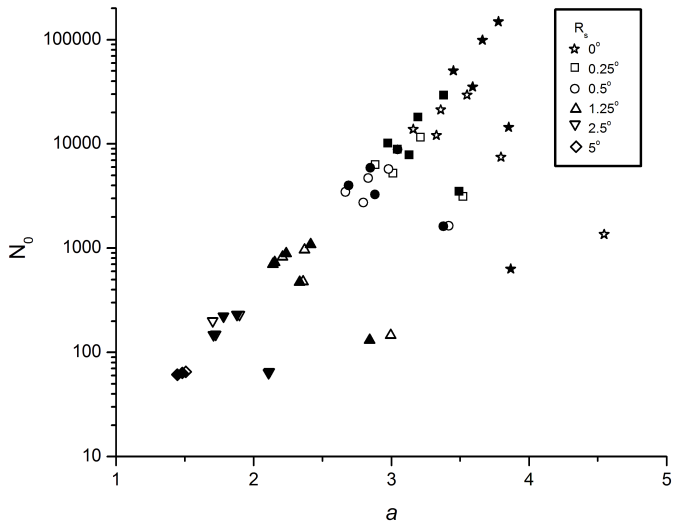


Figure 3.13: Log-linear graph of dependence of a and N_0 on spatial convolution radius R_s - no temporal convolution (no fill), with temporal convolution radius 3 hours (solid fill).

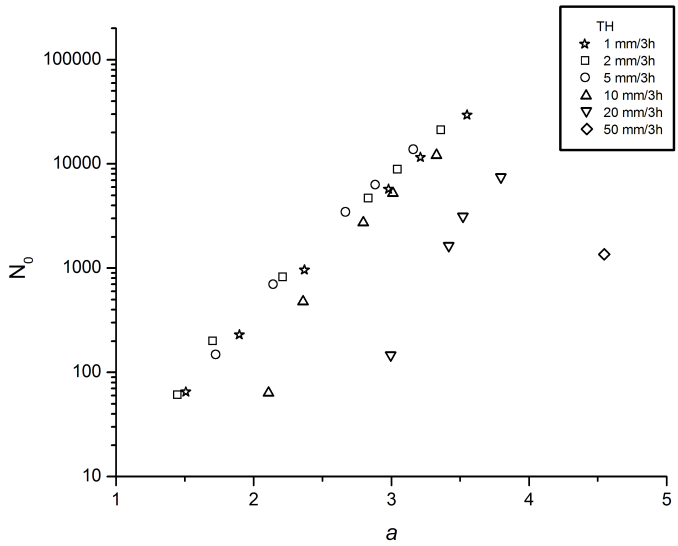


Figure 3.14: Log-linear graph of dependence of a and N_0 on threshold TH - no temporal convolution.

with a longer lifespan (smaller a). However, this behavior does not happen since a usually stays the same or increases when temporal convolution is introduced. Therefore, it can be said that while temporal convolution with a radius of three hours does somewhat influence the number of objects vs. lifespan relationship, the unwanted effects of "spill-over" make the use temporal convolution not worthwhile.

A larger threshold (e.g., greater than 5 mm/3h; Fig. 3.14) leads to a larger value of a. Higher thresholds are associated with more intense short-lived events (with larger a). Higher thresholds are also associated with a smaller number of events (i.e., smaller N_0).

3.6.3 The dependence of Number of objects on object lifespan for short-lived randomly placed storms

Since the relationship between the number of objects and the object lifespan can be represented reasonably well by a power law, an interesting question arises as to whether such a power law dependence would also appear if objects are based on randomly-placed short-lived storms. For example, at each time interval a number of randomly placed storms with a lifespan of one time interval would be generated. Since some storms would have an overlap with storms in the next time interval, the method of temporal object identification would incorrectly identify them as objects with a lifespan longer than one time interval.

For the purposes of this study a simplified model is assumed, where V is the domain area, and N is the number of self-enclosed objects at each time interval. All self-enclosed objects are the same circular size and shape, $S = \pi R^2$ is the area of every object (R is the object radius), and objects are randomly placed in the domain without overlapping in the same time interval. The number of time intervals T is large ($\gg 1$) and the assumption is made that $V \gg NS \gg S$.

The probability of a specific circular object overlapping any object in the next time interval is
A simplified model is assumed:

$$P_2 = \frac{4NS}{V}. \tag{3.8}$$

P_2 is obtained as follows. If the distance between the centers of the circular objects is smaller than $2R$ the objects overlap (Figure 3.15 left). To have an overlap, the center of a specific object has to be located inside an area with size

$$\pi(2R)^2 = 4\pi R^2 = 4S. \tag{3.9}$$

Since there are N objects in the next time interval the above area is multiplied by N. This is divided with domain area V to get the P_2 - probability of a specific object overlapping any object in the next time interval.

The probability that the temporal object will have a lifespan of minimum three time intervals (overlap again in the third time interval) is P_2 multiplied by P_2, so that $P_3 = P_2^2$. In general it can be written

$$P_L = \left(\frac{4NS}{V}\right)^{L-1} = (4k)^{L-1}, \tag{3.10}$$

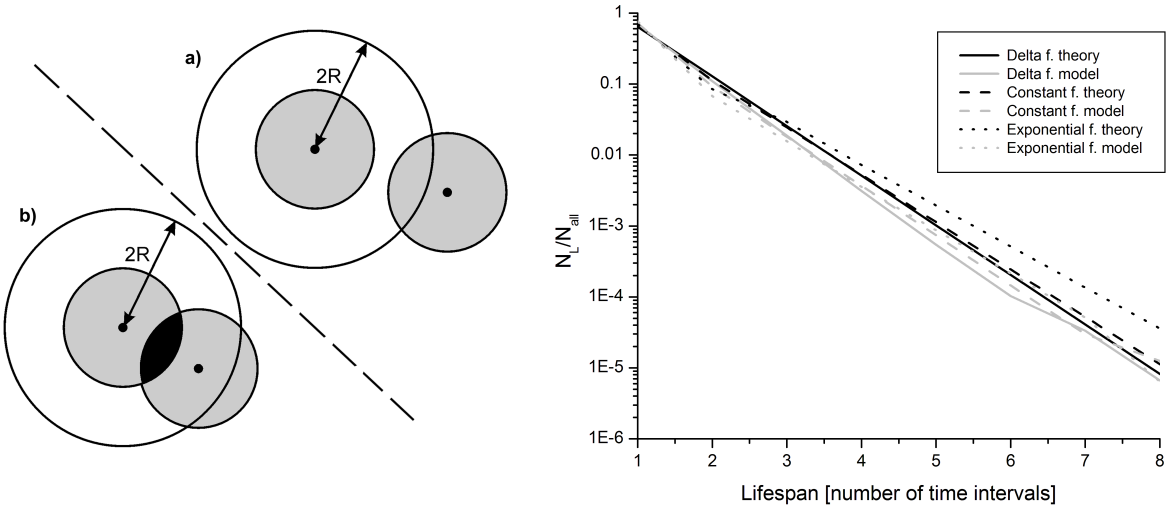


Figure 3.15: Left: Circle overlap when distance between centers is larger (a) or smaller (b) than $2R$. Right: Log-linear graph of number of objects vs. lifespan dependence for randomly distributed objects. Delta f. (case 1), constant f. (case 2) and exponential f. (case 3) represent three different number of spatial objects vs. radius functions. Theory and model indicate whether values were obtained with the use of theory or the use of model.

where k is ratio between area covered by all objects in a single time interval (NS) and domain area V. P_L is the probability that the lifespan of the object will be L or longer. The probability that the lifespan of object will be exactly L is $P_L - P_{L+1}$. In a finite (but large) number of time intervals $P_L - P_{L+1}$ can be estimated as:

$$P_L - P_{L+1} = \frac{N_L}{N_{3D}}, \tag{3.11}$$

where N_L is number of objects with lifespan L, and N_{3D} number of all temporal objects. N_L can be expressed from the above equation as

$$N_L = N_{3D} \cdot (P_L - P_{L+1}). \tag{3.12}$$

N_{3D} can be calculated from the number of spatial objects in all time intervals N_{all}. N_{all} equals NT, if there are N objects in each time interval and the number of time intervals is T. N_{all} can also be expressed as

$$N_{\text{all}} = \sum_{L=1}^{\infty} N_L \cdot L = N_{3D} \sum_{L=1}^{\infty} (P_L - P_{L+1}) \cdot L. \tag{3.13}$$

Thus,

$$N_{3D} = \frac{N_{all}}{\sum_{L=1}^{\infty}(P_L - P_{L+1}) \cdot L}, \tag{3.14}$$

and N_L can be expressed in general form as

$$N_L = \frac{(P_L - P_{L+1}) \cdot N_{all}}{\sum_{i=1}^{\infty}(P_i - P_{i+1}) \cdot i}. \tag{3.15}$$

Using the expression $P_L = (4k)^{L-1}$, N_L can be expressed as

$$N_L = (4k)^{L-1}\left[\frac{(1-4k) \cdot N_{all}}{\sum_{i=1}^{\infty}(4k)^{i-1}(1-4k) \cdot i}\right] = (4k)^{L-1}(1-4k)^2 \cdot N_{all}. \tag{3.16}$$

Since the only position where L is present in the above equation is the exponent and k and N_{all} are constants, the dependence of N_L on L is exponential.

Since in nature spatial objects do not all have the same size, we decided to analyze two additional cases. In both cases the objects are circular as in previous case, but their sizes can be different. Case 2 assumes that the number of objects vs. radius is a constant function, meaning that each object radius (up to a maximum allowable radius R_{max}) occurs with the same likelihood. Case 3 assumes that the number of objects vs. radius is a decaying exponential function, meaning that smaller objects are much more likely to occur than large ones. The number of objects vs. radius in Case 1 is a Dirac delta function since all objects are the same size. The probability that some randomly chosen objects will overlap any object in the next interval – P_2, can be obtained similarly as before with the addition that the probability has to be integrated over all possible sizes for objects in the current and next time interval, thereby obtaining

$$P_2 = A \cdot \int_0^{\infty} \frac{\mathrm{d}N}{\mathrm{d}R}(R_0)\left[\int_0^{\infty} \frac{\pi(R_0+R_1)^2}{V}\frac{\mathrm{d}N}{\mathrm{d}R}(R_1)\mathrm{d}R_1\right]\mathrm{d}R_0, \tag{3.17}$$

where the $\frac{\mathrm{d}N}{\mathrm{d}R}(R)$ is the number of objects vs. radius (R) function and A is a constant defined by $A = \left[\int_0^{\infty} \frac{\mathrm{d}N}{\mathrm{d}R}(R)\mathrm{d}R\right]^{-1}$.

P_L can be expressed in general form as L-order integral

$$P_L = \frac{A\pi^{L-1}}{V^{L-1}}\int_0^{\infty}\frac{\mathrm{d}N}{\mathrm{d}R}(R_0)\underbrace{\left[\int_0^{\infty}(R_0+R_1)^2\frac{\mathrm{d}N}{\mathrm{d}R}(R_1)\left[\int_0^{\infty}(R_1+R_2)^2\frac{\mathrm{d}N}{\mathrm{d}R}(R_2)\left[\ldots\right]\mathrm{d}R_2\right]\mathrm{d}R_1\right]}_{L-1}\mathrm{d}R_0.$$

For all three cases P_L can be expressed as

$$P_L = f(L) \cdot k^{L-1}, \tag{3.18}$$

where $f(L)$ is an numeric value independent of k and depends on the type of the number of objects vs. radius function. The above equation was checked to be true for all three cases for at least $L \leq 15$. For the first case where all objects had the same size, $f(L)$ equals 4^{L-1}. For the other two cases the expression for $f(L)$ can not be expressed in simple form and is obtained with the evaluation of the multiple integral presented in equation for P_L. While P_L for case two and three is not a solely exponential function, the expression is at least partly exponential due to the k^{L-1} part.

The above theory was tested with the identification algorithm for temporal objects. The analysis was done using a simple model where Pacific domain was used, 10000 time intervals were generated, k was chosen to be 0.05 (5% of domain covered with objects), 90 time intervals were kept in memory, circular objects were distributed inside domain in such a way that there was no overlap in the same time interval, random number generator ran2 (Press et al., 1992) was used.

The results for theoretical and model values of N_L for all three cases can be seen in Figure 3.15 right. The results show that for all cases, for both theoretical and model results, N_L decays with L very similar to exponential function. It can therefore be said that different choice for number of objects vs. radius function does not influence the number of objects vs. lifespan dependence enough to make it non exponential. That being said, there are some differences between different cases. The model values consistently show a smaller number of objects compared to theory for the same case (except for $L = 1$, where the number of objects in model is somewhat larger compared to theory). The discrepancy between the theoretical and model results can be attributed to the fact that some theoretical assumptions were not strongly enforced in the model. For example the assumption that $k \ll 1$, which was set to 0.05 in the model in order to reflect the average area covered by precipitation present in the dataset. Also the number of objects in each time interval was supposed to be large, which in reality was 100 or less to reflect the precipitation in the dataset.

When comparing theoretical values for the three cases, case 2 and 3 both show smaller number of objects for $L = 2$ and larger number for longer lifespan compared to case 1. The difference between case 2 and 1 is minimal, while the difference between case 3 and 1 is larger. Case 3 will have a considerably more objects with longer lifespan. This trend is also reflected in model results, although the difference between cases is smaller.

Since it is shown that in the case of short-lived randomly placed objects the relationship between the number of objects and the object lifespan would be similar to an exponential function and not a power-law function, it can be concluded that the precipitation systems in the TRMM dataset most likely do not consist of three-hourly storms since an exponential dependence was not found. The exponential function decreases much faster than the power law function; this comparison indicates that objects with a longer lifespan prevail in real-world precipitation (compared to randomly generated objects). We can conclude from this comparison that real-world precipitation does not consist of randomly placed short-lived storms and that the precipitation systems in the dataset either have

a longer lifespan than three hours, or their placement is not random (i.e., they are more likely to appear in some places than others). This is in agreement with the TRMM and PERSIANN datasets where the non-random placement is obvius from the non-random distribution of annual precipitation accumulation in Figure 3.4 (some areas like ITCZ contain large precipitation acumulation while some areas like south-east Pacific, contain alsom no precipitation). The finding that the precipitation systems last longer than three hours is in agreement with other studies such as Laing and Fritsch (1997) which state that the typical convective complex persists for about 10 hours.

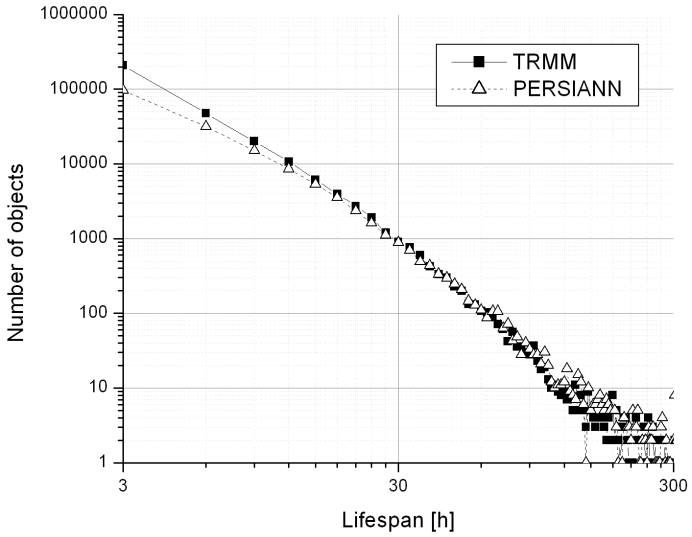


Figure 3.16: Log-log graph of average number of objects/lifespan distribution for TRMM and PERSIANN.

3.6.4 Object lifespan and movement in Pacific

In the previous section the influence of R_s, R_t and TH on the object distribution was studied using a sample of 100 randomly selected time intervals. In order to properly characterize the lifespan and the direction of movement of storm systems, the whole dataset period must be analyzed using specific values for R_s, R_t and TH. As in SubSection 3.5.3, a spatial convolution radius of 0.75 deg and a threshold 7 mm/3h were chosen. No temporal convolution was applied in order to avoid any unwanted spillover effect. The entire period from 1st March 2000 to 31st December 2005 was studied for both datasets across the Pacific domain; 100 previous time intervals were kept in memory during the processing.

Figure 3.16 shows that the relationship between the average number of objects and the object lifespan is linear-like (on a log-log graph) for both datasets. The PERSIANN dataset has fewer objects with a shorter lifespan (lifespan less than nine hours).

We defined an object trajectory as the movement of the center of the object. In Figures 3.17 and 3.18 all trajectories (annual distributions for year 2001) are obviously clustered in specific regions: their density (not surprisingly) looks similar to the distribution of annual precipitation accumulations. The highest density of objects with a longer life (red) is in the ICTZ and in the low-latitudes in the west. With somewhat lower density, long-lived objects also exist in the northwestern part of the domain and east of Australia. There is considerable difference between the two datasets in the eastern tip of the ITCZ – where there are many objects with a longer life in the PERSIANN dataset and almost none in the TRMM dataset.

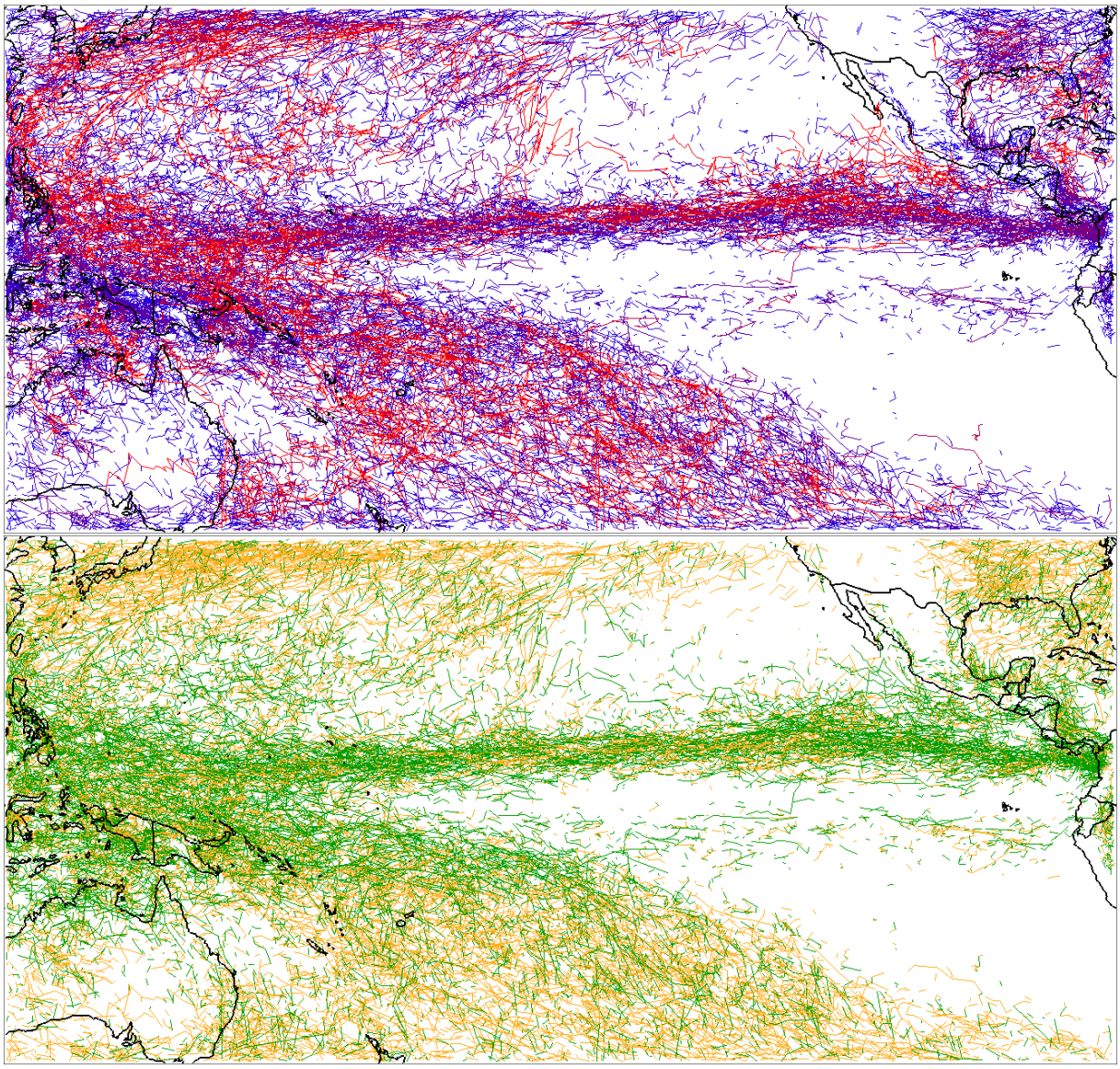


Figure 3.17: Trajectories of objects in year 2001 for TRMM. Top: trajectories are colored with respect to object lifespan - short lifespan (blue), long lifespan (red). Bottom: Trajectories are colored with respect to the westward/eastward movements. If movement in a certain time interval has a westward component the trajectory segment will be colored green. If the movement has an eastward component the color is orange.

Movements roughly follow that shown schematically in Figure 3.19. Movement in the northern and southern parts of domain is predominantly eastward (orange in Figures 3.17 and 3.18). In the ITCZ region, movement in both directions is present although westward movement (green) is more frequent. In the eastern and western part of the ITCZ the westward movement is clearly dominant.

The seasonal trajectories (Figure 3.20 for TRMM and Figure 3.21 for PERSIANN) also reflect the distribution of seasonal precipitation accumulations. Some properties of the seasonal trajectory distribution are similar to the annual distribution

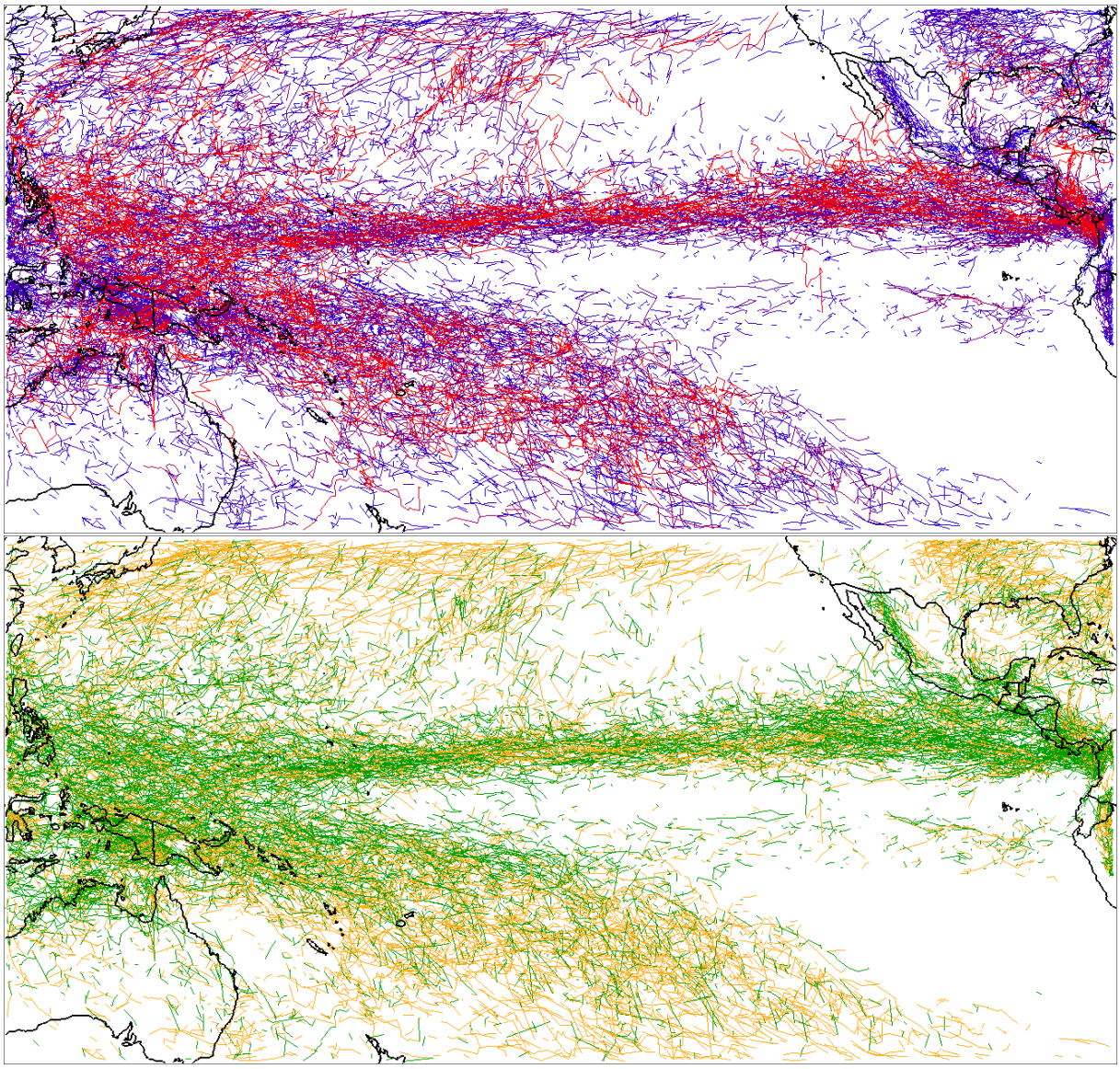


Figure 3.18: Trajectories of objects in year 2001 for PERSIANN. Top: trajectories are colored with respect to object lifespan - short lifespan (blue), long lifespan (red). Bottom: Trajectories are colored with respect to the westward/eastward movements. If movement in a certain time interval has a westward component the trajectory segment will be colored green. If the movement has an eastward component the color is orange.

(in the northern and southern parts of the domain, the movement is predominantly easterly). However there are also some differences, most notably the direction of movement in the ITCZ – in the DJF and MAM periods the movement is more eastward, which cannot be said for the JJA and SON periods. In MAM season there is considerably less precipitation with a longer storm lifespan in the ITCZ region compared to other seasons. The secondary ITCZ, which manifests itself south of the ITCZ during the MAM season, mainly contains precipitation areas with short lifespans. During the DJF season in the eastern part of the ICTZ region,

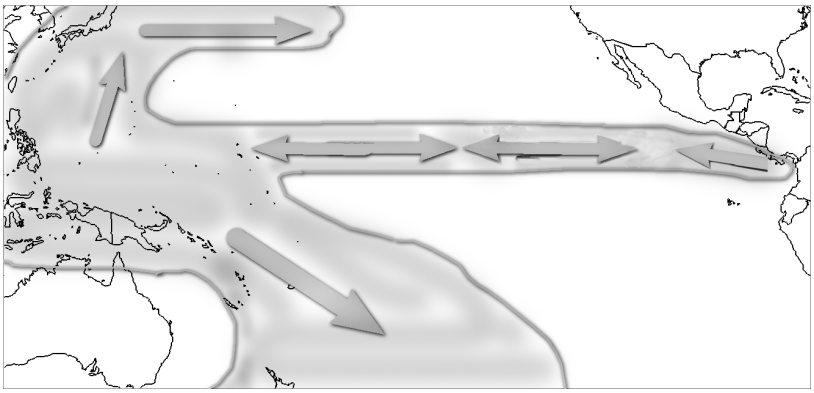


Figure 3.19: Schematic figure of movement for objects with longer lifespan (based on data from year 2001).

precipitation areas with short lifespans predominate. All of the above-mentioned properties are reflected in the results for both datasets. Both datasets also show a trail of objects moving along the western coast of Mexico towards the US border in the JJA period. These objects have a short lifespan and are more visible in the PERSIANN dataset. These objects are also present in the SON period, although their number is considerably smaller.

Figures 3.22 and 3.23 show the object frequencies in the TRMM and PERSIANN datasets, respectively. Although the frequencies in the datasets differ greatly, some properties are clearly captured by both datasets. Notably, the objects with a long lifespan are usually observed in a large area in the western part of the domain and in a smaller area in the eastern part of domain (west of the coast of Mexico). The frequencies are usually higher for objects in the PERSIANN dataset, which might be a result of higher precipitation accumulations in the PERSIANN dataset (Figure 3.4) which causes more objects being identified.

Figures AII-1 to AII-8 shown in APPENDIX II (page 95) show the average seasonal objects frequency for both datasets. The frequency of the objects with longest lifespan is shifting according to the seasons. In the SON and JJA seasons the long-lived objects are mostly located in the region north of Australia while a smaller number can be located in the region close to the west coast of Mexico. In JJA season these objects appear more frequently than in SON season. On the other hand in MAM season the long-lived objects can mainly be found only in the region east of Australia while in DJF season the objects can also be present north-east of Australia and in the northern part of Australia. In MAM season the frequency of these objects is extremely low (usually less than 1%). Although the frequencies are significantly different these properties are clearly captured by both datasets.

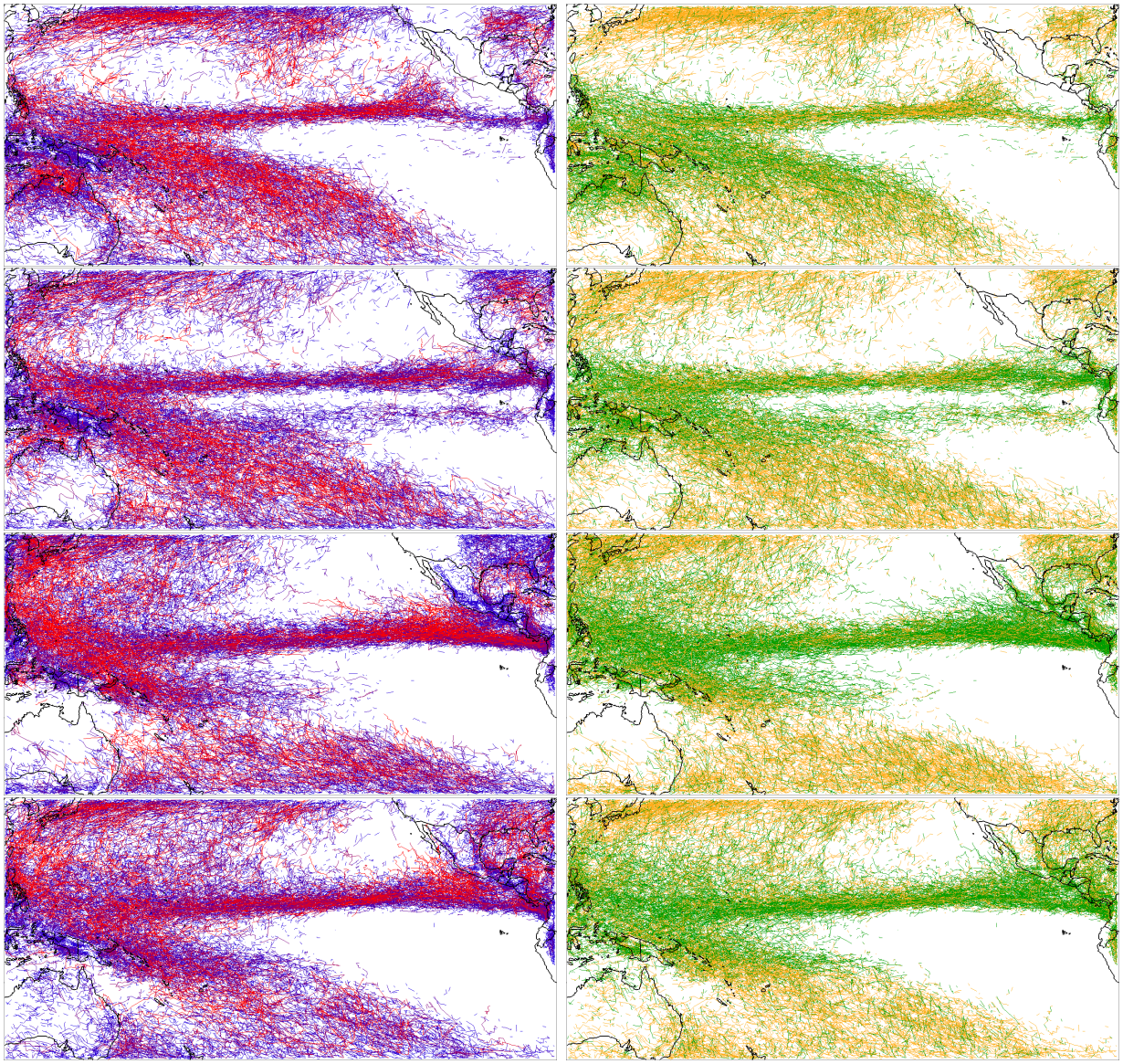


Figure 3.20: Seasonal trajectories for TRMM. Left: color by lifespan (longer lifespan more red), right: color by westward/ easterly direction of movement (green for westward and orange for eastward). From top to bottom: DJF, MAM, JJA and SON.

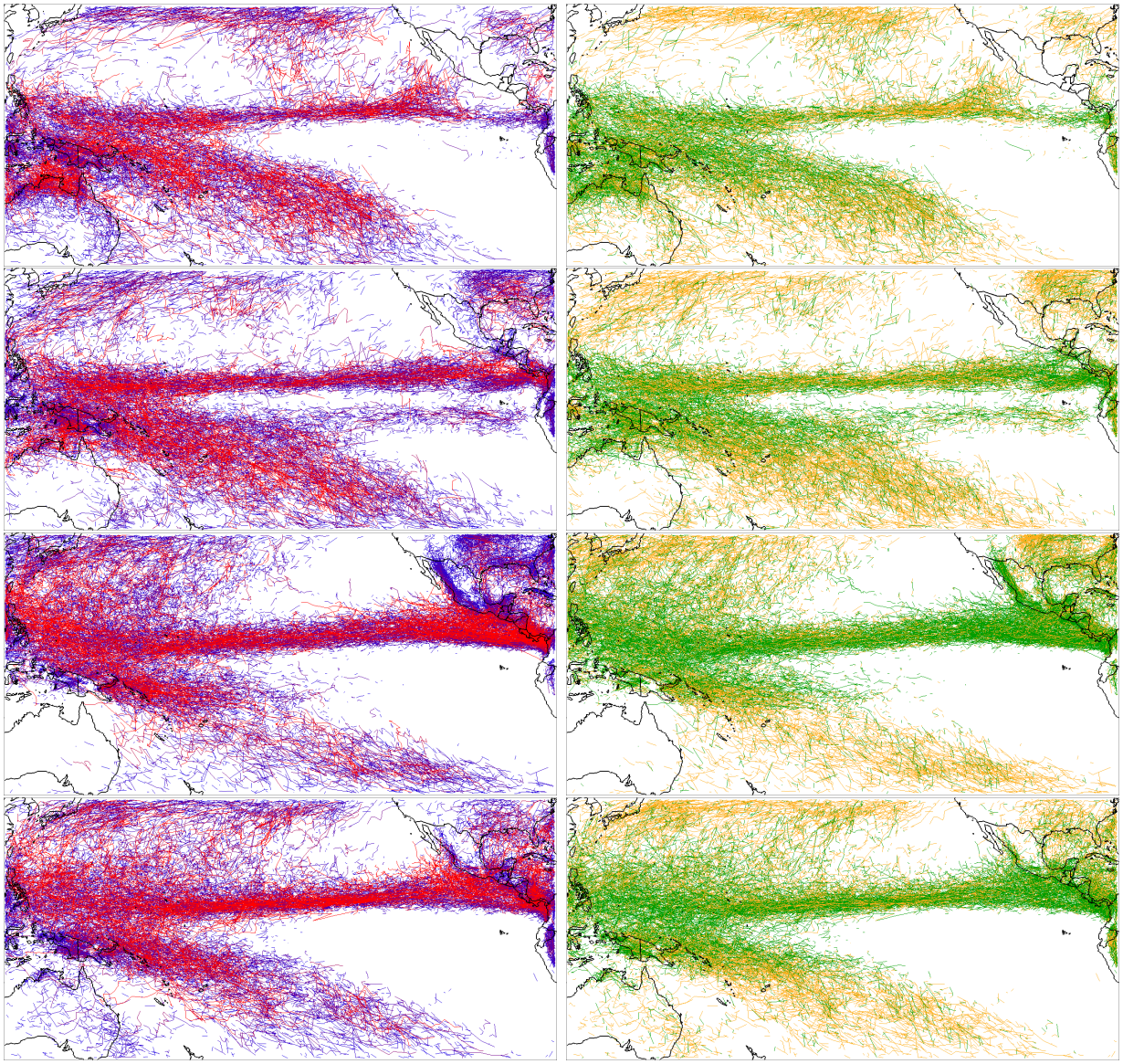


Figure 3.21: Seasonal trajectories for PERSIANN. Left: color by lifespan (longer lifespan more red), right: color by westward/ easterly direction of movement (green for westward and orange for eastward). From top to bottom: DJF, MAM, JJA and SON.

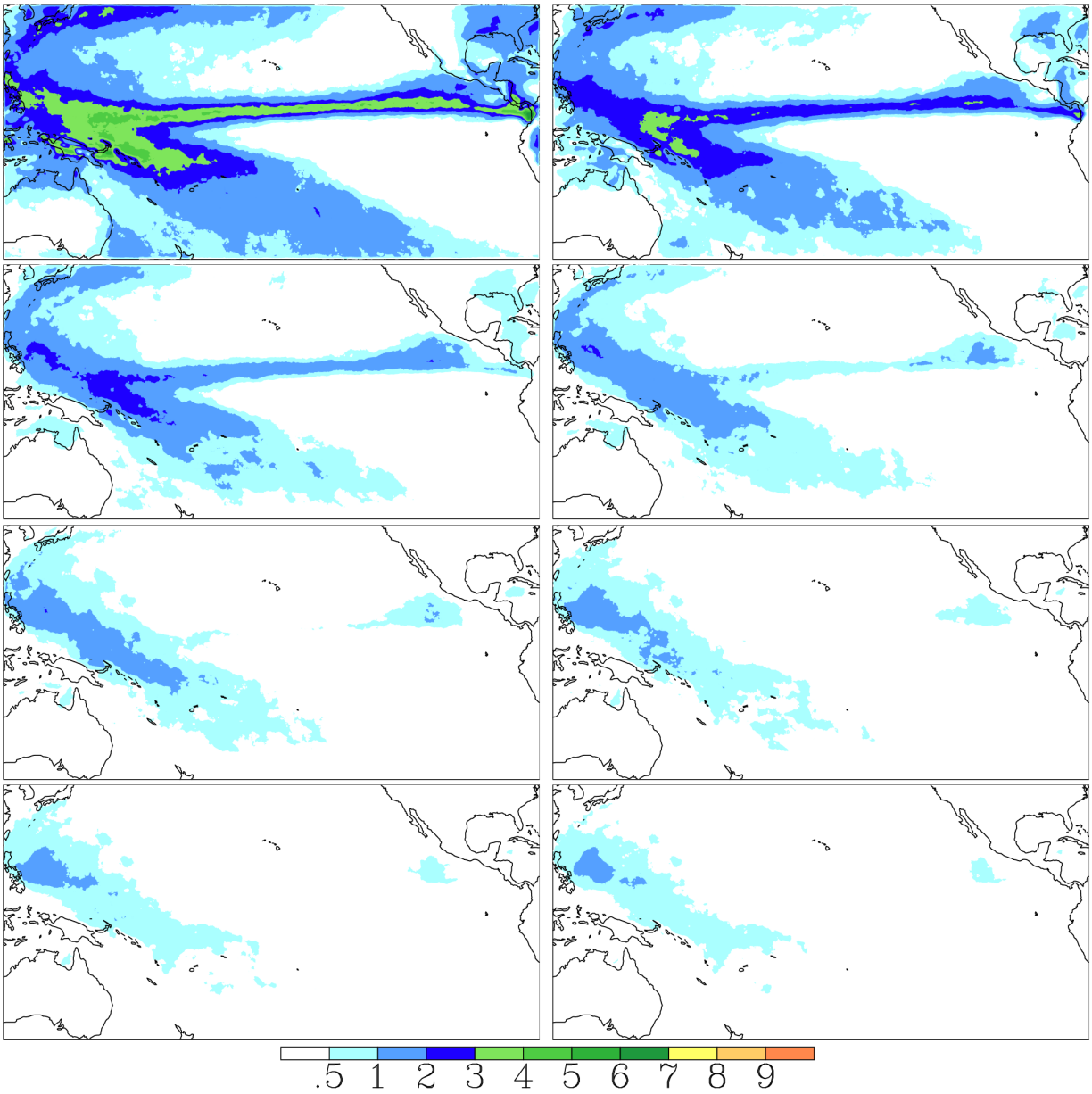


Figure 3.22: Average annual object frequency [%] considering lifespan for TRMM. Top left: all objects, top right: lifespan longer than 15 h. 2nd row left: lifespan longer than 30 h, 2nd row right: longer than 45 h. 3rd row left: lifespan longer than 60 h, 3rd row right: longer than 75 h (right). Bottom left: lifespan longer than 90 h, bottom left: longer than 105 h.

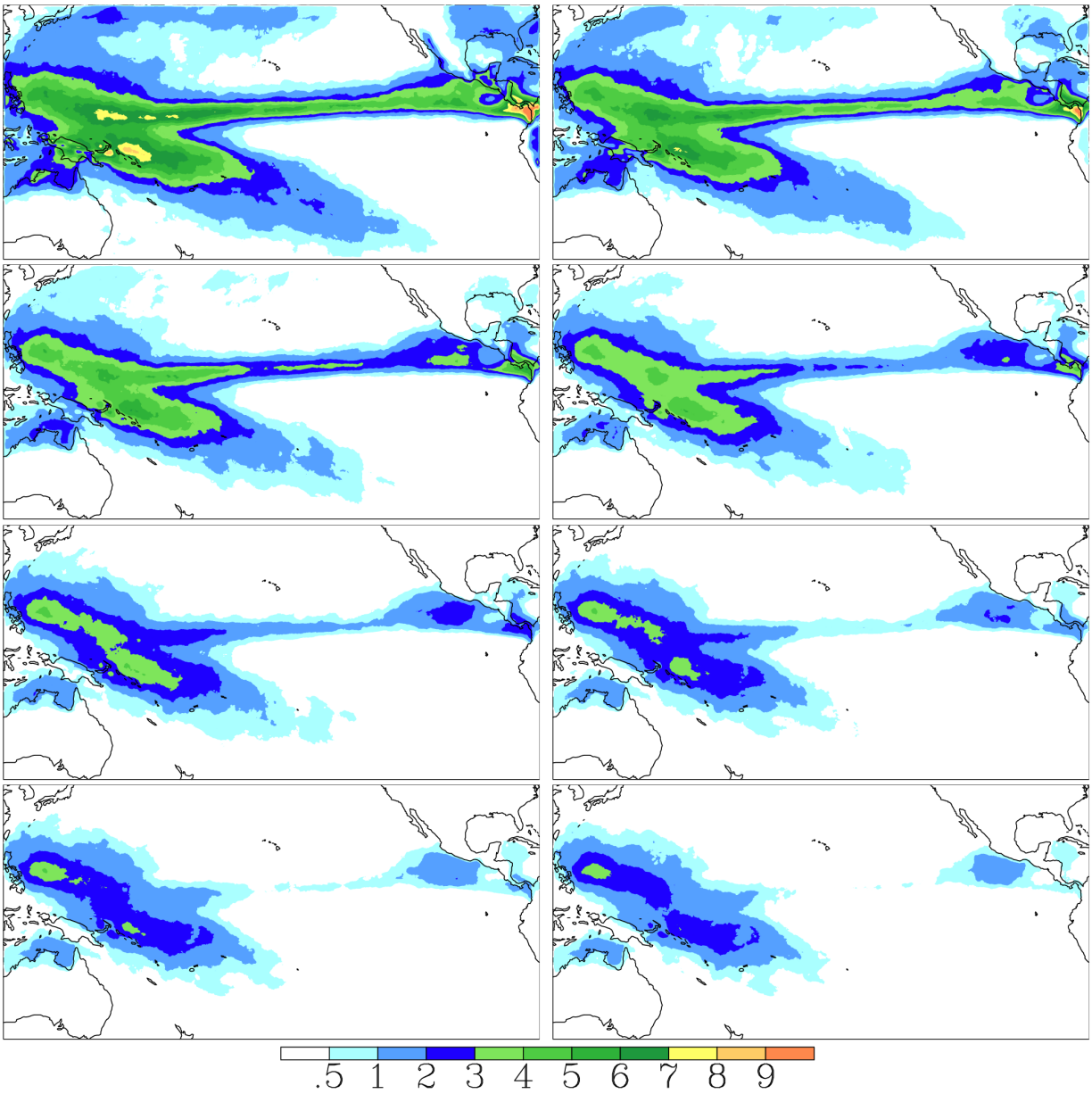


Figure 3.23: Average annual object frequency [%] considering lifespan for PERSIANN. Top left: all objects, top right: lifespan longer than 15 h. 2nd row left: lifespan longer than 30 h, 2nd row right: longer than 45 h. 3rd row left: lifespan longer than 60 h, 3rd row right: longer than 75 h (right). Bottom left: lifespan longer than 90 h, bottom left: longer than 105 h.

4 COMPARISON OF SATELLITE AND MODEL PRECIPITATION

The analysis and comparison of the TRMM and PERSIANN satellite precipitation datasets, which was presented in Chapter 3, enabled us to identify some important characteristics of the precipitation field as observed by the satellites. However, our motivation was also to make a comparison of measured precipitation (derived from satellite data) with precipitation provided by a model. In this way the object-based method developed in this thesis (presented on page 16 in Section 2.2) would be used as a verification tool on precipitation derived by a model simulation. We decided to verify the precipitation provided by WRF model Tropical Channel Simulation (performed by NCAR) by comparing it to the TRMM dataset. The WRF model Tropical Channel Simulation is a 5-year, high resolution, climate simulation of the tropics and surrounding regions. The precipitation data from the WRF simulation was provided by NCAR.

The WRF model and the WRF Tropical Channel Simulation are presented, missing data for the TRMM dataset is studied which is followed by a comparison of annual and seasonal accumulations. Then the sensitivity analysis is performed on a shorter time period, which provides the optimal values for parameters R_{s} and TH, and is followed by a full comparison of the two datasets in the Pacific domain for the 3-year period.

4.1 Overview of the WRF simulation

4.1.1 The WRF Tropical Channel Simulations

The overarching goal of Tropical Channel Simulations performed by NCAR[1], was to improve the understanding and simulation of the complex, 2-way scale interactions that are critical to climate and weather predictions. The main objectives were to improve downscaling from global climate simulations for accurate regional predictions and to improve upscaling of regional processes in global climate simulations.

In order to accomplish this goals a 5-year simulation (1996-2000) of the tropics was performed using the WRF model version 2.1. The domain encompassed

[1]The results and findings of the WRF Tropical Channel Simulations will be published in a special edition of a meteorological journal

latitudes from 30° S to 45° N and was uninterrupted in east-west direction. The horizontal resolution of the model was 36 km using a Mercator map projection. Early parts of the simulation run had a mix of 150/120 s time steps, while later part only 120 s time steps. In vertical direction 35 hybrid pressure-sigma levels were used. These levels are following terrain close to the surface and are transitioning to pressure levels at the model top. At north and south boundaries the NCER-NCAR reanalyses (6 hours at 2.5 deg) were used using a 10 grid-point buffer zone. At lower boundary the Atmospheric Model Intercomparison Project (AMIP) sea surface temperatures (SST) available in 0.5 degreee resolution and mean monthly vegetation fractions were used.

Furthermore, the model setup included the Comunity Atmoshere model (CAM) radiation scheme (detailed description available for the CAM web-page [2]) initialized each 30 minutes, the WSM-6 microphysics scheme, the Noah land suface model, YSU boundary layer and Kain Fritsch (new Eta) convection parametrization schemes.

4.1.2 A short overview of the WRF atmospheric model

The Weather Research and Forecasting (WRF) Model is a next-generation non-hydrostatic numerical weather prediction (NWP) system designed to serve both operational forecasting and atmospheric research needs. It features multiple dynamical solvers, a 3-dimensional variational (3DVAR) data assimilation system, and a software architecture allowing for computational parallelism and system extensibility. WRF is suitable for a broad spectrum of applications across scales ranging from meters to thousands of kilometers. Since the WRF is also meant for research, there are usually more schemes available for the same purpose. User has to make the choice which one to use beforehand. Detailed information about the model is availabe on the WRF model web-page[3].

The effort to develop WRF has been a collaborative partnership, principally among the National Center for Atmospheric Research (NCAR), the National Oceanic and Atmospheric Administration (the National Centers for Environmental Prediction (NCEP) and the Forecast Systems Laboratory (FSL), the Air Force Weather Agency (AFWA), the Naval Research Laboratory, Oklahoma University, and the Federal Aviation Administration (FAA). WRF allows researchers the ability to conduct simulations reflecting either real data or idealized configurations. WRF provides operational forecasting a model that is flexible and efficient computationally, while offering the advances in physics, numerics, and data assimilation contributed by the research community.

The current WRF software framework (WSF) supports two dynamical solvers: the Advanced Research WRF (ARW) (Skamarock et al., 2005) developed and maintained by the Mesoscale and Microscale Meteorology Division of NCAR, and the Nonhydrostatic Mesoscale Model (NMM) developed by the National Centers

[2]http://www.ccsm.ucar.edu/models/atm-cam/index.html
[3]www.wrf-model.org

for Environmental Prediction with user support provided by the Developmental Testbed Center. Both dynamical solvers are non-hydrostatic.

The dynamical solvers are separated from the physical package, which is shared by both solvers. The link between the dynamical solvers and physical package is the Standard Physics Interface, which is different for each solver. A list of major features of WRF-ARW (Skamarock et al., 2005) is summarized below:

- Vertical Coordinate: Terrain-following hydrostatic-pressure, with vertical grid stretching permitted. Top of the model is a constant pressure surface.
- Horizontal Grid: Arakawa C-grid staggering.
- Prognostic Variables: Velocity components u and v in Cartesian coordinate, vertical velocity w, perturbation potential temperature, perturbation geopotential, and perturbation surface pressure of dry air. Optionally, turbulent kinetic energy and any number of scalars such as water vapor mixing ratio, rain/snow mixing ratio, and cloud water/ice mixing ratio are prognostic variables as well.
- Equations: Fully compressible, Euler nonhydrostatic with a run-time hydrostatic option available. Conservative for scalar variables.
- Time Integration: Time-split integration using a 3rd order Runge-Kutta scheme with smaller time step for acoustic and gravity-wave modes.
- Spatial Discretization: 2nd to 6th order advection options in horizontal and vertical.
- Turbulent Mixing and Model Filters: Sub-grid scale turbulence formulation in both coordinate and physical space. Divergence damping, external-mode filtering, vertically implicit acoustic step off-centering. Explicit filter option is also available.
- Initial Conditions: Three dimensional for real-data, and one-, two- and three-dimensional using idealized data. A number of test cases are provided.
- Lateral Boundary Conditions: Periodic, open, symmetric, or specified options available.
- Top Boundary Conditions: Gravity wave absorbing (diffusion or Rayleigh damping). w = 0 top boundary condition at constant pressure level.
- Bottom Boundary Conditions: Physical or free-slip.
- Earth's Rotation: Full Coriolis terms included.
- Mapping to Sphere: Three map projections are supported for real-data simulation: polar stereographic, Lambert-conformal, and Mercator. Curvature terms included.
- Nesting: One-way, two-way, and moving nests.

- Microphysics: Bulk schemes ranging from simplified physics suitable for mesoscale modeling to sophisticated mixed-phase physics suitable for cloud-resolving modeling.
- Cumulus parameterizations: Adjustment and mass-flux schemes for mesoscale modeling including NWP.
- Surface physics: Multi-layer soil and land surface models ranging from a simple thermal model to full vegetation and soil moisture models, including snow cover and sea ice.
- Planetary boundary layer physics: Turbulent kinetic energy prediction or non-local K schemes.
- Atmospheric radiation physics: Longwave and shortwave schemes with multiple spectral bands and a simple shortwave scheme. Cloud effects and surface fluxes are included.

The WRF physics options fall into several categories, each containing several options. The physics categories are microphysics, cumulus parameterization, planetary boundary layer, landsurface model, and radiation.

The physics section is insulated from the rest of the dynamics solver by the use of physics drivers. These are between solver-dependent routines: a prephysics preparation and postphysics modifications of the tendencies. The physics preparation involves filling arrays with physics-required variables that include the temperature, pressure, heights, layer thickness, and other state variables in MKGS (Meter Kilogram Second) units at half-level grid points and on full levels. The velocities are also de-staggered so that the physics part is independent of the dynamical solver's velocity staggering. Physics packages compute tendencies for the velocity components, potential temperature, and moisture fields. The solver-dependent post-physics step will restagger these tendencies as necessary, couple tendencies with coordinate metrics, and convert to variables or units appropriate to the dynamics solver.

In the WRF-ARW solver in the first the Runge-Kutta step, prior to the acoustic steps, the tendencies are computed for radiation, surface, planetary boundary layer (PBL), and cumulus physics. These tendencies are then held fixed through the Runge-Kutta steps. Microphysics is computed after the last Runge-Kutta step in order to maintain proper saturation conditions at the end of the time-step.

The initialization of the physics is called prior to the first model step. This initialization may include reading in data files for physics tables or calculating look-up tables of functions. Each physics module includes an initialization routine for this purpose. Often physics packages will have many of their own constants that should also be included in their own module, while common physical constants are passed in from the physics drivers.

Since our main facus in this study was the precipitation, the precipitation schemes used in the WRF Tropical Channel Simulation will be presented in more detail in Appendix III.

Figure 4.1: Pacific domain used for comparing WRF simulation with TRMM dataset.

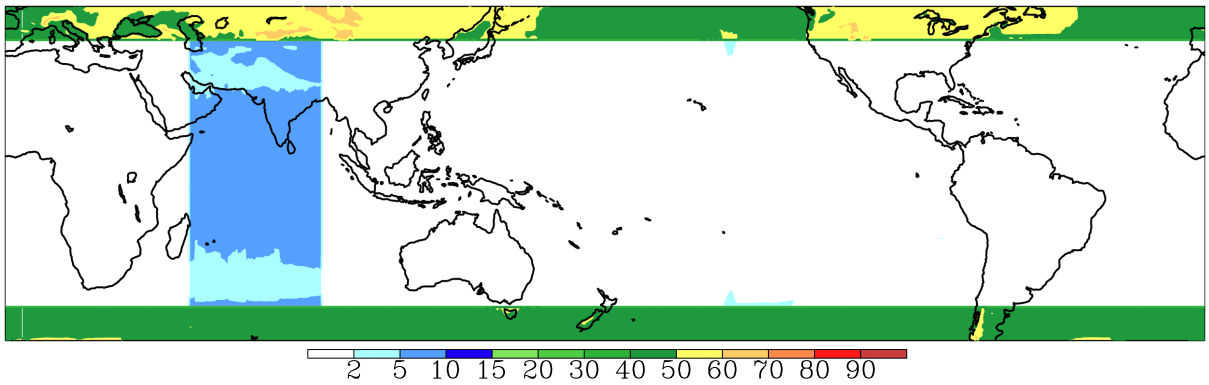


Figure 4.2: Percentage of missing data [%] for TRMM dataset over the whole time period from 1st January 1998 to 31st December 2000.

4.2 Domain, time period and missing data

A domain in tropical Pacific, similar to the one used in previous chapter, was chosen for comparison of precipitation datasets, the TRMM dataset and WRF simulation. However since the WRF simulation was only performed in region from 30° S to 45° N, the new domain encompasses somewhat smaller area compared to the Pacific domain used in previous chapter (120° E – 285° E and 30° S – 39° N and shown in Figure 4.1). The time period for comparison consists of only 3-years of data when both datasets are available: 1st January 1998 – 31th December 2000 (TRMM satellite data is not available before 1998 and WRF simulation was not performed for years after 2000). The domain consists of 661 × 273 for TRMM and 511 × 225 grid points for WRF in 8768 three-hourly time intervals.

Figure 4.2 shows the missing data in the TRMM dataset. The availability of data in the Pacific domain was very good, never falling below 95%.

4.3 Annual and seasonal precipitation accumulations

The average annual precipitation accumulations for TRMM dataset and WRF simulation can be seen in Figure 4.3. The overall spatial distribution of annual mean precipitation (i.e., the clear shape of the ITCZ with two maxima, one located in the east and one in the west) is similar for both datasets but there are also some considerable differences. The largest difference is that there is considerably more precipitation in the WRF simulation compared to TRMM dataset. Some of the other differences (marked with numbers in Figure 4.3) are; 1) less precipitation in WRF simulation over Indonesian island of Sulawesi, 2) more precipitation in WRF simulation in Gulf of Carpentaria, 3) some grid points with extreme precipitation over island of Papua in the WRF simulation, 4) more precipitation in WRF simulation in Philippine Sea, 5) less precipitation in WRF simulation in the middle of ITCZ, 6) less precipitation in TRMM dataset in Gulf of Honduras, 7) more precipitation over continental central America near Costa Rica in the WRF simulation.

The precipiation in WRF simulation and the TRMM dataset is also compared to the precipitation provided by GPCP. Figure 4.3 bottom shows the GPCP average annual accumulation. Although the resolution is lower than in the WRF simulation and the TRMM dataset, the precipitation is clearly much more similar to the TRMM dataset than to the WRF simulation. This may not be very surprising since the TRMM dataset uses similar satellite data as input. Similar to the TRMM dataset the GPCP precipiation accumulations are lower compared to the precipiation accumulations in the WRF simulation.

The yearly precipitation accumulations can be seen in Figure 4.4. The precipitation pattern in year 1998 is somewhat different than in years 1999 and 2000. Year 1998 had more precipitation around the equator, especially in the eastern part of the domain. The difference in the precipitation pattern can be linked to the fact that the period 1997-98 had a strong El Niño while period 1998-99 had a moderate La Niña (according to Oceanic Nino Index available on NOAA web-page[4]). The difference between years 1998 and 1999-2000 can be observed in both datasets. However there is also considerable difference between the datasets themselves, especially for accumulation is 1998. For example in WRF simulation the region with an increase in precipitation is more separated from the ICTZ while in TRMM the two are joined in a single region.

The average seasonal precipitation accumulations can be seen in Figure 4.5. Figure shown numerous differences between the datasets for different seasons (for example a lack of in precipitation in the middle of ICTZ in the WRF simulation), however the general shapes of the precipitation distribution are similar.

[4]http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml

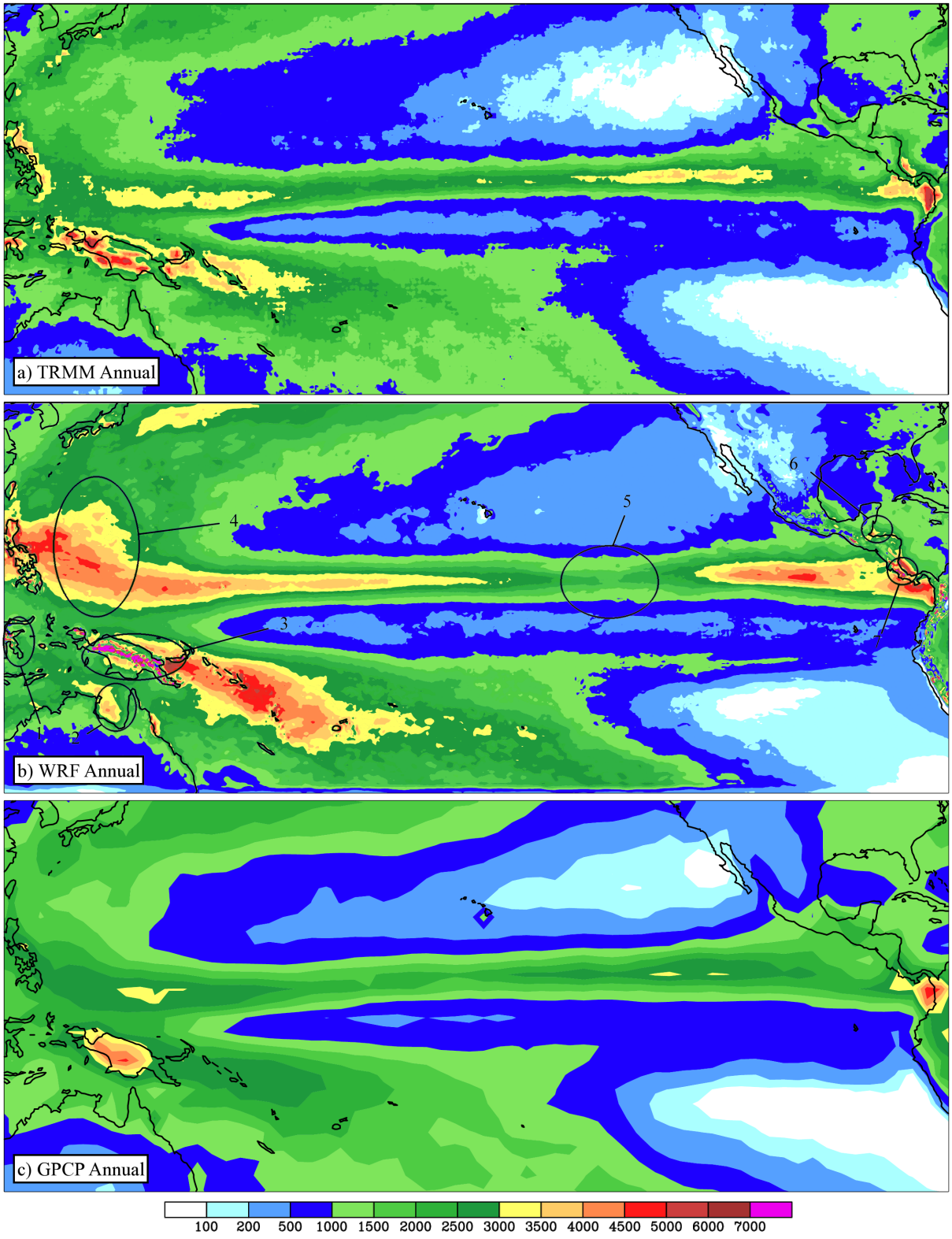


Figure 4.3: Average annual precipitation (mm) for TRMM dataset, WRF simulation and GPCP dataset for the period 1st January 1998 – 31th December 2000.

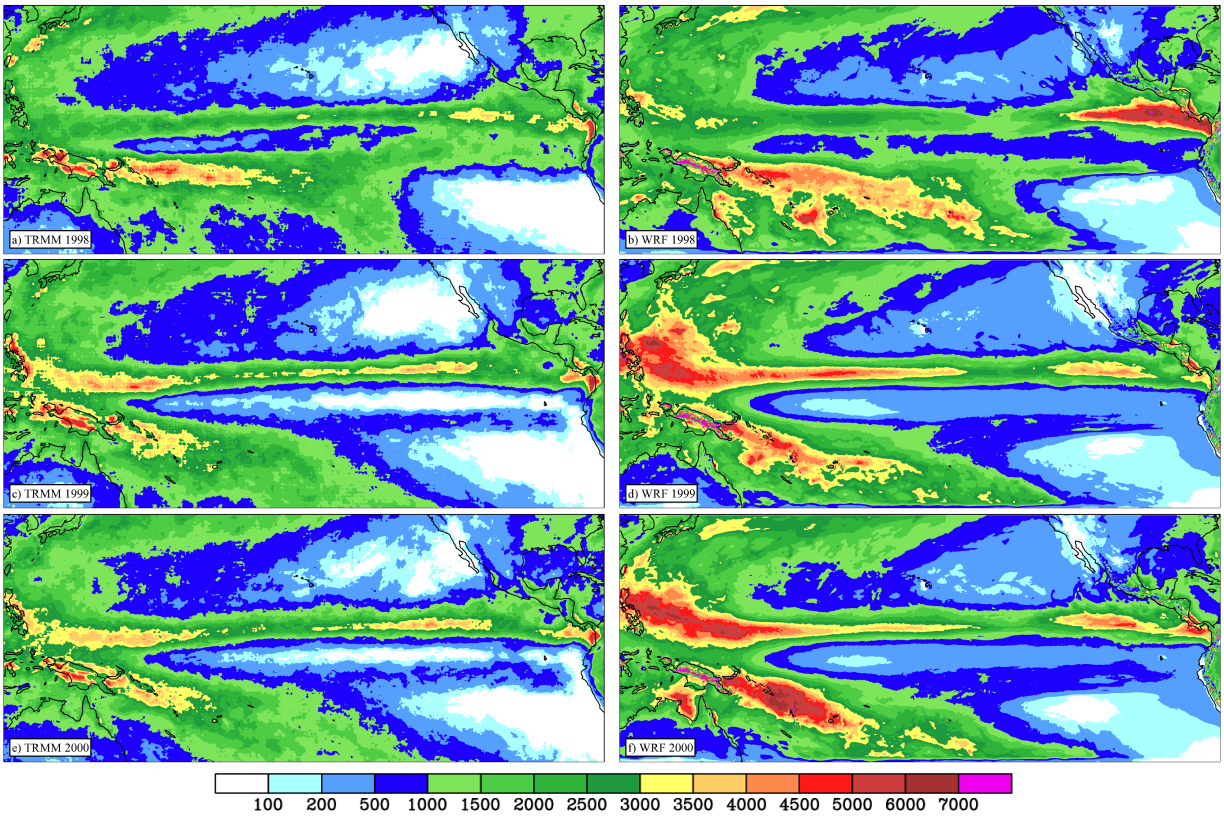


Figure 4.4: Yearly precipitation (mm) for TRMM dataset and WRF simulation.

4.4 Sensitivity analysis

Similar as before the analysis was performed in two steps. The first step used a wide range of parameter values, but due to the computational limitations only a simple propertie was studied - the portion of domain covered by objects. The second step used a smaller range of parameter values where analysis of lifespan was performed. In order to avoid the spill-over effect the temporal convolution was not applied.

4.4.1 Portion of domain covered by objects

The portion of domain covered by objects was analyzed for TRMM datset and WRF simulation using the entire 3-year period. The range of R_s and TH parameters can be seen in Table 4.1.

Figure 4.6 shows that the invariant is still visible at threshold values of 2-3 mm/3h. The WRF simulation has a very large portion of domain covered by objects for low values of precipitation thresholds (almost 50% at $TH = 0$ mm/3h), but when threshold is increased to 1 mm/3h, the portion falls below 10%. This indicates, that in the WRF simulation more than 40% of the domain is covered by

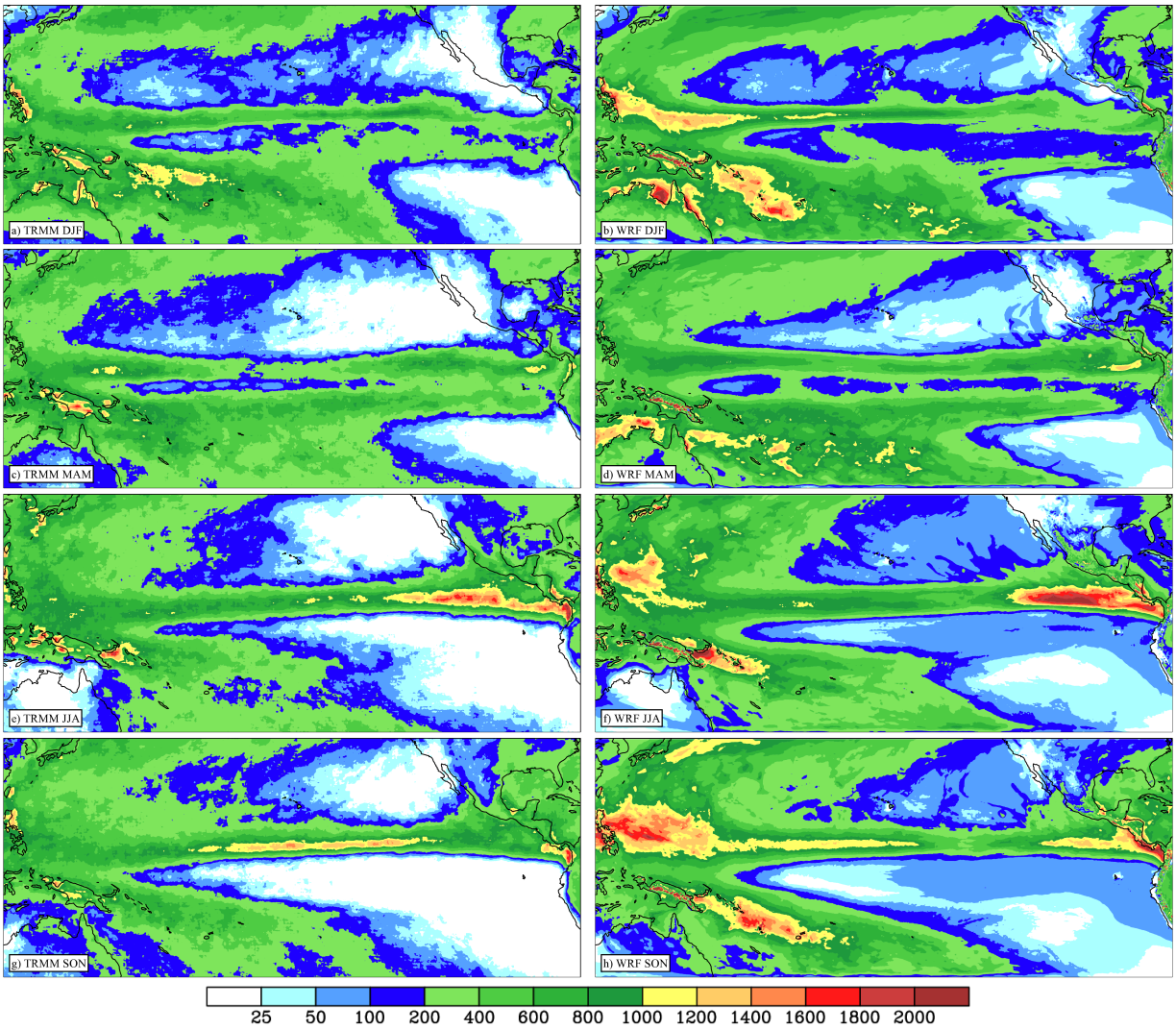


Figure 4.5: Average seasonal (three-month) precipitation (mm) for TRMM and WRF.

Table 4.1: Used values of convolution radius and threshold. All possible combinations of parameter values were analyzed.

Parameter	Values
R_s [deg]	0, 0.35, 0.7, 1.5, 3, 6
TH [mm/3h]	0, 0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1, 2, 5, 10, 20, 50, 100

precipitaion with intensity lower than 1 mm/3h - this is most likely contributed by stratiform precipitation which model treats seperately from convective precipitation. At thresholds greater than the invariant value, the portion of domain covered by objects in WRF simulation is comparable to portion in the TRMM dataset.

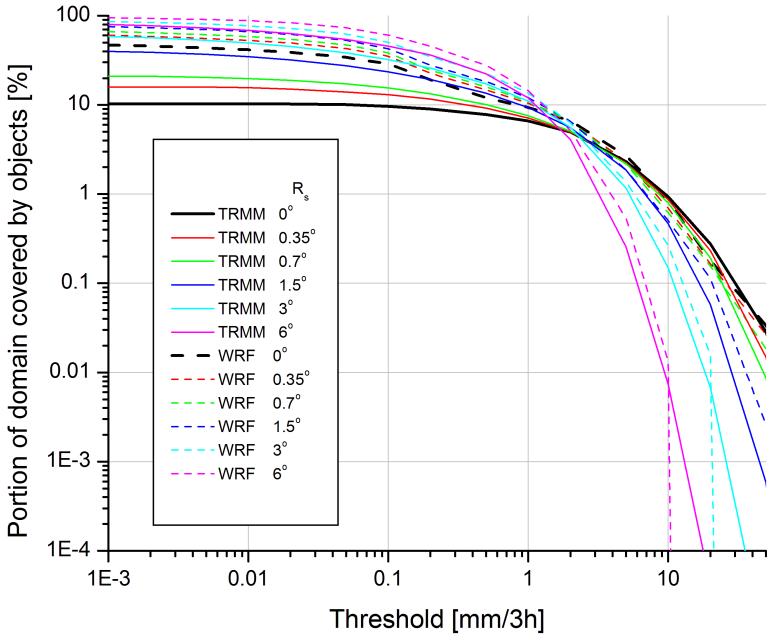


Figure 4.6: The dependence of the portion of domain covered by objects [%] on threshold.

Table 4.2: Used values of convolution radius and threshold. All possible combinations of parameter values were analyzed.

Parameter	Values
R_s [deg]	0, 0.35, 0.7, 1.5, 3, 6
TH [mm/3h]	1, 2, 5, 10, 20, 50, 100

4.4.2 Lifespan analysis

The dependence of number of objects on lifespan was analyzed for both datasets using the entire 3-year period. The range of R_s and TH parameters can be seen in Table 4.2.

Figure 4.7 shows that the dependence of number of objects on lifespan follows the power law in both datasets. In this respect the WRF simulation precipitation is similar to the precipitation in TRMM dataset. At longer lifespans the lines become noisy which can be contributed to small sampling rate of objects with very long lifespan. There is an increase in number of objects which have the longest possible lifespan (lifespan of 270 hours which equals 90 time intervals). This can be attributed to the fact that lifespans longer than 90 time intervals can not be represented well if only last 90 time intervals are kept in memory. Objects with lifespan longer than 90 time intervals are therefore split into two or more objects, where one of them usually has a maximum allowable lifespan.

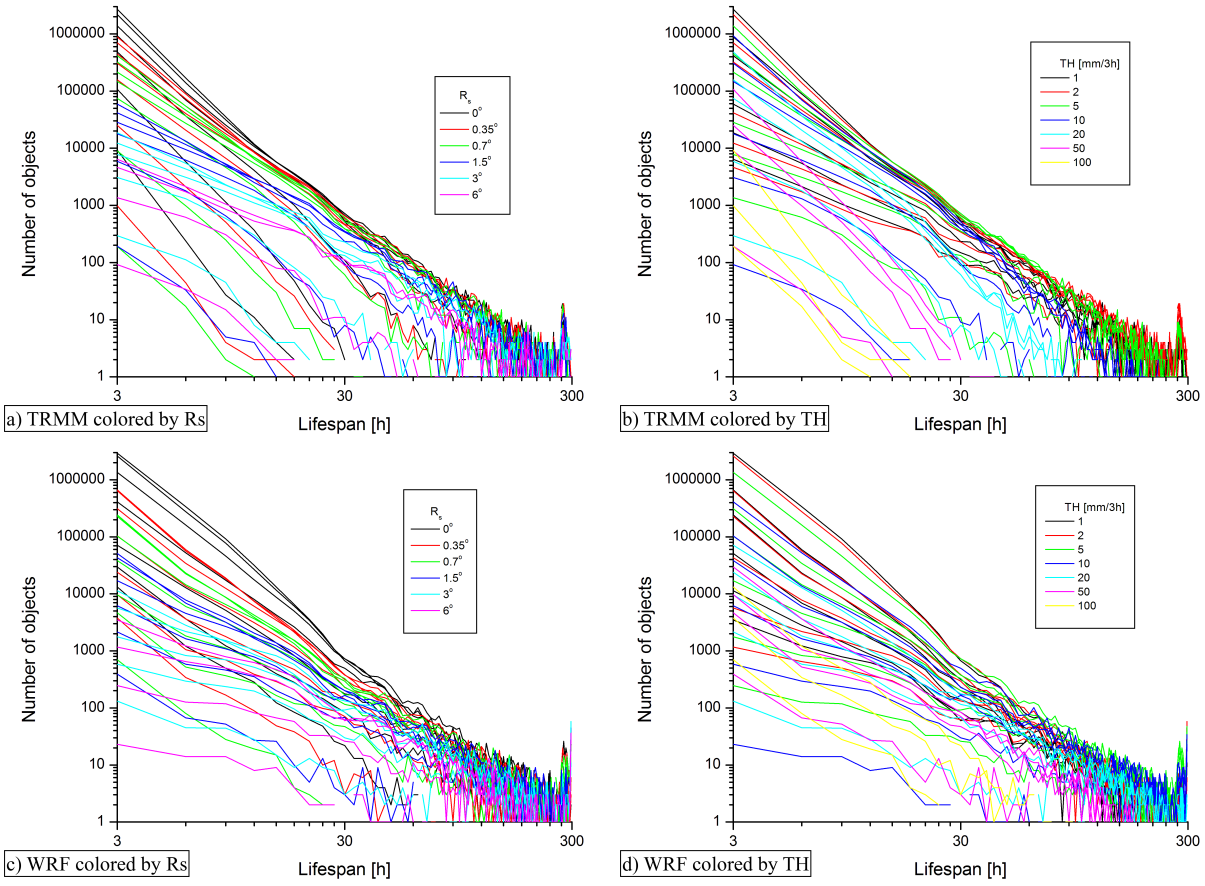


Figure 4.7: Number of objects in period from 1st January 1998 to 31st December 2000, with respect to Lifespan on log-log scale. On the left side the lines are colored with respect to R_s and on the right with respect to threshold.

4.5 Object lifespan and movement in Pacific

As in the previous Chapter, a spatial convolution radius of 0.75 deg and a threshold 7 mm/3h were chosen for the analysis. No temporal convolution was applied in order to avoid any unwanted spillover effect. The entire period from 1st January 1998 to 31st December 2000 was studied and 90 previous time intervals were kept in memory during the processing.

Figure 4.8 shows the number of objects vs. lifespan dependence for both datasets. Both display a power law like dependence and in this respect the precipitation in WRF simulation is similar to the one in TRMM dataset. Also, it seems that TRMM dataset has more shortlived objects (lifespan less than 30 h) and less objects with longer lifespan, when compared to WRF simulation.

Figures 4.9 and 4.10 show trajectories for different years for TRMM dataset and WRF simulation. As was the case with precipitation accumulations, year 1998 is considerably different compared to 1999 and 2000. In general there are more trajectories with longer lifespan (red) in WRF simulation than in TRMM dataset. This is in accordance with results in Figure 4.8. Also in the ITCZ

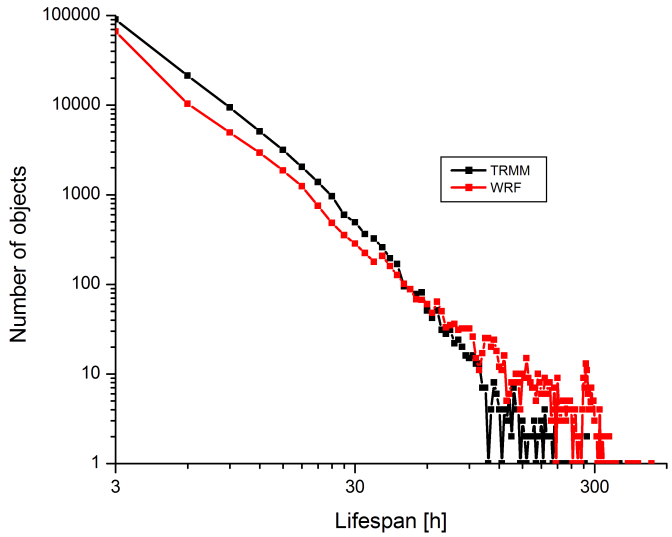


Figure 4.8: Log-log graph of average number of objects/lifespan distribution for TRMM dataset and WRF simulation.

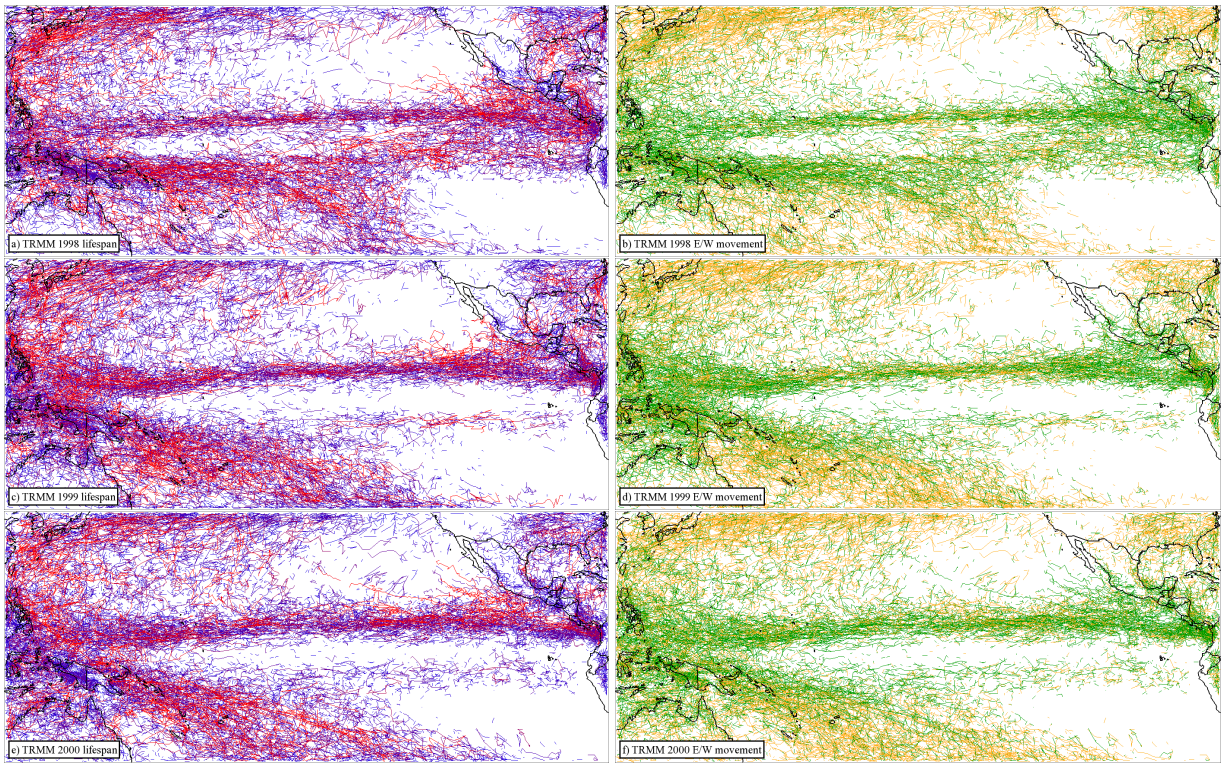


Figure 4.9: Trajectories of objects in different years for TRMM dataset. Left: trajectories are colored with respect to object lifespan - short lifespan (blue), long lifespan (red). Right: Trajectories are colored with respect to the eastward/ westward movements. If movement in a certain time interval has a westward component the trajectory segment will be colored green. If the movement has an eastward component the color is orange.

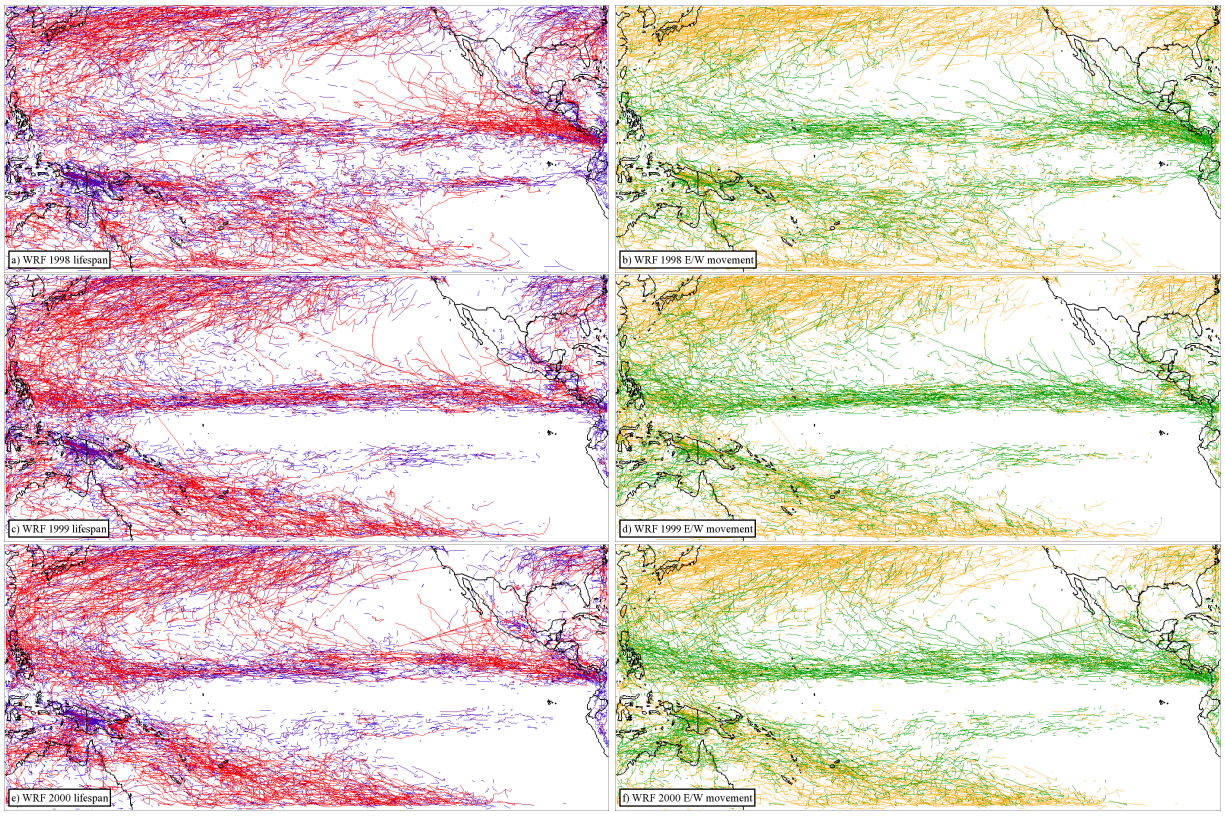


Figure 4.10: Trajectories of objects in different years for WRF simulation. Left: trajectories are colored with respect to object lifespan - short lifespan (blue), long lifespan (red). Right: Trajectories are colored with respect to the eastward/ westward movements. If movement in a certain time interval has a westward component the trajectory segment will be colored green. If the movement has an eastward component the color is orange.

and surrounding region the trajectories of objects in WRF simulation seem more straight compared to trajectories of objects in TRMM dataset. Trajectories of objects in WRF simulation are oftentimes spaning only in east-west direction, while trajectories of objects in the TRMM dataset almost always also include displacements in north-south direction. The movement at the top and bottom of domain is mostly eastward in both datasets, while the movement in the ITCZ is mostly both ways but can vary from year to year and there is also some difference between the datasets (for example the movement for year 2000 is mainly westward in WRF simulation and both ways in TRMM dataset). The seasonal trajectories can be seen in Figures 4.11 and 4.12.

Since the analysis of eastward/westward movement by visual examination of drawn trajectories is sometimes unreliable (for example in the ITCZ where a large number of overlapping trajectories make visual examination hard) we decided to introduce a simple Eastward/Westward movement Index (EWI). EWI is defined as,

$$EWI = \frac{N_{\mathrm{W}}}{N_{\mathrm{W}} + N_{\mathrm{E}}}, \tag{4.1}$$

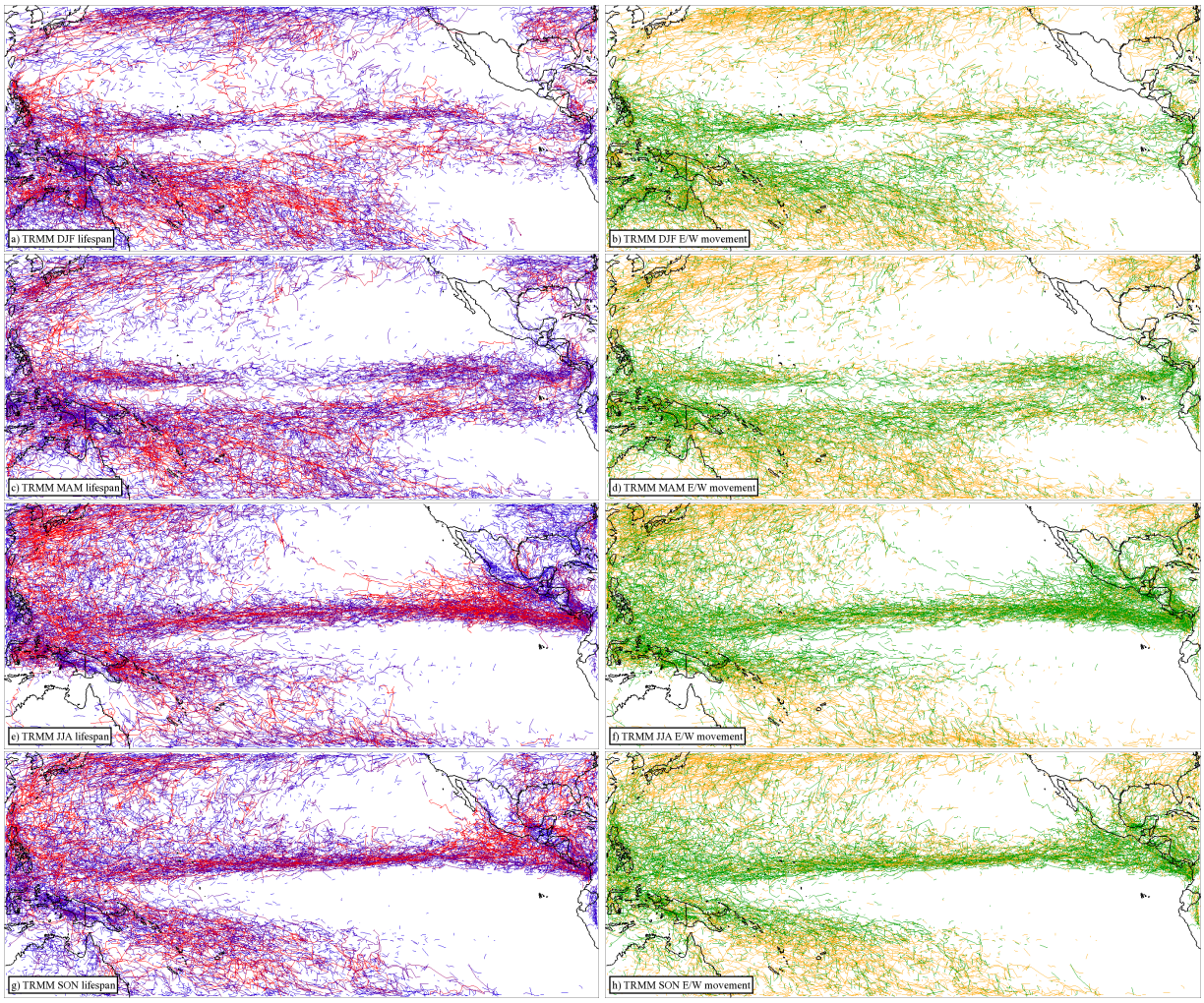


Figure 4.11: Seasonal trajectories for TRMM dataset for period 1. January 1998 – 31. December 2000. Left: color by lifespan (longer lifespan more red), right: color by easterly/ westward direction of movement (green for westward and orange for eastward)

where N_{W} is defined as the number of time intervals in the time period (whole three-year period, year or season) when westward moving object was detected over the grid proint. N_{E} is defined similarly to N_{W} except that it is related to eastward movement. If $EWI = 0$, only eastward movement was detected at the grid point, and if $EWI = 1$ only westward movement was detected. If $EWI = 0.5$ the same number of time intervals was detected with the eastward movement as with the westward movement. EWI is only reliable if there is enough time intervals with object detected at the grid point, therefore we decided to ignore all grid points where object frequency was less than 0.5% (objects were detected in less 0.5% of all time intervals).

Figure 4.13 shows EWI for the whole period and Figure 4.14 for different years. The EWI for the whole period is similar except in the central part of ICTZ. EWI indicates that in the TRMM dataset this region has both eastward and westward movement, while in WRF simulation the movement is clearly dominated

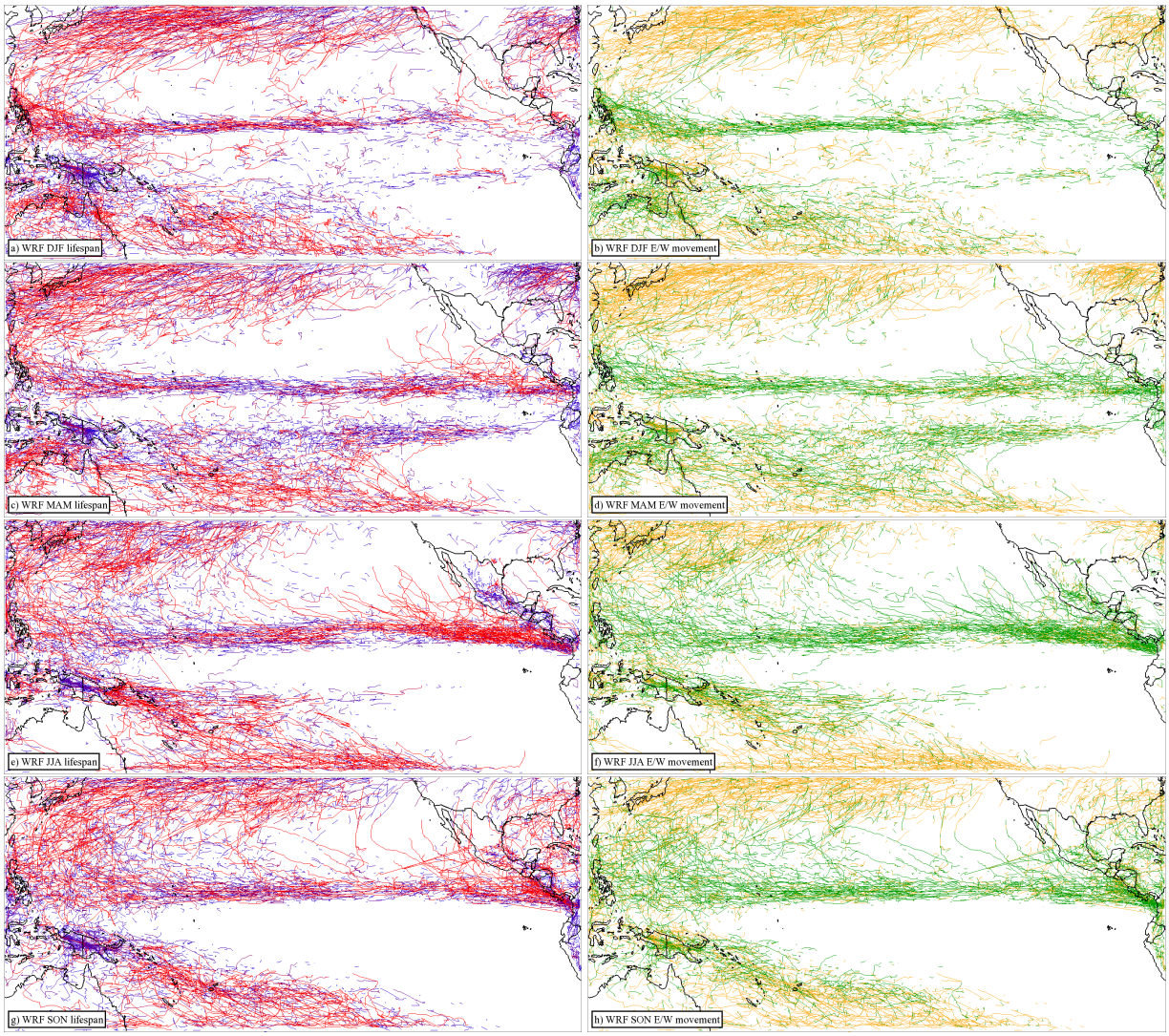


Figure 4.12: Seasonal trajectories for WRF simulation for period 1. January 1998 – 31. December 2000. Left: color by lifespan (longer lifespan more red), right: color by easterly/ westward direction of movement (green for westward and orange for eastward)

by westward moving objects. Similar observation can be made if movement is analyzed for a specific year. Every year there is more westward movement in the WRF dataset compared to TRMM satellite.

Figure 4.15 shows EWI for the different seasons. Similarly as before the biggest difference is observed in the central part of ITCZ where westward movement is dominant in the WRF simulation, in contrast to TRMM dataset where both eastward and westward movements are present. This is especially visible in DJF, SON and partly also in JJA.

Figure 4.16 shows average annual object frequency with regard to lifespan for TRMM dataset and WRF simulation. Frequencies are usually higher for WRF simulation, which is in agreement with previus findings that there is more precipitation and more longlived objects in WRF simulation. Although the frequencies in

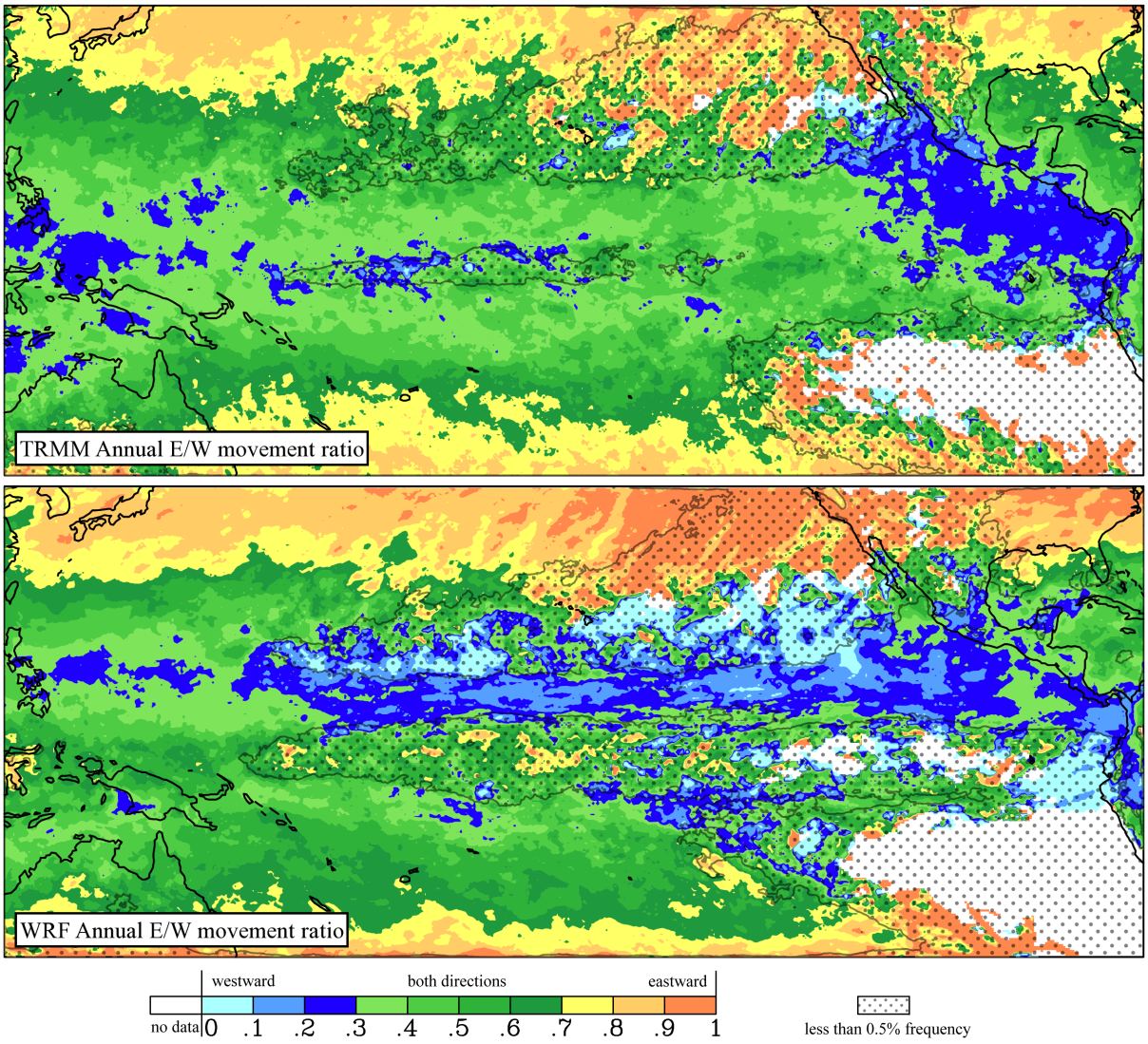


Figure 4.13: The eastward/westward movement index (EWI) for the 1st January 1998 – 31th December 2000 period for TRMM dataset (top) and WRF simulation (bottom). Areas marked with dots represent regions where reliable determination of EWI was not possible because not enough objects were detected (frequency of detected objects was less than 0.5%). Blue regions represent regions with dominant westward movement, red with eastward, and green with both westward and eastward movement.

the datasets differ greatly, some properties are clearly captured by both datasets. Notably, the objects with a long lifespan are usually observed in a two large areas in the western part of the domain and in a smaller area in the eastern part of domain (west of the coast of Mexico).

Figures in APPENDIX V show the yearly and average seasonal objects frequencies. The frequency of the objects with longest lifespan is shifting according to the year and the seasons.

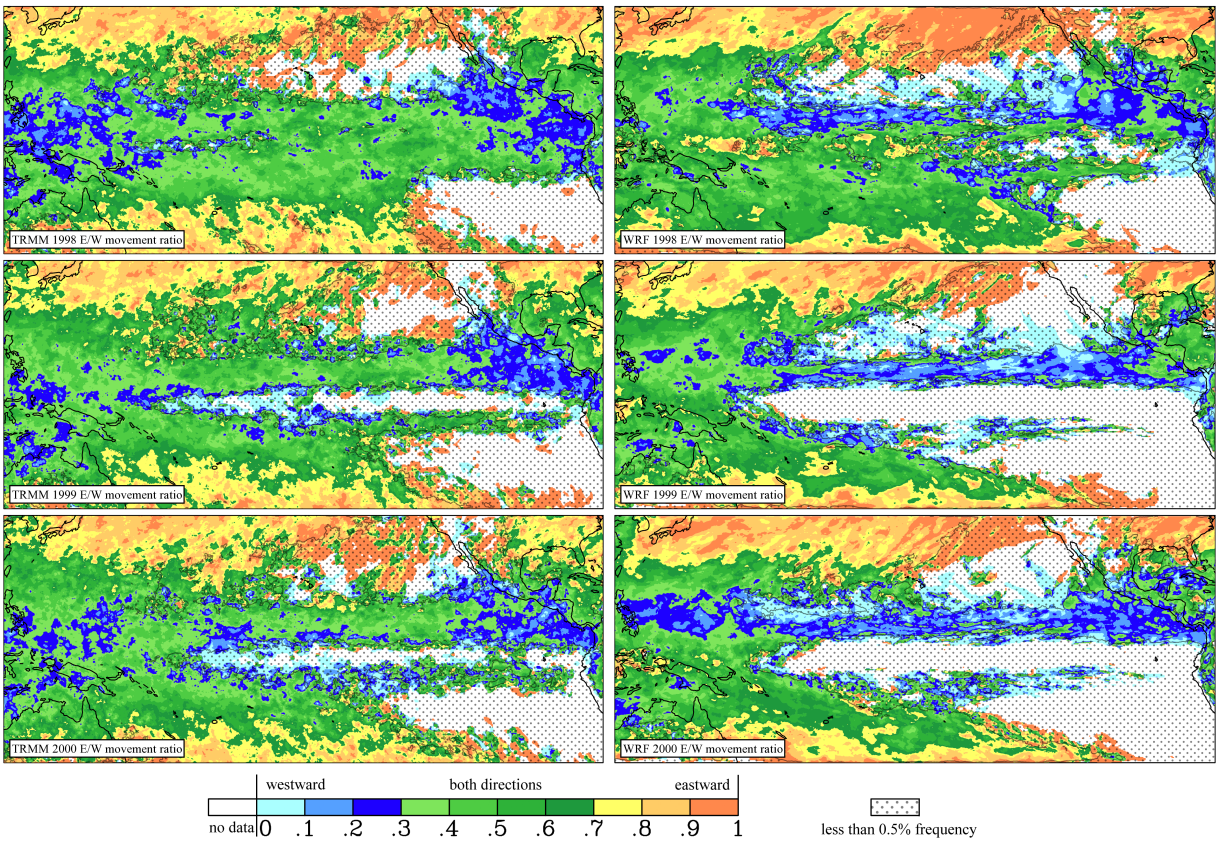


Figure 4.14: The eastward/westward movement index (EWI) for different years for TRMM dataset (left) and WRF simulation (right). Areas marked with dots represent regions where reliable determination of EWI was not possible because not enough objects were detected (frequency of detected objects was less than 0.5%). Blue regions represent regions with dominant westward movement, red with eastward, and green with both westward and eastward movement.

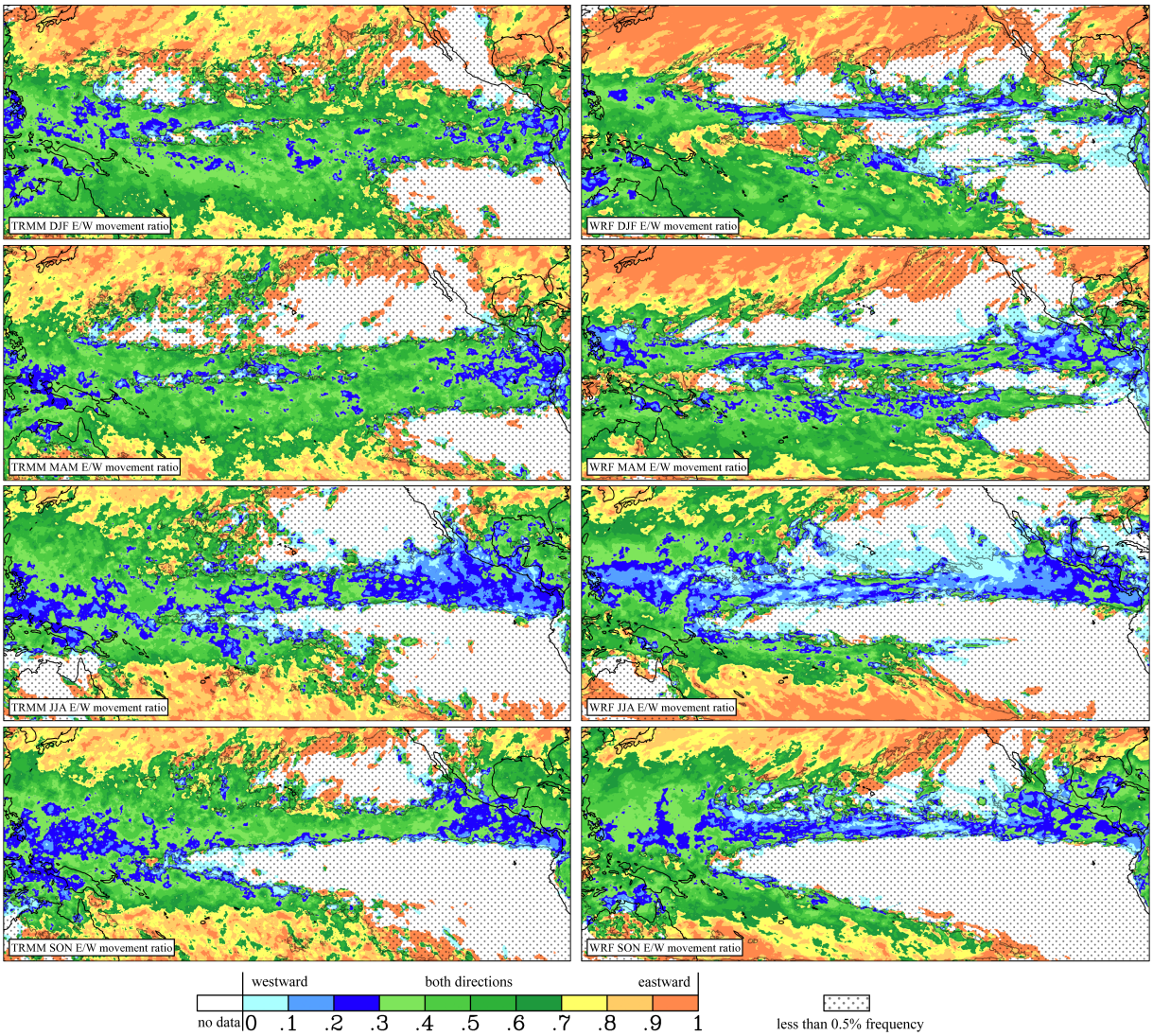


Figure 4.15: The eastward/westward movement index (EWI) for different seasons for TRMM dataset (left) and WRF simulation (right). Areas marked with dots represent regions where reliable determination of EWI was not possible because not enough objects were detected (frequency of detected objects was less than 0.5%). Blue regions represent regions with dominant westward movement, red with eastward, and green with both westward and eastward movement.

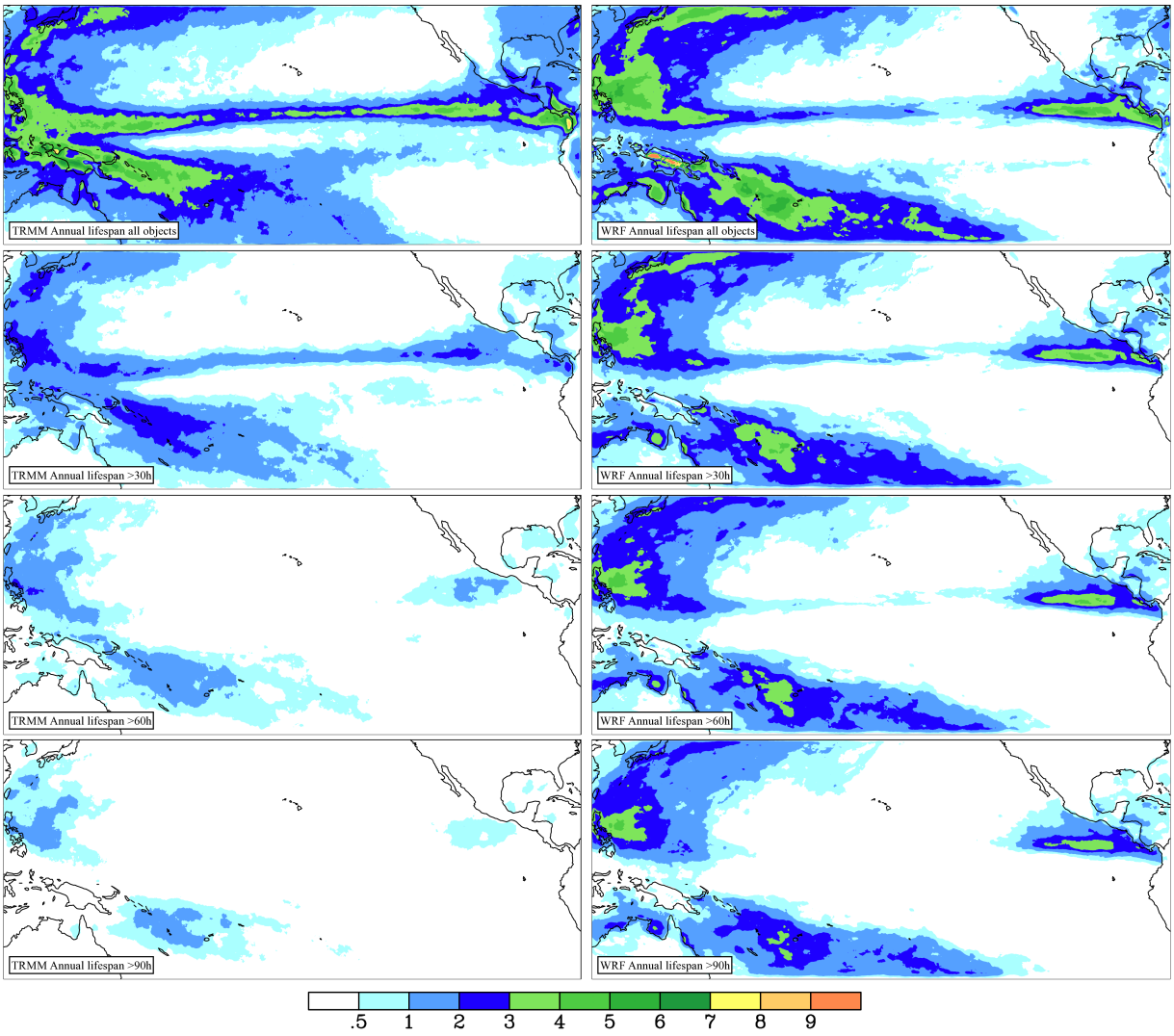


Figure 4.16: Object frequency [%] according to their lifespan for TRMM dataset (left) and WRF simulation (right) in period from 1st January 1998 to 31st December 2000. Top: all objects, 2nd row: lifespan longer than 30h, 3rd row: lifespan longer than 60h and bottom: lifespan longer than 90h.

5 CONCLUSIONS

In the course of this theses we applied the object-oriented analysis method to satellite derived precipitation data in the region of tropical Pacific. The tropics as a whole are play an immensely important role in the global hydrological cycle – it is estimated that two-thirds of global precipitation falls in the Tropics. The region of tropical Pacific is especially interesting to study since it hosts processes such as Madden-Julian oscillation, the Intertropical Convergence Zone and the El Niño-Southern Oscillation.

The object-oriented analysis was firstly done by the basic MODE object-based method developed by Davis et al. (2006a,b). However, since precipitation is a time-evolving process with inherent properties such as lifespan and direction of movement, we extended the basic MODE method to characterize the temporal behavior of objects. The analysis was first performed on two satellite derived datasets with high spatial-temporal resolution, namely TRMM and PERSIANN, for a six year period 2000-2005. Secondly, the object-based method was also used for verifying the WRF model Tropical Channel Simulation precipitation (performed by NCAR) against the TRMM satellite precipitation. The verification was done for a three year period 1998-2000, when the WRF simulation and the TRMM dataset precipitation were simultaneously available.

A sensitivity analysis permitted the choice of optimal threshold and convolution radius parameters. It showed that the use of temporal convolution is not desirable due to an unwanted spill-over effect and that a large spatial convolution radius produces too much smoothing which results in unrealistically large objects. The method enabled a quantitative analysis of the spatial distribution, lifespan, and trajectories of the precipitation features.

The analysis of annual precipitation accumulatios for TRMM and PERSIANN satellite datasets showed that the overall spatial distribution of precipitation (i.e., the clear shape of the ITCZ with two maxima, one located in the east and one in the west) is similar for both datasets, but the amount of precipitation in regions with high precipitation is considerably different. In these regions the amount of precipitation in the PERSIANN dataset seems to be higher by approximately 20-30%. It is not clear why there is such a large difference between the datasets. Perhaps it due to somewhat different input datasets that are used by the two products or due to fact that TRMM precipitation is adjusted with the rain gauge data while PERSIANN precipitation is not, or perhaps due to the differences in algorithms used by the products, or the fact that the products use different ground

validation data. However it is not valid to say that one product performs better than the other. All one might say that the uncertancy in estimating precipitation by satellites is of similar order.

The object-based analysis of the TRMM and PERSIANN satellite datasets indicated that the largest and most long-lived precipitation systems in the tropical Pacific are typically located in the western part while a smaller number of long-lived objects also exist in the northwestern part of the domain and east of Australia. There is considerable difference between the two datasets used in the study in the eastern tip of ITCZ (close to the eastern American coast) – where there are many objects with a long lifespan in the PERSIANN dataset and almost none in the TRMM dataset. The number of objects dependent on their lifespan was found to be a power law. It was also showed that randomly placed short-lived storms follow an exponential function. Thus, since the relationship between the number of objects and object lifespan follows a power-law function rather than an exponential function, the storms represented in the datasets are not simply randomly placed short-lived storms.

Movement of precipitation systems in the ITCZ is both westward and eastward, although westward movement is more frequent and in the eastern part of the Pacific ITCZ the westward movement is dominant. Movement of systems in the northern and southern parts of domain is predominantly eastward. The distribution of seasonal trajectories is consistent with the distributions of seasonal precipitation accumulations. In the northern and southern parts of the domain, the movement is predominantly eastward throughout the year. The direction of movement in the ITCZ is somewhat dependent on season — in the DJF period (and in smaller proportion in MAM) the movement is more eastward which can not be said for the JJA and SON periods.

In the MAM season in the ITCZ, there is considerably less precipitation and the precipitation systems have shorter lifespans compared to other seasons. The secondary ITCZ, which manifests itself south of the ITCZ during the MAM season, mainly contains precipitation areas with short lifespans. During the DJF season precipitation areas with short lifespans are predominant in the eastern part of the ICTZ region. These findings were common to both satellite products, despite the fact that the average rainfall accumulations can differ by 20–30% between the two datasets and the frequency of occurrence of systems with long lifespans can differ by 20%.

TRMM precipitation was also compared to precipitation provided by the WRF Tropical Channel Simulation[1]. The comparison, which was done for a three year period from 1998 to 2000, showed that the overall spatial distribution of annual mean precipitation (i.e., the clear shape of the ITCZ with two maxima, one located in the east and one in the west) is reproduced by the model but there are also numerable differences - the largest being that there is considerably more precipitation in the WRF simulation compared to the TRMM dataset. It is not clear why there is such a large difference between the precipitation accumulations.

[1]performed and provided by NCAR

It would had been interesting to compare WRF simulation to PERSIANN and see if PERSIANN is more similar to WRF since PERSIANN also produced more precipitation than TRMM, however the PERSIANN data is not available before year 2000 and therefore a proper comparison is not possible. There were also differences between datasets for yearly accumulations, especially for 1998 when strong El Niño was present (for example in WRF simulation the secondary ICTZ is separated from the primary ICTZ while in the TRMM dataset the two are joined in the eastern half of domain).

Sensitivity analysis was also performed, which showed that in the WRF simulation more than 40% of the domain is covered by weak precipitation with intensity lower than 1 mm/3h – this is most likely contributed by stratiform precipitation which model treats separately from convective precipitation. At thresholds greater than the invariant value of 2-3 mm/h, the portion of domain covered by objects in the WRF simulation is comparable to the portion in the TRMM dataset. The objects vs. lifespan dependence follows the power law in both datasets.

The object-based analysis showed that more shortlived objects (lifespan less than 30 h) and less objects with longer lifespan are present in the TRMM dataset compared to WRF simulation. The analysis of trajectories also showed that there are more trajectories with longer lifespan in the WRF simulation than in the TRMM dataset, which could be the consequence of higher accumulation in the WRF simulation. Also in the ITCZ and surrounding region the trajectories in WRF simulation seem straighter compared to trajectories in the TRMM dataset. Trajectories in WRF simulation are often spanning only in east-west direction, while the trajectories in the TRMM dataset almost always also include displacements in north-south direction, which is surprising. The reason for this does not seem to be clear, perhaps the lateral boundaries in the WRF simulation force the precipitation systems to move only in east-west direction, perhaps the real world precipitation systems evolve faster and change their outlines more often, thereby shifting their centers somewhat erratically, while the precipitation characterizations in the WRF model, do not allow such fast change of outlines or perhaps the uncertainty in the satellite measurements cause erratic, almost random, changes in object lines, causing the displacements of object centers.

The movement at the top and bottom of domain is mostly eastward in both datasets, while the movement in the ITCZ is mostly both ways but can vary from year to year and also there is some difference between the datasets. The analysis of datasets using eastward/westward movement index (EWI) further confirmed that the datasets are roughly similar with the exception of the central part of ICTZ where EWI indicated that both eastward and westward movements are present in the TRMM dataset, while westward movement is clearly dominating in the WRF simulation.

Analysis of object frequencies with regard to lifespan showed that frequencies are usually higher in WRF simulation, which is in agreement with previous findings that there is more precipitation and more longlived objects in WRF simulation. Although the frequencies in the datasets differ greatly, some properties are captured by both datasets. Notably, the objects with a long lifespan are usually

observed in a two large areas in the western part of the domain and in a smaller area in the eastern part of domain (west of the coast of Mexico).

6 BIBLIOGRAPHY

Adler, R.F., G.J. Huffman, A. Chang, R. Ferraro, P. Xie, J. Janowiak, B. Rudolf, U. Schneider, S. Curtis, D. Bolvin, A. Gruber, J. Susskind, P. Arkin, E. Nelkin 2003: The Version 2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979-Present). *J. Hydrometeor.*, **4**,1147-1167.

Arkin, P. A., and B. N. Meisner, 1987: The relationship between large-scale convective rainfall and cold cloud over the Western Hemisphere during 1982–84. *Mon. Wea. Rev.*, **115**, 51–74.

Bechtold, P., E. Bazile, F. Guichard, P. Mascart, and E. Richard, 2001: A mass-flux convection scheme for regional and global models. *Quart. J. Roy. Meteor. Soc.*, **127**, 869-886.

Dai A., X. Lin, KL Hsu, 2007: The frequency, intensity, and diurnal cycle of precipitation in surface and satellite observations over low- and mid-latitudes. *Climate Dynamics*, **29** (7-8), 727-744.

Davis, C., B. Brown, and R. Bullock, 2006a: Object-Based Verification of Precipitation Forecasts. Part I: Methodology and Application to Mesoscale Rain Areas. *Mon. Wea. Rev.*, **134**, 1772–1784.

Davis, C., B. Brown, and R. Bullock, 2006b: Object-Based Verification of Precipitation Forecasts. Part II: Application to Convective Rain Systems. *Mon. Wea. Rev.*, **134**, 1785–1795.

Doswell, C. A., R. Davies-Jones, and D. L. Keller, 1990: On summary measures of skill in rare event forecasting based on contingency tables. Wea. Forecasting, 5, 576–585.

Hong, S.-Y., H.-M. H. Juang, and Q. Zhao, 1998: Implementation of prognostic cloud scheme for a regional spectral model, *Mon. Wea. Rev.*, **126**, 2621-2639.

Hong, S.-Y., J. Dudhia, and S.-H. Chen, 2004: A Revised Approach to Ice Microphysical Processes for the Bulk Parameterization of Clouds and Precipitation, *Mon. Wea. Rev.*, **132**, 103-120.

Hsu, K., 1996: Rainfall estimation from satellite infrared imagery using artificial neural networks. Ph.D. dissertation, *University of Arizona*, 234 pp. [Available from Dept. of Hydrology and Water Resources, The University of Arizona, Tucson, AZ 85721.]

Hsu, K.l., X. Gao, S. Sorooshian, and H.V. Gupta, 1997: Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks. *J. Appl. Meteor.*, **36**, 1176–1190.

Hsu, K.l., H. V. Gupta, X. Gao, and S. Sorooshian, 1999: Estimation of physical variables from multichannel remotely sensed imagery using neural networks: Application to rainfall estimation. *Water Resour. Res.*, **35 (5)**, 1605–1618.

Huffman, G.J., R. F. Adler, M. Morrissey, D. T. Bolvin, S. Curtis, R. Joyce, B. McGavock, and J. Susskind, 2001: Global precipitation at one-degree daily resolution from multisatellite observations. *J. Hydrometeor.*, **2**, 36–50.

Haddad, Z. S., D. A. Short, S. L. Durden, E. Im, S. Hensley, M. B. Grable, and R. A. Black, 1997a: A new parameterization of the rain drop size distribution. IEEE Trans. *Geosci. Remote Sens.*, **35**, 532–539.

Haddad, Z. S., E. A. Smith, C. D. Kummerow, T. Iguchi, M. R. Farrar, S. L. Durden, M. Alves, and W. S. Olson, 1997b: The TRMM "day-1" radar/radiometer combined rain-profiling algorithm. *J. Meteor. Soc. Japan*, **75**, 799–809.

Huffman, G.J., and Coauthors, 1997: The Global Precipitation Climatology Project (GPCP) combined precipitation dataset. *Bull. Amer. Meteor. Soc.*, **78**, 5–20.

Huffman, G.J., R. F. Adler, M. Morrissey, D. T. Bolvin, S. Curtis, R. Joyce, B. McGavock, and J. Susskind, 2001: Global precipitation at one-degree daily resolution from multisatellite observations. *J. Hydrometeor.*, **2**, 36–50.

Huffman, G.J., R.F. Adler, D.T. Bolvin, G. Gu, E.J. Nelkin, K.P. Bowman, Y. Hong, E.F. Stocker, and D.B. Wolff, 2007: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales. *J. Hydrometeor.*, **8**, 38–55.

Janowiak, J. E., R. J. Joyce, and Y. Yarosh, 2001: A real-time global half-hourly pixel-resolution infrared dataset and its applications. *Bull. Amer. Meteor. Soc.*, **82**, 205–217.

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. Hydrometeor.*, **5**, 487–503.

Kain, J. S., and J. M. Fritsch, 1990: A one-dimensional entraining/ detraining plume model and its application in convective parameterization, *J. Atmos. Sci.*, **47**, 2784-2802.

Kain, J. S., and J. M. Fritsch, 1993: Convective parameterization for mesoscale models: The Kain-Fritcsh scheme, The representation of cumulus convection in numerical models, K. A. Emanuel and D.J. Raymond, Eds., *Amer. Meteor. Soc.*, 246 pp.

Kain, J.S., 2004: The Kain-Fritsch Convective Parameterization: An Update *J. Appl. Meteor.*, **43**, 170-181.

Kummerow, C., Barnes, W., T. Kozu, J. Shiue, and J. Simpson, 1998: The tropical rainfall measuring mission (TRMM) sensor package, *J. Atmos. Oceanic Technol.*, **15**, 809-817.

Laing A. K., and M. G. Fritsch, 1997: The global population of mesoscale convective complexes. *Quart. J. Roy. Meteor. Soc.*, **123**, 389–405.

Lin, Y.-L., R. D. Farley, and H. D. Orville, 1983: Bulk parameterization of the snow field in a cloud model. *J. Climate Appl. Meteor.*, **22**, 1065-1092.

Machado, L. A. T., W. B. Rossow, et al., 1998: Life cycle variations of mesoscale convective systems over the Americas, *Mon. Wea. Rev.*, **126(6)**: 1630-1654.

Madden, R. A., and P. R. Julian, 1971: Detection of a 40-50 day oscillation in the zonal wind in the tropical Pacific. *J. Atmos. Sci.*, **28**, 702-708.

Mahoney, J. L., B. G. Brown, J. E. Hart, and C. Fischer, 2002: Using verification techniques to evaluate differences among convective forecasts. Preprints, *16th Conf. on Probability and Statistics in the Atmospheric Sciences*, Orlando, FL, Amer. Meteor. Soc., 12–19.

Mapes, B.E., and R.A. Houze, 1993: Cloud Clusters and Superclusters over the Oceanic Warm Pool. *Mon. Wea. Rev.*, **121**, 1398–1416.

Murphy, A. H., 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting*, **8**, 281–293.

Press W., Teukolsky S., Vetterling W., Flannery B., 1992: Numerical Recipes in C - *Cambridge University Press*, 994p.

Rudolf, B., 1993: Management and analysis of precipitation data on a routine

basis. *Proceedings of International Symposium on Precipitation and Evaporation*, **Vol. 1**, B. Sevruk and M. Lapin, Eds., Slovak Hydrometeorology Institution, 69–76.

Skamarock, W.C., J.B. Klemp, J. Dudhia, D.O. Gill, D.M. Barker, W.Wang, J.G. Powers, 2005: A Description of the Advanced Research WRF Version 2. NCAR TECHNICAL NOTE. Available at http://wrf-model.org.

Weisstein, Eric W. "Convolution." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/Convolution.html)

Welch, R., K. Kuo, B. Wielicki, S. Sengupta, and L. Parker, 1988: Marine Stratocumulus Cloud Fields off the Coast of Southern California Observed Using LANDSAT Imagery. Part I: Structural Characteristics. *J. Appl. Meteor.*, **27**, 341–362.

Wilcox, E. M., 2003: Spatial and temporal scales of precipitating tropical cloud systems in satellite imagery and the NCAR CCM3, *J. of Climate*, **16(22)**: 3545-3559.

Wilks, D. S., 1995: Statistical Methods in the Atmospheric Sciences. Academic Press, 467 pp.

Xie, P., and P. A. Arkin, 1996: Gauge-based monthly analysis of global land precipitation from 1971 to 1994. *J. Geophys. Res.*, **101**, 19 023–19 034.

Zhang, D.-L., and J.M. Fritsch, 1986: Numerical simulation of the meso-β scale structure and evolution of the 1977 Johnstown flood. Part I: Model description and verification. *J. Atmos. Sci.*, **43**, 1913-1943.

APPENDIX I: Seasonal object frequency considering object size for TRMM and PERSIANN datasets

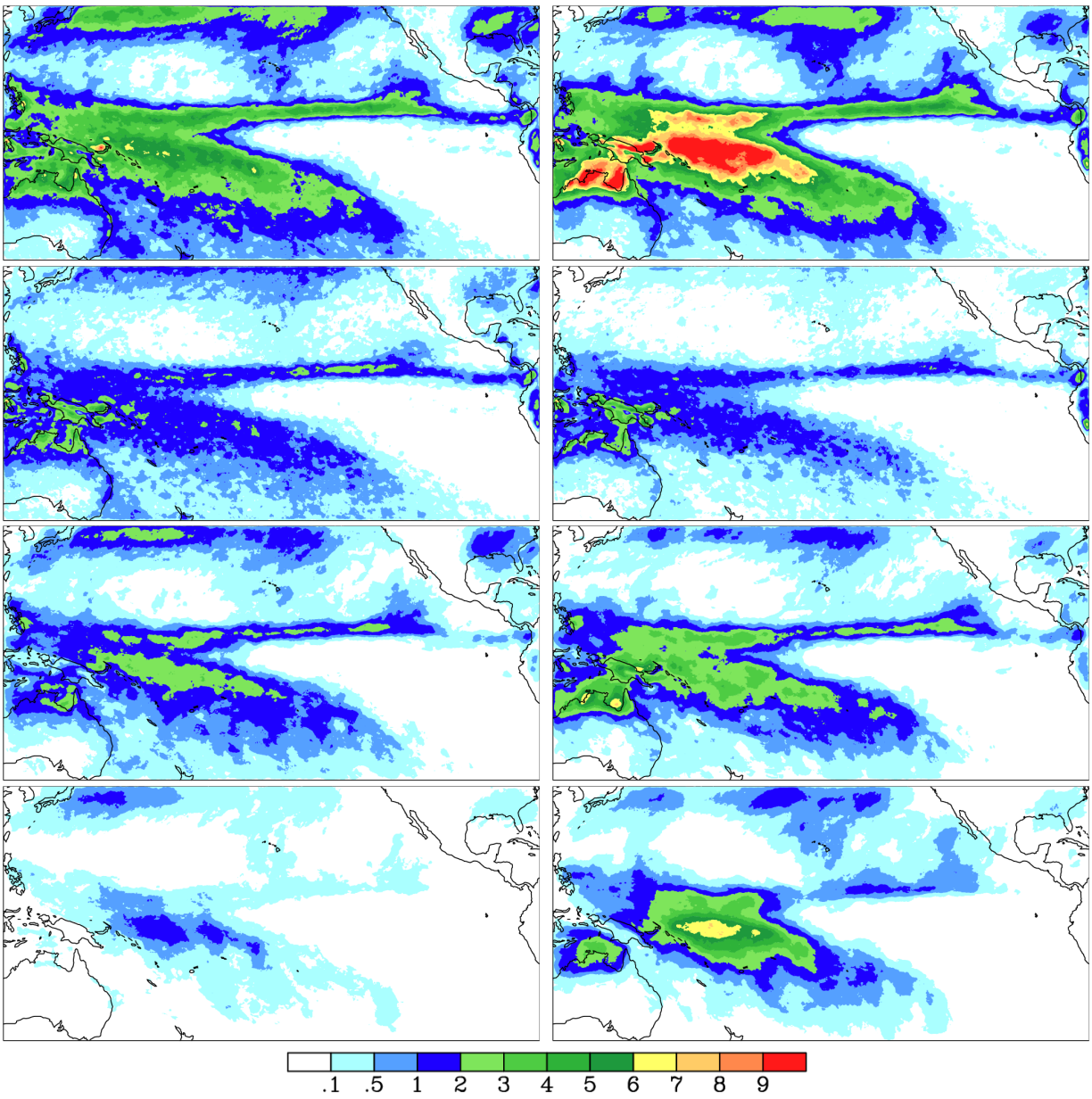


Figure AI-1: DJF object frequency [%] for TRMM (left) and PERSIANN (right): top - all objects regardless of size, 2nd row - objects with area $0 - 87,198$ km^2, 3rd row - 87,198 – 348,792 km^2 and bottom - larger than 348,792 km^2.

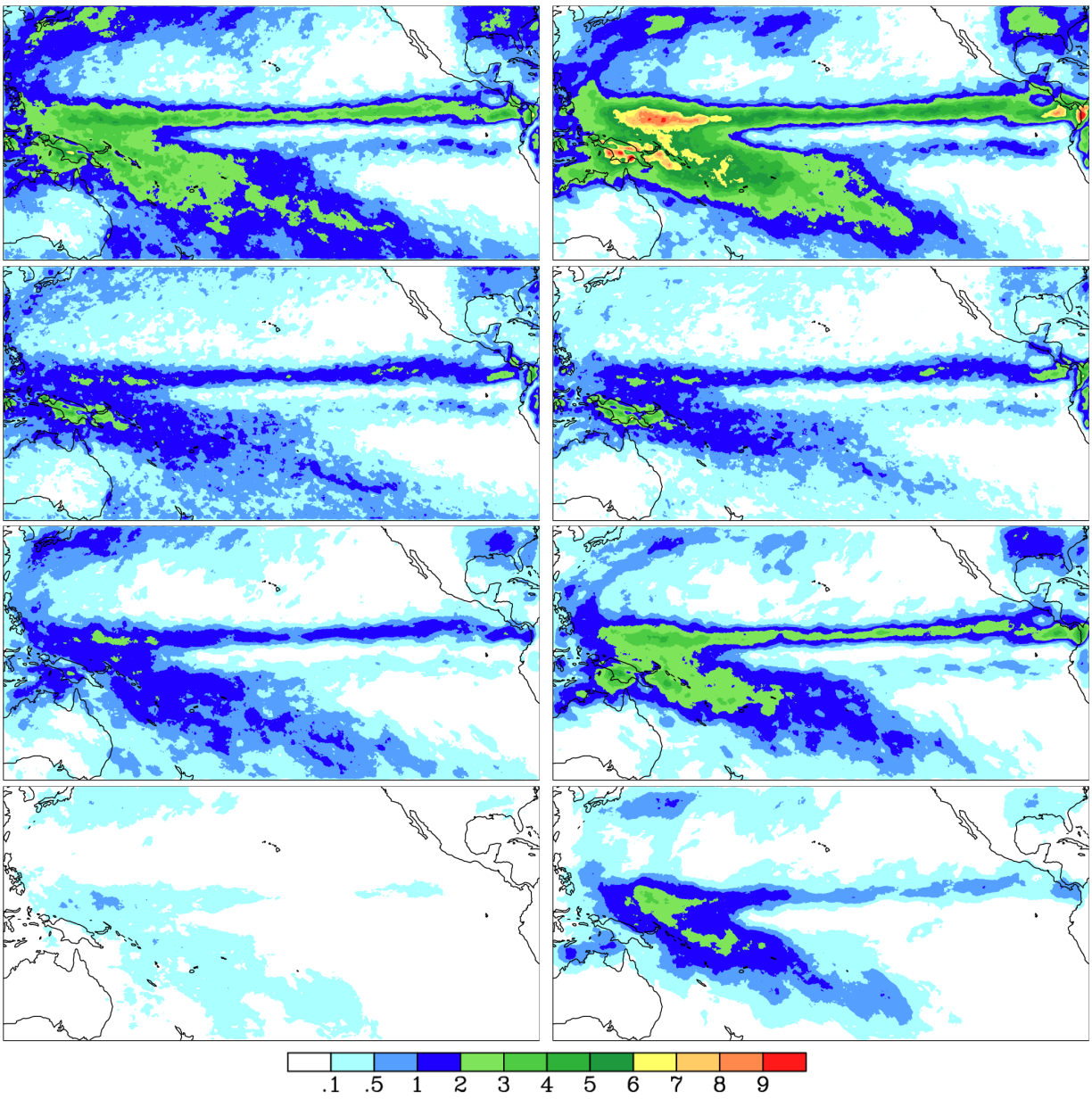


Figure AI-2: MAM object frequency [%] for TRMM (left) and PERSIANN (right): top - all objects regardless of size, 2nd row - objects with area 0 – 87,198 km^2, 3rd row - 87,198 – 348,792 km^2 and bottom - larger than 348,792 km^2.

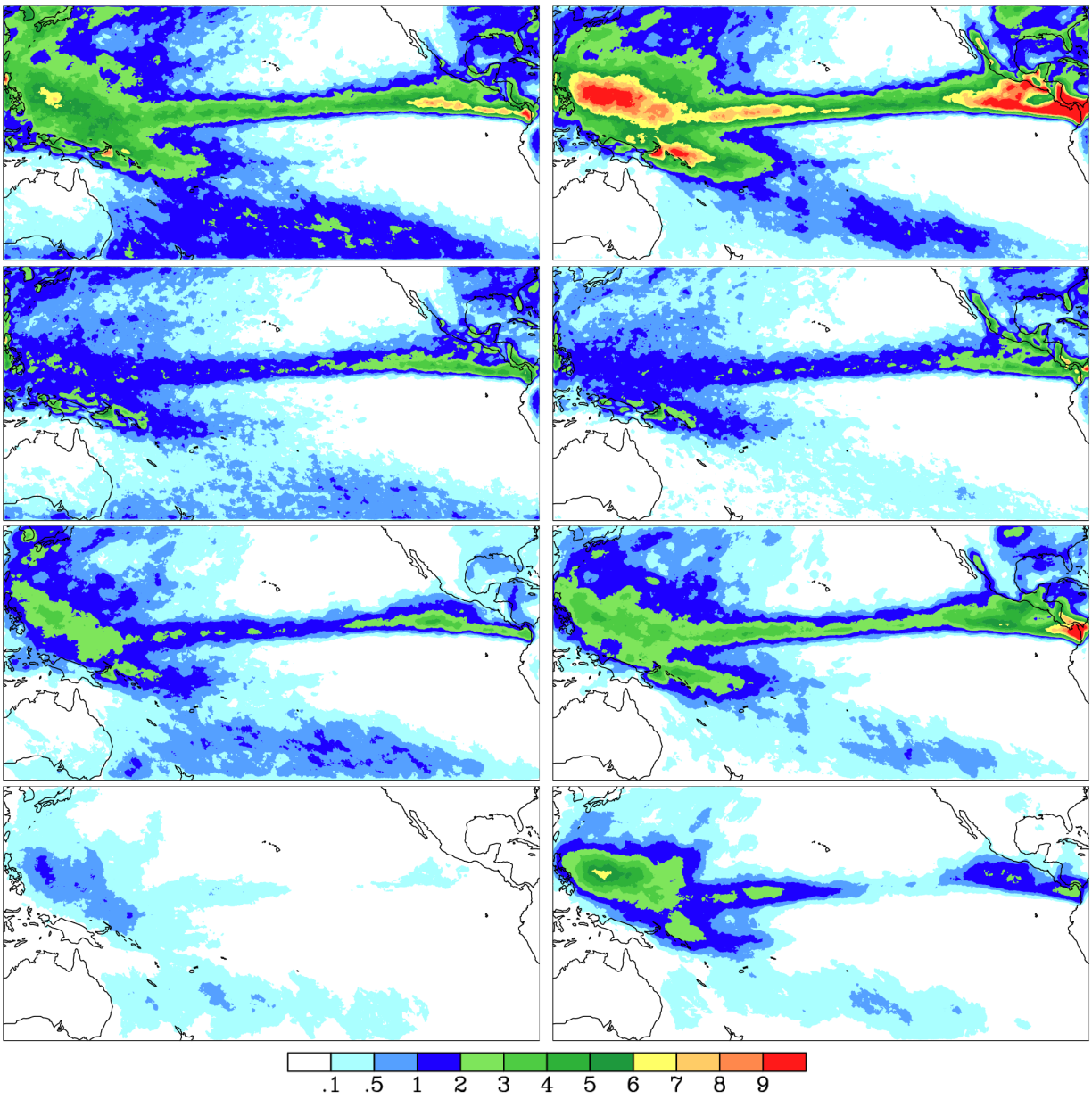


Figure AI-3: JJA object frequency [%] for TRMM (left) and PERSIANN (right): top - all objects regardless of size, 2nd row - objects with area 0 – 87,198 km^2, 3rd row - 87,198 – 348,792 km^2 and bottom - larger than 348,792 km^2.

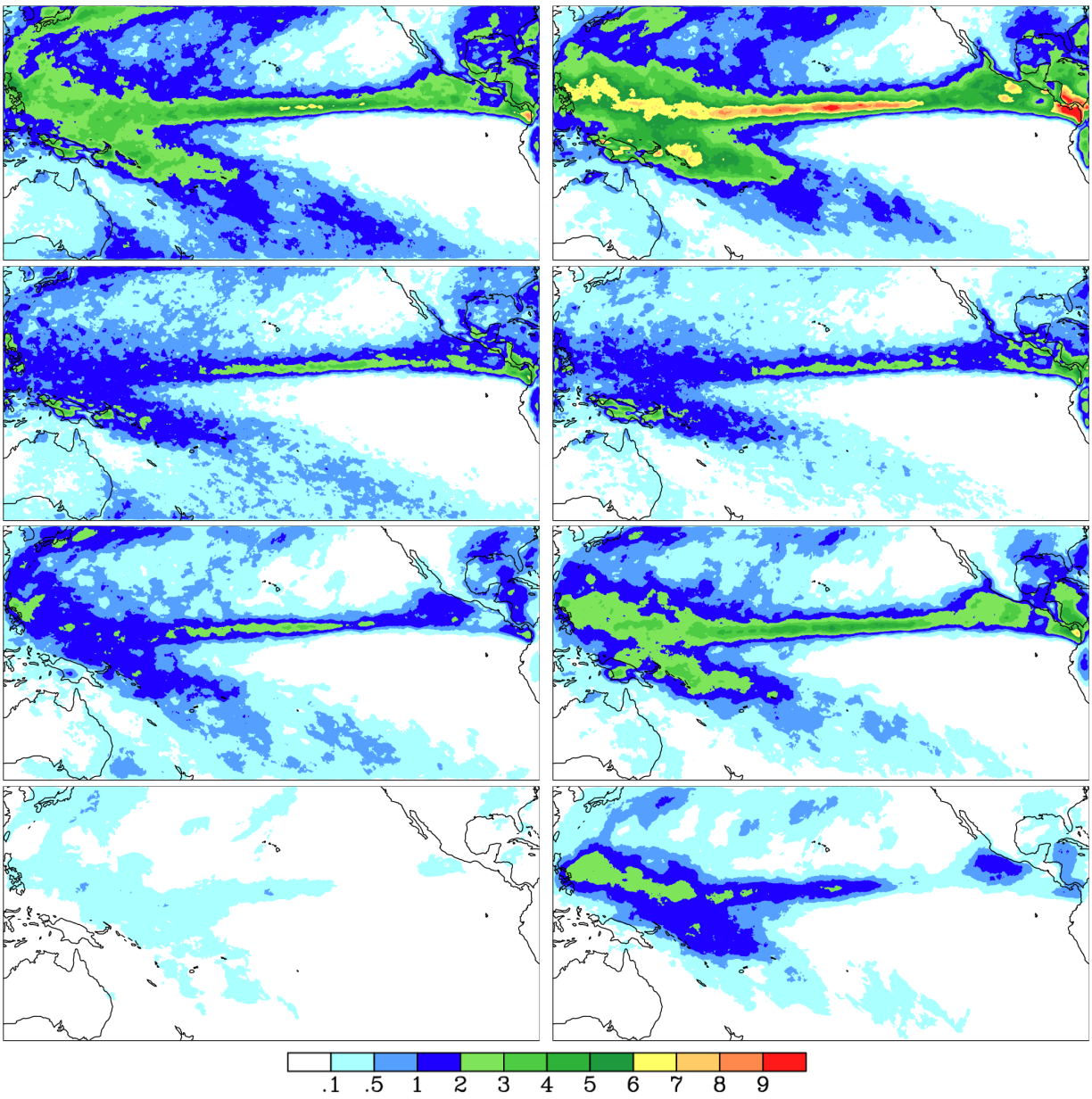


Figure AI-4: SON object frequency [%] for TRMM (left) and PERSIANN (right): top - all objects regardless of size, 2nd row - objects with area 0 – 87,198 km^2, 3rd row - 87,198 – 348,792 km^2 and bottom - larger than 348,792 km^2.

APPENDIX II: Seasonal object frequency considering their lifespan for TRMM and PERSIANN datasets

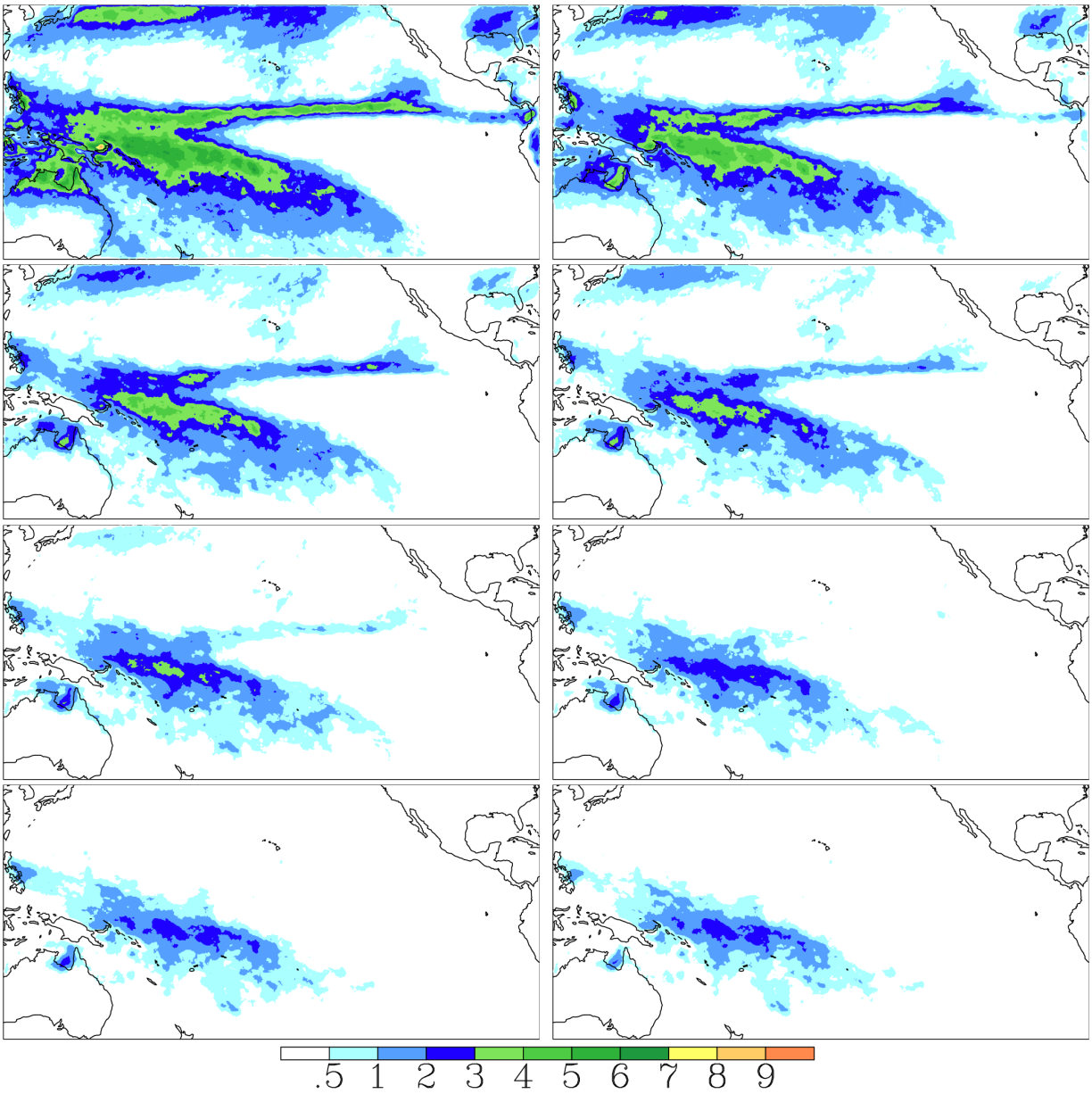


Figure AII-1: Average DJF object frequency [%] considering lifespan for TRMM. Top left: all objects, top right: lifespan longer than 15 h. 2nd row left: lifespan longer than 30 h, 2nd row right: longer than 45 h. 3rd row left: lifespan longer than 60 h, 3rd row right: longer than 75 h (right). Bottom left: lifespan longer than 90 h, bottom left: longer than 105 h.

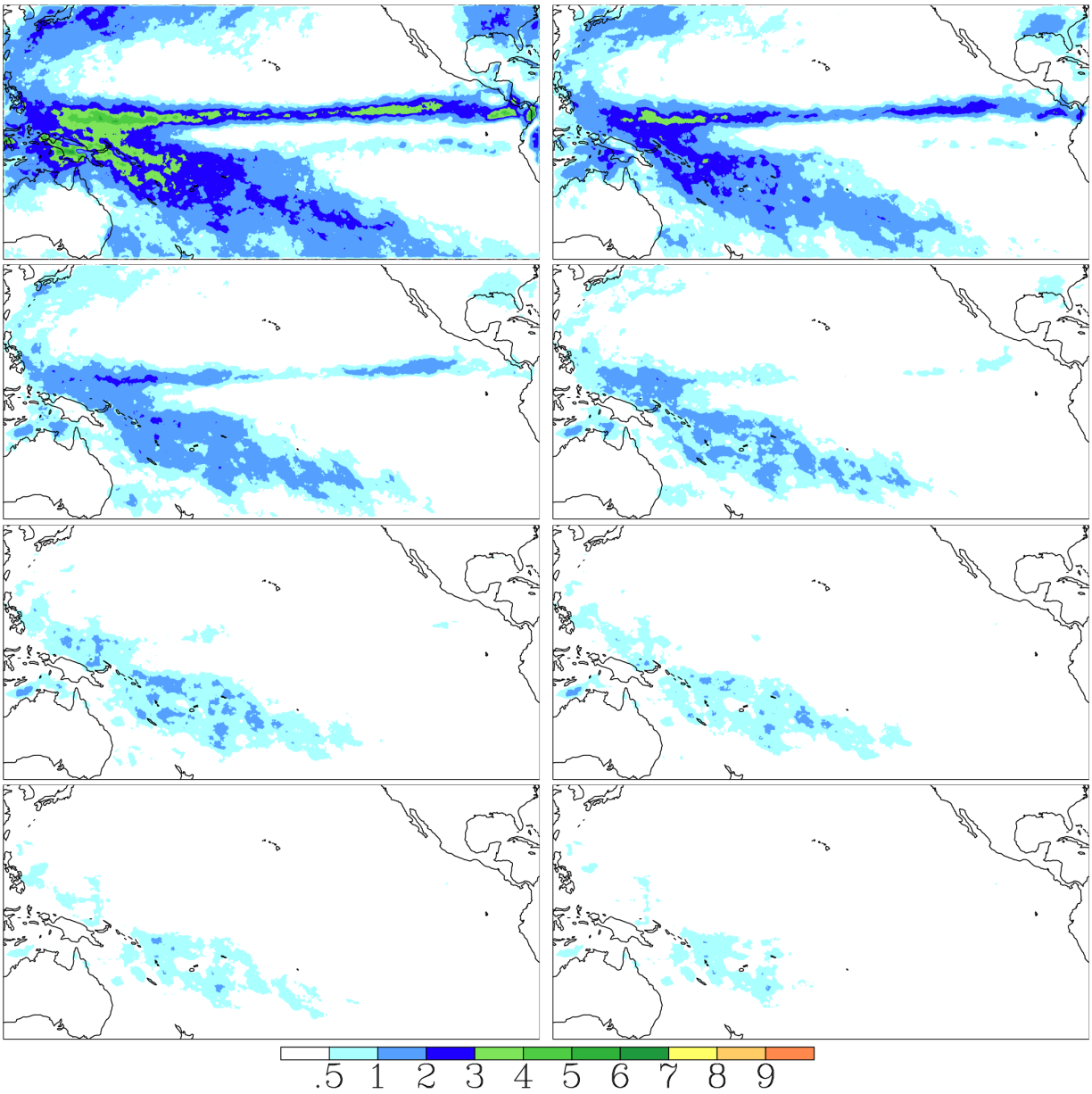


Figure AII-2: Average MAM object frequency [%] considering lifespan for TRMM. Top left: all objects, top right: lifespan longer than 15 h. 2nd row left: lifespan longer than 30 h, 2nd row right: longer than 45 h. 3rd row left: lifespan longer than 60 h, 3rd row right: longer than 75 h (right). Bottom left: lifespan longer than 90 h, bottom left: longer than 105 h.

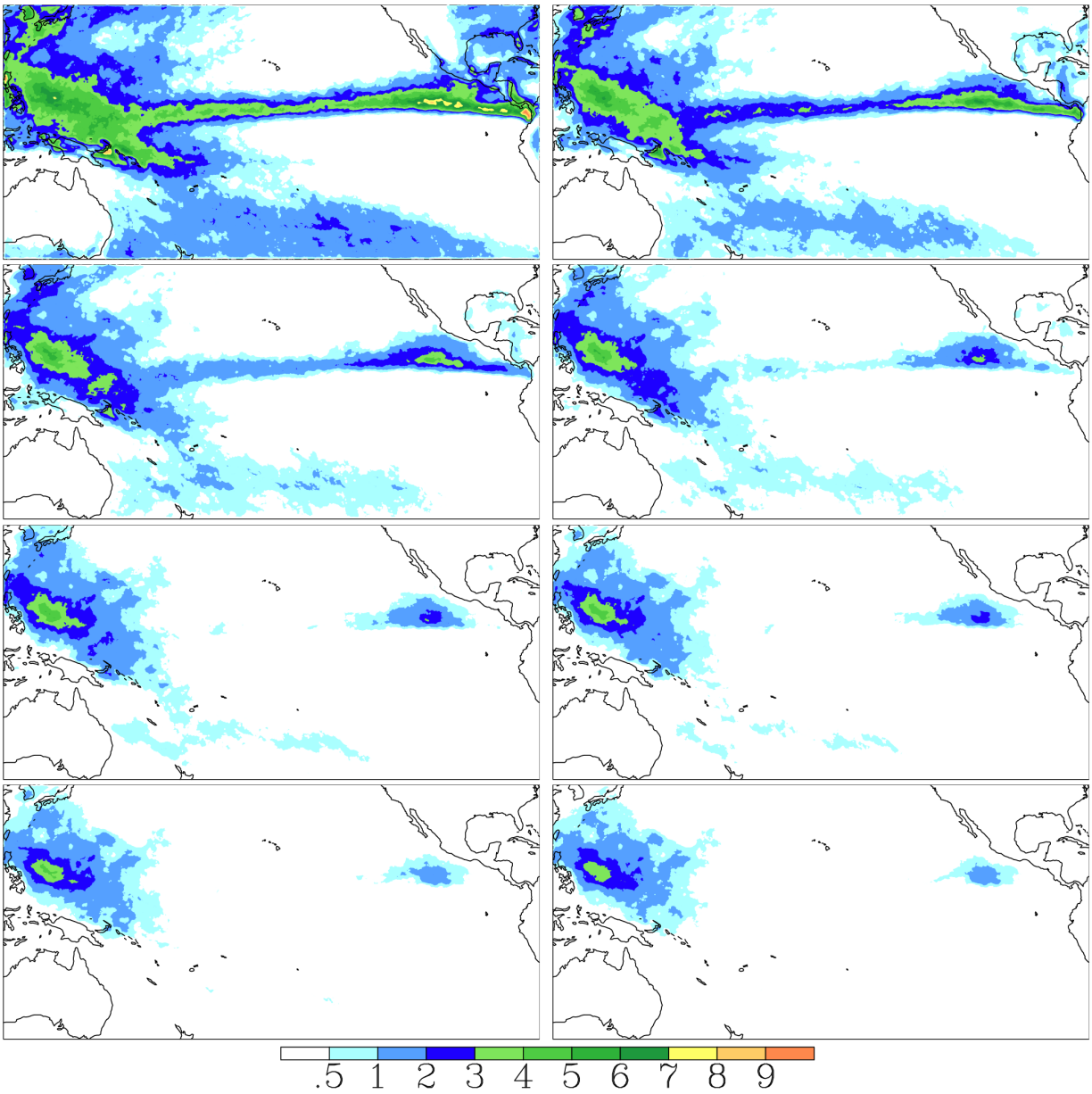


Figure AII-3: Average JJA object frequency [%] considering lifespan for TRMM. Top left: all objects, top right: lifespan longer than 15 h. 2nd row left: lifespan longer than 30 h, 2nd row right: longer than 45 h. 3rd row left: lifespan longer than 60 h, 3rd row right: longer than 75 h (right). Bottom left: lifespan longer than 90 h, bottom left: longer than 105 h.

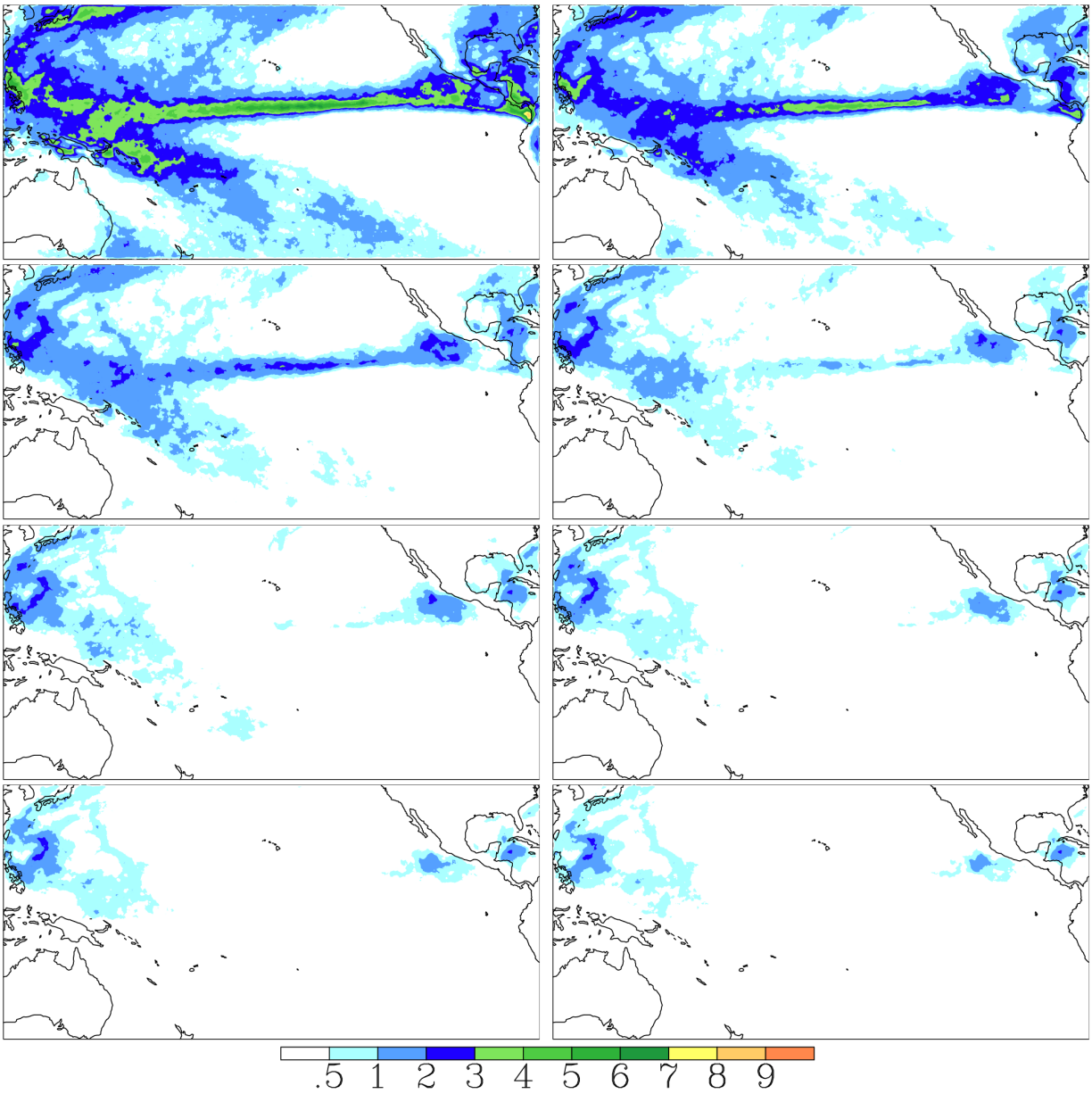


Figure AII-4: Average SON object frequency [%] considering lifespan for TRMM. Top left: all objects, top right: lifespan longer than 15 h. 2nd row left: lifespan longer than 30 h, 2nd row right: longer than 45 h. 3rd row left: lifespan longer than 60 h, 3rd row right: longer than 75 h (right). Bottom left: lifespan longer than 90 h, bottom left: longer than 105 h.

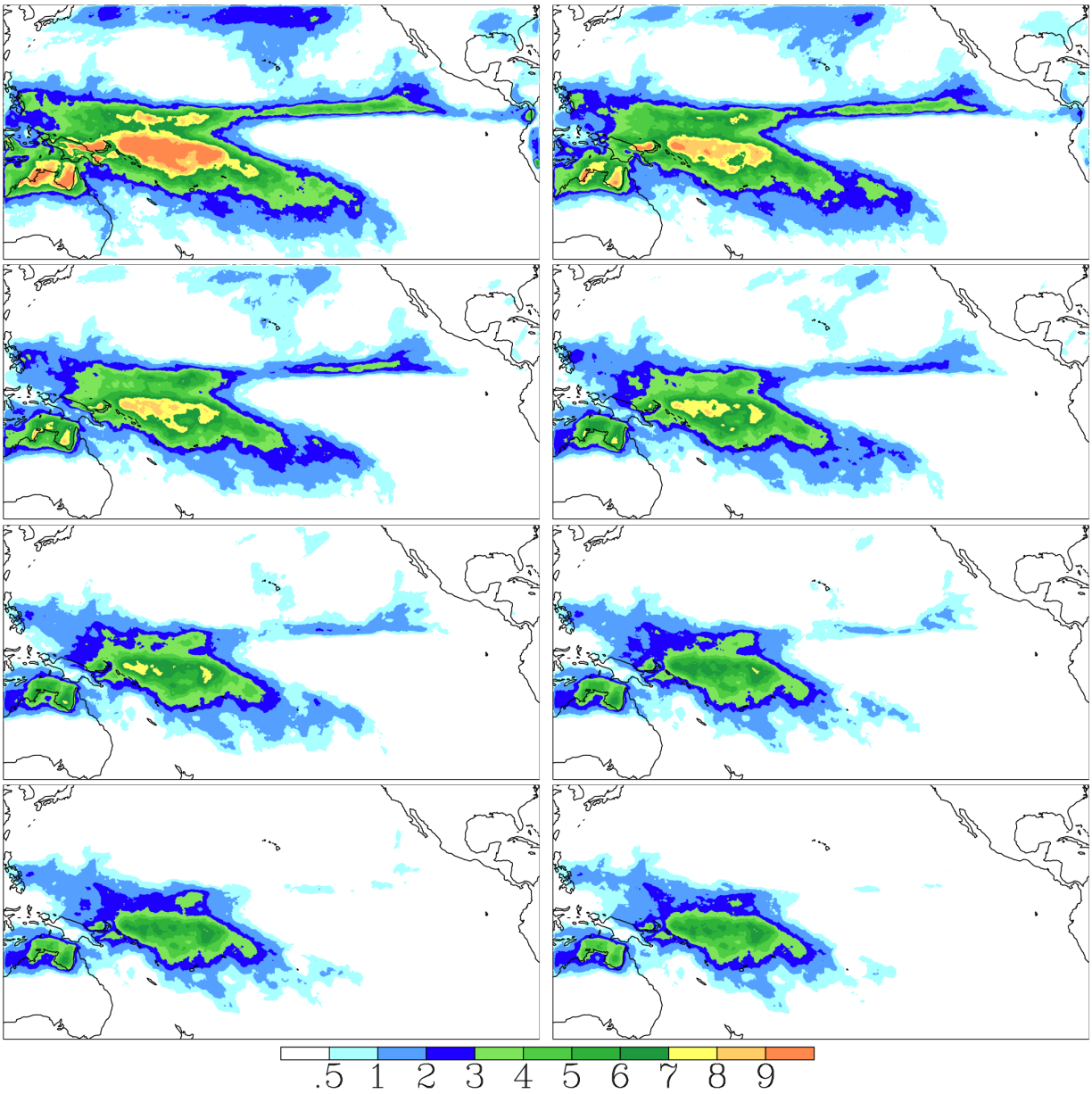


Figure AII-5: Average DJF object frequency [%] considering lifespan for PERSIANN. Top left: all objects, top right: lifespan longer than 15 h. 2nd row left: lifespan longer than 30 h, 2nd row right: longer than 45 h. 3rd row left: lifespan longer than 60 h, 3rd row right: longer than 75 h (right). Bottom left: lifespan longer than 90 h, bottom left: longer than 105 h.

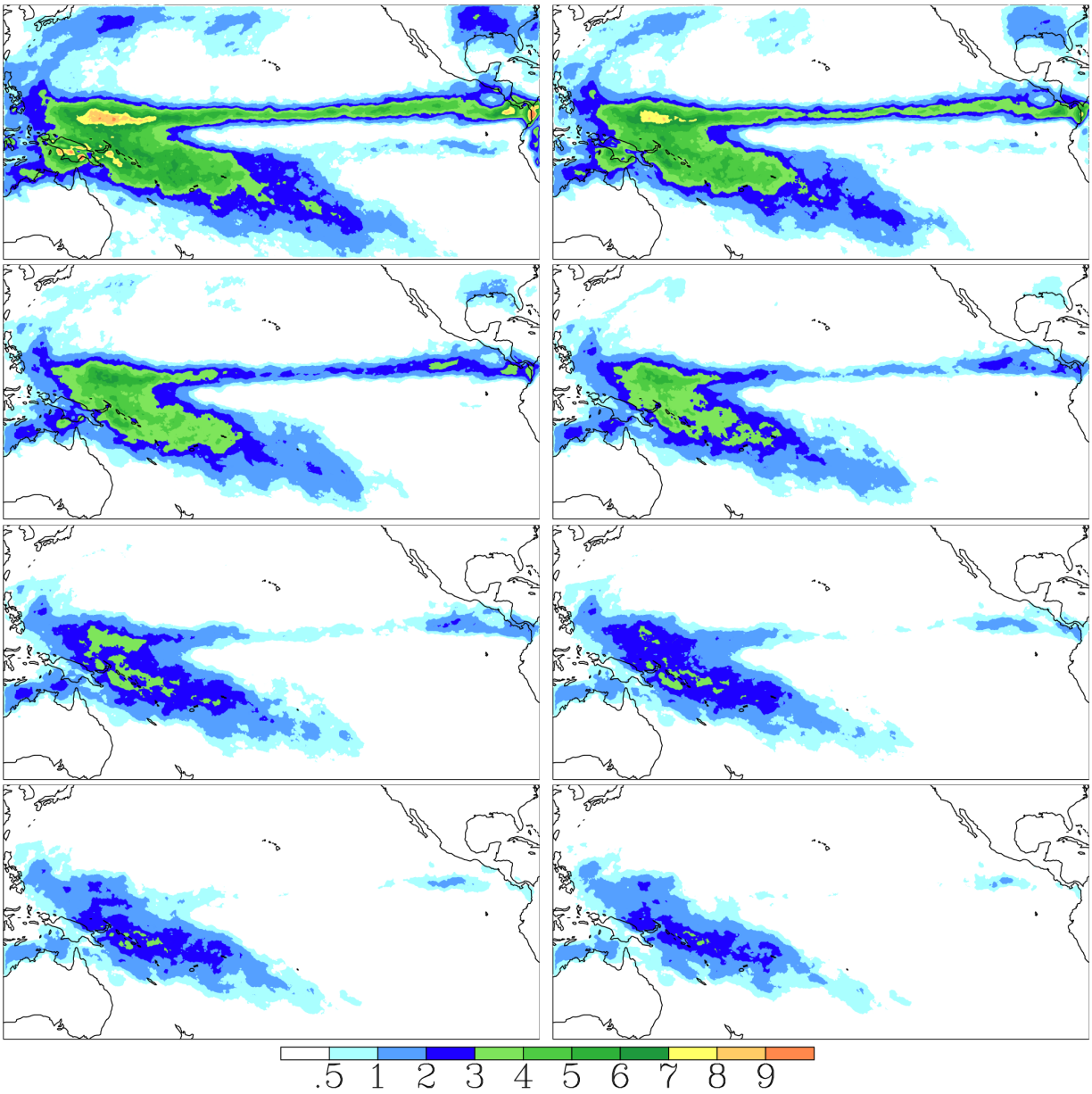


Figure AII-6: Average MAM object frequency [%] considering lifespan for PERSIANN. Top left: all objects, top right: lifespan longer than 15 h. 2nd row left: lifespan longer than 30 h, 2nd row right: longer than 45 h. 3rd row left: lifespan longer than 60 h, 3rd row right: longer than 75 h (right). Bottom left: lifespan longer than 90 h, bottom left: longer than 105 h.

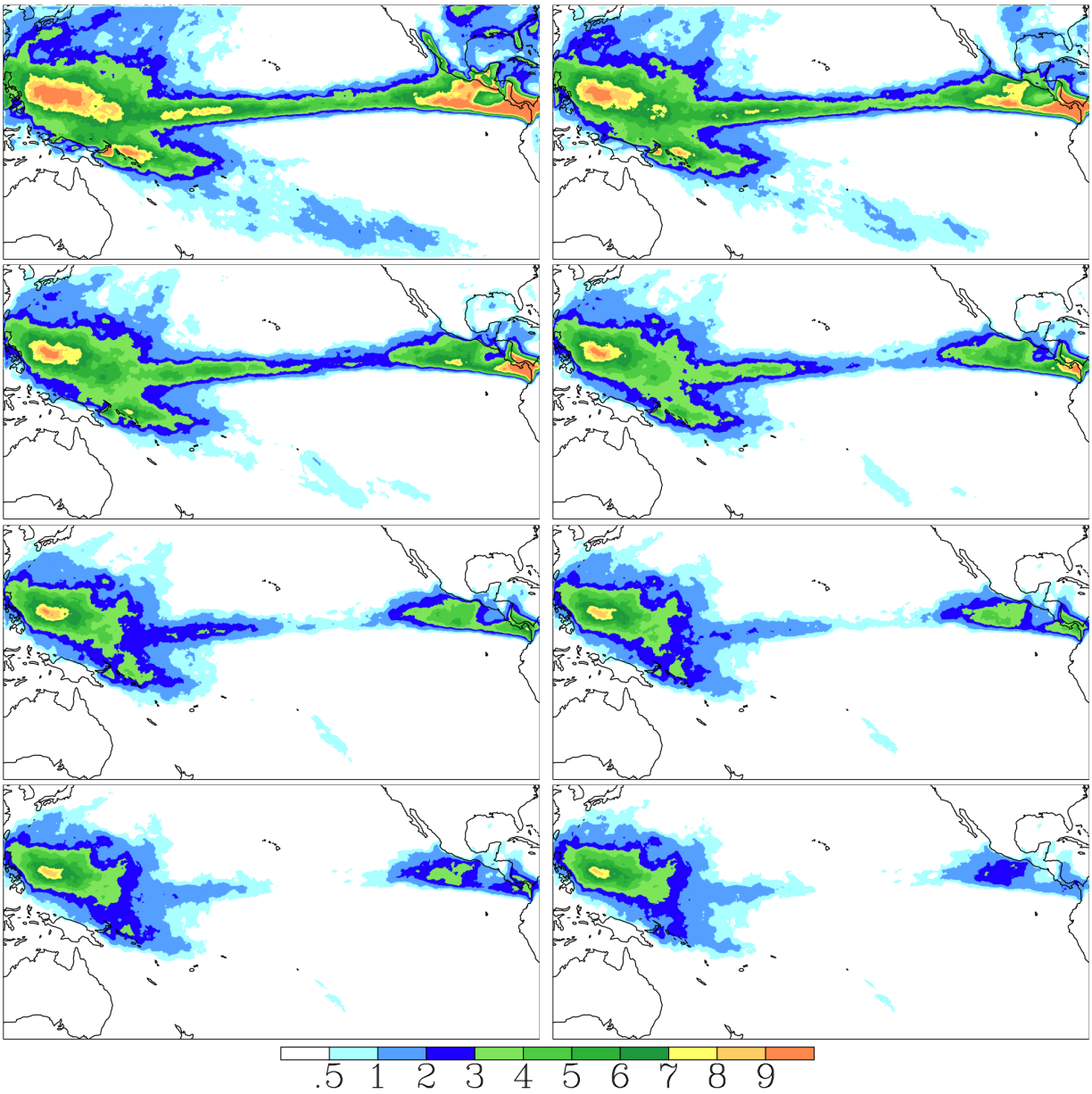


Figure AII-7: Average JJA object frequency [%] considering lifespan for PERSIANN. Top left: all objects, top right: lifespan longer than 15 h. 2nd row left: lifespan longer than 30 h, 2nd row right: longer than 45 h. 3rd row left: lifespan longer than 60 h, 3rd row right: longer than 75 h (right). Bottom left: lifespan longer than 90 h, bottom left: longer than 105 h.

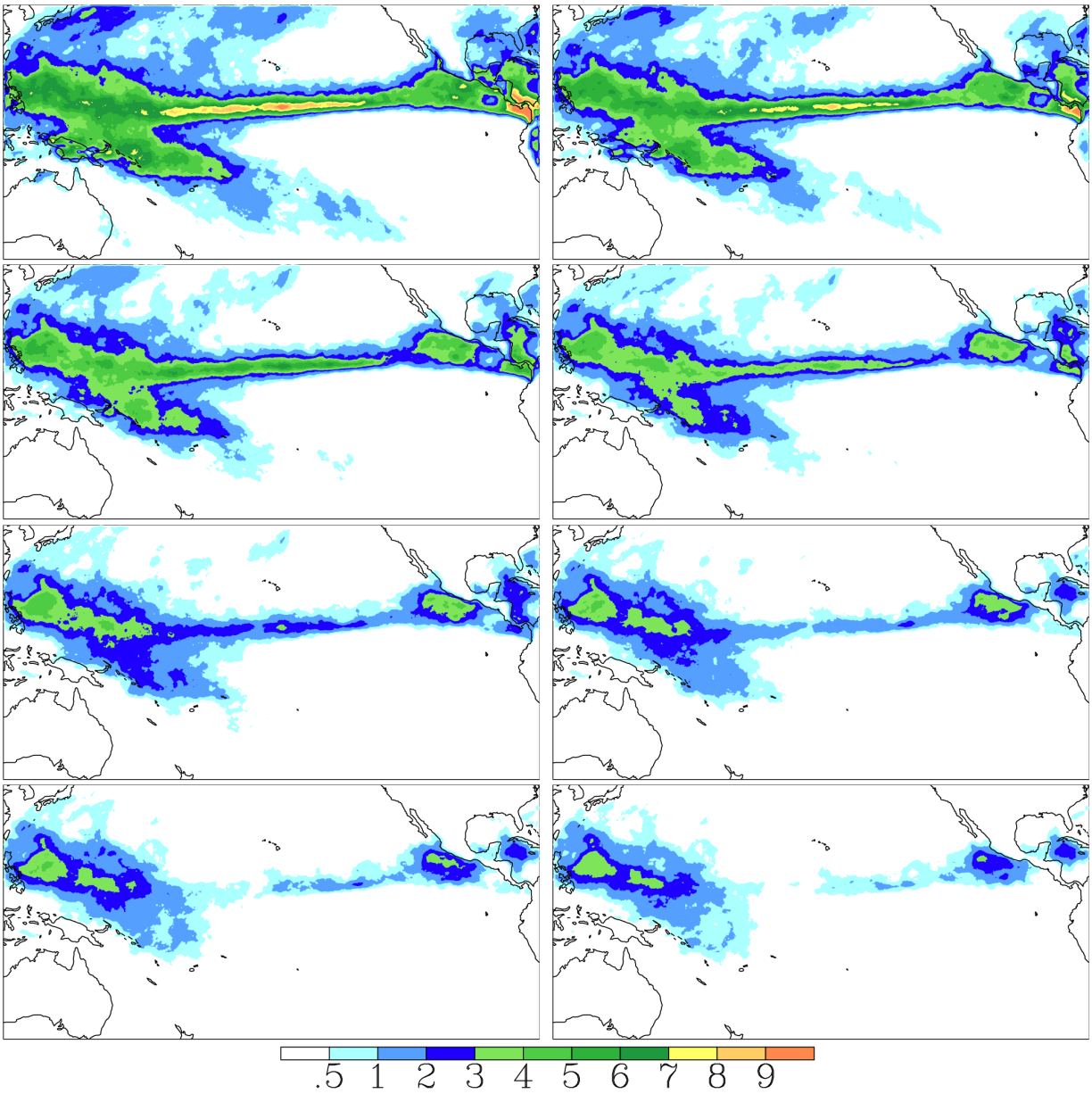


Figure AII-8: Average SON object frequency [%] considering lifespan for PERSIANN. Top left: all objects, top right: lifespan longer than 15 h. 2nd row left: lifespan longer than 30 h, 2nd row right: longer than 45 h. 3rd row left: lifespan longer than 60 h, 3rd row right: longer than 75 h (right). Bottom left: lifespan longer than 90 h, bottom left: longer than 105 h.

APPENDIX III: Description of precipitation processes in WRF model

The description of WSM-6 microphysics scheme

The WSM-6 microphysics scheme is an upgraded WSM-3 scheme. The WSM-3 scheme follows Hong et al. (2004) including ice sedimentation and other new ice-phase parameterizations revised from the older NCEP3 scheme (Hong et al., 1998) that was in WRF Version 1. A major difference from other schemes available in WRF is that a diagnostic relation is used for ice number concentration that is based on ice mass content rather than temperature. Three categories of hydrometers are included: vapor, cloud water/ice, and rain/snow. This is a so-called simple-ice scheme wherein the cloud ice and cloud water are counted as the same category. They are distinguished by temperature: namely, cloud ice can only exist when the temperature is less than or equal to the freezing point; otherwise, cloud water can exist. The same condition is applied to rain and snow. Though the ice phase is not included, it is considered efficient enough for using in operational models.

Below a detailed overview of the old scheme (Hong et al., 1998) followed by changes introduced in WSM-6 scheme will be presented. The scheme introduces three prognostic moisture variables for the mixing ratios of water vapor $q_{\rm v}$, cloud water/ice $q_{\rm ci}$ and rain/snow $q_{\rm rs}$. The thermodynamic equation and prognostic moisture equations are described as:

$$\frac{\partial q'_{\rm v}}{\partial t} = -m^2\left(u^*\frac{\partial q_{\rm v}}{\partial x}+v^*\frac{\partial q_{\rm v}}{\partial y}\right)-\dot\sigma\frac{\partial q_{\rm v}}{\partial\sigma}+F_{\rm v}^{\rm vdif}+F_{\rm v}^{\rm hdif}+F_{\rm v}^{\rm impl}+F_{\rm v}^{\rm expl}-\frac{\partial q_{\rm v}^{\rm base}}{\partial t} \quad (6.1)$$

$$\frac{\partial q'_{\rm ci}}{\partial t} = -m^2\left(u^*\frac{\partial q_{\rm ci}}{\partial x}+v^*\frac{\partial q_{\rm ci}}{\partial y}\right)-\dot\sigma\frac{\partial q_{\rm ci}}{\partial\sigma}+F_{\rm ci}^{\rm hdif}+F_{\rm ci}^{\rm expl}-\frac{\partial q_{\rm ci}^{\rm base}}{\partial t}, \quad (6.2)$$

$$\frac{\partial q'_{\rm rs}}{\partial t} = -m^2\left(u^*\frac{\partial q_{\rm rs}}{\partial x}+v^*\frac{\partial q_{\rm rs}}{\partial y}\right)-\dot\sigma\frac{\partial q_{\rm rs}}{\partial\sigma}+F_{\rm rs}^{\rm hdif}+F_{\rm rs}^{\rm expl}-\frac{\partial q_{\rm rs}^{\rm base}}{\partial t}-\frac{g}{p_s}\frac{\partial\rho q_{\rm rs}V_{\rm t}}{\partial\sigma} \quad (6.3)$$

and the corresponding thermodynamic equation

$$\frac{\partial T'_\rho}{\partial t} = -m^2\left(u^*\frac{\partial T_\rho}{\partial x}+v^*\frac{\partial T_\rho}{\partial y}\right)-\sigma^\kappa\dot\sigma\frac{\partial T_\rho\sigma^{-\kappa}}{\partial\sigma}+\kappa T_\rho\left[\frac{\partial Q}{\partial t}+m^2\left(u^*\frac{\partial Q}{\partial x}+v^*\frac{\partial Q}{\partial y}\right)\right]+$$
$$+ \quad F_T^{\rm vdif}+F_T^{\rm hdif}+F_T^{\rm impl}+F_T^{\rm expl}+F_T^{\rm rad}-\frac{\partial T_\rho^{\rm base}}{\partial t} \quad (6.4)$$

where the density-corrected temperature, T_ρ, is given by

$$T_\rho = T_\mathrm{v} \frac{1 + q_\mathrm{v}}{1 + q_\mathrm{v} + q_\mathrm{ci} + q_\mathrm{rs}} \tag{6.5}$$

and the virtual temperature, $T_\mathrm{v} = (1 + 0.608\, q_\mathrm{v})T$.

All symbols in Equations 6.1 -6.5 follow the conventional notation. m is the map factor. The prime indicates perturbation, and the superscript "base" indicates base field values. The superscript * in variables u^* and v^* are u/m and v/m. Here Q is $\ln(p_\mathrm{s})$ and $\kappa = R_\mathrm{d}/c_\mathrm{p}$. The F's represent tendency terms that are provided by physics as indicated by the superscript: vdif (vertical diffusion), hdif (horizontal diffusion), impl (heating due to subgridscale precipitation physics), expl (heating due to grid-resolvable precipitation physics) and rad (radiation). The microphysical processes in the scheme contain condensation of water vapor into cloud water (ice) at water saturation, accretion of cloud by rain (ice by snow), autoconversion of cloud to rain (ice to snow), evaporation (sublimation) of rain (snow), initiation of ice crystals, and sublimation or deposition of ice crystals. The last term in Eq. 6.3 represents the removal of hydrometeors due to settling, where V_t in Eq. 6.3 designates the mass-weighted fall speed of the precipitates. This term is computed in gridpoint space following the calculation of the microphysical processes. For numerical stability, a split time step for each model column is applied on computing the fallout terms, so that precipitate does not cross over any vertical grid within a single loop of the calculation.

Implementation of the prognostic cloud scheme necessitates several additional considerations. First, negative values of cloud water/ice crystals and snow/rain are set to zero in gridpoint space (physics space) before microphysical processes are computed. In other words, negative values in gridpoint space due to the vertical and horizontal advection, and horizontal diffusion calculations are eliminated.

If the model employs a fairly large time step at a given grid size due to the integration scheme use, the microphysical processes could produce an unreasonable evolution of clouds even though maintaining mass conservation for all the prognostic liquid species. This probably arises due to the fact that, with a relatively large time step for microphysics calculation, relatively fast processes (such as condensation and ice nucleation) are represented as slow processes while at the same time slow processes (such as sublimation/deposition, evaporation of raindrops) become relatively fast processes.

The cloud fraction formula needed for the radiative transfer of both long and shortwaves in the atmosphere is used:

$$CF = RH \left[1 - \exp\left(\frac{-1000\, q_\mathrm{l}}{1 - RH}\right)\right],$$

where RH is the realtive humidity, q_l is the mixing ratio for total liquid species.

The WSM-5 scheme is similar to the WSM-3 simple ice scheme. However, the five clases of cloud/precipitation water are held in five different arrays. Thus, it allows supercooled water to exist, and a gradual melting of snow as it falls below the melting layer. Details can be found in Hong et al. (2004).

The WSM-6 six-class scheme extends the WSM-5 scheme to include graupel and its associated processes. Many of these processes are parameterized similarly to Lin et al. (1983), but there are differences for the accretion calculation and in some other parameters. The freezing/melting processes are computed during the fall-term sub-steps to increase accuracy in the vertical heating profile of these processes. The order of the processes is also optimized to decrease the sensitivity of the scheme to the time step of the model. As with WSM-3 and WSM-5, saturation adjustment treats separately ice and water saturation processes, rather than a combined saturation.

The description of Kain-Fritsch convection scheme

The Kain and Fritsch scheme (Kain and Fritsch, 1990, 1993) provides mass flux parameterization of subgrid convection. It uses the Lagrangian parcel method, including vertical momentum dynamics, to estimate whether instability exists and, if so, what the properties of convective clouds will be. It is convenient to compartmentalize the scheme into three parts: a) the convective trigger function, b) the mass flux formulation, and c) the closure assumptions. Each of these is discussed briefly below.

a) The trigger function. The first task of the scheme is to identify potential source layers for convective clouds, i.e., updraft source layers (USLs). Beginning at the surface, vertically adjacent layers in the host model are mixed until the depth of mixing reaches at least 60 hPa. This combination of adjacent model layers comprises the first potential USL. The mean thermodynamic characteristics of this mixing are computed, along with the temperature and height of this "parcel" at its lifting condensation level (LCL). As a first guess at the likelihood of convective initiation, parcel temperature, $T_{\rm LCL}$, is compared to the ambient temperature, $T_{\rm ENV}$ at the parcel LCL. Typically, the parcel will be colder than its environment, i.e., negatively buoyant. Based on observations suggesting that convective development tends to be favored by background vertical motion, the parcel is assigned a temperature perturbation linked to the magnitude of grid-resolved vertical motion. The specific formula for this perturbation, $\delta T_{\rm vv}$, is

$$\delta T_{\rm vv} = k(w_{\rm g} - c(z))^{1/3} \tag{6.6}$$

where k is a unit number with dimensions K/(cm/s)$^{1/3}$, $w_{\rm g}$ is the grid-resolved vertical velocity (cm/s), and $c(z)$ is a threshold vertical velocity given by

$$c(z) = \begin{cases} w_0(Z_{\rm LCL}/2000) & , Z_{\rm LCL} \leq 2000 \text{ m} \\ w_0 & , Z_{\rm LCL} > 2000 \text{ m} \end{cases} \tag{6.7}$$

where $w_0 = 2$ cm/s, $Z_{\rm LCL}$ is the height of the LCL above the ground (m). For example, this equation yields a temperature perturbation of 1 K for a background vertical velocity of 1 cm/s above the threshold value, and just over 2 K when the vertical velocity is 10 cm/s above the threshold value.

The use of this perturbation term allows us to effectively eliminate most parcels as candidates for deep convection, which is important for computational efficiency. The elimination process involves adding the computed temperature perturbation (typically 1-2 K, for example, in environments with weak to moderate upward motion) to the parcel temperature at the LCL. If the resulting temperature is still less than the environmental value (i.e., $T_{\mathrm{LCL}} + \delta T_{\mathrm{vv}} < T_{\mathrm{ENV}}$), this parcel is eliminated from consideration. If, however, the perturbed parcel is warmer than its environment, it is allowed to proceed as a candidate for deep convection. At this stage, the parcel is released at its LCL with its original (unperturbed) temperature and moisture content and a vertical velocity derived from the perturbation temperature. Specifically, its initial vertical velocity, w_{p0}, is loosely based on the parcel buoyancy equation and is given by

$$w_{\mathrm{p0}} = 1 + 1.1 \left[(Z_{\mathrm{LCL}} - Z_{\mathrm{USL}}) \cdot \frac{\delta T_{\mathrm{vv}}}{T_{\mathrm{ENV}}} \right]^{1/2} \tag{6.8}$$

where Z_{USL} is the height at the base of the USL. This formula yields starting vertical velocities of up to several m/s.

Above the LCL parcel vertical velocity is estimated at each model level using the Lagrangian parcel method, including the effects of entrainment, detrainment, and water loading. If vertical velocity remains positive over a depth exceeding a specified minimum cloud depth (typically 3-4 km), deep convection is activated using this USL. If not, the base of the potential USL is moved up one model layer and the procedure is repeated. This process continues until either the first suitable source layer is found, or the sequential search has moved up above the lowest 300 hPa of the atmosphere, where the search is terminated. This complete set of criteria comprises the trigger function, but note that the updraft model described in the next subsection plays an important role in determining cloud depth.

b) Mass flux formulation. Convective updrafts in the KF scheme are represented using a steadystate entraining/detraining plume model, as described in detail in Kain and Fritsch (1990). In this model, entrainment and detrainment rates are inversely proportional, with high entrainment (detrainment) rates being favored by high (low) parcel buoyancy and moist (dry) environments. In practice, the distinction between the updraft and the trigger function can become blurred because the specific formulation of the updraft can determine whether or not the specified minimum cloud depth for deep convection is achieved.

Convective downdrafts are fueled by evaporation of condensate that is generated within the updraft. A fraction of this total condensate is made available for evaporation within the downdraft, based on empirical formulas for precipitation efficiency as a function of vertical wind shear and cloud-base height (Zhang and Fritsch 1986). This fraction effectively dictates the relative magnitudes between downdraft and updraft mass fluxes once other critical downdraft parameters are specified. These other parameters include the downdraft starting and ending levels, its relative humidity profile, and the characteristics and amounts of air that get entrained into the downdraft. The downdraft is specified to start at the level of minimum saturation equivalent potential temperature, θ_{es}, with a mixture

of updraft and environmental air. It is moved downward in a Lagrangian sense, with a specified entrainment rate (entraining environmental air only) and a fixed relative humidity of 100 % above cloud base and 90 % below cloud base. The downdraft is terminated if it becomes warmer than its environment or it reaches the surface. It is forced to detrain into the environment within and immediately above the termination level, such that the minimum depth of the detrainment layer is the same as the minimum depth of the USL, 60 mbar.

Environmental mass fluxes are required to compensate for the upward and downward transports in the updrafts and downdrafts, so that the net mass flux at any level in the column is zero. Formulation of these fluxes is described in Kain and Fritsch (1993).

c) Closure assumptions. The method by which the KF scheme satisfies its closure assumptions is described in Bechtold et al. (2001). Fundamentally, the KF scheme rearranges the mass in a column using the updraft, downdraft, and environmental mass fluxes until at least 90 % of the CAPE (convective available potential energy) is removed. CAPE is computed in the traditional way, using undilute parcel ascent, with the "parcel" characteristics being those of the USL. CAPE is removed by the combined effects of lowering θ_e in the USL and warming the environment aloft.

The scheme provides convective tendencies of temperature, water vapor mixing ratio, and cloud water mixing ratio. By default, convective precipitation particles simply accumulate at the surface rather than being introduced aloft, but the code has a "switch" to activate feedback of precipitation at the level it is formed. The switch can be set to any value between 0 (no feedback) to 1 (100 % feedback).

The modified version of the Kain-Fritsch scheme (KF-Eta) is based on old scheme, but has been modified based on testing within the Eta model (Kain, 2004). As with the original KF scheme, it utilizes a simple cloud model with moist updrafts and downdrafts, including the effects of detrainment, entrainment, and relatively simple microphysics. It differs from the original KF scheme in the following ways:

- A minimum entrainment rate is imposed to suppress widespread convection in marginally unstable, relatively dry environments.
- Shallow (non precipitating) convection is allowed for any updraft that does not reach minimum cloud depth for precipitating clouds; this minimum depth varies as a function of cloud-base temperature.
- The entrainment rate is allowed to vary as a function of low-level convergence.
- Downdraft changes:
  - Source layer is the entire 150–200 mbar deep layer just above cloud base.
  - Mass flux is specified as a fraction of updraft mass flux at cloud base. Fraction is a function of source layer RH rather than wind shear or other parameters, i.e., old precipitation efficiency relationship not used.
  - Detrainment is specified to occur in updraft source layer and below.

APPENDIX IV: Yearly and seasonal TRMM dataset and WRF simulation object frequencies considering their lifespan

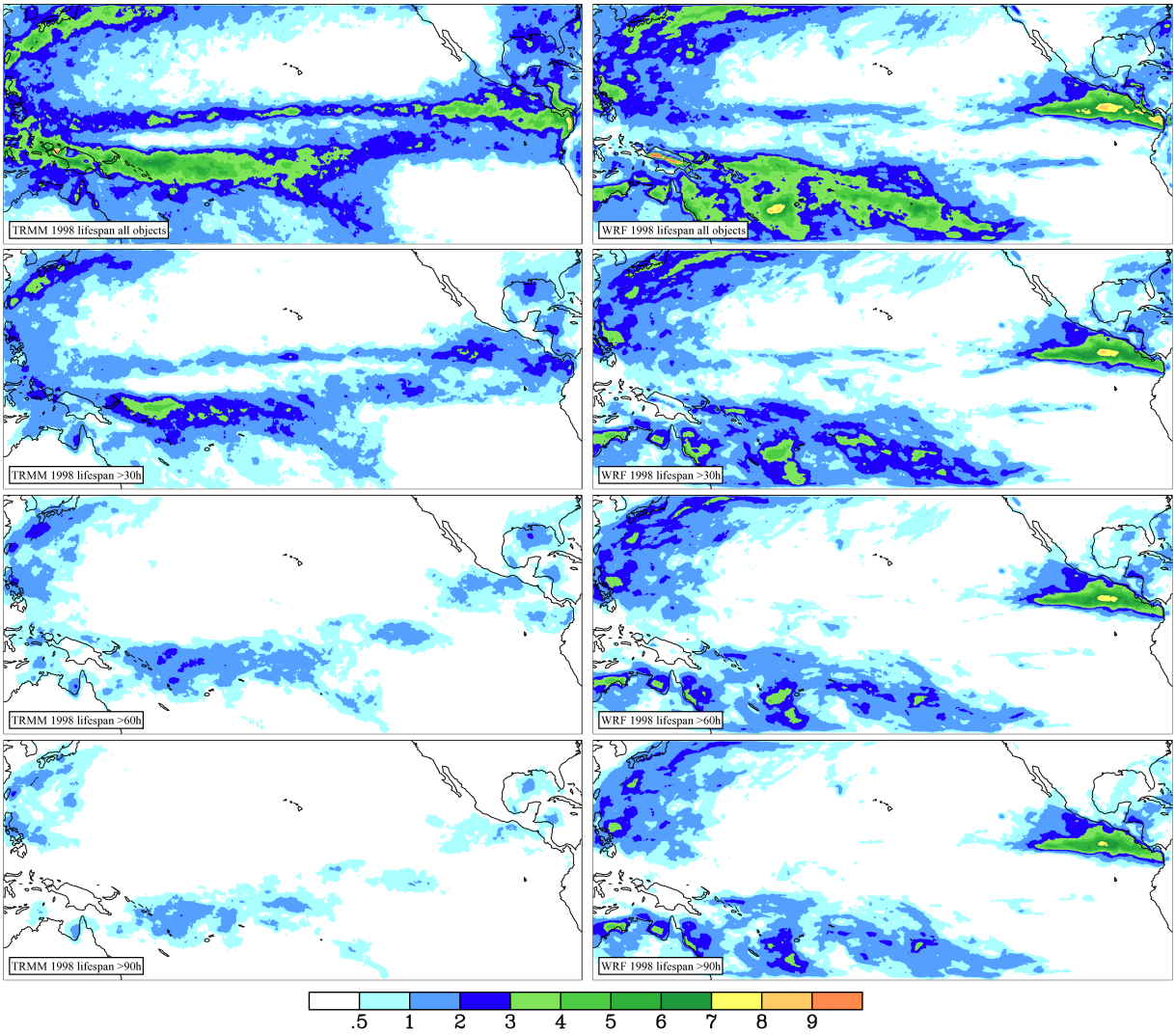


Figure AIV-1: Object frequency [%] considering lifespan for TRMM (left) and PESIANN (right) in year 1998. Top: all objects, 2nd row: lifespan longer than 30h, 3rd row: lifespan longer than 60h and bottom: lifespan longer than 90h.

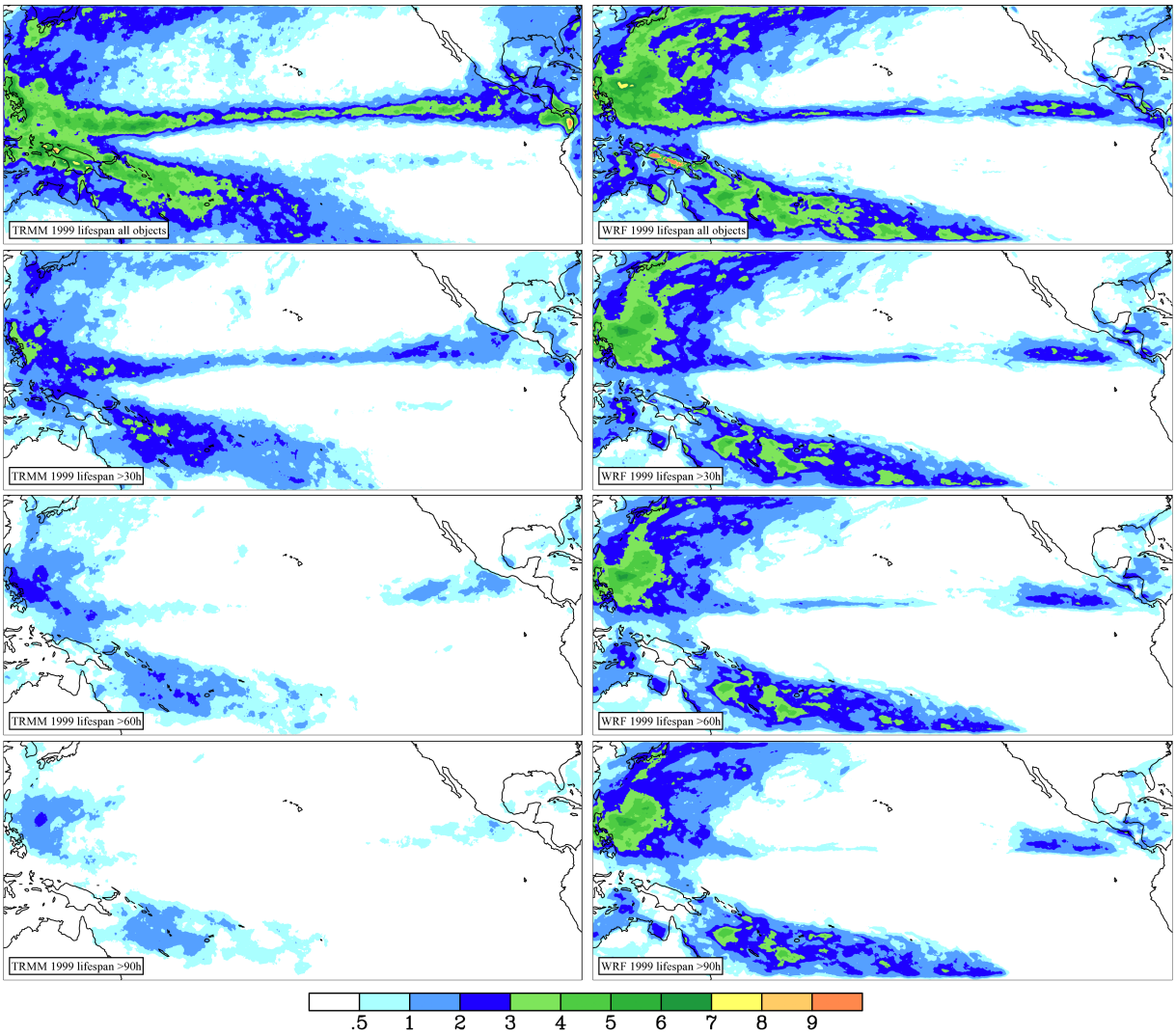


Figure AIV-2: Object frequency [%] considering lifespan for TRMM (left) and PESIANN (right) in year 1999. Top: all objects, 2nd row: lifespan longer than 30h, 3rd row: lifespan longer than 60h and bottom: lifespan longer than 90h.

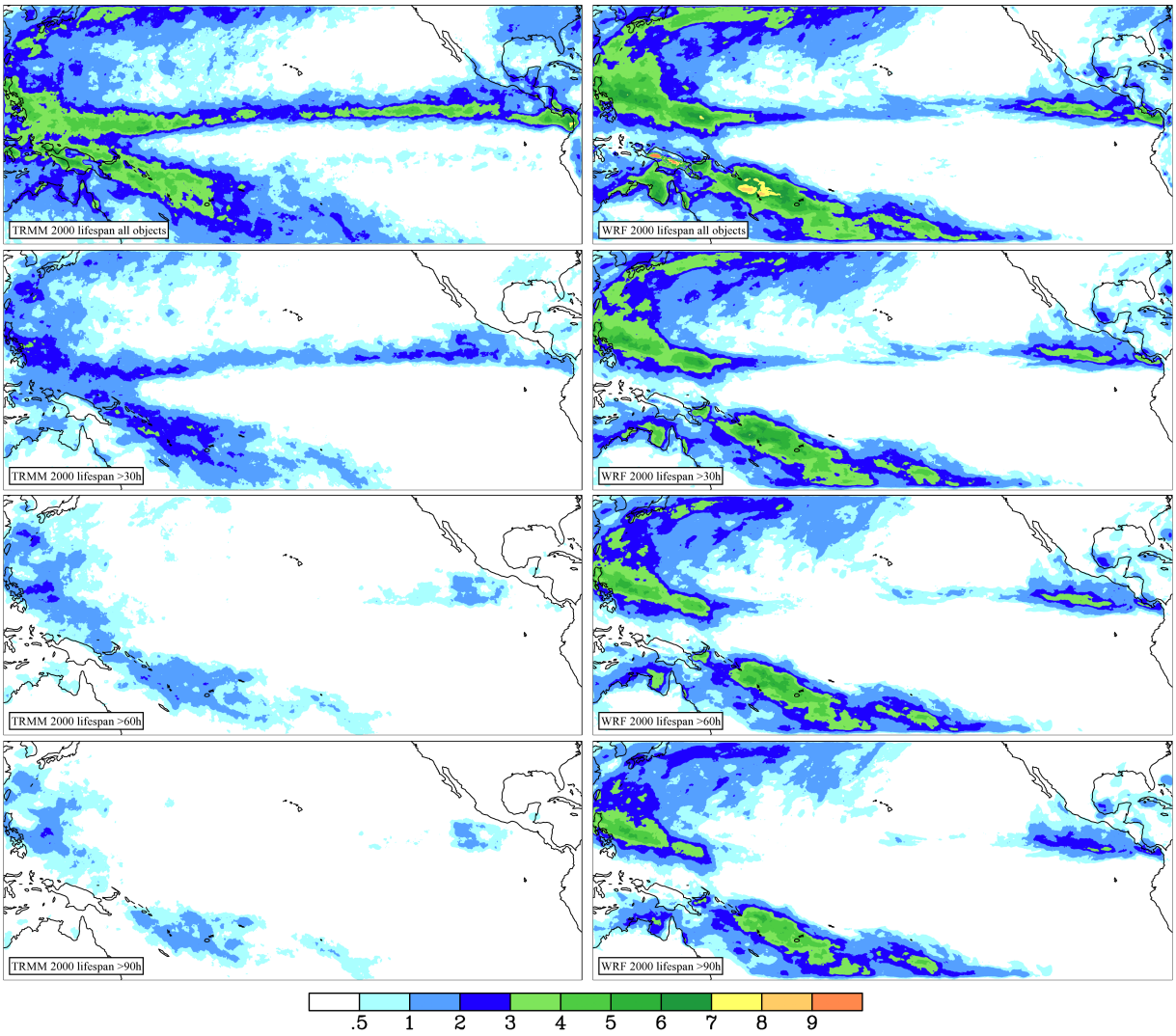


Figure AIV-3: Object frequency [%] considering lifespan for TRMM (left) and PESIANN (right) in year 2000. Top: all objects, 2nd row: lifespan longer than 30h, 3rd row: lifespan longer than 60h and bottom: lifespan longer than 90h.

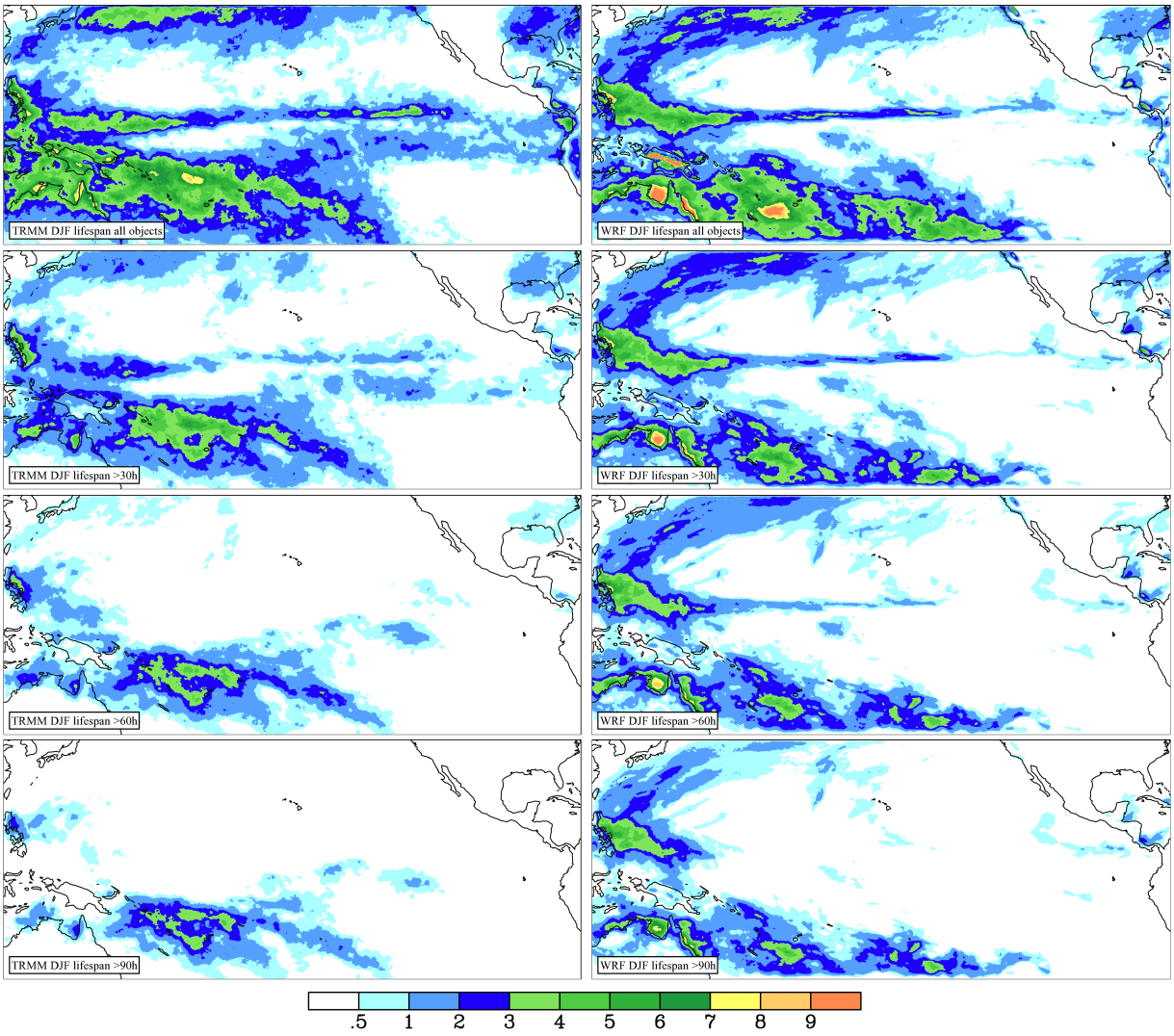


Figure AIV-4: Object frequency [%] considering lifespan for TRMM (left) and PESIANN (right) for DJF season in period from 1st January 1998 to 31st December 2000. Top: all objects, 2nd row: lifespan longer than 30h, 3rd row: lifespan longer than 60h and bottom: lifespan longer than 90h.

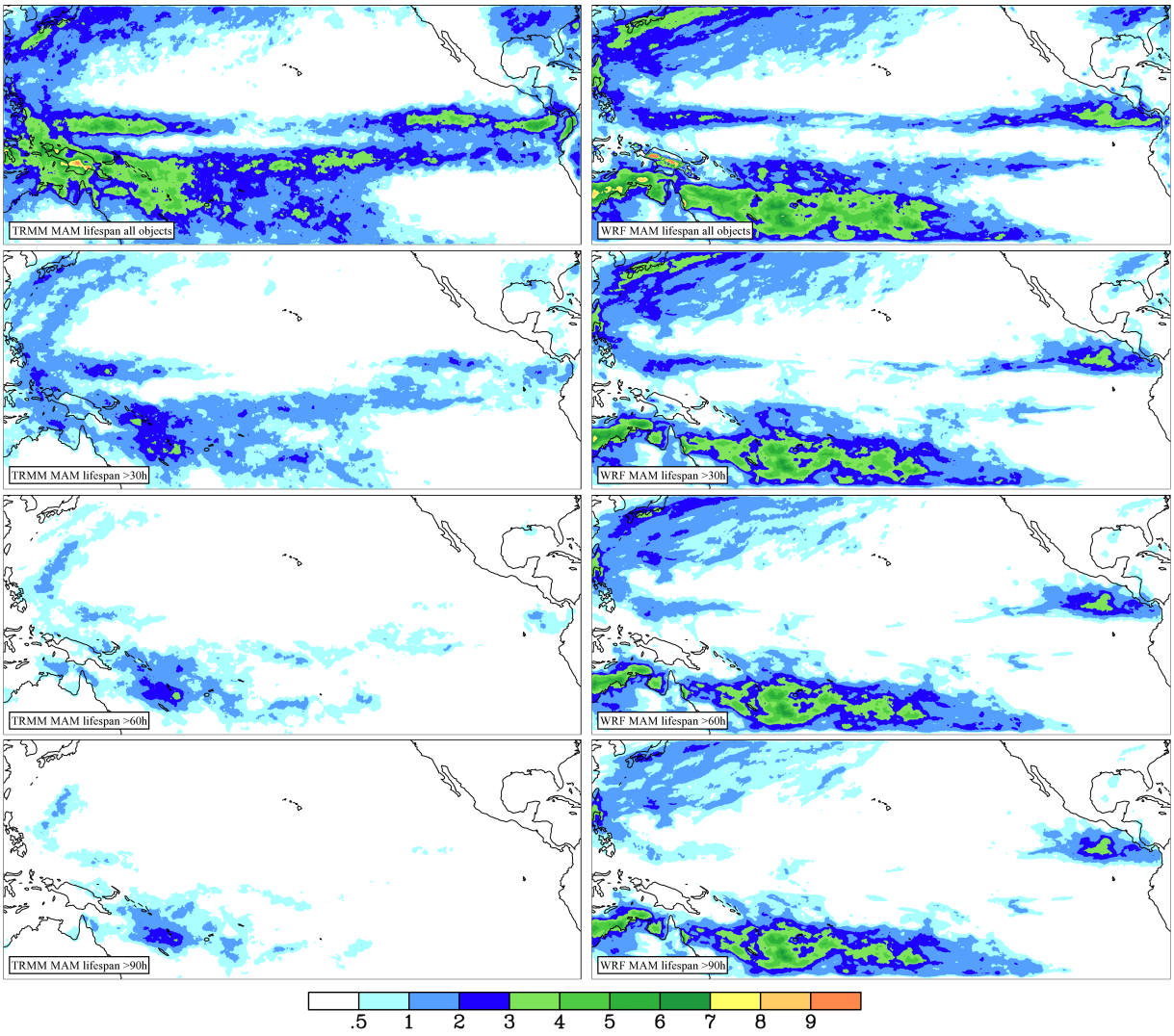


Figure AIV-5: Object frequency [%] considering lifespan for TRMM (left) and PESIANN (right) for MAM season in period from 1st January 1998 to 31st December 2000. Top: all objects, 2nd row: lifespan longer than 30h, 3rd row: lifespan longer than 60h and bottom: lifespan longer than 90h.

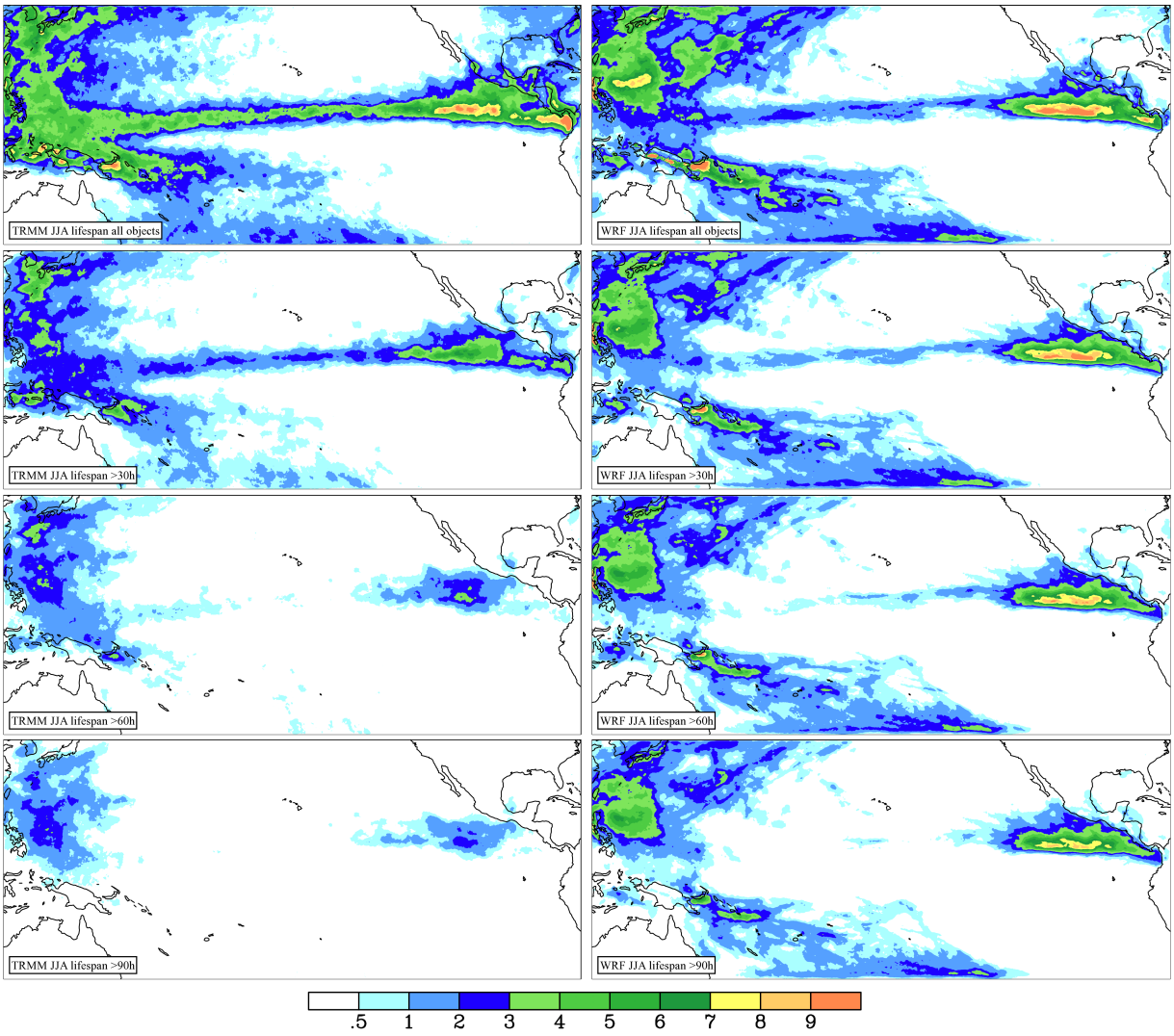


Figure AIV-6: Object frequency [%] considering lifespan for TRMM (left) and PESIANN (right) for JJA season in period from 1st January 1998 to 31st December 2000. Top: all objects, 2nd row: lifespan longer than 30h, 3rd row: lifespan longer than 60h and bottom: lifespan longer than 90h.

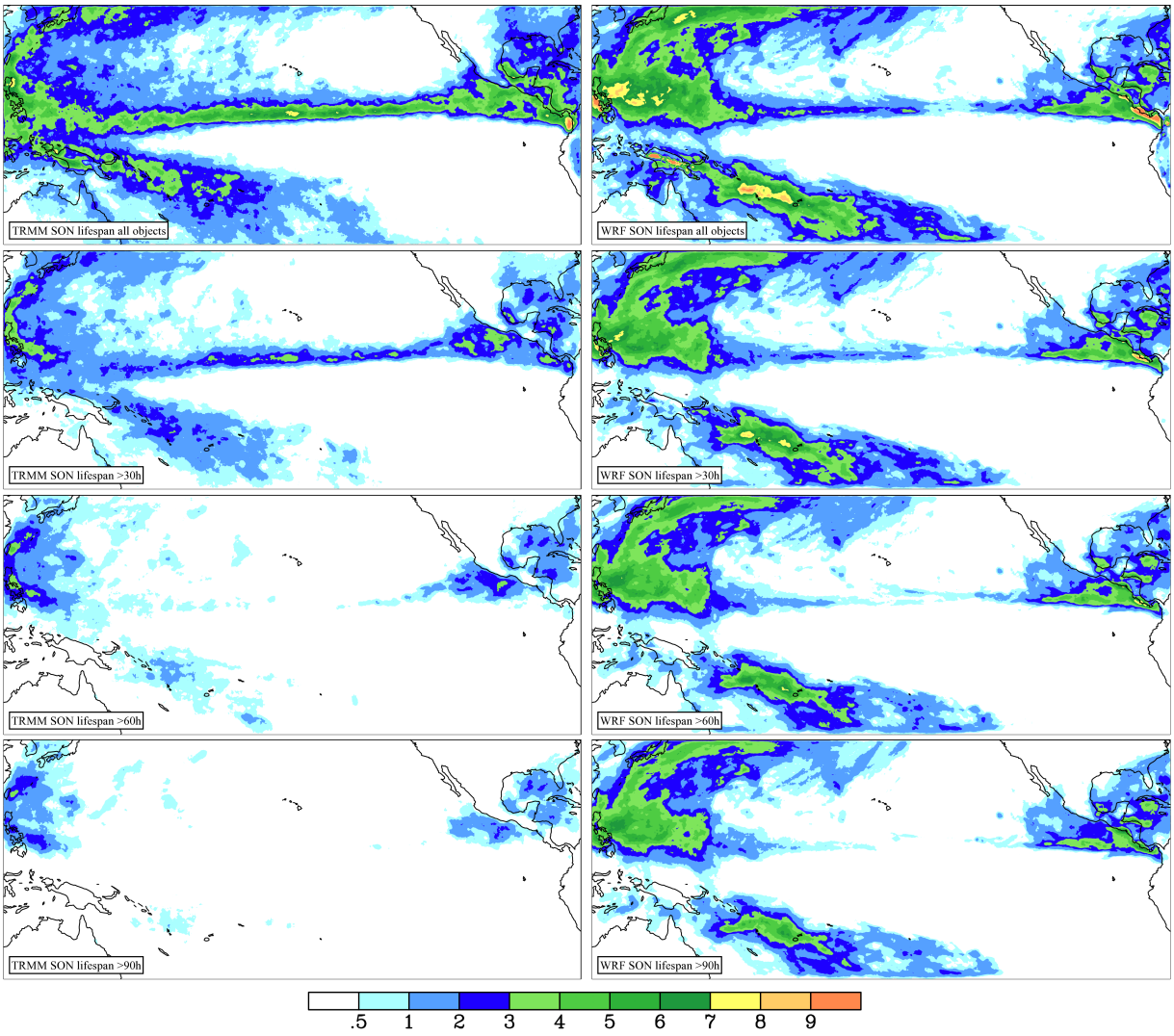


Figure AIV-7: Object frequency [%] considering lifespan for TRMM (left) and PESIANN (right) for SON season in period from 1st January 1998 to 31st December 2000. Top: all objects, 2nd row: lifespan longer than 30h, 3rd row: lifespan longer than 60h and bottom: lifespan longer than 90h.

7 DALJŠI SLOVENSKI POVZETEK

7.1 Uvod

Diagnostična analiza je osnovno orodje pri razvoju sistemov za opazovanje in napovedovanje vremena in služi za vrednotenje kvalitete vremenskih simulacij. Tradicionalna sinoptična analiza poteka preko identifikacije različnih dobro definiranih struktur, kot so fronte in konvektivni sistemi. Kvantitativno vrednotenje vremenskih simulacij na takšen način pa se je izkazalo za težavno. V zadnjem času so se razvile metode, ki temeljijo na uporabi objektov in omogočajo kvalitativno in kvantitativno vrednotenje prostorske organiziranosti in oblike posameznih vremenskih struktur. Eno takšnih metode je vpeljal Davis (2006a,b) – v splošnem znano pod imenom MODE (Method for Object-Based Diagnostic Evaluation; Davis 2009). MODE definira objekt kot zaključeno horizontalno območje v katerem količina akumuliranih padavin presega določen prag. Z namenom, da se oblike objektov čim bolj ujemajo z oblikami, kot bi jih narisal človek, se meje objektov izgladijo s pomočjo konvolucije. Na ta način poizkuša metoda simulirati, kako bi analizo vizualno opravil človek.

Šele v zadnjem času je možno izvesti klimatološko statistiko padavin z uporabo objektov, saj prej ni bilo na voljo dovolj kvalitetnih podatkov. Sedaj so na voljo različni satelitski padavinski produkti v visoki prostorski in časovni ločljivosti (0,25 stopinje, tri ure), ki pokrivajo trope in srednje geografske širine (40° S do 40° N). To so TRMM 3B42 (Huffman 2007), PERSIANN (Hsu 1997) in CMORPH (Joyce 2004). Kar nekaj študij satelitskih podatkov z uporabo objektov je bilo že narejenih v preteklosti, na primer Welch (1988), Machado (1998), Mapes in Houze (1993) in Wilcox (2003), vendar so te študije uporabljale podatke iz enega samega satelita, medtem ko so vsi novejši satelitski produkti pridobljeni na podlagi podatkov iz več satelitov ter zaradi tega omogočajo boljšo časovno in prostorsko ločljivost kot tudi pokrivajo večje območje.

Tropski predeli Zemlje igrajo ključno vlogo pri globalnem hidrološkem ciklu, saj v tropih pade 75 % celotne količine padavin na Zemlji. Zato je razumevanje procesov v teh predelih bistvenega pomena. Kljub temu, da je bilo v to smer narejenih že veliko študij, pa natančno razumevanje nekaterih procesov v tropih še ni znano, na primer razlaga oscilacije Madden-Julian (Madden and Julian 1971), intertropske konvergenčne cone ali El Niña.

7.2 Metodologija

7.2.1 Osnovna metoda MODE

Osnovna ideja metode, ki sem jo razvil, temelji na metodi MODE (Davis 2006a,b, Davis 2009). Ta metoda je sestavljena iz treh korakov: i) konvolucije, ii) zanemarjanja padavin pod določenim pragom in iii) identifikacije objektov. Pri tem osnovna metoda ne poizkuša najti povezave med objekti v različnih časovnih intervalih (vsak časovni interval se obravnava posebej).

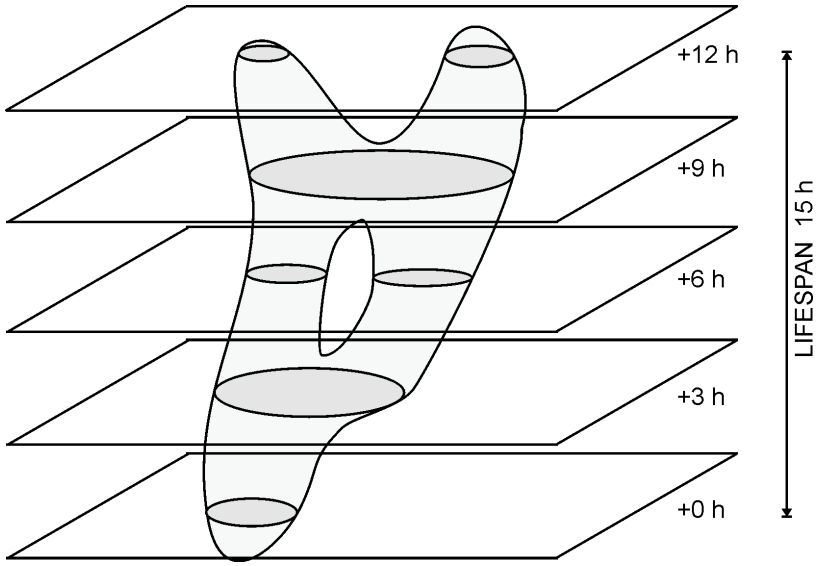
Konvolucija se izvede kot drseče povprečje v dveh dimenzijah z radijem R_{s} (ki ga imenujemo tudi radij konvolucije). Vse vrednosti padavin v mrežnih točkah $Z_{\mathrm{x},y}$, ki se nahajajo znotraj radija R_{s}, se povprečijo:

$$\widetilde{Z} = \frac{1}{N_{\mathrm{ell}}} \sum_{-R_{\mathrm{s}}}^{R_{\mathrm{s}}} \sum_{-R_{\mathrm{s}}}^{R_{\mathrm{s}}} Z_{x,y} \times \begin{cases} 1; & \text{če je znotraj radija} \\ 0; & \text{če je zunaj radija} \end{cases} , \tag{7.1}$$

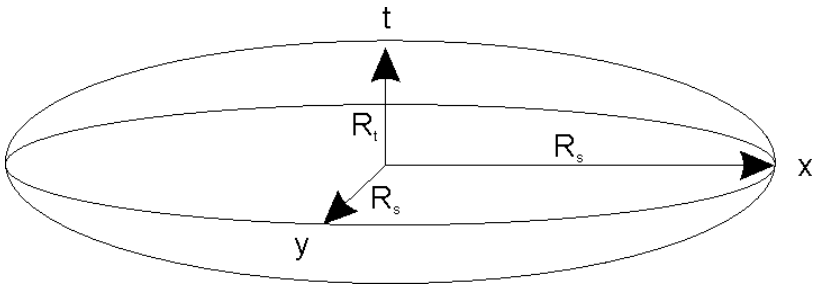
kjer je N_{ell} število točk znotraj krožnega območja z radijem R_{s} in $Z_{x,y}$ količina padavin na lokaciji (x,y). Uporaba konvolucije povzroči, da postane začetno polje padavin bolj gladko – zato dobijo tudi objekti bolj gladke meje. Drugi korak je zanemarjanje padavin pod določenim pragom TH, kar se preprosto doseže s tem, da se za vsako mrežno točko preveri ali je prag dosežen ali ne, in v primeru, če prag ni dosežen, se vrednost padavin spremeni na 0 mm. Če je vrednost TH visoka, bodo objekti predstavljali le območja z veliko padavinami, medtem ko bodo, če je vrednost TH manjša, objekti vsebovali tudi območja z manj padavinami. Zadnji korak je identifikacija objektov, ki se izvede z algoritmom, ki najde vsa vase zaključena območja padavin. Algoritem za vse mrežne točke sekvenčno preveri, ali je katera od sosednjih točk že del kakšnega prej identificiranega objekta, ali del kakšnega novega objekta ali ni del nobenega objekta. Ko so vsi objekti identificirani, se lahko izračunajo njihove karakteristike, kot so lokacija centra ter velikost.

7.2.2 Nadgradnja metode MODE

Moj prispevek je bila nadgradnja osnovne MODE metode tako, da omogoča spremljanje časovne evolucije objektov. Takšne objekte smo poimenovali časovni objekti (ČO). Časovni objekt sem definiral kot časovno zaporedje navadnih objektov, za katere se smatra, da so del enega padavinskega sistema (nevihte, konvektivnega sistema, fronte ...). Če se za trenutek zanemari razliko med prostorsko in časovno dimenzijo, si lahko časovni objekt predstavljamo kot zaključeno območje padavin nad določenim pragom v treh dimenzijah (dveh horizontalnih prostorskih ter eni časovni dimenziji). Oblika ČO je lahko zapletena – npr. objekt se lahko razdeli, se združi ali pa ima luknje kot objekt na Sliki S–1. Življenjska doba objekta je definirana kot čas od njegove prve pojave do izginotja. Če je ČO prisoten le v enem časovnem intervalu je njegova življenjska doba enaka trajanju časovnega intervala.



Slika S–1: Primer časovnega objekta z življenjsko dobo 15 ur.



Slika S–2: Konvolucijski elipsoid

Gibanje časovnega objekta smo definirali kot gibanje njegovega geometrijskega središča.

Z vpeljavo časovnih objektov se drugače izvaja tudi konvolucija ter identifikacija objektov. Konvolucija se sedaj izvaja v treh dimenzijah, v dveh prostorskih z radijem R_s ter v eni časovni dimenziji z radijem R_t. Tako imamo namesto enega radija konvolucije sedaj dva, ki nam definirata konvolucijski elipsoid (Slika S–2):

$$\frac{x^2}{R_\mathrm{s}^2} + \frac{y^2}{R_\mathrm{s}^2} + \frac{t^2}{R_\mathrm{t}^2} = 1, \tag{7.2}$$

kjer sta x, y prostorski in t časovna dimenzija. Za izvedbo konvolucije je zopet potrebno povprečiti vse vrednosti znotraj konvolucijskega elipsoida:

$$\widetilde{Z} = \frac{1}{N_\mathrm{ell}} \sum_{-R_\mathrm{t}}^{R_\mathrm{t}} \sum_{-R_\mathrm{s}}^{R_\mathrm{s}} \sum_{-R_\mathrm{s}}^{R_\mathrm{s}} Z_{x,y,t} \times \begin{cases} 1; & \text{če je znotraj elipsoida} \\ 0; & \text{če je zunaj} \end{cases}, \tag{7.3}$$

kjer je N_ell število vseh mrežnih točk znotraj elipsoida (izračuna se lahko vnaprej) in $Z_{x,y,t}$ količina padavin na lokaciji (x, y) in v času t. Računska za-

htevnost konvolucije se lahko izboljša za faktor $6/\pi$ (razmerje med volumnom kocke in krogle), če se sešteje le tiste mrežne točke, ki so znotraj elipsoida.

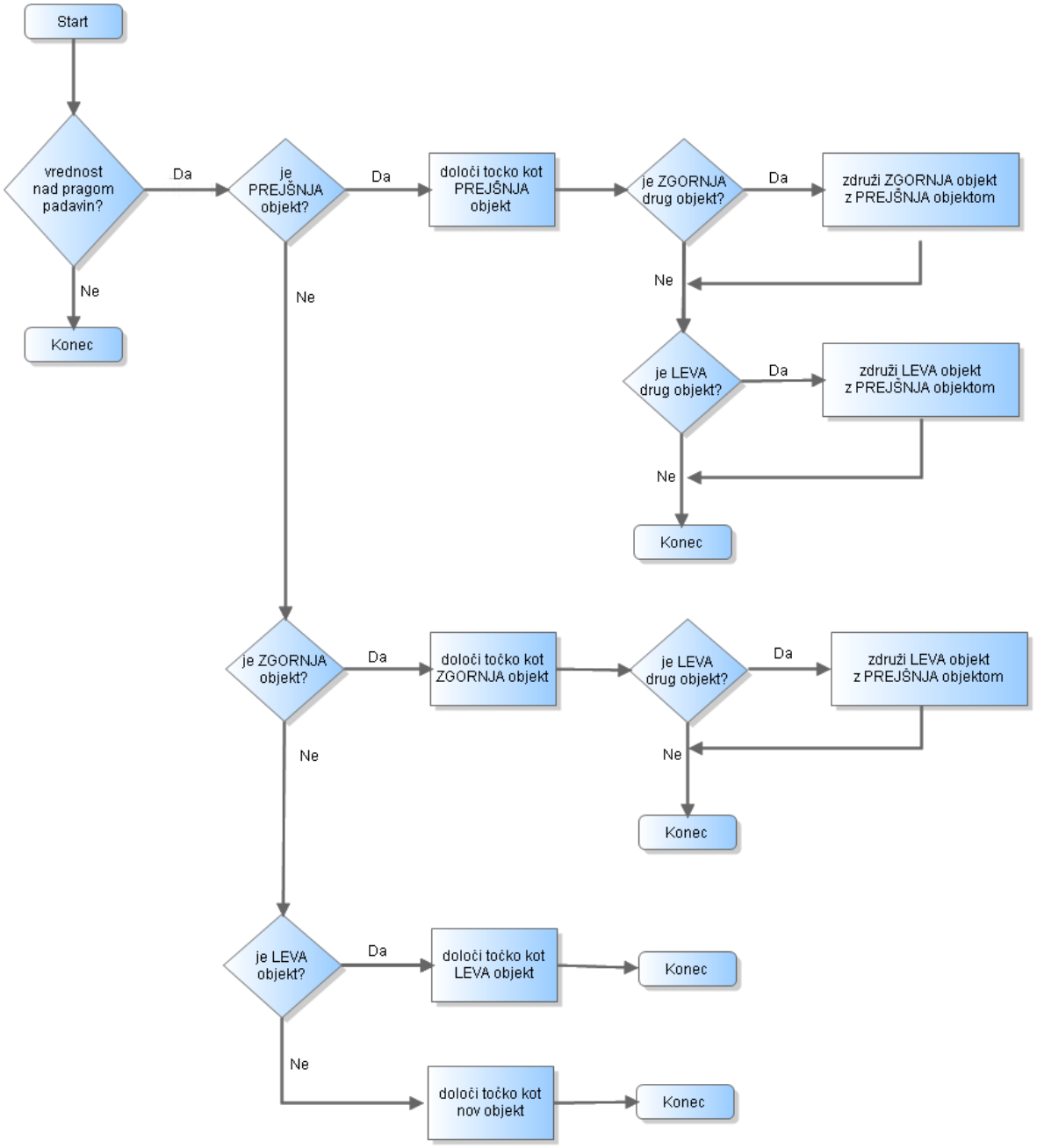
Po konvoluciji se zopet zanemari vse padavine, ki ne dosežejo določenega praga. To se doseže z zaporednim preverjanem vseh mrežnih točk, pri čemer se v primeru, da količina padavin ne doseže praga, mrežna točka deklarira kot ne-objekt.

Identifikacija objektov sedaj poteka v treh dimenzijah, kar pomeni, da moramo polja iz vseh (predhodnih) časovnih intervalov obdržati v spominu računalnika. Če je število časovnih intervalov zelo veliko, to ni možno zaradi fizičnih omejitev spomina računalnika. Problem lahko rešimo tako, da časovne intervale obravnavamo sekvenčno in hkrati v spominu obdržimo le nekaj predhodnih intervalov. Število zaporednih časovnih intervalov, ki jih moramo naenkrat obdržati v računalniškem spominu, je odvisno od tega, kakšna je maksimalna življenjska doba objekta v podatkovni bazi, ki je odvisna tudi od parametrov R_{s}, R_{t} in TH.

Identifikacija časovnih objektov poteka tako, da se združuje mrežne točke, ki po konvoluciji presežejo prag padavin in so hkrati sosednje, bodisi v času ali prostoru (ali obeh). Za vsako točko v trenutnem časovnem intervalu se izvede procedura prikazana v programski shemi na Sliki S–3.

Združevanje časovnega objekta A z B se izvede tako, da se pri trenutnem in predhodnih časovnih intervali preveri, če je objekt A že obstajal in se pri vseh mrežnih točkah, ki so del objekta A, indeks spremeni v indeks objekta B. Časovna zahtevnost združevanja objektov je lahko velika, vendar se lahko izboljša s tem, da se pregleda le toliko časovnih intervalov v preteklost, v katerih je objekt A še obstajal. Npr. če je objekt A že obstajal v časovnem intervalu -6 h, v časovnem intervalu -9 h pa ne, ni potrebno preverjati časovnega intervala -12 h (ali še bolj nazaj v preteklost), saj objekt ne more biti vase zaključeno območje padavin, če ne obstaja v vseh vmesnih časovnih intervalih.

Ko se za trenuten časovni interval zaključi procedura identifikacije objektov, se lahko ugotovi število ČO. To se lahko izvede na več načinov. Najpreprostejši način je, da se sekvenčno preveri vse mrežne točke, pri čemer se pri vsaki točki preveri ali je na tej lokaciji nov objekt, star objekt (objekt, ki smo ga že našli pri kateri od mrežnih točk) ali pa ni objekta. Izkaže se, da je časovna zahtevnost te metode velika, če imamo v časovnem intervalu veliko število objektov (npr. 1000), saj je za vsako mrežno točko, kjer je bil najden objekt, potrebno preveriti listo že najdenih objektov. Časovno zahtevnost se lahko zelo omili s tem, da si algoritem zapomni indeks objekta pri prejšnji mrežni točki. Na ta način ni potrebno preverjati liste vseh najdenih objektov pri mrežnih točkah, ki so del objekta (razen pri mrežnih točkah, ki so na mejah objektov).



Slika S–3: Programska shema za identifikacijo objektov, ki se izvede v vsaki mrežni točki. Izraz PREJŠNJA se nanaša na mrežno točko na isti lokaciji vendar v predhodnem časovnem intervalu. Izraza ZGORNJA in LEVA pa se nanašata na severno in zahodno sosednjo točko v trenutnem časovnem intervalu.

Tabela S–1: Izbrane karakteristike analiziranih satelitskih padavinskih podatkovnih baz.

	TRMM	**PERSIANN**
Ime (angleško)	TRMM Multi-satellite precipitation analysis (TMPA, a.k.a. 3B42)	Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN)
Inštitucija	GSFC (G. Huffman)	UC Irvine (K.-L. Hsu)
Vhodni podatki	infrardeče sevanje iz geostacionarnih satelitov, mikrovalovno sevanje iz SSM/I, TRMM, AMSU, AMSR inštrumentov na polarno-orbitalnih satelitih ter podatki iz dežemerov	infrardeče sevanje iz geostacionarnih satelitov in mikrovalovno sevanje iz TMI inštrumenta na TRMM polarno orbitalnem satelitu
Metoda	Algoritem, ki iz vhodnih podatkov oceni padavine	Nevronska mreža, ki iz vhodnih podatkov oceni padavine
Območje	50° S – 50° N	60° S – 60° N
Začetni datum	1. januar 1998	1. marec 2000
Končni datum	31. december 2005	31. december 2005
Prostorska ločljivost	0,25°	0.,5°
Časovna ločljivost	3 h	1 h

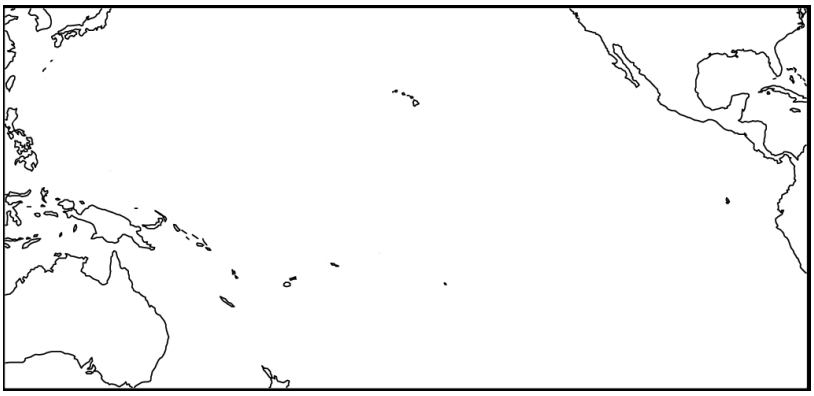
7.3 Analiza in Primerjava padavin dveh satelitskih produktov

V tem poglavju bom predstavil analizo in primerjavo padavinskih satelitskih produktov TRMM in PERSIANN v tropskem Pacifiku. Najprej je bila narejena analiza manjkajočih podatkov ter primerjava povprečnih letnih in sezonskih akumulacij padavin. Temu je sledila analiza padavin z uporabo osnovne metode MODE (sekcija 7.2.1) ter nato še nadgrajene metode (sekcija 7.2.2). Pri obeh metodah je bila najprej narejena analiza občutljivosti, ki je omogočila izbiro optimalnih parametrov R_s, R_t in TH, potem pa je bila izvedena še analiza celotnega obdobja padavinskih podatkov v tropskem Pacifiku.

7.3.1 Predstavitev satelitskih podatkov in območja, analiza manjkajočih vrednosti in analiza padavinskih akumulacij

Za analizo in primerjavo padavin v tropskem Pacifiku smo si izbrali dve izmed treh podatkovnih baz, ki smo jih omenili v Uvodu. To sta TRMM 3B42 in PERSIANN, ki smo ju izbrali zaradi njune dobre razpoložljivosti meritev v daljšem časovnem obdobju. Nekaj karakteristik podatkovnih nizov je naštetih v Tabeli S–1.

Za analizo padavin smo si izbrali območje tropskega Pacifika, ki je zelo zanimivo področje zaradi razsežnih padavinskih pojavov, povezanih z intertropsko



Slika S–4: Območje analize v tropskem Pacifiku.

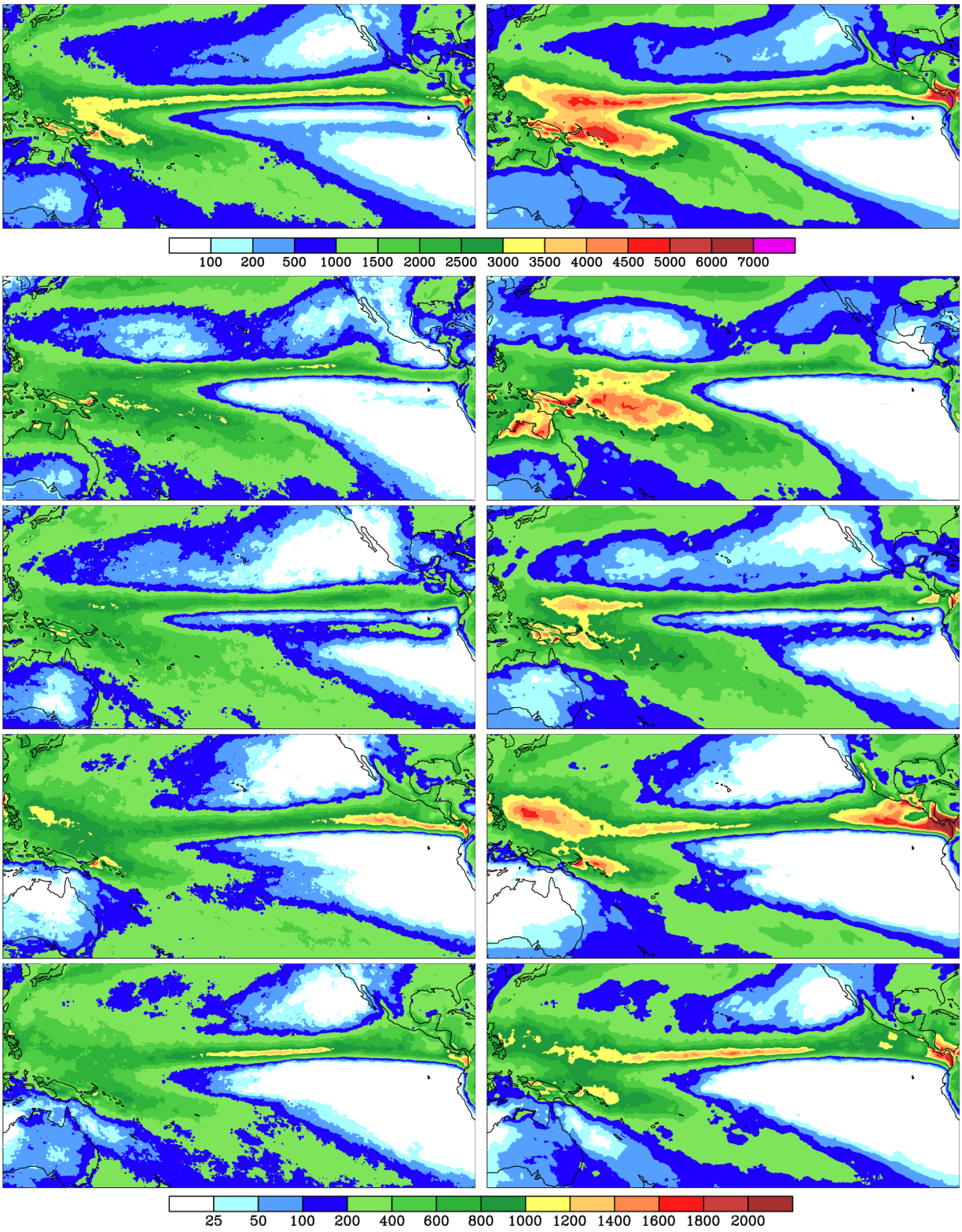
konvergenčno cono (ITCZ) ter El-Niñom/La-Niñjo. Območje analize obsega: 120° E – 285° E in 39° S – 39° N (Slika S–4). Primerjavo padavinskih produktov smo naredili za obdobje, ko sta bila oba produkta na voljo: 1. marec 2000 – 31. december 2005. Območje sestavljeno iz 622 × 313 mrežnih točk v 17,048 triurnih intervalih.

Manjkajoče vrednosti so predstavljale kar velik problem, saj jih je bilo na nekaterih območjih veliko. Tak primer je območje na sredini Pacifika, kjer je bilo manjkajočih tudi do 70 % enournih vrednosti v podatkovni bazi PERSIANN. V podatkovni bazi TRMM je bilo manjkajočih vrednosti malo in niso predstavljale večjega problema (manjkajočih vrednosti ni bilo nikjer več kot 5 %). Problem s pogostimi manjkajočimi vrednosti v podatkovni bazi PERSIANN smo rešili tako, da smo na določen način sešteli tri enourne akumulacije v eno triurno akumulacijo (kar je bilo potrebno narediti, da smo jih lahko primerjali s podatkovno bazo TRMM, ki je imela triurne akumulacije padavin). Namreč, če je vsaj ena izmed treh enournih akumulacij bila med podatki, smo njeno vrednost privzeli za vse tri ure in rezultat pomnožili z tri, da smo dobili triurne akumulacije. Poleg tega pa smo manjkajoče vrednosti v obeh podatkovnih bazah še uskladili. To je bilo narejeno tako, da če je imela vsaj ena izmed podatkovnih baz manjkajočo vrednost smo manjkajočo vrednost predpisali tudi drugi bazi. Npr. če je vrednost manjkala v bazi TRMM, smo tudi v bazi PERSIANN za ta podatek definirali manjkajočo vrednost. Na ta način je bilo manjkajočih podatkov v obeh bazah enako in nikjer niso presegali pogostosti 10 %.

Za boljše razumevanje razporeditve padavin smo najprej analizirali povprečne letne in sezonske akumulacije padavin. Analizo smo naredili za celotno obdobje od 1. marca 2000 do 31. decembra 2005 in rezultati so predstavljeni na Sliki S–5.

7.3.2 Analiza občutljivosti

Analiza občutljivosti smo izvedli predvsem za parametre R_{s}, R_{t} in TH. Vrednosti teh parametrov drastično vplivajo na število, porazdelitev in karakteristike



Slika S–5: Povprečne letne in sezonske padavine (mm) za TRMM (levo) in PERSIANN (desno). Zgoraj: povprečne letne padavine. Od druge vrste navzdol: povprečne sezonske padavine: DJF, MAM, JJA in SON.

identificiranih objektov. Analiza smo naredili tako, da smo iz celotnega obdobja naključno izbrali 100 triurnih časovnih intervalov, nato pa smo za teh 100 izbranih intervalov identificirali objekti, pri čemer smo se testirali različne kombinacije vrednosti parametrov. Zaloga vrednosti za parameter R_s je segala od 0 do 5 geografskih stopinj, za parameter R_t od 0 do 15 h ter za parameter TH od 0 do 100 mm/3 h.

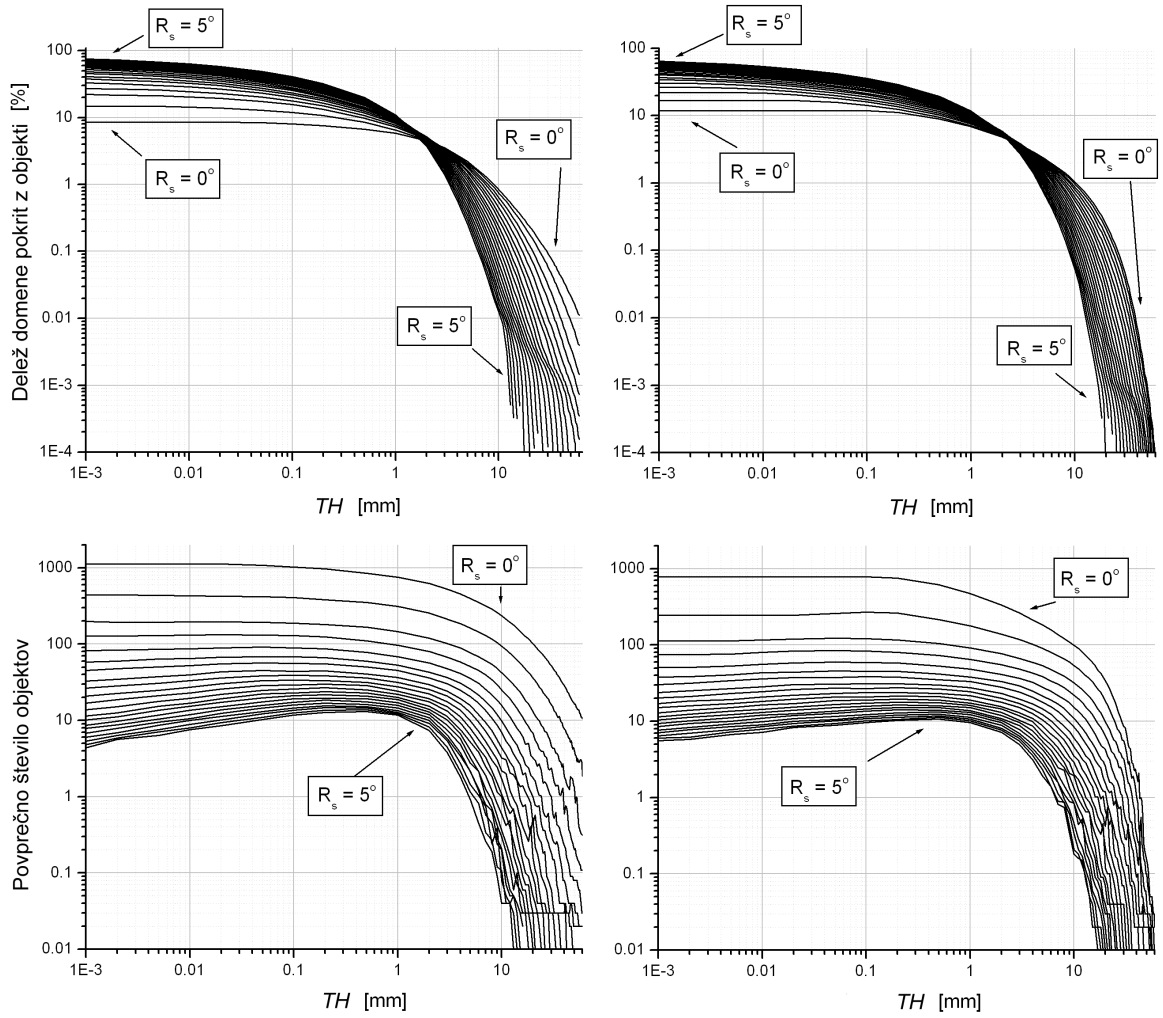
Konvolucija na padavinsko polje deluje tako, da ga izgladi. Po konvoluciji se velikost padavinskih sistemov nekoliko poveča, hkrati pa se njihov padavinski maksimum zmanjša (čim večji je R_s, večji bodo sistemi in manjši bo maksimum). Obenem pa se skupna količina padavin na celotnem območju ohranja (razen blizu robov območja), saj konvolucija ne proizvede ali zmanjša količine skupnih padavin na območju.

Najprej smo analizirali pokritost z objekti (delež površine območja). Rezultati so prikazani na Sliki S–6 zgoraj, kjer sta oba zgornja grafa podobne oblike. Pri TH okoli 2–3 mm/3 h je invariantna vrednost, kjer izbira R_s ne vpliva na pokritost območja z objekti. Invariantna vrednosti TH razdeli graf na dve različni območji. Če je vrednost TH manjša kot invariantna vrednost, se bo pri povečanju R_s vedno povečala tudi pokritost območja z objekti. Nasprotno velja če je vrednost TH večja kot invarantna vrednost.

Pri $TH = 0$ mm/3 h in $R_\mathrm{s} = R_\mathrm{t} = 0$ (brez praga za padavine in konvolucije) objekti v povprečju pokrivajo okoli 10 % območja. To pomeni, da v povprečju padavine v naravi ne pokrivajo več kot 10 % območja, zato ne bi bilo smotrno dovoliti metodi, da preseže to vrednost. Tako zavrnemo vse kombinacije parametrov R_s, R_t in TH, ki prekrivajo več kot 10 % območja. V teh primerih so padavinski sistemi preveč povečani zaradi močnega vpliva konvolucije. Tako so v primeru na Sliki S–6 neuporabne vrednosti TH od 0 do 0,25 mm/3 h, vrednosti od 0,25 mm/3 h do invariantne vrednosti so pogojno uporabne (v kombinacijami s pravimi vrednostmi R_s). Če pa je vrednost TH večja kot invariantna vrednost pa je pokritost območja z objekti vedno pod 10 %.

V primeru ko ni konvolucije, je pokritost območja z objekti večja v podatkovni bazi PERSIANN kot v TRMM, če je prag padavin majhen (< 1 mm/3 h). Slednje nakazuje, da na splošno padavine zavzemajo večje območje v podatkovni bazi PERSIANN (12 % v primerjavi z 9 % v podatkovni bazi TRMM). Če je prag padavin višji (od 1 do 30 mm/3 h), imata obe podatkovni bazi skorajda enako pokritost. Če pa je prag še višji (od 30 do 60 mm/3 h) pa padavine v podatkovni bazi TRMM zavzamejo veliko večje območje kot v podatkovni bazi PERSIANN. To kaže na to, da imamo v podatkovni bazi TRMM večja območja z velikimi akumulacijami.

Slika S–6 spodaj prikazuje povprečno število objektov. Pri $TH = 0$ mm/3 h in $R_\mathrm{s} = R_\mathrm{t} = 0$ (brez praga za padavine in konvolucije) je povprečno število objektov blizu 1000 v obeh podatkovnih bazah. Pri drugih vrednostih parametrov TH, R_s in R_t pa je običajno več objektov v podatkovni bazi TRMM. To kaže na to, da so padavinska polja v podatkovni bazi PERSIANN v povprečju bolj gladka, saj se identificira manj objektov kot v podatkovni bazi TRMM. Če je število objektov zelo majhno (< 3), črte na Sliki S–6 spodaj niso več gladke, kar je posledica



Slika S–6: Pokritost območja z objekti [%] (zgoraj) in povprečno število objektov (spodaj): za TRMM (levo) in PERSIANN (desno). Interval vrednosti R_s sega od 0 do 5° v razmakih velikosti 0,25° (ena mrežna točka). Slike so za primer ko je $R_\mathrm{t} = 0$ (brez konvolucije v časovni dimenziji).

majhnega vzorca (izbrali smo le 100 triurnih intervalov). Prav tako je iz Slike S–6 spodaj razvidno, da večji R_s vedno pomeni manjše število objektov. Razlog je v tem, da večji R_s pomeni močnejši efekt glajenja s konvolucijo, ki ima za posledico več združevanja bližnjih objektov, kar se kaže v zmanjšanju števila objektov. V naši raziskavi smo se odločili omejiti na kombinacije parametrov R_s, R_t in TH, kjer povprečno število objektov ne presega 100 objektov. To smo naredili zato, da simuliramo subjektivno identifikacijo objektov, ki bi jo izvedel človek, ki ponavadi v padavinskem polju ne bi našel več kot 100 objektov.

Če poleg konovlucije v prostorskih dimenzijah izvedemo še konvolucijo v časovni dimenziji, dobimo tako imenovan efekt »spillover«, ki povzroči, da se objekti iz sosednjih časovnih intervalov razširijo tudi v tiste časovne intervale, v katerih

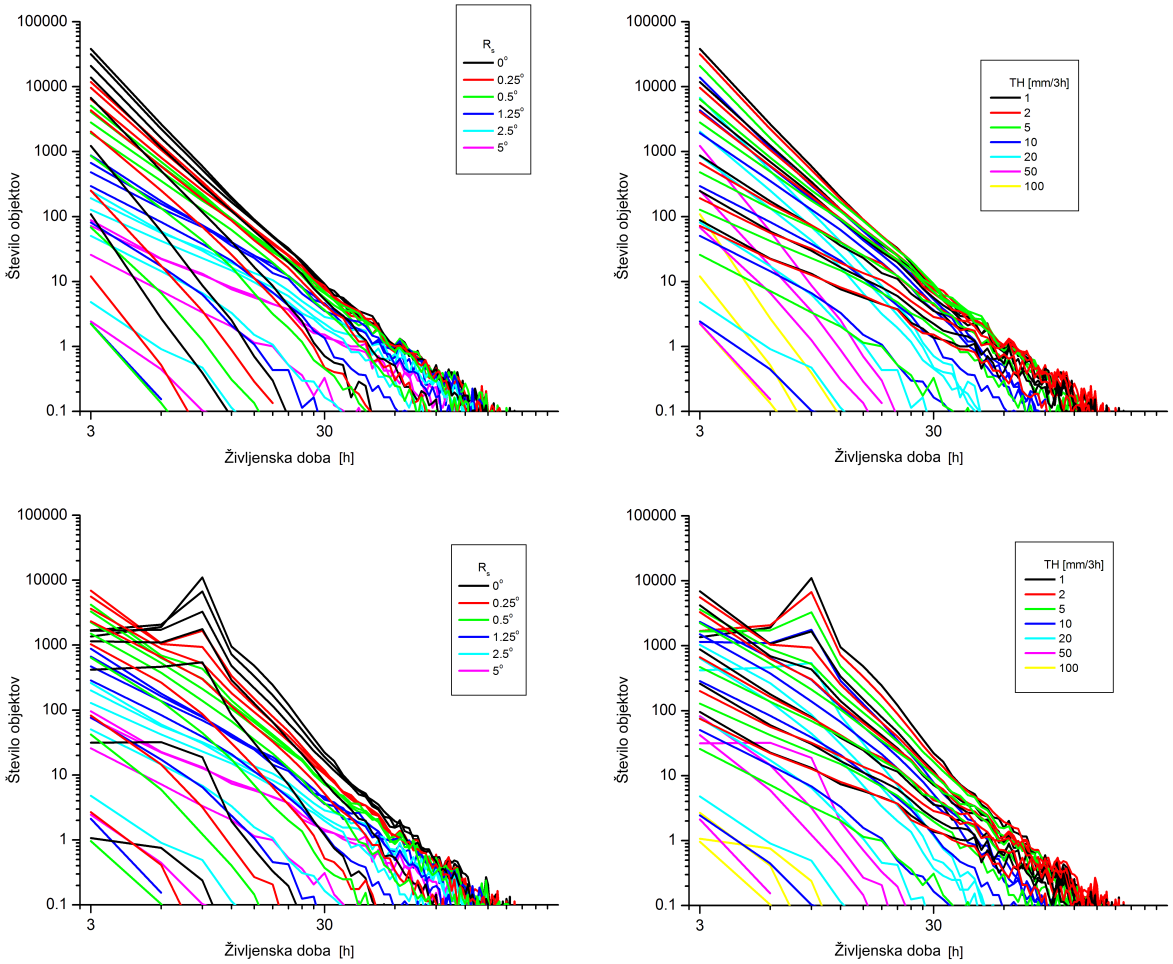
tam prej ni bilo padavin. To je zagotovo neželen efekt, saj umetno podaljšuje življenjsko dobo objektov. Primer tega je prikazan na Sliki S–7, kjer je prikazano povprečno število objektov glede na življenjsko dobo in kjer se vidi velika razlika med spodnjima in zgornjima slikama. Na spodnjih slikah ($R_\mathrm{t} = 3$ h) se izrazito vidi maksimum pri življenjski dobi 9 h, ki ga na zgornjih slikah ($R_\mathrm{t} = 0$ h) ni bilo, in je posledica efekta »spillover«, ko se objekti z življenjsko dobo 3 h razširijo na dva sosednja časovna intervala tako da dobijo življenjsko dobo 9 h. Efekt »spillover« je prisoten vedno, ko imamo konvolucijo v časovni dimenziji, vendar je močnejši, če je padavinski prag manjši kot invariantna vrednost. Velikost R_t direktno vpliva na moč efekta »spillover«, saj večja vrednost pomeni tudi daljši časovni doseg efekta. Ker je efekt prisoten vedno, ko uporabimo konvolucjo v časovni dimenziji, smo se odločili, da konvolucija v časovni dimenziji ni smiselna. Tako smo za nadaljnjo analizo vedno uporabili $R_\mathrm{t} = 0$ h.

7.3.3 Analiza objektov v Pacifiku

Na podlagi analize občutljivosti smo izbrali vrednosti parametrov R_s, R_t in TH. Postopek izbire je potekal po naslednjih korakih:

1. Da bi se v celoti izognili efektu »spillover« nismo uporabili konvolucije v časovni dimenziji ($R_\mathrm{t} = 0$ h).
2. Želeli smo analizirati predvsem konvektivne padavinske sisteme, še posebej organizirano konvekcijo v tropih, ki jo pogosto spremljajo intenzivne padavine. Tako smo se odločili, da nas zanimajo le padavine s količino nad 8 mm/3 h.
3. Nato smo iz zgornjih grafov na Sliki S–6 odčitali, da padavine nad 8 mm/3 h (brez konvolucije) v povprečju pokrivajo okoli 1,12 % območja v podatkovni bazi TRMM in 1,41 % območja v podatkovni bazi PERSIANN.
4. Nato smo zopet is Slike S–6 našli kombinacijo vrednosti R_s in TH, ki odraža prej omenjene odstotke pokritosti območja ter hkrati v povprečju nima več kot 100 objektov. Tako smo ugotovili, da kombinacija vrednosti $R_\mathrm{s} = 0,75°$ in $TH = 7$ mm/3 h ustreza tem merilom, saj je povprečna pokritost območja z objekti 1,22 % za podatkovno bazo TRMM ter 1,44 % za podatkovno bazo PERSIANN, hkrati pa je povprečno število objektov v obeh podatkovnih bazah manjše od 40.

Analizo smo izvedli za celotno obdobje marec 2000 - december 2005. Najprej smo analizirali prostorsko porazdelitev objektov, kjer smo ugotovili, da je porazdelitev podobna kot akumulacije padavin, saj so področja kjer so akumulacije večje, tudi območja kjer so objekti bolj pogosti. Analizirali smo tudi prostorsko porazdelitev objektov glede na velikost. Ugotovili smo, da so manjši in srednje veliki objekti porazdeljeni podobno kot akumulacije padavin, področja kjer so

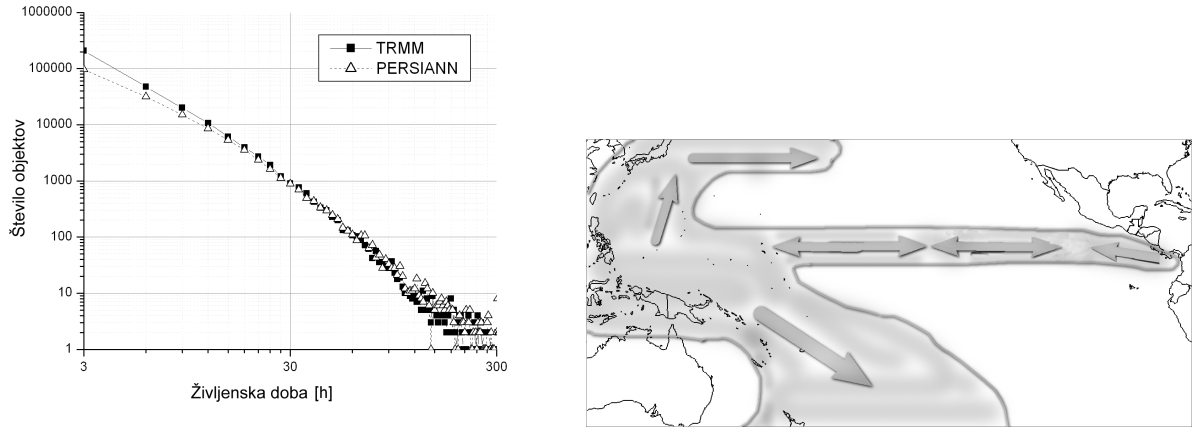


Slika S–7: Povprečno število objektov glede na življenjsko dobo na log-log grafu za $R_\mathrm{t} = 0$ h (zgoraj) in $R_\mathrm{t} = 3$ h (spodaj). Na levi strani so črte pobarvane glede na vrednost R_s medtem ko so na desni strani črte pobarvane glede na TH.

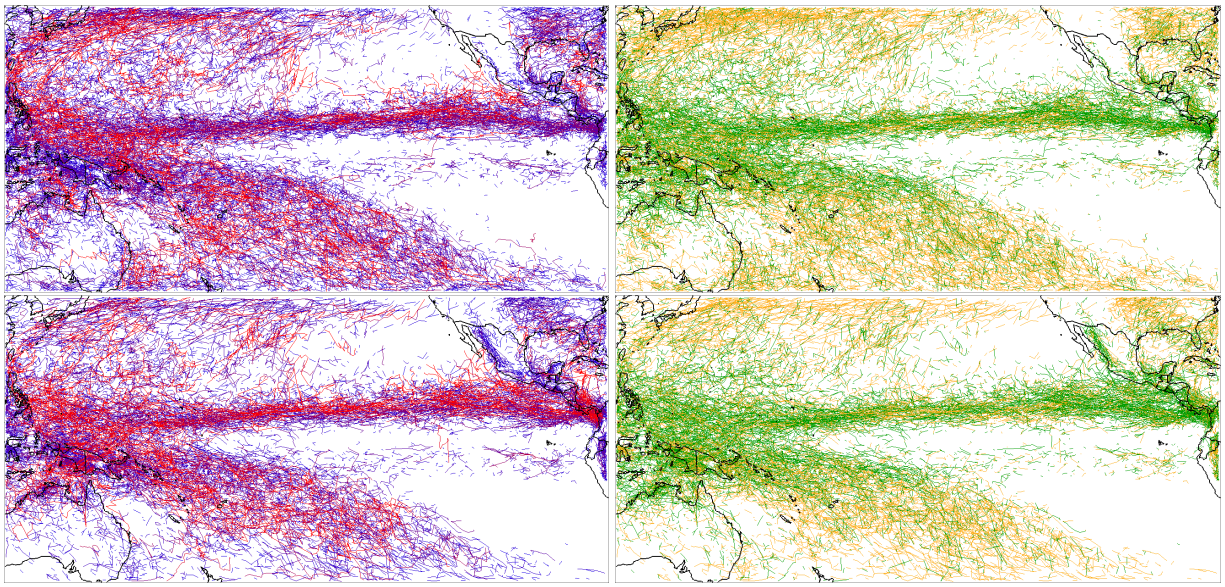
akumulacije večje, so tudi območja kjer so ti objekti bolj pogosti (slika ni prikazana v slovenskem povzetku). Ugotovili smo tudi, da se veliki objekti (večji kot ~ 348 tisoč km^2) nahajajo pretežno na zahodnem delu območja ter deloma tudi na vzhodnem.

Sledila je analiza števila objektov glede na življenjsko dobo, kar kaže Slika S–8 levo, iz katere je razvidno da je povezava linearna v log-log grafu, kar pomeni, da je povezava potenčna. Vidno je tudi, da je v podatkovni bazi PERSIANN manj objektov s krajšo življenjsko dobo (življenjska doba manj kot 9 h).

Izvedli smo tudi analizo trajektorij objektov. Slika S–9 kaže trajektorije za leto 2001. Trajektorije so očitno koncentrirane v določenih področjih. Porazdelitev trajektorij izgleda podobno kot porazdelitev akumulacij padavin. Največja koncentracija trajektorij dolgoživih objektov (rdeče trajektorije) je v ITCZ in v zmernih



Slika S–8: Levo: Število objektov glede na življenjsko dobo v log-log grafu. Desno: Shematski prikaz gibanja za dolgožive objekte (leto 2001).



Slika S–9: Trajektorije objektov za leto 2001 za TRMM (zgoraj) in PERSIANN (spodaj). Levo: trajektorije so obarvane glede na trajanje življenjske dobe (modro za krajšo in rdeče za daljšo). Desno: trajektorije so obarvane glede na gibanje v smeri zahod/vzhod. Če je gibanje imelo komponento v smeri zahoda, je segment trajektorije obarvan zeleno. Če je gibanje imelo komponento v smeri vzhoda, je segment trajektorije obarvan oranžno.

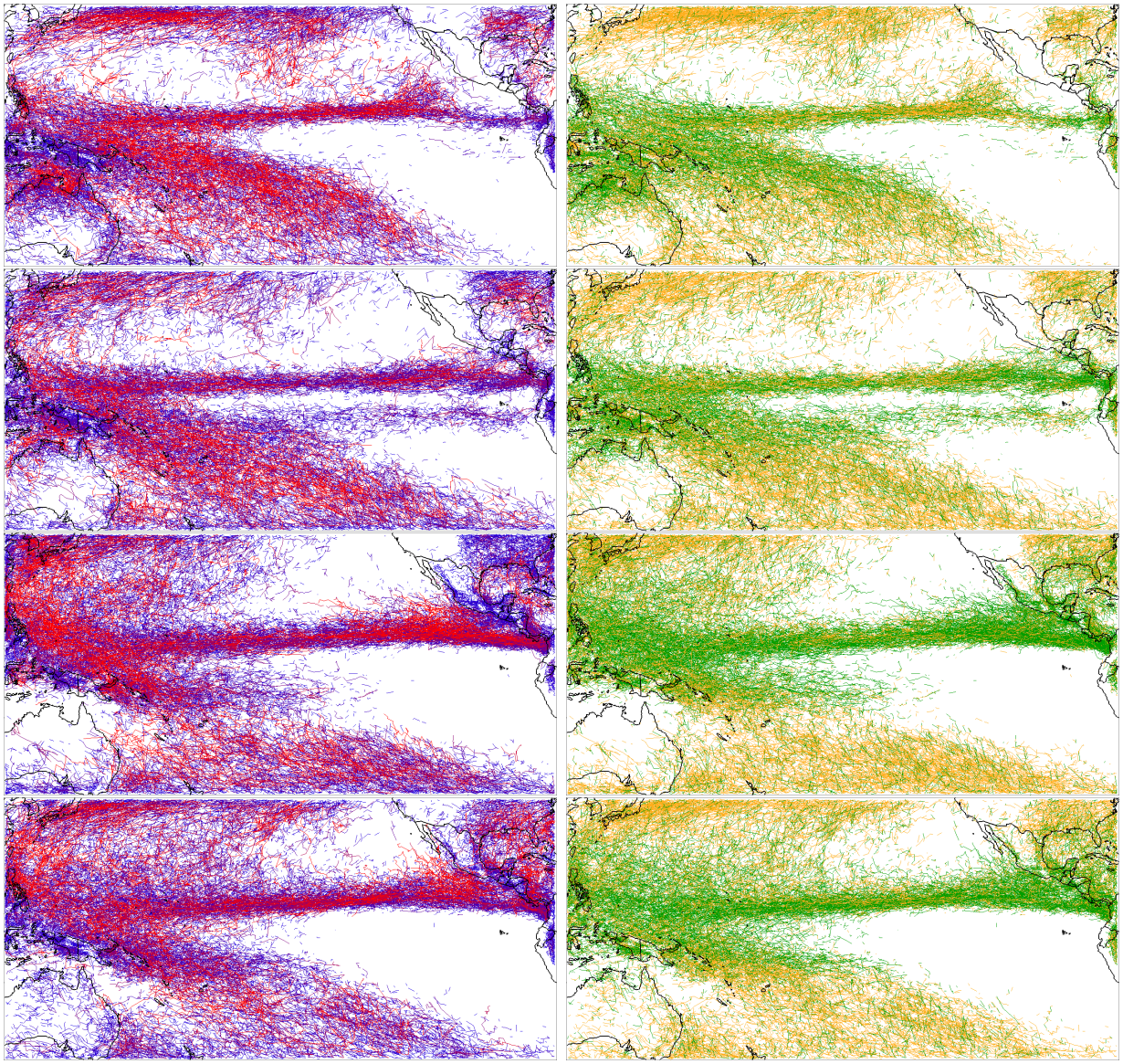
širinah na zahodni strani območja. Z nekoliko manjšo pogostnostjo se trajektorije dolgoživih objektov nahajajo tudi v severozahodnem delu območja in vzhodno od Avstralije. Med podatkovnima bazama je opazna razlika na vzhodnem koncu ITCZ, kjer je v podatkovni bazi PERSIANN precejšnje število dolgoživih objektov in skoraj nič dolgoživih objektov v podatkovni bazi TRMM.

Gibanje objektov vzdolž trajektorij je shematsko prikazano na Sliki S–8 desno. Gibanje v severnih in južnih predelih območja je predvsem proti vzhodu, medtem ko je gibanje v ITCZ proti vzhodu in proti zahodu, čeprav je gibanje proti zahodu bolj pogosto. V zahodnem in vzhodnem delu ITCZ pa je gibanje proti zahodu tudi prevladujoče.

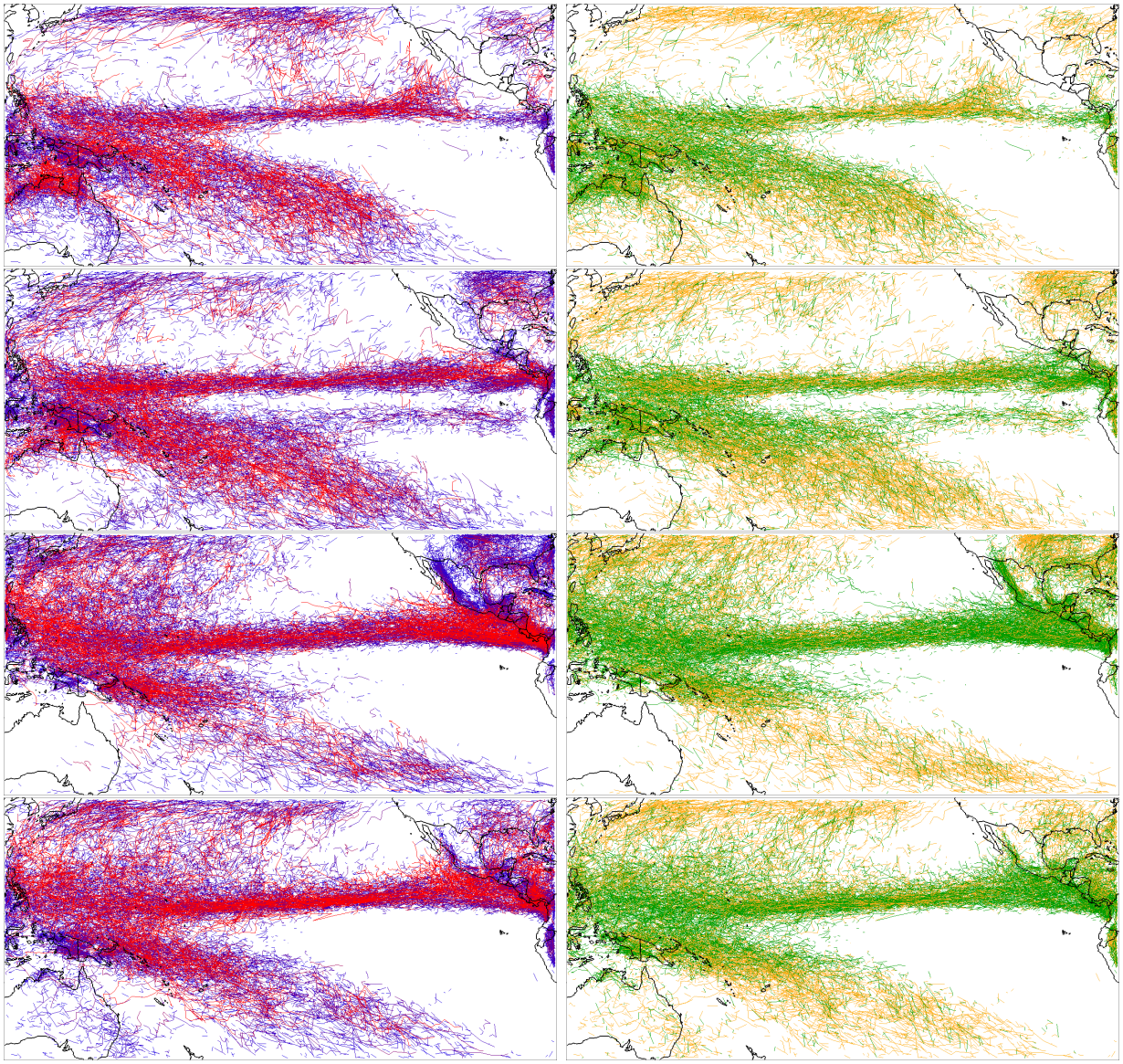
Podobno smo analizirali tudi sezonske trajektorije, ki so prikazane na Slikah S–10 (podatkovna baza TRMM) in S–11 (podatkovna baza PERSIANN). Sezonske trajektorije odražajo porazdelitev akumulacij padavin. V splošnem so podobne letnim trajektorijam, na primer v tem, da je v severnem in južnem delu območja gibanje pretežno v smeti proti vzhodu. Vendar pa obstaja tudi nekaj razlik, med katerimi je najbolj očitna da je gibanje v ITCZ pozimi in spomladi (DJF in MAM) bolj proti vzhodu, kar za poletje in jesen (JJA in SON) ne bi mogli trditi. V spomladanskem obdobju je v ITCZ prisotno precej manj objektov z dolgo življenjsko dobo. Sekundarni ITCZ, ki se spomladi pojavi južno od ekvatorja, je pretežno sestavljen iz objektov s kratko življenjsko dobo. Pozimi v vzhodnem delu ITCZ prevladujejo objekti s kratko življenjsko dobo. Vse zgoraj omenjene ugotovitve so enake za obe podatkovni bazi.

Na koncu smo analizirali še porazdelitev objektov v območju glede na življenjsko dobo (ni prikazano na slikah v slovenskem povzetku). Čeprav se med podatkovnima bazama pogostosti pojavljanja objektov na določenih območjih precej razlikujejo, smo ugotovili nekaj značilnosti, ki so skupne obema. Najbolj očitna je ta, da se običajno objekti z dolgo življenjsko dobo nahajajo v večjem območju na zahodnem delu območja in na manjšem območju v vzhodnem delu območja (zahodno od obale Mehike). V splošnem se objekti bolj pogosto pojavljajo v podatkovni bazi PERSIANN, kar bi lahko bil rezultat večjih akumulacij padavin prisotnih v podatkovni bazi PERSIANN, kar povzroči, da se identificira večje število objektov.

Analizirali smo tudi porazdelitev objektov v območju glede na sezone (ni prikazano na slikah v slovenskem povzetku). Po pričakovanjih se področje, kjer se pojavlja največ objektov, spreminja glede na sezono. Poleti in jeseni se največ objektov z dolgo življenjsko dobo nahaja na območju severno od Avstralije ter na manjšem območju blizu zahodne obale Mehike. Poleti se ti objekti pojavljajo bolj pogosto kot jeseni. Po drugi strani pa se v spomladanskem obdobju objekti z dolgo življenjsko dobo nahajajo predvsem na področjuvzhodno od Avstralije, medtem ko se v zimskem obdobju objekti nahajajo severo-vzhodno od Avstralije in v severnem delu Avstralije. Spomladi je pogostost teh objektov zelo nizka (pojavljajo se v manj ko 1 % časovnih intervalov).



Slika S–10: Sezonske trajektorije za podatkovno bazo TRMM. Levo: trajektorije so obarvane glede na trajanje življenjske dobe (modro za krajšo in rdeče za daljšo). Desno: trajektorije so obarvane glede na gibanje v smeri zahod/vzhod. Če je gibanje imelo komponento v smeri zahoda, je segment trajektorije obarvan zeleno. Če je gibanje imelo komponento v smeri vzhoda, je segment trajektorije obarvan oranžno. Od zgoraj navzdol: DJF (zima), MAM (pomlad), JJA (poletje) in SON (jesen).



Slika S–11: Sezonske trajektorije za podatkovno bazo PERSIANN. Levo: trajektorije so obarvane glede na trajanje življenjske dobe (modro za krajšo in rdeče za daljšo). Desno: trajektorije so obarvane glede na gibanje v smeri zahod/vzhod. Če je gibanje imelo komponento v smeri zahoda, je segment trajektorije obarvan zeleno. Če je gibanje imelo komponento v smeri vzhoda, je segment trajektorije obarvan oranžno. Od zgoraj navzdol: DJF (zima), MAM (pomlad), JJA (poletje) in SON (jesen).

7.4 Primerjava med satelitskimi in modelskimi padavinami

Naš osnovni namen ni bila samo primerjava in analiza padavin izmerjenih s satelitov, ampak tudi primerjava teh padavin s padavinami, ki jih dobimo s pomočjo numeričnega modela za napovedovanje vremena. Na ta način bi se naša objektna metoda uporabila kot orodje za verifikacijo kakovosti modelskih padavin. Odločili smo se, da bomo satelitske podatke TRMM primerjali s padavinami, ki so jih dobili v modelski simulaciji tropskega kanala z WRF (angleško: WRF model Tropical Channel Simulation), ki jo je izvedel National Center for Atmospheric Research (NCAR), s sedežem v Boulderju, ZDA. Modelska simulacija tropskega kanala z WRF je petletna simulacija klime v visoki ločljivosti na območju tropov in bližnje okolice.

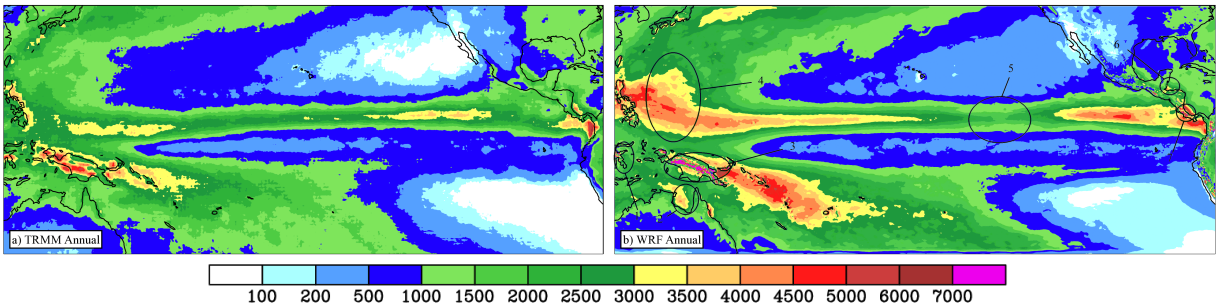
7.4.1 Kratka predstavitev simulacije na območju tropskega kanala z modelom WRF

Glavni namen simulacije je bila izboljšava v razumevanju kompleksne dvosmerne interakcije med različnimi skalami (globalno in regionalno skalo), ki igra zelo pomembno vlogo pri napovedovanju klime in vremena. Glavna cilja sta bila izboljšava klimatskih napovedi pri prehodu iz globalnih na regionalne skale ter izboljšava v razumevanju vpliva regionalnih procesov na globalne klimatske simulacije.

Za ta namen je bila narejena petletna simulacija (1996-2000) tropskega predela zemlje z uporabo modela WRF verzije 2.1. Območje analize je zavzemalo geografske širine od 30° S do 45° N in je bilo globalno v smeri vzhod-zahod. Horizontalna ločljivost modela je bila 36 km pri čemer je model uporabljal Merkatorjevo kartografsko projekcijo. Časovni korak je bil kombinacija korakov 150 s in 120 s. V vertikalni smeri je model imel 35 hibridnih tlačnih-sigma nivojev. Ti nivoji pri tleh sledijo površju, na veliki višini pa sledijo nivojem konstantnega zračnega tlaka. Za robne pogoje so se na severni in južni meji območja uporabile reanalize NCER-NCAR (časovna ločljivost 6 h in horizontalna ločljivost 2,5°). Na severni in južni meji območja se je uporabila prehodna cona široka 10 mrežnih točk. Za robne podatke pri tleh so uporabile temperature na površini morja (Atmospheric Model Intercomparison Project sea surface temperatures (AMIP-SST)), ki so na voljo v ločljivosti 0,5 geografskih stopinj in pa mesečne vrednosti pokritosti z vegetacijo.

7.4.2 Območje, časovno obdobje, manjkajoči podatki in akumulacije padavin

Za primerjavo med podatkovno bazo TRMM in simulacijo z WRF smo izbrali podobno območje kot v Poglavju 7.3.1. Ker se je simulacija z WRF izvajala le v



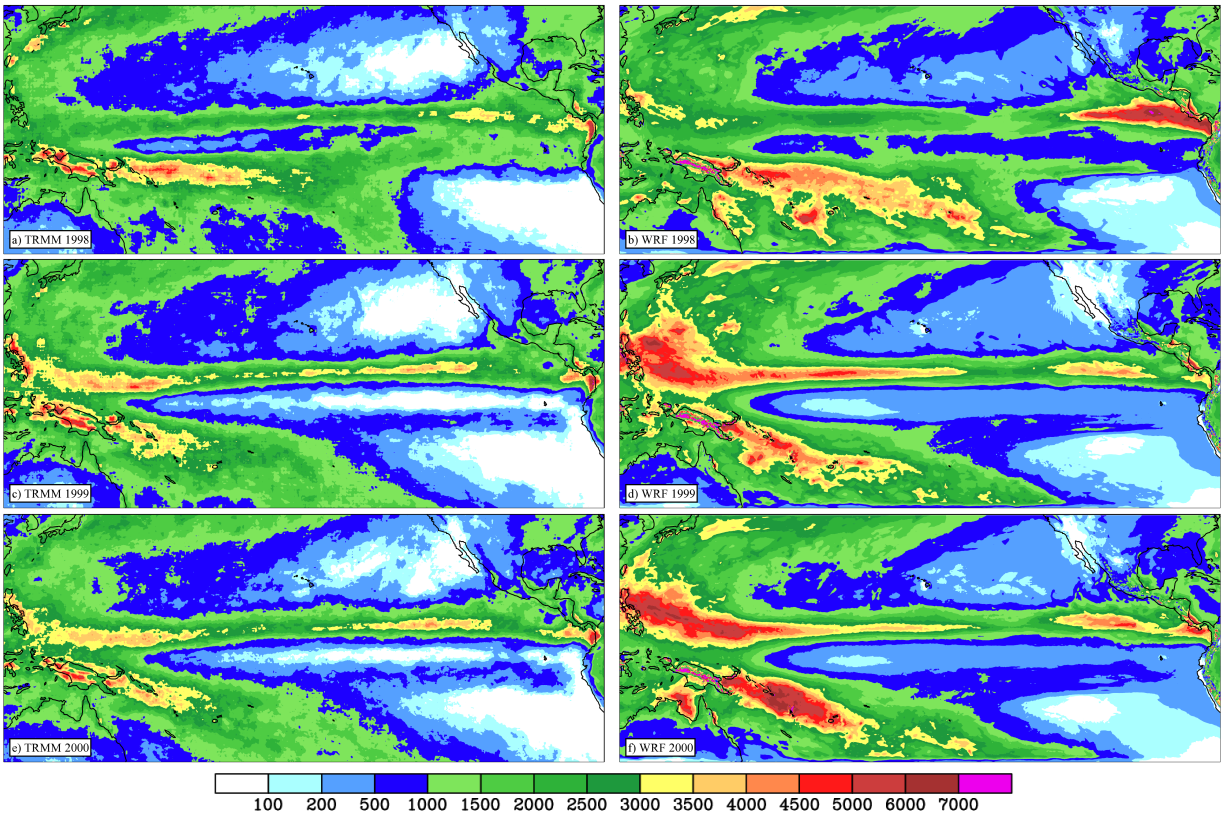
Slika S–12: Povprečne letne padavine (mm) za podatkovno bazo TRMM (levo) in simulacijo z WRF (desno) za obdobje 1. januar 1998 – 31. december 2000. Desna slika ima označena nekatera področja, kjer so opazne razlike.

območju od 30° S do 45° N je novo območje nekoliko manjše kot prejšnje. Časovno obdobje, ko sta bila na voljo oba produkta, je bilo tri leta, od 1. januarja 1998 do 31. decembra 2000 (podatkovna baza TRMM ni na voljo pred 1998 in simulacija z WRF se ni izvajala za leta po 2000). Območje sestavlja 661×273 mrežnih točk za podatkovno bazo TRMM ter 511×225 mrežnih točka za simulacijo z WRF, ki so na voljo v 8768 triurnih časovnih intervalih. Manjkajoči podatki v tem primeru niso bili velik problem, saj jih v podatkovni bazi TRMM ni nikjer bilo več kot 5 % (simulacija z WRF pa seveda ni imela manjkajočih podatkov).

Primerjava povprečnih letnih akumulacij med podatkovno bazo TRMM in simulacijo z WRF (Slika S–12) je pokazala, da je splošna porazdelitev padavin na območju podobna (npr. jasno opazna oblika ITCZ z dvema maksimuma, enem na zahodu in drugem na vzhodu območja), vendar pa je v simulaciji z WRF količina padavin občutno večja. Poleg tega se najde tudi mnogo drugih opaznih razlik med podatkovnima bazama.

Akumulacije padavin po posameznih letih so prikazane na Sliki S–13. Porazdelitev padavin v letu 1998 se nekoliko razlikuje od porazdelitve v letih 1999 in 2000. V letu 1998 je bilo več padavin v neposredni okolici ekvatorja, še posebej v vzhodnem delu območja. Razlika v akumulacijah med leti se lahko pripiše dejstvu, da je bil v letih 1997-98 močan El Niño, medtem ko je bil v letih 1998-1999 srednje močna La Niña (glede na Oceanic Nino indeks, ki je objavljen na spletni strani NOAA[1]). Razlika med leti 1998 in 1999-2000 se opazi v obeh podatkovnih bazah, vendar je tudi med bazama nekaj razlik, še posebej za leto 1998. Npr. v simulaciji z WRF je območje, kjer so padavine povečane, bolj ločeno od ITCZ, medtem ko sta v podatkovni bazi TRMM obe območji združeni v eno območje.

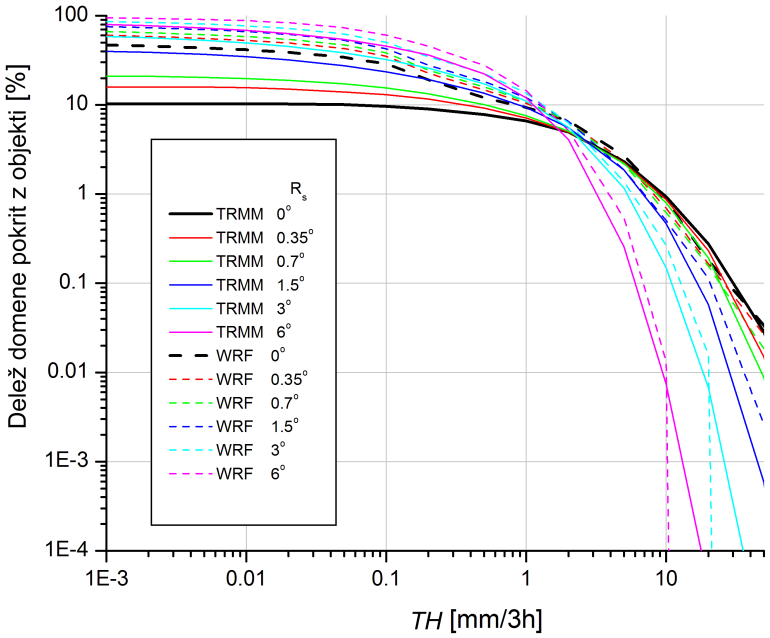
[1] http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml



Slika S–13: Letne padavine (mm) za podatkovno bazo TRMM (levo) in simulacijo z WRF (desno) za leta 1998 (zgoraj), 1999 (sredina) in 2000 (spodaj).

7.4.3 Analiza občutljivosti

Analiza občutljivosti smo izvedli podobno kot v Poglavju 7.3.2. Naključno smo izbrali 100 triurnih časovnih intervalov na katerih smo izvedli identifikacija objektov z uporabo različnih vrednosti parametrov R_s in TH. Pokritost območja z objekti (Slika S–14) za simulacijo z WRF je bila skoraj 50 % pri nizkih vrednostih TH, pri $TH = 1$ mm/3 h pa je bila pod 10 %, kar je podobno kot v podatkovni bazi TRMM. Visok delež pokritosti pri nizkem pragu padavin kaže na to, da v simulaciji z WRF okoli 40 % območja prekrivajo padavine z nizko intenziteto, za kar so verjetno odgovorne stratiformne padavine, ki se v modelu WRF obravnavajo ločeno od konvektivnih padavin. Invariantna vrednost TH je zopet vidna. Pri vrednostih TH, ki so večje kot invariantna vrednost, je pokritost v obeh podatkovnih bazah podobna. Pokazalo se je tudi, da zopet velja potenčna povezava med številom objektom in njihovo življenjsko dobo (ni prikazano na sliki v slovenskem povzetku).



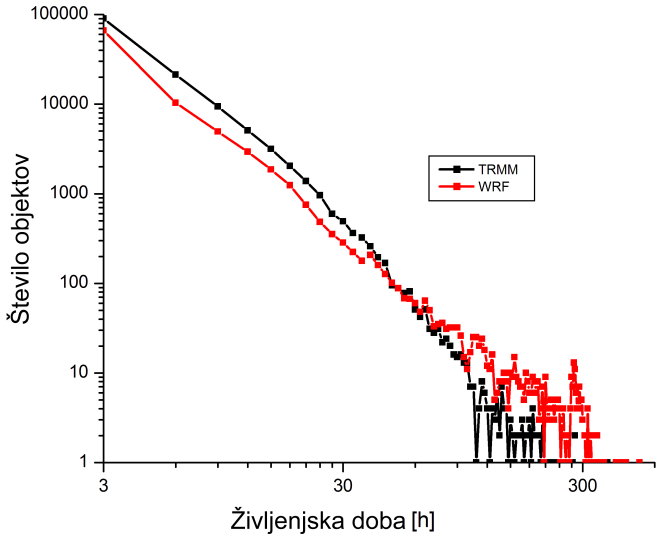
Slika S–14: Pokritost območja z objekti [%] v odvisnosti od padavinskega praga.

7.4.4 Analiza objektov v Pacifiku

Enako kot v Poglavju 7.3.2 smo izbrali vrednosti $R_s = 0,75°$ in $TH = 7$ mm/3 h. Da bi se izognili vsem vplivom efekta »spillover« ni bilo konvolucije v časovni dimenziji ($R_t = 0$ h). Analiza zajema celoten časovni interval od 1. januarja 1998 do 31. decembra 2000. Zopet se je pokazalo, da med številom objektov in življensko dobo velja potenčna povezava (Slika S–15). Vidi se tudi, da je v podatkovni bazi TRMM več objektov s kratko življenjsko dobo (manj kot 30 h) ter manj objektov z daljšo življenjsko dobo v primerjavi z simulacijo z WRF.

Slike S–16 in S–17 kažjo trajektorije objektov za različna leta. Kot je bilo že opaziti pri akumulacijah padavin, se leto 1998 precej razlikuje od let 1999 in 2000. V splošnem je v simulaciji z WRF opaziti več trajektorij objektov z daljšo življenjsko dobo (rdeče črte) kot v podatkovni bazi TRMM, ker se ujema z ugotovitvami iz Slike S–15. Poleg tega pa so trajektorije v območju ITCZ bolj ravne v simulaciji z WRF kot v podatkovni bazi TRMM. Trajektorije objektov v simulaciji z WRF pogosto potekajo le v smeri vzhod-zahod, medtem ko trajektorije objektov v podatkovni bazi TRMM vsebujejo tudi premike v smeri sever-jug, kar naredi trajektorije manj ravne. Premikanje na severnem in južnem delu območja je v obeh podatkovnih bazah predvsem v smeri proti vzhodu, medtem ko se smer premikanja objektov v ITCZ spreminja glede na leto, poleg tega pa so razlike tudi med podatkovnima bazama (npr. premikanje objektov v letu 2000 je v simulaciji z WRF predvsem v smeri proti zahodu, medtem ko se v podatkovni bazi TRMM objekti premikajo tudi v smeri proti zahodu in proti vzhodu).

Ker je vizualna subjektivna analiza slik trajektorij precej nezanesljiva, še posebej v območju ITCZ, kjer se nahaja večje število trajektorij, ki se med se-



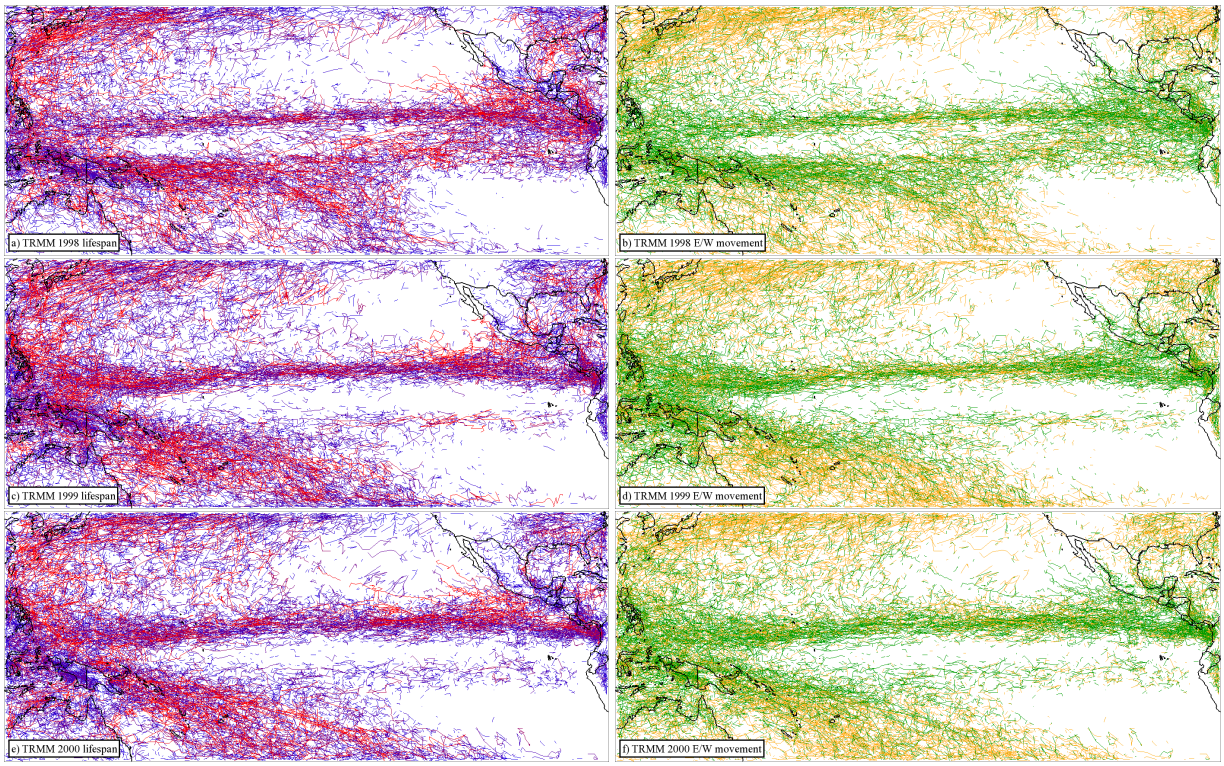
Slika S–15: Število objektov glede na življenjsko dobo v log-log grafu za podatkovno bazo TRMM in simulacijo z WRF.

boj prekrivajo, smo se odločili uvesti poseben indeks, ki bi zadovoljivo opisal količino gibanja v smeri vzhoda in zahoda. Tako smo uvedli indeks EWI (Eastward/Westward movement Index), ki je definiran kot

$$EWI = \frac{N_{\mathrm{W}}}{N_{\mathrm{W}} + N_{\mathrm{E}}}, \tag{7.4}$$

kjer je N_{W} definiran kot število časovnih intervalov, ko je bil nad mrežno točko identificiran objekt, ki se je premikal proti zahodu. Podobno je definiran N_{E}, le da je povezan z gibanjem proti vzhodu. Če je $EWI = 0$, potem so bili nad mrežno točko prisotni le objekti, ki so se premikali proti vzhodu in obratno, če je $EWI = 1$ so bili nad mrežno točko prisotni le objekti, ki so se premikali proti zahodu. Če je $EWI = 0.5$ potem so se nad mrežno točko enako pogosto pojavljali tako objekti, ki so se gibali proti vzhodu kot objekti, ki so se gibali to proti zahodu. EWI je zanesljiv samo v tistih mrežnih točkah, kjer se je pojavilo dovolj objektov, zato smo pri analizi zanemarili vse mrežne točke, kjer so se objekti pojavljali v manj kot 0.5 % časovnih intervalih.

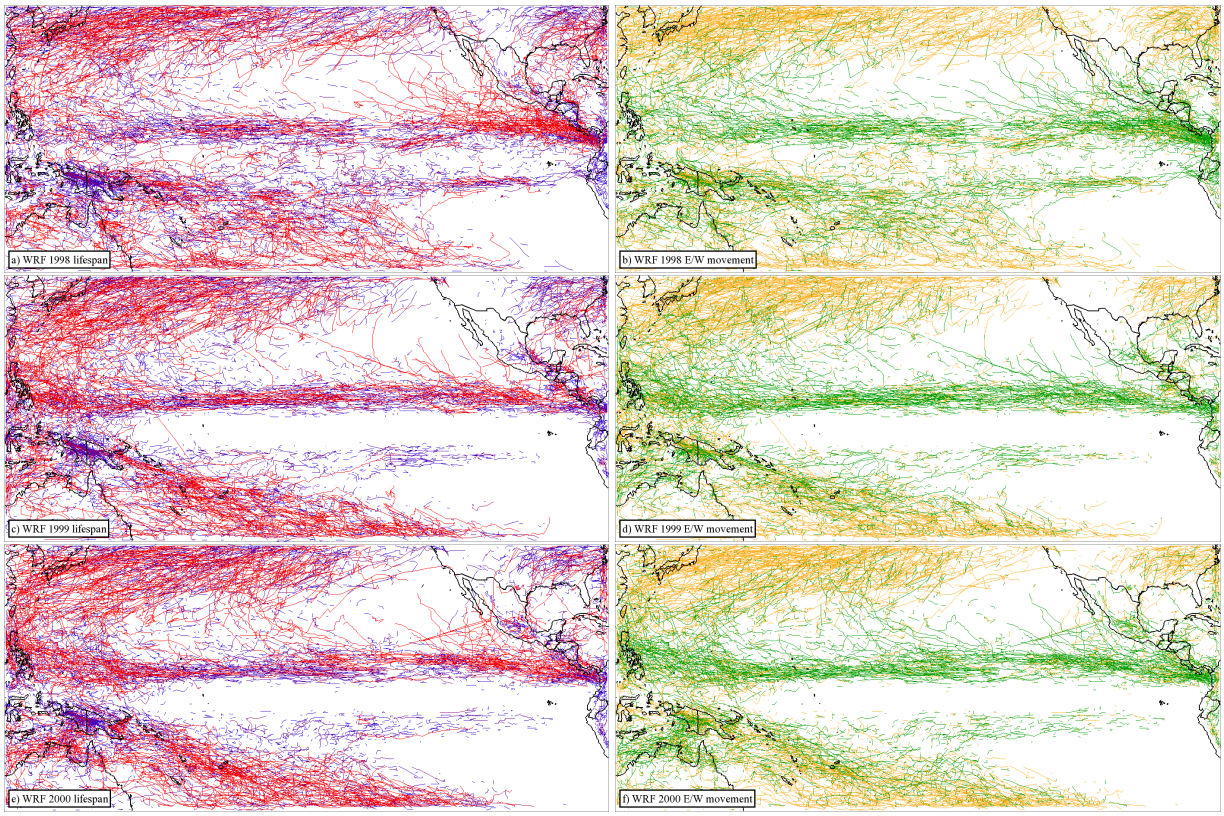
Slika S–18 prikazije indeks EWI za celotno obdobje 1998-2000 ter za različna leta. Indeks EWI za celotno obdobje je podoben v obeh podatkovnih bazah, razen v osrednjem delu ITCZ, kjer so v podatkovni bazi TRMM prisotni tako objekti, ki se gibljejo proti vzhodu kot objekti, ki se gibljejo proti zahodu, medtem ko se objekti v simulaciji z WRF gibljejo predvsem v smeri proti zahodu. Podobna ugotovitev velja tudi za posamezna leta. Vsa leta imamo v ITCZ več gibanja v smeri proti zahodu v simulaciji z WRF kot v podatkovni bazi TRMM. Ta razlika je še posebej očitna v obdobjih DJF, SON in deloma tudi JJA (slika ni prikazana



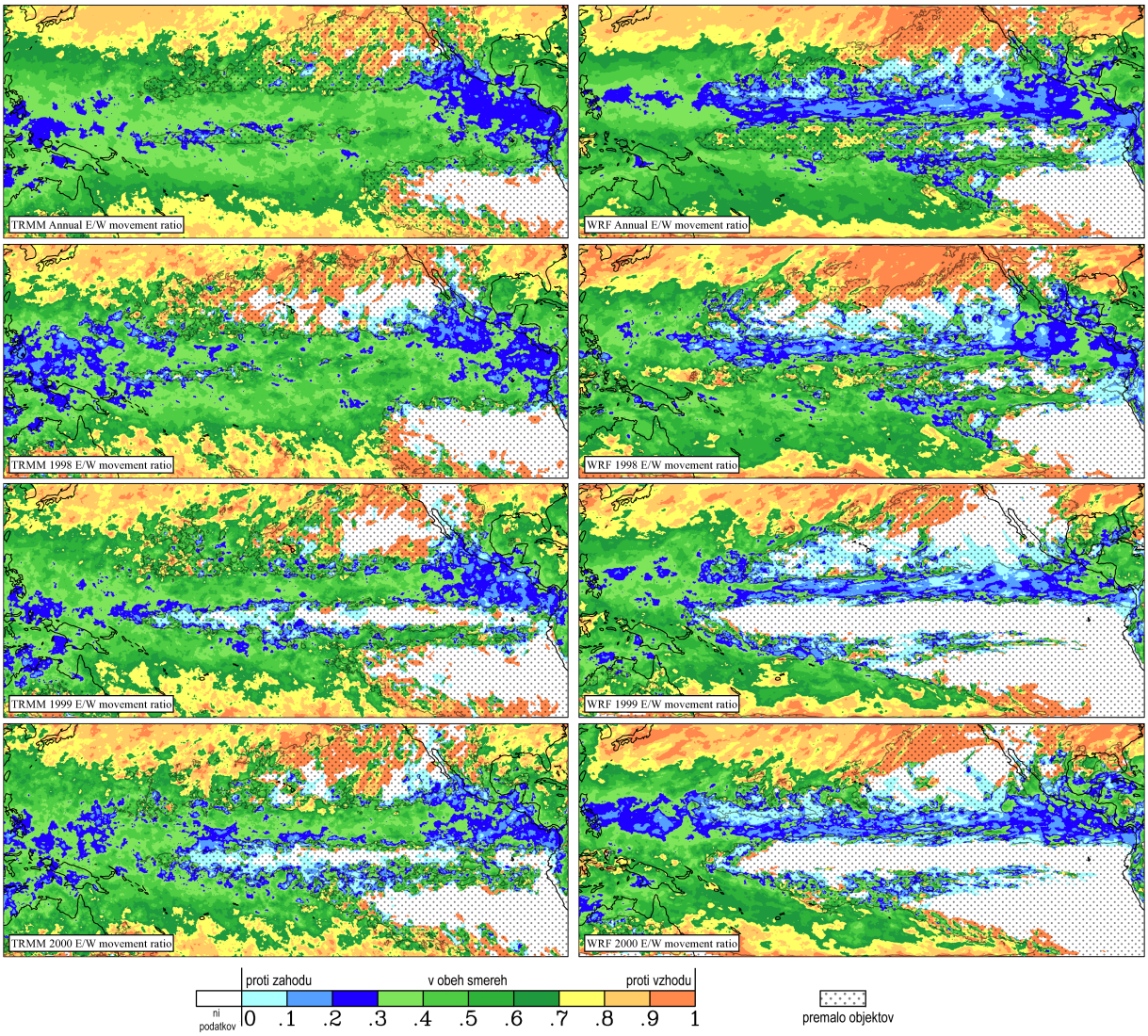
Slika S–16: Trajektorije objektov v podatkovni bazi TRMM za leta 1998, 1999 in 2000 (od zgoraj navzdol) Levo: trajektorije so obarvane glede na trajanje življenjske dobe (modro za krajšo in rdeče za daljšo). Desno: trajektorije so obarvane glede na gibanje v smeri zahod/vzhod. Če je gibanje imelo komponento v smeri zahoda, je segment trajektorije obarvan zeleno. Če je gibanje imelo komponento v smeri vzhoda, je segment trajektorije obarvan oranžno.

v slovenskem povzetku).

Naredila se je tudi primerjava porazdelitve objektov na območju glede na življenjsko dobo (slika ni prikazana v slovenskem povzetku). Pogostost objektov je večja v simulaciji z WRF, kar je v skladu s predhodnimi ugotovitvami, da so v simulaciji z WRF prisotne večje akumulacije padavin in več objektov z dolgo življenjsko dobo. Čeprav so razlike med podatkovnima bazama velike, je obema skupno to, da se objekti z najdaljšo življenjsko dobo običajno pojavljajo v dveh večjih območjih na zahodni strani območja ter manjšem območju na vzhodu območja (blizu zahodne obali Mehike).



Slika S–17: Trajektorije objektov v simulaciji z WRF za leta 1998, 1999 in 2000. Levo: trajektorije so obarvane glede na trajanje življenjske dobe (modro za krajšo in rdeče za daljšo). Desno: trajektorije so obarvane glede na gibanje v smeri zahod/vzhod. Če je gibanje imelo komponento v smeri zahoda, je segment trajektorije obarvan zeleno. Če je gibanje imelo komponento v smeri vzhoda, je segment trajektorije obarvan oranžno.



Slika S–18: *EWI* indeks za celotno obdobje (zgoraj) in različna leta (naslednje vrstice) za podatkovno bazo TRMM (levo) in simulacijo z WRF (desno). Na območjih, ki so označena s pikami, zaradi majhnega števila objektov ni bilo mogoče zanesljivo oceniti vrednosti indeksa *EWI*. Modra področja predstavljajo območja, kjer je prevladujoče gibanje v smeri zahoda, rdeča kjer je prevladujoče gibanje v smeri vzhoda, ter zelena območja, kjer je bilo gibanje v smeri proti zahodu primerljivo z gibanjem proti vzhodu.

7.5 Zaključki

V okviru doktorata smo nadgradili osnovno metodo MODE , tako da omogoča tudi analizo časovnega razvoja padavinskih objektov. S pomočjo metode smo najprej analizirali in primerjali padavinske podatke dveh satelitskih podatkovnih nizov TRMM in PERSIANN. Nato smo metodo uporabili še za analizo in verifikacijo modelskih padavin pridobljenih v okviru simulacije tropskega kanala z modelom WRF.

Primerjava satelitskih podatkovnih nizov TRMM in PERSIANN je najprej pokazala, da se že akumulacije povprečnih letnih padavin razlikujejo tudi do 30 %. Kljub precejšnjim razlikam pa je analiza objektov pri obeh podatkovnih bazah pripeljala do nekaterih podobnih ugotovitev. Npr. da se največji objekti in objekti z najdaljšo življenjsko dobo ponavadi nahajajo na zahodnem delu območja, v manjšem številu pa tudi na vzhodnem delu blizu obale Mehike. Izkazalo se je tudi, da je povezava med številom objektov in življenjsko dobo potenčna funkcija. Gibanje objektov na širšem območju ITCZ običajno poteka v obeh smereh (proti vzhodu in proti zahodu), v vzhodnem delu ITCZ pa je gibanje proti zahodu bolj pogosto, medtem ko je v zahodnem delu ICTZ gibanje proti zahodu celo dominantno. Gibanje objektov na severnem in južnem delu območja poteka predvsem v smeri proti vzhodu. Vse zgoraj omenjene ugotovite se spreminjajo glede na sezono. Npr. spomladi se pojavi sekundarni ITCZ, ki ga večinoma sestavljajo objekti s krajšo življenjsko dobo. Bistvena je bila tudi analiza občutljivosti preko katere smo lahko določili najprimernejše vrednosti parametrov metode R_{s} in TH, hkrati pa ugotovili tudi, da konvolucija v časovni dimenziji ni primerna zaradi efekta »spillover«.

Naredili smo tudi primerjavo padavin podatkovnega niza TRMM s padavinami simulacije tropskega kanala z modelom WRF[2]. Primerjava povprečnih letnih akumulacij padavin je pokazala, da obstaja veliko razlik med podatkovnima nizoma. Najpomembnejša razlika je, da je v splošnem količina padavin v simulaciji z WRF precej večja kot v podatkovni bazi TRMM. Velike razlike so tudi med posameznimi leti, saj je bil v letu 1998 prisoten močan El Niño, v letu 1999 pa šibka La Niña. Analiza občutljivosti je pokazala, da v simulaciji z WRF okoli 40 % območja pokrivajo zelo šibke stratiformne padavine. Povezava med številom objektov in življenjsko dobo je potenčna funkcija. Pokazalo se je tudi, da je v podatkovni bazi TRMM manjše število objektov s krajšo življenjsko dobo (manj kot 30 h). Analiza trajektorij je pokazala, da so na območju ITCZ trajektorije precej bolj ravne v simulaciji z WRF . No območju ITCZ trajektorije objektov v simulaciji z WRF pogosto potekajo le v smeri vzhod-zahod, medtem ko trajektorije objektov v podatkovni bazi TRMM vsebujejo tudi premike v smeri sever-jug, kar naredi trajektorije manj ravne. Gibanje objektov na severnem in južnem delu območja je pretežno v smeri proti vzhodu pri obeh podatkovnih bazah. Na območju ITCZ pa je prisotno gibanje objektov v obe smeri (proti vzhodu in proti zahodu), vendar se spreminja glede na leto in je tudi nekoliko

[2]Simulacijo je izvedel NCAR

različno med podatkovnima bazama. Analiza z indeksom EWI je potrdila, da so gibanja podobna, razen na osrednjem območju ITCZ, kjer so v podatkovni bazi TRMM prisotna gibanja v obeh smereh medtem ko je v simulaciji z WRF dominantno gibanje v smeri proti zahodu. V obeh podatkovnih nizih se objekti z dolgo življenjsko dobo nahajajo predvsem v dveh večjih področjih na zahodu območja ter v manjšem področju na vzhodu območja (blizu obale Mehike).

Izjava

Podpisani Gregor Skok izjavljam, da predložena disertacija predstavlja rezultate lastnega znanstveno-raziskovalnega dela.

Ljubljana, 1. avgust 2009

Gregor Skok