A multi-state Markov chain model to assess drought risks in rainfed agriculture: a case study in the Nineveh Plains of Northern Iraq

Rasha M. Fadhil[a], Koichi Unami[b,1]

[a] *College of Engineering, University of Mosul, Mosul, Iraq*

[b] *Graduate School of Agriculture, Kyoto University, Kyoto 606-8502, Japan*

Abstract

The occurrence of prolonged dry spells and the shortage of precipitation are two different hazardous factors affecting rainfed agriculture. This study investigates a multi-state Markov chain model with the states of dry spell length coupled with a probability distribution of positive rainfall depths. The Nineveh Plains of Northern Iraq is chosen as the study site, where the rainfed farmers are inevitably exposed to drought risks, for demonstration of applicability to real-time drought risk assessment. The model is operated on historical data of daily rainfall depths observed at the city Mosul bordering the Nineveh Plains during the period 1975-2018. The methodology is developed in the context of contemporary probability theory. Firstly, the Kolmogorov-Smirnov tests are applied to extracting two sub-periods where the positive rainfall depths obey to respective distinct gamma distributions. Then, empirical estimation of transition probabilities determining a multi-state Markov chain results in spurious oscillations, which are regularized in the minimizing total variation flow solving a singular diffusion equation with a degenerating coefficient that controls extreme values of 0 and 1. Finally, the model yields the statistical moments of the dry spell length in the future and the total rainfall depth until a specified terminal day. Those statistical moments, termed hazard futures, can quantify drought risks based on the information of the dry spell length up to the current day. The newly defined

[1] Corresponding author. *E-mail address:* unami.koichi.6v@kyoto-u.ac.jp *Phone:* +81-75-753-6161 (K. Unami)

hazard futures are utilized to explore measures to avert drought risks intensifying these decades, aiming to establish sustainable rainfed agriculture in the Nineveh Plains.

Keywords: Dry spell length, Rainfall depth, Multi-state Markov chain model, Northern Iraq, Hazard futures, Minimizing total variation flow

1. Introduction

The occurrence of prolonged dry spells and the shortage of precipitation are two different hazardous factors affecting rainfed agriculture, whose measures of risk aversion, such as irrigation facility or weather insurance, are vulnerable. This study shows that a multi-state Markov chain model with the states of dry spell length (DSL) coupled with a probability distribution of positive rainfall depths, referred to as a multi-state Markov chain model, has overwhelming advantages in application to drought risk assessment in rainfed agriculture. The Nineveh Plains of Northern Iraq, where the rainfed farmers are inevitably exposed to drought risks, is chosen as the study site for demonstration.

The shortage of precipitation in the future is traded as futures contracts since the introduction of rainfall derivatives at the Chicago Mercantile Exchange (CME) in 2011, motivating slow but steady development of studies on stochastic processes modeling time series of hydrological phenomena (Tong and Liu 2021). Turvey (2001) was one of the earliest studies arguing the applicability of rainfall derivatives to risk hedges in agriculture. Leobacher and Ngare (2011) introduced a discrete-time Markovian model for pricing rainfall derivatives, with numerical illustrations of Monte Carlo simulations. Masala (2014) proposed a semi-Markov model defined on the probability space to price some rainfall contracts issued by CME. These two papers in the 2010s cited above did not explicitly consider the filtration of information. As can be found in textbooks of probability theory such as Williams (1991), most objects in such a financial engineering context involve the filtered probability space to rigorously deal with a stochastic process whose future behavior depends on the information available up to the current time. Cabrera et al. (2013) stated that the standard approach of pricing a weather futures contract at the current time is to calculate the risk-neutral expectation of an index, which is an accumulated value of the weather variable during a given period of time, based on the filtration, requiring a model for the index or the underlying weather variable.

Unfortunately, most methodologies developed by hydrologists lack such a mathematical perspective on the filtered probability space. A range of statistical methods and descriptors of drought characteristics are utilized for a-posteriori evaluation of drought severity in the sense of return period (Wilby et al. 2015). The World Meteorological Organization (2012) patronizes the standardized precipitation index (SPI), which is the accumulated precipitation depth during a given period standardized with its statistical moments. SPI is widely applied to drought risk assessment, linked with different probability distribution functions (Angelidis et al. 2012), spatial characterization (Cavus and Aksoy 2019), and intensity-duration-frequency curves (Cavus and Aksoy 2020).

On the other hand, financial engineers and actuary scientists have not been pursuing the DSL, which is defined as the number of consecutive days without rainfall of a specified threshold depth (Anagnostopoulou et al. 2003; Vicente-Serrano and Beguería-Portugués 2003). Unlike the other indexes depending on the weather variables during a given period of time, the DSL is given as a variable period of time. A prolonged dry spell cannot afford to keep the soil moisture in the root zone readily available for water consumption by rainfed crops, as it mostly stems from infiltrated local rainfall (Sharifi et al. 2016). Therefore, the temporal distribution of DSL significantly affects the structure of drought that may lead to famine. Knowledge of DSL coupled with the probability distribution of precipitation depths can aid in drought prediction and hence drought disaster preparedness. A primitive method for such purposes is an analysis using the weather generator (WGEN) model of Richardson and Wright (1984), which considers a first-order Markov chain with two states: dry day and wet day. WGEN and its variants are still in use for drought research with practical applications (Fischer et al. 2013; Yadeta et al. 2020), although the first-order Markov chain with two states generates DSL with incorrect autocorrelation structures. Markov chains of higher orders or with more states may better represent the persistence of drought. Martin-Vide and Gomez (1999) attempted to apply Markov chains of higher orders up to the tenth for distribution of DSL over Peninsular Spain under mostly semi-arid Mediterranean climate. However, explicitly estimating the transition probabilities in Markov chains of higher orders is not an easy task, as their number is equal to the states' number to the power of the order (Gao et al. 2020). Al-Khayat and Al-Sulaiman (2013) proposed a method to predict the situation of rain in the next day based on the information actually available by today, using Markov chains with four states: lack of rain, light rain, moderate rain, and heavy rain. The method was applied to historical data recorded in the city Mosul, where this study focuses on, with two algorithms of determining transitions among the states. Another innovative approach is based on a stochastic differential equation model which takes the cumulative rainfall depth, not the time, as the principal independent variable, to comprehensively assess drought and flood risks (Unami et al. 2010).

This study provides a more straightforward approach to drought risk assessment. Firstly, the set of DSL itself is taken as the space of states to constitute a first-order Markov chain. The DSL of zero represents a wet day. Astonishingly, DSL as such a state variable has been paid the least attention since Tatano et al. (1992), which might be attributed to the dissemination of the Poisson processes (Onof et al. 2000; Sirangelo et al. 2015; Sirangelo et al. 2017), including the fractional ones (Yang et al. 2020), applied to the arrivals of rainfall events without utilizing the benefit of filtration for decision making. Then, the probability distribution of positive rainfall depths, or rainfall depths on wet days, is considered. In application to historical data observed at Mosul during the period 1975-2018, a preliminary analysis showed better fitting to the gamma distribution when the data throughout the whole year is treated as a single population, rather than monthly treatment (Fadhil 2018). The nonparametric Kolmogorov-Smirnov (K-S) tests are used for comparing empirical time-homogeneous probability distributions of positive rainfall depths for distinct two sub-periods, as well as for comparing an empirical time-homogeneous probability distribution of positive rainfall depths for a sub-period with a gamma distribution. Coupling a fitted gamma distribution of positive rainfall depths to the first-order Markov chain with the multiple states of DSL results in a multi-state Markov chain model, which can be applied to drought risk assessment. This multi-state Markov chain model is essentially different from any of the earlier models by the authors, including Sharifi et al. (2016), Unami and Mohawesh (2018), and Nop et al. (2021). Sharifi et al. (2016) used a time-continuous Markov process with the continuous states of soil moisture. Unami and Mohawesh (2018) developed a time-continuous Markov process with the continuous states of a water flow index. Nop et al. (2021) considered a multi-state Markov chain with the states of discretized rainfall depth ranges. The overwhelming advantages of this multi-state Markov chain model are the first-order Markovian properties and the ability to capture the memory effect of sequential dry days. The transition probabilities of the multi-state Markov chain are identified from observed rainfall data as functions of the time and the states. However, due to the scarcity of recorded wet days, spurious oscillations occur in the empirical transition probabilities. Jimoh and Webster (1999) applied the conventional Fourier fitting technique to smoothing the transition probabilities of Nigerian rainfall. Still, it does not work well for Iraqi cases where abrupt alternation between dry and wet seasons is intrinsic. Therefore, a novel regularization technique is introduced to avoid the spurious oscillations without spoiling true abrupt variations, inspired by the minimizing total variation flow (MTVF), which is called the ROF model (Rudin et al. 1992). The ROF model is applied initially to image denoising, which has the same requirement as the regularization of the transition probabilities here: avoiding spurious oscillations without spoiling true abrupt variations. The success of the ROF model is attributed to working in the space of functions of bounded total variation rather than the disappointing space of square-integrable functions. The mathematical difficulty is that functions of the MTVF with bounded total variation may not be smooth, similarly to value functions appearing in optimal control problems (Unami and Mohawesh 2018; Unami et al. 2019). There is a tremendous number of papers dealing with the MTVF applied to image denoising. However, the regularization technique developed here has an advantage over the conventional ones that a special degenerating coefficient skillfully distinguishes the spurious oscillations from true abrupt variations in controlling extreme values of 0 and 1 as transition probabilities. Then, it is shown that statistical moments of the DSL in the future and the total rainfall depth until a specified terminal day, which are termed hazard futures in this study, can be calculated with recursive formulae. Finally, the multi-state Markov chain model is applied to drought risk assessment in the Nineveh Plains of Northern Iraq in terms of the hazard futures defined for each of the states at each time. The rainfed farmers observing the DSL up to the current day can make decisions at stopping times in the context of the filtered probability space, as implicitly suggested in Ojara et al. (2020). The results of detailed drought risk assessment warn of failure in rainfed agriculture practiced in the Nineveh Plains, implying a regime shift of DSL and positive rainfall depth from the first sub-period to the second, and thus measures to avert the risks are explored.

Mathematical preliminaries required in this study are as follows. The pair of a *set* of possible outcomes and a *σ-algebra* on it is called a *measurable space*. When a *probability measure* is defined on a measurable space, the triple consisting of the set, the σ-algebra, and the probability measure is called a *probability space*. A family of sub-σ-algebras ordered non-decreasingly is called a *filtration* of the σ-algebra. A probability space equipped with a filtration is a *filtered probability space*. A *random variable* is a real-valued measurable function on the set. A *stochastic process* is a collection of random variables parameterized with time. A *stopping time* is a random variable with specific properties defined on a filtered probability space. The first-order Markov chain with the multiple states of DSL in this study is a stochastic process defined on a filtered probability space.

2. Overview of the study site

The land of the Republic of Iraq lies within the subtropical and the temperate zones of the northern hemisphere, located in the western part of Asia. Average annual precipitation ranges from less than 100 mm in the arid deserts covering over 60 % of Iraq in the south up to 1,200 mm in the north and north-eastern mountain regions of hot-summer Mediterranean climate (Al-Ansari 2013), and precipitation in Iraq is generally seasonal and mostly occurs in the winter from November to April. It should be noticed that the record highest temperature in Iraq was renewed to 52.0 °C in 2010, which might be attributed to the global warming (Al-Ansari 2013), and to 53.6 °C in 2016 according to media reports. Kadim (2013) confirmed clear trends in air temperature and rainfall at three weather stations in Iraq: Mosul, Baghdad, and Basra, using the records for the longest possible periods. Basra showed the most increasing air temperature, followed by Mosul and then Baghdad. Baghdad showed the most decreasing rainfall, followed by Mosul and then Basra. Azooz and Talal (2015) applied nonlinear regression to compiled historical data of mean monthly temperature and precipitation for four main cities of Iraq, namely, Baghdad, Mosul, Basra, and Kirkuk with the observation periods of 1887-2013, 1900-2013, 1923-2013, and 1935-2013, respectively. The results show a significant increase in temperature and a decrease in precipitation, which are considered as two manifestations of climate change. Their extrapolation to future predictions for temperature agreed well with conclusions of the Intergovernmental Panel for Climate Change 2007 (IPCC 2007) report on greenhouse effect warming. Robaa and AL-Barazanji (2013) also showed rising trends of annual mean surface air temperature in 11 Iraqi stations. Furthermore, the lengths of dry seasons in Iraq increased by two months in the late century (Evans 2009). A more comprehensive statistical analysis by Salman et al. (2018b) suggested unidirectional trends in rainfall and rainfall-related extremes in Iraq. Agha and Şarlak (2016) analyzed climate variables observed at 28 Iraqi meteorological stations in the period of 1980-2011 and concluded that such climatic impacts were spatially uniform. However, Salman et al. (2018a) selected an ensemble of general circulation models to project higher increases in temperatures in the north and northeast of Iraq in the 21st century.

In addition to the harsh climates, Iraq has geopolitical disadvantages in water resources as a downstream country. Owing to the Tigris and Euphrates Rivers, Iraq was lavish in its water resources compared to other countries. However, dams constructed on those rivers and their tributaries outside the border of Iraq are negatively affecting the flow regimes. Paradoxically, reliance on those major rivers is making Iraq more vulnerable to both flood and drought risks.

The Nineveh Plains refers to the region extending over Tel Kaif, Al-Hamdaniya, and Shekhan Districts of Nineveh Governorate, bordered by the Tigris River flowing through the city Mosul to the southwest, the Great Zab River to the southeast, and fold mountains continuing to Dohuk Governorate to the north. It is under a semi-arid environment, which is transitional from the arid deserts to the mountain regions. The topography of the Nineveh Plains and the vicinities is shown in Figure 1, depicted with the SRTM digital elevation data (Farr et al. 2007). There are four main meteorological stations in the vicinities of the Nineveh Plains, namely, Mosul, Sinjar, Telafer, and Rabea, whose positions are shown in Figure 1 as well. Mustafa (2012) calculated basic statistics of monthly rainfall depths observed at those four stations during the period 1974-2002, suggesting similar declining trends at all the stations. Taha (2014) showed that annual rainfall depths in Mosul and Sinjar obey to respective normal distributions. In the Nineveh Plains, July and August are the hottest months of the summer season, where the mean maximum temperatures are 39–43 °C and often reach nearly 50 °C though the night temperatures may drop down to 20 °C, while the mean maximum temperatures are 7–16 °C and the mean minimum temperatures are 2–7 °C with a possibility of frost during the coldest months (Awchi and Kalyana 2017). Zakaria et al. (2013) analyzed historical records of temperature and rainfall in Mosul for the period 1900-2009, showing significant fluctuations in their average monthly values during the sub-periods 1900-1930, 1930-1960, 1960-1990, and 1990-2009. An impression at a glance is that the extremes in temperature amplified after 1990, involving delays in the onset of wet seasons. The Nineveh Plains is more prone to drought than the other parts of Iraq, as rainfed agriculture for winter grain crops is widely practiced.

Figure 1: The topography of the Nineveh Plains and the vicinities with the locations of the four main meteorological stations.

Under the above-mentioned peculiar circumstances, drought risk assessment attracts more attention to sustain rainfed agriculture in the Nineveh Plains. It is generally known that the erratic occurrence of rainfall with the uneven spatio-temporal distribution of rainfall amounts leads to unsuccessful agricultural production. Variability of rainfall in the growing seasons, rather than the total annual precipitation, can severely affect productivity (Barron et al. 2003; Rockström et al. 2010). In this context, Al-Najafee and Rashad (2012) conducted a sensitivity analysis of rainfall distribution impacting on wheat productivity in the Nineveh Plains. Rasheed (2010) revealed that 56% of the years 1941-2002 were drought years in terms of SPI at nine metrological stations in the northern part of Iraq. Furthermore, as a follow-up study of Kalyan and Awchi (2015) using the deciles method, Awchi and Kalyana (2017) employed SPI at different time scales of 3, 6, 12, and 24 months to analyze the meteorological drought in the northern part of Iraq, based on monthly rainfall data during the period 1937–2010. Results showed that severe drought events occurred every decade, but the severest ones clustered during the years 1997–2001 and 2006–2010.

This study examines time series data of daily rainfall observed at the meteorological station in Mosul during the period from January 1st, 1975 through December 31st, 2018, as the observation periods at the other three meteorological stations in the vicinity of the Nineveh Plains are much shorter. Figure 2 shows the temporal accumulation of rainfall depths observed in each Gregorian year without missing data, visualizing the alternation of dry and wet seasons as well as the variability of the annual rainfall depths.

Figure 2: Accumulated rainfall depths observed at the meteorological station in Mosul in each Gregorian year without missing data.

3. Methodology

Figure 3 presents a flowchart of the methodology developed in this study. The time series data of daily rainfall mostly shown in Figure 2 and the multi-state Markov chain to be defined in Equation (9) are the two sources of the multi-state Markov chain model. The procedures involving positive rainfall depths and transition probabilities for DSL are described in subsection *3.1* and subsection *3.2*, respectively. The resulting multi-state Markov chain model is applied to drought risk assessment with the hazard futures calculated by the methods in subsection *3.3*. Measures of risk aversion are discussed in subsection *4.3* of the next section.

Figure 3: Flowchart of the methodology.

3.1 Probability distributions of positive rainfall depths

Considering the severe decline in rainfall during that 44 years period, identification and operation of the multi-state Markov chain model are based on the data for distinct two sub-periods. The statistical methods described below, which can be found in standard textbooks of hydrology such as Loucks and van Beek (2005), are employed to extract sub-periods having time-homogeneous rainfall regimes. The well-known K-S tests are firstly applied to comparing empirical probability distribution of positive rainfall depths for distinct two sub-periods. The empirical cumulative distribution function (ECDF) F_E for data set E of observed positive rainfall depths r_i sorted in ascending order is given by

$$F_E\left(r\right)=\frac{1}{n_E}\sum_{i=0}^{i<n_E} I_{[-\infty,r]}\left(r_i\right) \tag{1}$$

where n_E is the number of observations in the data set E, and $I_{[-\infty,r]}\left(r_i\right)$ is the indicator function, which is equal to 1 if $r_i \leq r$ and equal to 0 otherwise. The Kolmogorov-Smirnov statistic for the ECDFs of two data sets E_A and E_B is calculated as

$$D_{E_A,E_B}=\sup_r\left|F_{E_A}\left(r\right)-F_{E_B}\left(r\right)\right| \tag{2}$$

which must satisfy the inequality

$$D_{E_A,E_B}>\sqrt{-\frac{1}{2}\ln\alpha_L}\sqrt{\frac{n_{E_A}+n_{E_B}}{n_{E_A}n_{E_B}}} \tag{3}$$

to reject, at a significance level α_L, the null hypothesis that the observed positive rainfall depths in the two data sets are drawn from the same distribution. After extracting data sets obeying to single distributions, the K-S test is applied to examining whether each of them fits to a gamma distribution or not. The probability density function (PDF) $f_{\text{gam}}(r)$ of the gamma distribution is given by

$$f_{\text{gam}}(r)=\frac{\beta^{\alpha}r^{\alpha-1}\exp\left(-\beta r\right)}{\Gamma\left(\alpha\right)} \tag{4}$$

with the two parameters α and β having the properties

$$\alpha=\frac{\left(\mathrm{E}\left[r\middle|r>0\right]\right)^2}{\mathrm{Var}\left[r\middle|r>0\right]},\quad \beta=\frac{\mathrm{E}\left[r\middle|r>0\right]}{\mathrm{Var}\left[r\middle|r>0\right]} \tag{5}$$

where E and Var represent the mean and the variance, respectively. The Kolmogorov-Smirnov statistic for the ECDF of a data sets E and a given cumulative distribution function (CDF) F is calculated as

$$D_E = \sup_r \left| F_E(r) - F(r) \right| \tag{6}$$

which must satisfy the inequality

$$\sqrt{n_E} D_E > K_\alpha \tag{7}$$

for a criterion K_α, to reject, at a significance level α_L, the null hypothesis that the observed positive rainfall depths in the data set are drawn from the given distribution. The criterion K_α solves the equation

$$\frac{\sqrt{2\pi}}{K_\alpha} \sum_{k=1}^{\infty} \exp\left(-\frac{(2k-1)^2 \pi^2}{8K_\alpha^2} \right) = 1 - \alpha_L . \tag{8}$$

3.2 Multi-state Markov chain model

Let $\mathbb{N}$ denote the set of non-negative integers (0, 1, 2, …), which is countably infinite. Let R_k represent the rainfall depth on the day $k \in \mathbb{N}$. The DSL up to the day k is chosen as the state variable X_k, which is rigorously defined as

$$X_k = \inf_{\kappa \in \{ l | r_\theta \le R_l, l \le k \}} (k - \kappa) \tag{9}$$

where $l \in \mathbb{N}$, and r_θ is the threshold of rainfall depth such that the day l is regarded as a dry day if $R_l < r_\theta$. Then, the space of the state variables is $\mathbb{N}$. Regarding the state variables X_k as random variables, the series of X_k parameterized by the day k becomes a first-order Markov chain defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P : t \in \mathbb{N})$, where Ω is the set of all possible outcomes, $\mathcal{F} = \bigcup_{t \in \mathbb{N}} \mathcal{F}_t$ is the σ-algebra with the sub-σ-algebras $\mathcal{F}_t$ generated by X_k

where $0 \le k \le t$, $\{\mathcal{F}_t\}$ is the filtration, and P is the probability measure on Ω. The probability measure P is fully determined if transition probabilities

$$\begin{cases} P_{i0} = P\left(X_{k+1} = 0 \middle| X_k = i\right) = \Pr\left(R_{k+1} \ge r_\theta \middle| X_k = i\right) \\ P_{i1} = P\left(X_{k+1} = i+1 \middle| X_k = i\right) = \Pr\left(R_{k+1} < r_\theta \middle| X_k = i\right) = 1 - P_{i0} \end{cases} \tag{10}$$

are given for each $i \in \mathbb{N}$, achieving the Markovian property $P\left(X_t \in U \middle| \mathcal{F}_s\right) = P\left(X_t \in U \middle| X_s\right)$ for any Borel set U on $\mathbb{R}$ if $0 \le s < t$. We assume that P_{i0} for each i is a year-periodic function of the time t, which is compatible with the day of the Gregorian year as

$$t = D_{\text{year}} \frac{\text{The day of the Gregorian year} - 1/2}{\text{The number of day in the Gregorian year}}, \tag{11}$$

where $D_{\text{year}} = 365.25$, and P_{i0} at the time t is denoted by $u(t,i)$. With a specified time range δt, an empirical estimate $\hat{u}(t,i)$ for $u(t,i)$ from historical data is given by

$$\hat{u}(t,i) = \frac{N_{i0}^t}{N_{i0}^t + N_{i1}^t} \tag{12}$$

where N_{i0}^t is the number of days k such that $|t-k| < \delta t$ and $X_k = i$ and $R_{k+1} \ge r_\theta$, and N_{i1}^t is the number of days k such that $|t-k| < \delta t$ and $X_k = i$ and $R_{k+1} < r_\theta$. An unconditional expectation of the dry day is estimated as

$$\hat{u}(t,\infty) = \frac{\sum_i N_{i0}^t}{\sum_i N_{i0}^t + \sum_i N_{i1}^t}. \tag{13}$$

3.3 Regularization of transition probabilities

The number of data available for empirical estimation of a transition probability is inversely proportional to the time range δt, implying the tradeoff between resolution and accuracy. Here, we propose a novel regularization technique, which reduces the total variation in the function $u(t,i)$ defined on the set $\left[0, D_{\text{year}}\right) \times \mathbb{N}$. Firstly, the function $u(t,i)$ is embedded into $u = u(t,x)$ defined on the set $\left[0, D_{\text{year}}\right) \times [0,\infty)$, using the piecewise linear interpolation in the x-direction. Then, a novel singular diffusion equation with a degenerating coefficient is introduced as

$$\frac{\partial u}{\partial \tau} = u(1-u)\nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right) \tag{14}$$

where τ is a virtual time, and ∇ is the del operator in the t-x-plane. The scope of (14) is outlined in Appendix 1, explaining how singular diffusion equations are derived in the context of the variational calculus and how the degenerating coefficient operates. The MTVF is the solution to the initial value problem of (14) with

$$u = u(t,x) = \hat{u}(t,x) \text{ at } \tau = 0 \tag{15}$$

where $\hat{u}(t,x)$ for $x = i \in \mathbb{N}$ is set as $\hat{u}(t,i)$ in (12) if the denominator is large enough and as $\hat{u}(t,\infty)$ in (13) otherwise, and the piecewise linear interpolation is applied to $\hat{u}(t,x)$ for $x \notin \mathbb{N}$. A numerical method is developed here to approximately solve (14), considering the values of u on the grids $(m\Delta t, i\Delta x)$ for $m = 0,1,\cdots,n_t - 1$, $i = 0,1,\cdots,n_x - 1$, $\Delta t = D_{\text{year}}/n_t$, and $\Delta x = 1$, where n_t and n_x are finite positive integers. The flux $\nabla u/|\nabla u|$ in the right-hand side of (14) at a generic grid $(m\Delta t, i\Delta x)$ is approximated by signum functions $\sigma_t^{m,i}$ and $\sigma_x^{m,i}$ defined as

$$\begin{pmatrix} \sigma_t^{m,i} \\ \sigma_x^{m,i} \end{pmatrix} = \begin{pmatrix} \begin{cases} \dfrac{u((m+1)\Delta t,i) - u(m\Delta t,i)}{\left|u((m+1)\Delta t,i) - u(m\Delta t,i)\right|} & \text{if } u((m+1)\Delta t,i) \neq u(m\Delta t,i) \\ 0 & \text{if } u((m+1)\Delta t,i) = u(m\Delta t,i) \end{cases} \\ \begin{cases} \dfrac{u(m\Delta t,i+1) - u(m\Delta t,i)}{\left|u(m\Delta t,i+1) - u(m\Delta t,i)\right|} & \text{if } u(m\Delta t,i+1) \neq u(m\Delta t,i) \\ 0 & \text{if } u(m\Delta t,i+1) = u(m\Delta t,i) \end{cases} \end{pmatrix} \tag{16}$$

where the periodic boundary condition $u(n_t \Delta t, i) = u(0, i)$ and the Neumann boundary condition $u(m\Delta t, n_x) = u(m\Delta t, n_x - 1)$ are imposed. This Neumann boundary condition stems from an assumption that the occurrence probability of rain is indifferent whether the conditional DSL is $n_x - 1$ or n_x if n_x is large enough. The coefficient $u(1-u)$, which degenerates when u approaches to 0 or 1, is estimated as $u_{\max}(1-u_{\min})$ using

$$u_{\max} = \max\left(u(m\Delta t, i), u_{ave}\right) \tag{17}$$

and

$$u_{\min} = \min\left(u(m\Delta t, i), u_{ave}\right) \tag{18}$$

where

$$u_{ave} = \frac{u((m+1)\Delta t, i) + u((m-1)\Delta t, i) + u(m\Delta t, i+1) + u(m\Delta t, i-1)}{4}. \tag{19}$$

Then, (14) is finally discretized as a system of the ordinary differential equations

$$\frac{\mathrm{d}}{\mathrm{d}\tau} u(m\Delta t, i) = u_{\max}(1-u_{\min})\left(\frac{\sigma_t^{m,i} - \sigma_t^{m-1,i}}{\Delta t} + \frac{\sigma_x^{m,i} - \sigma_x^{m,i-1}}{\Delta x}\right), \tag{20}$$

imposing a boundary condition $\sigma_x^{m,-1} = \sigma_x^{m,0}$ on $i = 0$ for each *m*. The standard Runge-Kutta method is employed for the integration of (20) in the τ-direction. A critical issue with the dynamical system of (20) is that it does not necessarily converge to a steady state because of its singular diffusivity. Osher et al. (2005) proposed a reasonable stopping criterion for the ROF model using a discrepancy principle. However, an ad hoc value of τ is selected to stop the numerical integration so that acceptable regularization results are obtained.

3.4 Hazard futures

The multi-state Markov chain model constructed above is now applicable to the real-time assessment of drought risks based on the information of the state X_t, the DSL up to the current day *t*. This implies that the state X_t of the multi-state Markov chain is utilized as the index for decision making. Classic weather indexes such as SPI are not suitable for that purpose because they are not Markovian. Two critical quantities are considered here: the DSL in the future; and the total rainfall depth until a specified terminal day *N*. The last day the dry spell continuing from the current day *t* is denoted by *T*. The first and the second statistical moments of the DSL in the future, which is $T-t$ and represented by L, satisfy

$$\mathrm{E}\left[T-t \middle| X_t = i\right] = 1 + P_{i1}\mathrm{E}\left[T-(t+1)\middle| X_{t+1} = i+1\right] \tag{21}$$

and

$$\mathrm{E}\left[(T-t)^2 \middle| X_t = i\right] = 1 + P_{i1}\left(2\mathrm{E}\left[T-(t+1)\middle| X_{t+1} = i+1\right] + \mathrm{E}\left[(T-(t+1))^2 \middle| X_{t+1} = i+1\right]\right), \tag{22}$$

respectively, with the Neumann boundary conditions

$$\mathrm{E}\left[T-t \middle| X_t = n_x - 1\right] = \mathrm{E}\left[T-t \middle| X_t = n_x\right] \tag{23}$$

and

$$\mathrm{E}\left[(T-t)^2 \middle| X_t = n_x - 1\right] = \mathrm{E}\left[(T-t)^2 \middle| X_t = n_x\right]. \tag{24}$$

While, the first and the second statistical moments of the total rainfall depth until a specified terminal day *N*, which is $\sum_{k=t}^{k<N} r_{k+1}$ and represented by S, satisfy

$$\mathrm{E}\left[\sum_{k=t}^{k<N} r_{k+1} \middle| X_t = i\right] = P_{i0}\left(\frac{\alpha}{\beta} + \mathrm{E}\left[\sum_{k=t+1}^{k<N} r_{k+1} \middle| X_{t+1} = 0\right]\right) + P_{i1}\mathrm{E}\left[\sum_{k=t+1}^{k<N} r_{k+1} \middle| X_{t+1} = i+1\right] \tag{25}$$

and

$$\begin{aligned} &\mathrm{E}\left[\left(\sum_{k=t}^{k<N} r_{k+1}\right)^2 \middle| X_t = i\right] \\ &= \frac{\alpha(\alpha+1)}{\beta^2}P_{i0} + \frac{2\alpha}{\beta}P_{i0}\mathrm{E}\left[\sum_{k=t+1}^{k<N} r_{k+1} \middle| X_{t+1} = 0\right] \\ &+ P_{i0}\mathrm{E}\left[\left(\sum_{k=t+1}^{k<N} r_{k+1}\right)^2 \middle| X_{t+1} = 0\right] + P_{i1}\mathrm{E}\left[\left(\sum_{k=t+1}^{k<N} r_{k+1}\right)^2 \middle| X_{t+1} = i+1\right] \end{aligned}, \tag{26}$$

respectively. Proofs of (21), (22), (25), and (26) are provided in Appendix 2. As those (21), (22), (25), and (26) are recursive formulae, all the statistical moments $\mathrm{E}[L]$, $\mathrm{SD}[L]$, $\mathrm{E}[S]$, and $\mathrm{SD}[S]$ distributed on the grids over the set $[0, D_{\text{year}}) \times [0, \infty)$ can be routinely computed and serve as the hazard futures. Indeed, those hazard futures can be utilized to prescribe a critical level λ_t of the state X_t to take measures to avert the risks for each day t, and the day τ firstly attaining $X_\tau = \lambda_\tau$ is a hitting time in the mathematical context of the filtered probability space.

4. Results and discussions

4.1 Extraction of sub-periods having time-homogeneous rainfall regimes

In order to establish the multi-state Markov chain model whose dynamics in terms of positive rainfall depths is invariant, sub-periods are extracted from the whole period of 44 years. A water year is defined as the period of one year from August 1st through July 31st of the next year so that the wet season should not be split. Durations of 10 and 15 water years are examined as potential sub-periods of statistical homogeneity. Figure 4 shows the results of the K-S tests in terms of the significance levels α_L, indicating that the empirical distributions of positive rainfall depths well fit to the gamma distributions in the sub-periods including the 1980s and the 2010s and that there is a statistically significant difference between those two eras; the significance level α_L = 0.00367 in comparison of the two sub-periods 1984-1994 and 2002-2012 is the minimum among the cases of 10 water years, and the significance level α_L = 0.0123 in comparison of the two sub-periods 1982-1997 and 2003-2018 is the minimum among the cases of 15 water years. While, the null hypothesis that the observed positive rainfall depths in the data set of the sub-period 1977-1992 is drawn from the gamma distribution with parameters of (5) is scarcely rejected because of the significance level α_L = 0.999 (marked with a yellow frame in Figure 4). The significance level α_L = 0.951 in comparison of the sub-period 2004-2019 with the corresponding gamma distribution is the maximum among the cases of 15 water years (marked with an orange frame in Figure 4). The null hypothesis that the observed positive rainfall depths in the two data sets of those subperiods 1977-1992 and 2004-2019 are drawn from the same distribution is possibly rejected at the significance level α_L = 0.207 (marked with a red frame in Figure 4). Hereinafter, we focus on those two sub-periods of 15 water years 1977-1992 and 2004-2019. Figure 5 shows the empirical and the gamma distributions for the two sub-periods in terms of CDFs, implying a shift from the statistically homogeneous rainfall regime of 1977-1992 to the other of 2004-2019, which is much drier. Other basic statics of the data sets are summarized in Table 1. Occurrence of wet days is steadily decreasing, while the trend of positive rainfall depths is not definite.

Figure 4: Results of Kolmogorov-Smirnov tests in terms of the significance levels α_L to extract the sub-periods 1977-1992 and 2004-2019 of time-homogeneous rainfall regimes.

Figure 5: CFDs of empirical and gamma distributions for the sub-periods 1977-1992 and 2004-2019.

Table 1: Basic statistics of the data sets for the whole period and for the disjoint sub-periods including the selected ones.

4.2 The multi-state Markov chain model with regularized transition probabilities

The multi-state Markov chain model is constructed for each of the sub-periods 1977-1992 and 2004-2019, determining the transition probabilities. The time range δt is preferred to be small to achieve better resolution; however, it must be large enough to make (12) valid. By the trial-and-error method, $\delta t = 5$ days is chosen for all cases. To consider statistical moments of the DSL in the future and the total rainfall depth until a specified terminal day in the next subsection, two thresholds $r_\theta = 5$ mm, which is a typical threshold of effective rainfall, and $r_\theta = 0$ mm, which must coincide with the threshold of rainfall depths obeying to the gamma distributions, are respectively prescribed. The empirical transition probabilities are regularized according to the procedure described in the subsection 3.3. The computational grids are defined with $n_t = 365$ and $n_x = 200$. Strictly speaking, systematic errors occur in all the computed quantities below, associated with the difference between $D_{\text{year}} = 365.25$ and $n_t = 365$. However, for the sake of simplicity, we would be indifferent to those errors at the order of $(365.25 - 365) \times 100/365$ = 0.068 %. The step in the τ-direction is set as small as 10^{-4} for stable implementation of the Runge-Kutta method. Significant effects of regularization can be seen in the computed MTVF at $\tau = 20$. Figure 6 and Figure 7 compare the empirical and the regularized transition probabilities with the threshold $r_\theta = 5$ mm for the sub-periods 1977-1992. Figure 8 depicts the regularized transition probabilities with the threshold $r_\theta = 5$ mm for the sub-period 2004-2019 to compare with Figure 7 for the sub-period 1977-1992. In those figures, the value of $u(t, x)$ at each grid is plotted with the monotone colors on the t-x plane, where the state x refers to the DSL up to t. Due to the scarcity of recorded wet days, extreme values of 0 and 1, which are represented by white and black grids in the figures, respectively, often occur for empirical transition probabilities in adjacent states. Those spurious oscillations are successfully regularized in the MTVF without spoiling true extreme values of 0 and 1 identified from enough number of data. A significant difference between Figure 7 and Figure 8 can be seen in the distributions of white grids, representing the transition to the dry day almost surely. This indicates that the substantial onset of the dry seasons shifted from mid-April in the sub-period 1977-1992 to mid-March in the sub-period 2004-2019. The local people well perceive such a regime shift of DSL. Analogous comparisons between the empirical and the regularized transition probabilities with the threshold $r_\theta = 5$ mm for the sub-periods 2004-2019 and among the cases of the threshold $r_\theta = 0$ mm lead to similar observations.

Figure 6: Empirical transition probabilities P_{i0} with the range $\delta t = 5$ days and the threshold $r_\theta = 5$ mm for the sub-period 1977-1992.

Figure 7: Regularized transition probabilities P_{i0} with the range $\delta t = 5$ days and the threshold $r_\theta = 5$ mm for the sub-period 1977-1992.

Figure 8: Regularized transition probabilities P_{i0} with the range $\delta t = 5$ days and the threshold $r_\theta = 5$ mm for the sub-period 2004-2019.

The most significant technical novelty in constructing the multi-state Markov chain model from historical data is the introduction of the MTVF to regularize the distribution of the transition probability $u = u(t, x)$ as a function of the time and the state. Although fundamental properties of (14) are not well explored in the context of singular diffusion equations with mathematical rigor, it has successfully regularized the spurious oscillations indeed.

4.3 Application to real-time drought risk assessment

The multi-state Markov chain model is applied to real-time drought risk assessment, quantifying the different statistical moments as the hazard futures based on the information of the DSL up to t. The regularized transition probabilities are used for evaluation of statistical moments of L, the DSL in the future, and S, the total rainfall depth until the terminal day *N*. All the variable at $t = n_t$ are regarded as identical with those at $t = 0$ during the temporally backward computation of the recursive formulae (21), (22), (25), and (26). The fitted gamma distributions are assumed for positive rainfall depths. Dependence of those statistical moments of L and S on the states representing the information of the DSL up to t enables the farmers or their agricultural cooperative associations rationally assessing drought risks and thus tight decision making on crop management. Major annual rainfed crops in Nineveh Governorate, including the Nineveh Plains, are wheat, barley, clover, and flax (Hajim et al. 1996). Those crops are sown during the autumn months of October and November and harvested in May of the next Gregorian year. Therefore, the terminal day N is specified as 130 (May 10$^{\text{th}}$). Provided that supplementary irrigation is feasible, the standard irrigation water requirements of those crops and few other annual crops in a water year are summarized in Table 2, which has been adapted from Hajim et al. (1996). Water requirements under the rainfed condition can be inferred from Table 2 as well. Soil moisture ahead of sowing is needed for leaching and land preparation.

Table 2: The depths and numbers of irrigation (Depth (mm) : Number (times)) in a water year for major annual crops in Nineveh Governorate, adapted from Hajim et al. (1996).

For the DSL in the future, the recursive formulae (21) and (22) with (23) and (24) are computed until achieving steady year-periodic states to obtain the expectation $\mathrm{E}[L]$ and the standard deviation $\mathrm{SD}[L]$. The results are depicted in Figures 9 and 10 for the sub-period 1977-1992 and in Figures 11 and 12 for the sub-period 2004-2019. Those figures present $\mathrm{E}[L]$ and $\mathrm{SD}[L]$, which are also the functions of t and x, in the same way as Figures 6-8 but using the different colors indicated in the legend. In general, the variations of $\mathrm{E}[L]$ and $\mathrm{SD}[L]$ over the t-x plane are more intense in the sub-period 2004-2019 than in the sub-period 1977-1992. The maximum $\mathrm{E}[L]$ in the sub-period 1977-1992 is 230 days on the condition that DSL up to $t = 96$ (April 6$^{\text{th}}$) is 11 days, while that in the sub-period 2004-2019 is 269 days on the condition that DSL up to $t = 79$ (March 20th) is 11 days. As a validation of the computed values of $\mathrm{E}[L]$ and $\mathrm{SD}[L]$, statistical analysis is made for each dry season, assuming the Gumbel distribution for its length L defined as the maximum DSL in each Gregorian year. The Gumbel distribution, whose CDF $F_{\mathrm{Gum}}(L)$ is given by

$$F_{\mathrm{Gum}}(L) = \exp\left(\exp-\left(L - \mu_{\mathrm{Gum}}\right)/\beta_{\mathrm{Gum}}\right) \tag{27}$$

with the two parameters β_{Gum} and μ_{Gum}, is commonly used for statistically modeling extreme values such as the annual maximum DSL (Vicente-Serrano and Beguería-Portugués 2003). It is also an advantage of using the Gumbel distribution in this case that the two parameters β_{Gum} and μ_{Gum} are uniquely estimated from $\mathrm{E}[L]$ and $\mathrm{SD}[L]$ as $\beta_{\mathrm{Gum}} = \sqrt{6}\,\mathrm{SD}[L]/\pi$ and $\mu_{\mathrm{Gum}} = \mathrm{E}[L] - \beta_{\mathrm{Gum}}\gamma$, where γ is the Euler-Mascheroni constant, approximately equal to 0.577216. The mode of the Gumbel distribution is equal to the value of the parameter μ_{Gum}. Table 3 and Table 4 present different parameter values for each year in the sub-period 1977-1992 and in the sub-period 2004-2019, respectively, including the day of the onset of the dry season, $\mathrm{E}[L]$ and $\mathrm{SD}[L]$ on that day with the state $x = 1$, the values of the parameters of the Gumbel distribution, the observed length L of the dry season, the CDF of the Gumbel distribution at that observed L, the return period ($= 1/(1 - \mathrm{CDF})$) (year), and the significance level of the K-S test with $n_E = 1$ at which the null hypothesis that the observed L is drawn from the Gumbel distribution. The onsets of the dry seasons fall on the spring months from March to May, except for the year 1989, where variations in $\mathrm{E}[L]$ and $\mathrm{SD}[L]$ are the most intense, implying that this validation examines the most sensitive cases of the multi-state Markov chain. For the sub-period 1977-1992, the averages of the mode μ_{Gum} and the observed L are 153.923 days and 185.667 days, respectively. For the sub-period 1977-1992, the averages of the mode μ_{Gum} and the observed L are 172.748 days and 194.500 days, respectively. The null hypothesis cannot be rejected in all the years during the two sub-periods at a significance level not less than 0.27. The return periods are quite normal in the sub-period 1977-1992, while anomalous 91.735 years in the year 2007 and 1.000 year in the year 2018 can be seen in the sub-period 2004-2019. This is comparable with the analysis of Awchi and Kalyana (2017) based on SPI, as mentioned in Section 2. The local people have perceived these anomalies as well. From the statistical analysis above, we consider that the computed values of $\mathrm{E}[L]$ and $\mathrm{SD}[L]$ comprehensively represent the statistical behavior of the lengths of the dry seasons including the anomalies. A significant change from the sub-period 1977-1992 to the sub-period 2004-2019 is that occurrence of a long dry spell ($\mathrm{E}[L] >$ 50 days as white patches surrounded by blue zones in Figure 11, $\mathrm{SD}[L]$ decreasing with increasing DSL up to t as colored changing from dodger blue to white and then to blue in Figure 12) has become evident on the condition of a prolonged (5-10 days) DSL up to t in December or January but is not tangible at the time of sowing. It is another serious concern emerged in the sub-period 2004-2019 that the onset of the dry season ($\mathrm{E}[L] \approx$ 100 days as colored light blue in Figure 11, $\mathrm{SD}[L] \approx$ 120 days as colored green yellow in Figure 12) has become very likely on the condition of a prolonged (5-10 days) DSL up to t in early-March. However, there is not much difference between the two sub-periods in the values of $\mathrm{E}[L]$ and $\mathrm{SD}[L]$ if a long dry spell of 50 days is actually observed on April 1st. To summarize the assessment above, it can be said that the occurrence of a long dry spell in the middle of the growing season, as well as the early onset of the dry season, have become common drought risks in recent decades. Introducing supplementary irrigation, if feasible, is an effective measure to avert such risks. More illustratively, an example of the critical level λ_t for each day t, such that supplementary irrigation should be performed at a hitting time, is prescribed as

$$\lambda_t = \inf\left\{ \xi \,\middle|\, x + \mu_{\mathrm{Gum}}(t, x) > \Lambda \text{ for } \forall x \geq \xi \right\} \tag{28}$$

where $\xi, x \in \mathbb{N}$, $\mu_{\mathrm{Gum}}(t,x)$ is the parameter μ_{Gum} of the Gumbel distribution estimated from $\mathrm{E}[L]$ and $\mathrm{SD}[L]$ on the condition that DSL up to t is x days, and Λ is a specified critical length of the mode of the total dry spell (= $x + \mu_{\mathrm{Gum}}(t,x)$). The hitting time of X_τ to λ_t is a stopping time relative to the filtration $\{\mathcal{F}_t\}$ such that $\{\tau = t\} \in \mathcal{F}_t$ for any t. Figure 13 shows the critical levels λ_t for the two sub-periods with Λ = 50 days, clearly verifying the assessment about the risk in the middle of the growing season with the most noticeable difference between the two sub-periods seen in the month of January. According to Hajim et al. (1996) based on the survey before 1996, supplementary irrigation was supposed to be unnecessary in January as being consistent with the large critical levels λ_t in January for the sub-period 1977-1992. However, the critical levels λ_t in January for the sub-period 2004-2019 are as small as 11 days and very likely achieved.

Figure 9: Expected DSL in the future based on the regularized model for the sub-period 1977-1992.

Figure 10: Standard deviation of DSL in the future based on the regularized model for the sub-period 1977-1992.

Figure 11: Expected DSL in the future based on the regularized model for the sub-period 2004-2019.

Figure 12: Standard deviation of DSL in the future based on the regularized model for the sub-period 2004-2019.

Table 3: Statistical analysis of each dry season during the sub-period 1977-1992, assuming the Gumbel distribution for the length *L*.

Table 4: Statistical analysis of each dry season during the sub-period 2004-2019, assuming the Gumbel distribution for the length *L*.

Figure 13: Critical levels λ_t for the two sub-periods 1977-1992 and 2004-2019 to alert the occurrence of a long dry spell persisting over Λ = 50 days.

The total rainfall depth until a specified terminal day is another main concern in rainfed agriculture. The recursive formulae (25) and (26) are computed from the trivial terminal condition $\mathrm{E}\left[0 \middle| X_N = i\right] = 0$ for one year to obtain the expectation $\mathrm{E}\left[S\right]$ and the standard deviation $\mathrm{SD}\left[S\right]$ of the total rainfall depth until May 10$^{\mathrm{th}}$. The results are depicted in Figures 14 and 15 for the sub-period 1977-1992 and in Figures 16 and 17 for the sub-period 2004-2019. Those figures have the same structure as Figures 9-13. The values of $\mathrm{E}\left[S\right]$ on the day $N+1=131$ (May 11$^{\mathrm{th}}$) range from 374.8 mm to 375.3 mm for the sub-period 1977-1992 and from 312.8 mm to 314.3 mm for the sub-period 2004-2019, while the observed average annual rainfall depths were 364.3 mm and 314.7 mm for those two respective sub-periods. Taking the seasonally biased missing data due to the Iran-Iraq War (1980-1988) and the Iraqi Civil War (2014-2017) into account, those comparable values are well proving the consistency of the regularization technique. In general, $\mathrm{E}\left[S\right]$ and $\mathrm{SD}\left[S\right]$ are less dependent on the DSL up to *t*, but exceptions can been seen during the months of October and November; there are depressions in $\mathrm{E}\left[S\right]$ for both sub-periods if the DSL up to t lasts for several weeks. The severity of those depressions in $\mathrm{E}\left[S\right]$ is higher in the sub-period 2004-2019, especially in conjunction with the evident occurrence of dry spells during the growing months of December and January. Nevertheless, observing the DSL up to t in those sowing months of October and November, the farmers or their agricultural cooperative associations who assess the drought risks in terms of $\mathrm{E}\left[S\right]$ and $\mathrm{SD}\left[S\right]$ still can make decisions on species and varieties of the annual rainfed crops to cultivate. Such a time t of decision making is also a stopping time relative to the filtration $\left\{\mathcal{F}_t\right\}$. For instance, if clover is sown in early October with the provision of soil moisture brought by exceptionally early rain in September, and if a dry spell lasts for several weeks in October, then it is recommendable to abandon the clover and to substitute the other crops which can be sown after the dry spell ends in early November with less water requirement in total. This measure of risk aversion may be more feasible than supplementary irrigation in the Nineveh Plains.

Figure 14: Expected total rainfall depth until May 10th based on the regularized model for the sub-period 1977-1992.

Figure 15: Standard deviation of total rainfall depth until May 10th based on the regularized model for the sub-period 1977-1992.

Figure 16: Expected total rainfall depth until May 10th based on the regularized model for the sub-period 2004-2019.

Figure 17: Standard deviation of total rainfall depth until May 10th based on the regularized model for the sub-period 2004-2019.

5. Conclusions

The multi-state Markov chain model constructed in this study is more straightforward than conventional weather generation models and drought indices in the sense that the set of DSL itself is taken as the space of states. The transition probability from one state to another is year-periodically varying, while the probability law for positive rainfall depths is time-homogeneous. That sophisticated structure of the multi-state Markov chain model enables different types of real-time drought risk assessment.

Mathematical insight shall be given to that regularization technique in future studies, as this study merely provides a formal derivation of the technique for environmental scientists and engineers. The MTVF can be considered in a generalized framework of negative Sobolev spaces (Giga et al. 2019) to discuss more advanced approaches to the regularization of transition probabilities.

The K-S tests on the probability distribution of positive rainfall depths provided more quantitative and detailed information than conventional trend analyses did. The results revealed the clear regime shift from the 1980s to the 2010s, though the statistical analysis conducted here cannot scientifically attribute it to any causal linkage from the climate change. Then, the multi-state Markov chain model with regularized transition probabilities was constructed for each of the sub-periods having time-homogeneous rainfall regimes, to be applied to real-time drought risk assessment evaluating the statistical moments of L and S. The computed $\mathrm{E}\left[L\right]$ and $\mathrm{SD}\left[L\right]$ implied drought risks which cannot be anticipated in the sowing months; the occurrence of a long dry spell in the middle of the growing season became more frequent, and the onset of substantial dry seasons shifted significantly earlier, which might negatively affect the annual rainfed crops at the late growth stages. The illustrative example suggested the validity of supplementary irrigation, which is unfortunately not feasible due to the lack of infrastructure and security as of 2020. The computed $\mathrm{E}\left[S\right]$ and $\mathrm{SD}\left[S\right]$ quantified the information on the total rainfall depth which the annual rainfed crops could receive until their harvest time. The values of computed $\mathrm{E}\left[S\right]$ were used for verifying the regularization technique developed in this study as well. Crop management in terms of choosing species and varieties may be the only feasible measure of risk aversion in the current situation of the Nineveh Plains. Other indices involving DSL and rainfall depth can be defined to assess different aspects of drought risks. Statistical moments of higher order can be calculated as well.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study is based on the discussions at 2018 IMI Joint Use Research Program Workshop (II) "Modeling and Analysis of Time Series Data in Math-Agro Sciences" at Kyushu University, Japan. The authors thank Iraqi authorities for the provision of meteorological data. The authors acknowledge grants-in-aid for scientific research No.16KT0018 and No.19KK0167 from the Japan Society for the Promotion of Science (JSPS) and ISHIZUE 2020 from Kyoto University Research Development Program.

Appendix 1: Scope of the singular diffusion equation with the degenerating coefficient

The purpose of regularization in general is to remove spurious oscillations appearing in a function. Let $u = u(t, x)$ be such a function defined in a domain Ω included in the *t*-*x*-plane. The magnitude of oscillations in u is evaluated with the functional

$$J = \int_{\Omega} |\nabla u| \, \mathrm{d}\Omega \tag{29}$$

which is referred to as the total variation of u. The Euler-Lagrange equation in the context of the variational calculus to minimize the functional J in (29) formally becomes

$$\nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right) = 0 . \tag{30}$$

The flux $\nabla u / |\nabla u|$ in the left hand side of (30) is a unit vector if $|\nabla u| \neq 0$ and is not well defined if $|\nabla u| = 0$, resulting in the singularity of (30). The proposed approximation of the flux with (16) is a basic method to overcome such singularity. On the other hand, the practical difficulty encountered in the application to the transition probabilities is that there are true abrupt variations in the neighborhoods of the points achieving extreme values of 0 and 1. The idea employed here is to multiply the degenerating coefficient $u(1-u)$ to both sides of (30) as

$$u(1-u)\nabla\cdot\left(\frac{\nabla u}{|\nabla u|}\right)=0 \qquad (31)$$

where the removal of oscillations is inactivated if u is equal to 0 or 1. Using the estimate $u_{\max}(1-u_{\min})$ defined with (17), (18) , and (19) is to detect the appropriate points of inactivation. However, the singularity of (30) still remains in (31), and its direct solution is difficult to implement. Inspired by the celebrated ROF model, the unsteady term $\partial u/\partial\tau$ is added to (31) in order to obtain the singular diffusion equation (14), from which the desired MTVF is successfully computed.

Appendix 2: Proofs

Proofs of recursive formulae of (21), (22), (25), and (26) are provided as below. The relations given in (5) should be referred to as well.

Proof of (21):

$$\begin{aligned}
&\mathrm{E}\left[T-t\middle|X_t=i\right]\\
&=P_{i0}\left(1+\mathrm{E}\left[T-(t+1)\middle|X_{t+1}=0\right]\right)\\
&\quad+P_{i1}\left(1+\mathrm{E}\left[T-(t+1)\middle|X_{t+1}=i+1\right]\right)\\
&=P_{i0}+\left(1-P_{i0}\right)\left(1+\mathrm{E}\left[T-(t+1)\middle|X_{t+1}=i+1\right]\right)\\
&=1+\left(1-P_{i0}\right)\mathrm{E}\left[T-(t+1)\middle|X_{t+1}=i+1\right]\\
&=1+P_{i1}\mathrm{E}\left[T-(t+1)\middle|X_{t+1}=i+1\right]
\end{aligned} \qquad (32)$$

Proof of (22):

$$\begin{aligned}
&\mathrm{E}\left[\left(T-t\right)^2 \middle| X_t = i\right] \\
&= \mathrm{E}\left[\left(1+T-\left(t+1\right)\right)^2 \middle| X_t = i\right] \\
&= \mathrm{E}\left[1^2 + 2\left(T-\left(t+1\right)\right)+\left(T-\left(t+1\right)\right)^2 \middle| X_t = i\right] \\
&= 1 + P_{i1}\left(2\mathrm{E}\left[T-\left(t+1\right) \middle| X_{t+1} = i+1\right] + \mathrm{E}\left[\left(T-\left(t+1\right)\right)^2 \middle| X_{t+1} = i+1\right]\right)
\end{aligned} \tag{33}$$

Proof of (25):

$$\begin{aligned}
&\mathrm{E}\left[\sum_{k=t}^{k<N} r_{k+1} \middle| X_t = i\right] \\
&= P_{i0}\left(\mathrm{E}\left[r_{t+1} \middle| X_{t+1} = 0\right] + \mathrm{E}\left[\sum_{k=t+1}^{k<N} r_{k+1} \middle| X_{t+1} = 0\right]\right) + P_{i1}\left(0 + \mathrm{E}\left[\sum_{k=t+1}^{k<N} r_{k+1} \middle| X_{t+1} = i+1\right]\right) \\
&= P_{i0}\left(\mathrm{E}\left[r_{t+1} \middle| r_{t+1} > 0\right] + \mathrm{E}\left[\sum_{k=t+1}^{k<N} r_{k+1} \middle| X_{t+1} = 0\right]\right) + P_{i1}\mathrm{E}\left[\sum_{k=t+1}^{k<N} r_{k+1} \middle| X_{t+1} = i+1\right] \\
&= P_{i0}\left(\frac{\alpha}{\beta} + \mathrm{E}\left[\sum_{k=t+1}^{k<N} r_{k+1} \middle| X_{t+1} = 0\right]\right) + P_{i1}\mathrm{E}\left[\sum_{k=t+1}^{k<N} r_{k+1} \middle| X_{t+1} = i+1\right]
\end{aligned} \tag{34}$$

with

$$\begin{aligned}
&\mathrm{E}\left[\sum_{k=t}^{k<t+1} r_k \middle| X_t = i\right] \\
&= \mathrm{E}\left[r_t \middle| X_t = i\right] = \begin{cases} 0 & \text{if } i > 0 \\ \mathrm{E}\left[r_t \middle| X_t = 0\right] & \text{if } i = 0 \end{cases} \\
&= \begin{cases} 0 & \text{if } i > 0 \\ \mathrm{E}\left[r_t \middle| r_t > 0\right] & \text{if } i = 0 \end{cases} \\
&= \begin{cases} 0 & \text{if } i > 0 \\ \dfrac{\alpha}{\beta} & \text{if } i = 0 \end{cases}
\end{aligned} \tag{35}$$

Proof of (26):

$$
\begin{aligned}
&\mathrm{E}\left[\left(\sum_{k=t}^{k<N} r_{k+1}\right)^2 \middle| X_t = i\right] \\
&= \mathrm{E}\left[\left(r_{t+1} + \sum_{k=t+1}^{k<N} r_{k+1}\right)^2 \middle| X_t = i\right] \\
&= \mathrm{E}\left[r_{t+1}^2 + 2r_{t+1}\sum_{k=t+1}^{k<N} r_{k+1} + \left(\sum_{k=t+1}^{k<N} r_{k+1}\right)^2 \middle| X_t = i\right] \\
&= \mathrm{E}\left[r_{t+1}^2 \middle| X_t = i\right] + 2\mathrm{E}\left[r_{t+1}\sum_{k=t+1}^{k<N} r_{k+1} \middle| X_t = i\right] + \mathrm{E}\left[\left(\sum_{k=t+1}^{k<N} r_{k+1}\right)^2 \middle| X_t = i\right] \\
&= P_{i0}\mathrm{E}\left[r_{t+1}^2 \middle| X_{t+1} = 0\right] + 2P_{i0}\mathrm{E}\left[r_{t+1}\sum_{k=t+1}^{k<N} r_{k+1} \middle| X_{t+1} = 0\right] \\
&\quad + P_{i0}\mathrm{E}\left[\left(\sum_{k=t+1}^{k<N} r_{k+1}\right)^2 \middle| X_{t+1} = 0\right] + P_{i1}\mathrm{E}\left[\left(\sum_{k=t+1}^{k<N} r_{k+1}\right)^2 \middle| X_{t+1} = i+1\right] \\
&= P_{i0}\mathrm{E}\left[r_{t+1}^2 \middle| r_{t+1} > 0\right] + 2P_{i0}\left(\mathrm{E}\left[r_{t+1} \middle| r_{t+1} > 0\right]\mathrm{E}\left[\sum_{k=t+1}^{k<N} r_{k+1} \middle| X_{t+1} = 0\right]\right) \\
&\quad + P_{i0}\mathrm{E}\left[\left(\sum_{k=t+1}^{k<N} r_{k+1}\right)^2 \middle| X_{t+1} = 0\right] + P_{i1}\mathrm{E}\left[\left(\sum_{k=t+1}^{k<N} r_{k+1}\right)^2 \middle| X_{t+1} = i+1\right] \\
&= \frac{\alpha(\alpha+1)}{\beta^2}P_{i0} + \frac{2\alpha}{\beta}P_{i0}\mathrm{E}\left[\sum_{k=t+1}^{k<N} r_{k+1} \middle| X_{t+1} = 0\right] \\
&\quad + P_{i0}\mathrm{E}\left[\left(\sum_{k=t+1}^{k<N} r_{k+1}\right)^2 \middle| X_{t+1} = 0\right] + P_{i1}\mathrm{E}\left[\left(\sum_{k=t+1}^{k<N} r_{k+1}\right)^2 \middle| X_{t+1} = i+1\right]
\end{aligned} \tag{36}
$$

References

Agha OMM, Şarlak N (2016) Spatial and temporal patterns of climate variables in Iraq. Arab J Geosci 9:302 doi:10.1007/s12517-016-2324-y

Al-Ansari N (2013) Management of water resources in Iraq: perspectives and prognoses. Engineering 5:667-684 doi:10.4236/eng.2013.58080

Al-Khayat BYT, Al-Sulaiman MSS (2013) Forecasting of rainy conditions in Mosul city. Iraqi Journal of Statistical Sciences 13:19-32

Al-Najafee EH, Rashad EM (2012) Rainfall levels and their impact on wheat productivity: a comparative study between rainfall assured and semi-assured regions in Nineveh. Tanmyat al-Rafidain 34:161-170

Anagnostopoulou C, Maheras P, Karacostas T, Vafiadis M (2003) Spatial and temporal analysis of dry spells in Greece. Theor Appl Climatol 74:77-91 doi:10.1007/s00704-002-0713-5

Angelidis P, Maris F, Kotsovinos N, Hrissanthou V (2012) Computation of drought index SPI with alternative distribution functions. Water Resour Manage 26:2453-2473 doi:10.1007/s11269-012-0026-0

Awchi TA, Kalyana MM (2017) Meteorological drought analysis in northern Iraq using SPI and GIS. Sustain Water Resour Manag 3:451-463 doi:10.1007/s40899-017-0111-x

Azooz A, Talal S (2015) Evidence of climate change in Iraq. Journal of Environment Protection and Sustainable Development 1:66-73

Barron J, Rockstrom J, Gichuki F, Hatibu N (2003) Dry spell analysis and maize yields for two semi-arid locations in east Africa. Agr Forest Meteorol 117:23-37 doi:10.1016/S0168-1923(03)00037-6

Cabrera BL, Odening M, Ritter M (2013) Pricing rainfall futures at the CME. J Bank Financ 37:4286-4298 doi:10.1016/j.jbankfin.2013.07.042

Cavus Y, Aksoy H (2019) Spatial drought characterization for Seyhan River basin in the Mediterranean region of Turkey. Water 11 doi:10.3390/w11071331

Cavus Y, Aksoy H (2020) Critical drought severity/intensity-duration-frequency curves based on precipitation deficit. J Hydrol 584 doi:10.1016/j.jhydrol.2019.124312

Evans JP (2009) 21st century climate change in the Middle East. Clim Change 92:417-432 doi:10.1007/s10584-008-9438-5

Fadhil RM Markovian properties and distribution of daily rainfall in northern Iraq. In: Al-Zyoud F, Abdel-Ghani, A. (ed) The Eighth Scientific Agricultural Conference ESAC-2018, Karak, Jordan, 2018. National Agricultural Research Center, Baqa', Jordan, p 57

Farr TG et al. (2007) The shuttle radar topography mission. Rev Geophys 45:RG2004 doi:10.1029/2005RG000183

Fischer BMC, Mul ML, Savenije HHG (2013) Determining spatial variability of dry spells: a Markov-based method, applied to the Makanya catchment, Tanzania. Hydrol Earth Syst Sci 17:2161-2170 doi:10.5194/hess-17-2161-2013

Gao C, Booij MJ, Xu YP (2020) Development and hydrometeorological evaluation of a new stochastic daily rainfall model: Coupling Markov chain with rainfall event model. J Hydrol 589 doi:10.1016/j.jhydrol.2020.125337

Giga Y, Muszkieta M, Rybka P (2019) A duality based approach to the minimizing total variation flow in the space H^{-s}. Jpn J Ind Appl Math 36:261-286 doi:10.1007/s13160-018-00340-4

Hajim AY, Al-Dabagh AY, Yaseen HI, Dawud AF, Shayth AH (1996) Analysis of irrigation in fields and orchards in Nineveh. Department of Irrigation and Drainage, College of Engineering, University of Mosul, Mosul, Iraq (in Arabic)

IPCC (2007) Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor, and H.L. Miller (eds.).

Jimoh OD, Webster P (1999) Stochastic modelling of daily rainfall in Nigeria: intra-annual variation of model parameters. J Hydrol 222:1-17 doi:10.1016/S0022-1694(99)00088-8

Kadim AA (2013) The negative and positive divergence of the heat and the rain refraction from their common average at Mosul, Baghdad and Basra stations. Adab al-Basrah 67:309-334

Kalyan MM, Awchi TA (2015) Investigating the meteorological drought in Northern Iraq using deciles method. Al-Rafdain Engineering 23:12-21

Leobacher G, Ngare P (2011) On modelling and pricing rainfall derivatives with seasonality. Appl Math Finance 18:71-91 doi:10.1080/13504861003795167

Loucks DP, van Beek E (2005) Water Resource Systems Planning and Management: An Introduction to Methods, Models and Applications. Studies and Reports in Hydrology. UNESCO PUBLISHING, Paris

Martin-Vide J, Gomez L (1999) Regionalization of peninsular Spain based on the length of dry spells. Int J Climatol 19:537-555 doi:10.1002/(Sici)1097-0088(199904)19:5<537::Aid-Joc371>3.0.Co;2-X

Masala G (2014) Rainfall derivatives pricing with an underlying semi-Markov model for precipitation occurrences. Stoch Environ Res Risk Assess 28:717-727 doi:10.1007/s00477-013-0784-0

Mustafa LMF (2012) Spatial and Temporal Variation of Rainfall in Ninava Governorate. Journal of Education and Science 25:98-114 doi:10.33899/edusj.2012.66773

Nop C, Fadhil RM, Unami K (2021) A multi-state Markov chain model for rainfall to be used in optimal operation of rainwater harvesting systems. J Clean Prod 285:124912 doi:10.1016/j.jclepro.2020.124912

Ojara MA, Lou YS, Aribo L, Namumbya S, Uddin MJ (2020) Dry spells and probability of rainfall occurrence for Lake Kyoga Basin in Uganda, East Africa. Nat Hazards 100:493-514 doi:10.1007/s11069-019-03822-x

Onof C, Chandler RE, Kakou A, Northrop P, Wheater HS, Isham V (2000) Rainfall modelling using Poisson-cluster processes: a review of developments. Stoch Environ Res Risk Assess 14:384-411 doi:DOI 10.1007/s004770000043

Osher S, Burger M, Goldfarb D, Xu JJ, Yin WT (2005) An iterative regularization method for total variation-based image restoration. Multiscale Model Sim 4:460-489 doi:10.1137/040605412

Rasheed AMM (2010) Analysis of rainfall drought periods in the North of Iraq using standard precipitation index (SPI). Al-Rafdain Engineering 18 60-72

Richardson CW, Wright DA (1984) WGEN: A model for generating daily weather variables. US Department of Agriculture, Agricultural Research Service Washington, DC, USA,

Robaa SM, AL-Barazanji ZJ (2013) Trends of annual mean surface air temperature over Iraq. Nature and Science 11:138-145

Rockström J et al. (2010) Managing water in rainfed agriculture—The need for a paradigm shift. Agric Water Manage 97:543-550 doi:10.1016/j.agwat.2009.09.009

Rudin LI, Osher S, Fatemi E (1992) Nonlinear total variation based noise removal algorithms. Physica D 60:259-268 doi:10.1016/0167-2789(92)90242-F

Salman SA, Shahid S, Ismail T, Ahmed K, Wang XJ (2018a) Selection of climate models for projection of spatiotemporal changes in temperature of Iraq with uncertainties. Atmos Res 213:509-522 doi:10.1016/j.atmosres.2018.07.008

Salman SA, Shahid S, Ismail T, bin Abd Rahman N, Wang XJ, Chung ES (2018b) Unidirectional trends in daily rainfall extremes of Iraq. Theor Appl Climatol 134:1165-1177 doi:10.1007/s00704-017-2336-x

Sharifi E, Unami K, Yangyuoru M, Fujihara M (2016) Verifying optimality of rainfed agriculture using a stochastic model for drought occurrence. Stoch Environ Res Risk Assess 30:1503-1514 doi:10.1007/s00477-015-1129-y

Sirangelo B, Caloiero T, Coscarelli R, Ferrari E (2015) A stochastic model for the analysis of the temporal change of dry spells. Stoch Environ Res Risk Assess 29:143-155 doi:10.1007/s00477-014-0904-5

Sirangelo B, Caloiero T, Coscarelli R, Ferrari E (2017) Stochastic analysis of long dry spells in Calabria (Southern Italy). Theor Appl Climatol 127:711-724 doi:10.1007/s00704-015-1662-0

Taha MAQ (2014) Selecting the best models in calculating the amount of rainfall in Sinjar and Mosul stations. Journal of University of Babylon for Pure and Applied Sciences 22:2015-2022

Tatano H, Okada N, Kawai H (1992) Optimal operation model of a single reservoir with drought duration explicitly concerned. Stoch Hydrol Hydraul 6:123-134 doi:10.1007/BF01591334

Tong Z, Liu A (2021) A censored Ornstein-Uhlenbeck process for rainfall modeling and derivatives pricing. Physica A 566:125619 doi:10.1016/j.physa.2020.125619

Turvey CG (2001) Weather derivatives for specific event risks in agriculture. Rev Agric Econ 23:333-351

Unami K, Abagale F, Yangyuoru M, Alam A, Kranjac-Berisavljevic G (2010) A stochastic differential equation model for assessing drought and flood risks. Stoch Environ Res Risk Assess 24:725-733 doi:10.1007/s00477-009-0359-2

Unami K, Mohawesh O (2018) A unique value function for an optimal control problem of irrigation water intake from a reservoir harvesting flash floods. Stoch Environ Res Risk Assess 32:3169-3182 doi:10.1007/s00477-018-1527-z

Unami K, Mohawesh O, Fadhil RM (2019) Time periodic optimal policy for operation of a water storage tank using the dynamic programming approach. Appl Math Comput 353:418-431 doi:10.1016/j.amc.2019.02.005

Vicente-Serrano SM, Beguería-Portugués S (2003) Estimating extreme dry-spell risk in the middle Ebro valley (northeastern Spain): a comparative analysis of partial duration series with a general Pareto distribution and annual maxima series with a Gumbel distribution. International Journal of Climatology: A Journal of the Royal Meteorological Society 23:1103-1118 doi:10.1002/joc.934

Wilby RL, Prudhomme C, Parry S, Muchan K (2015) Persistence of hydrometeorological droughts in the United Kingdom: A regional analysis of multi-season rainfall and river flow anomalies. J Extreme Events 2:1550006 doi:10.1142/S2345737615500062

Williams D (1991) Probability with Martingales. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, UK

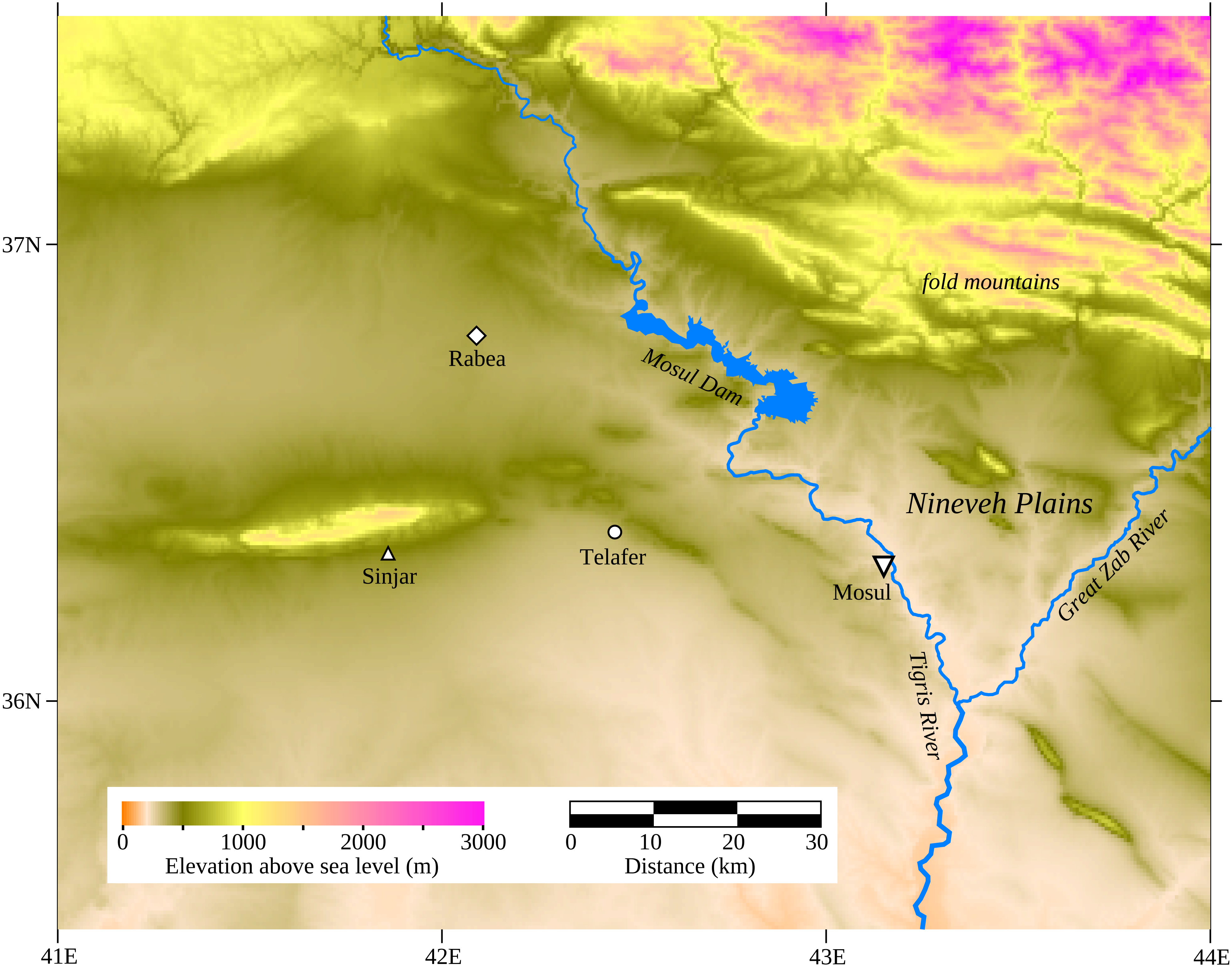
World Meteorological Organization (2012) Standardized Precipitation Index User Guide vol WMO-No.1090. WMO, Geneva

Yadeta D, Kebede A, Tessema N (2020) Climate change posed agricultural drought and potential of rainy season for effective agricultural water management, Kesem sub-basin, Awash Basin, Ethiopia. Theor Appl Climatol 140:653-666 doi:10.1007/s00704-020-03113-7

Yang LC, Franzke CLE, Fu ZT (2020) Power-law behaviour of hourly precipitation intensity and dry spell duration over the United States. Int J Climatol 40:2429-2444 doi:10.1002/joc.6343

Zakaria S, Al-Ansari N, Knutsson S (2013) Historical and future climatic change scenarios for temperature and rainfall for Iraq. Journal of Civil Engineering and Architecture 7:1574-1594

Fig. 1



37N
36N
41E
42E
43E
44E
fold mountains
Rabea
Mosul Dam
Nineveh Plains
Great Zab River
Telafer
Sinjar
Mosul
Tigris River
0
1000
2000
3000
Elevation above sea level (m)
0
10
20
30
Distance (km)

Fig. 2

Click here to access/download;colour figure;Fig2-RainChart.eps

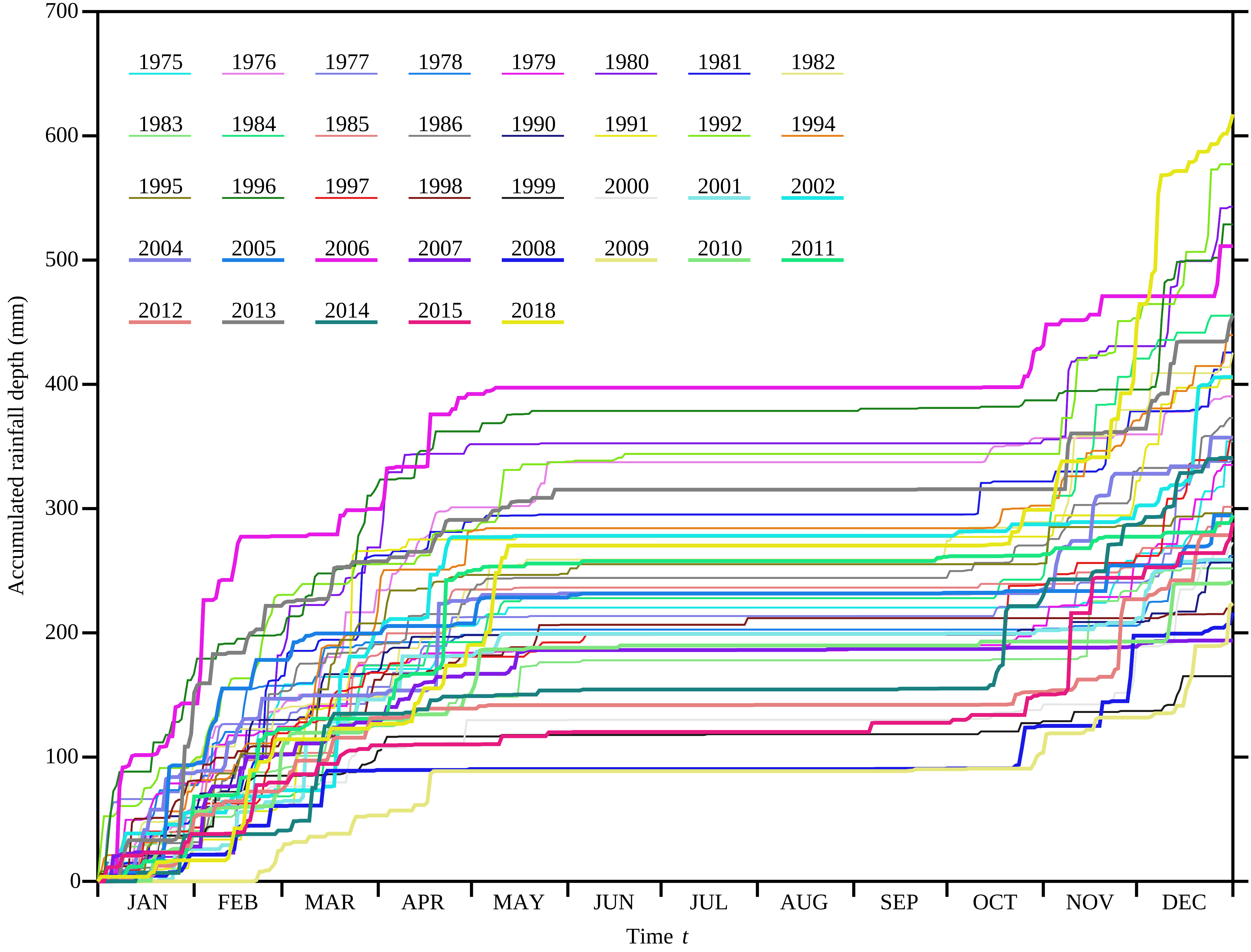


Fig. 3

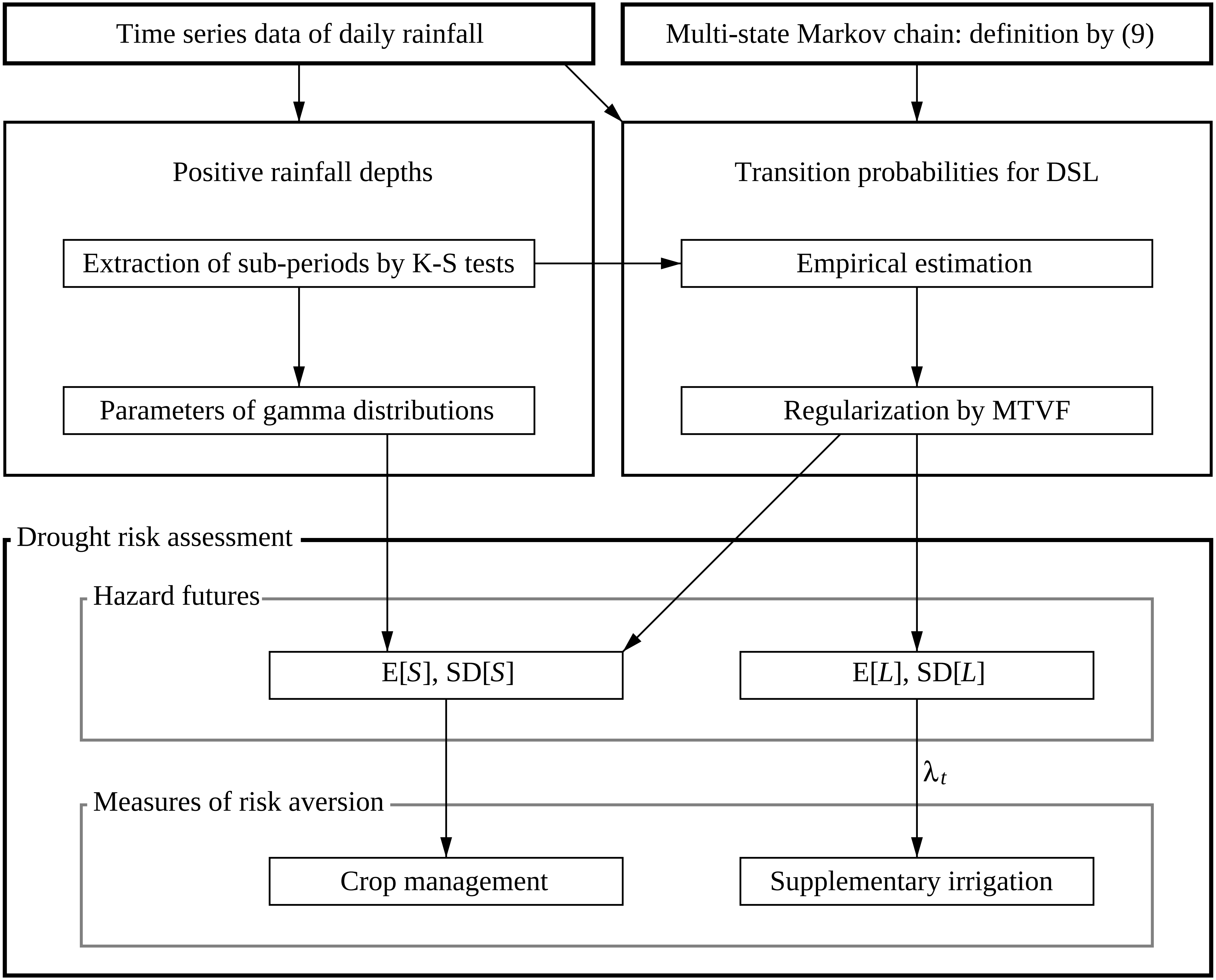

Time series data of daily rainfall
Multi-state Markov chain: definition by (9)
Positive rainfall depths
Transition probabilities for DSL
Extraction of sub-periods by K-S tests
Empirical estimation
Parameters of gamma distributions
Regularization by MTVF
Drought risk assessment
Hazard futures
E[S], SD[S]
E[L], SD[L]
λt
Measures of risk aversion
Crop management
Supplementary irrigation

Fig. 4

Click here to access/download;colour figure;Fig4-alphaL.eps

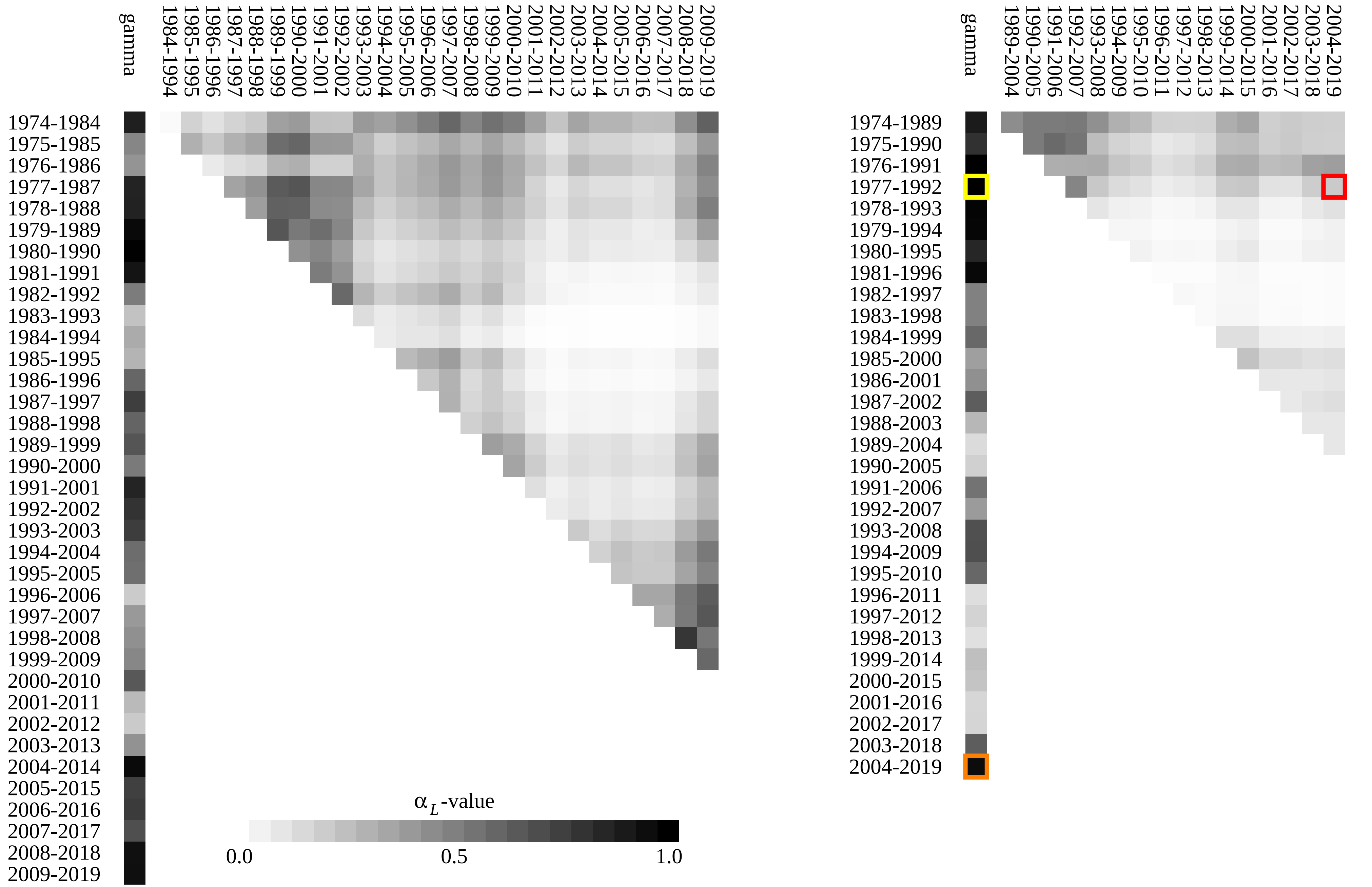


Fig. 5

Click here to access/download;colour figure;Fig5-mosulCDF.eps

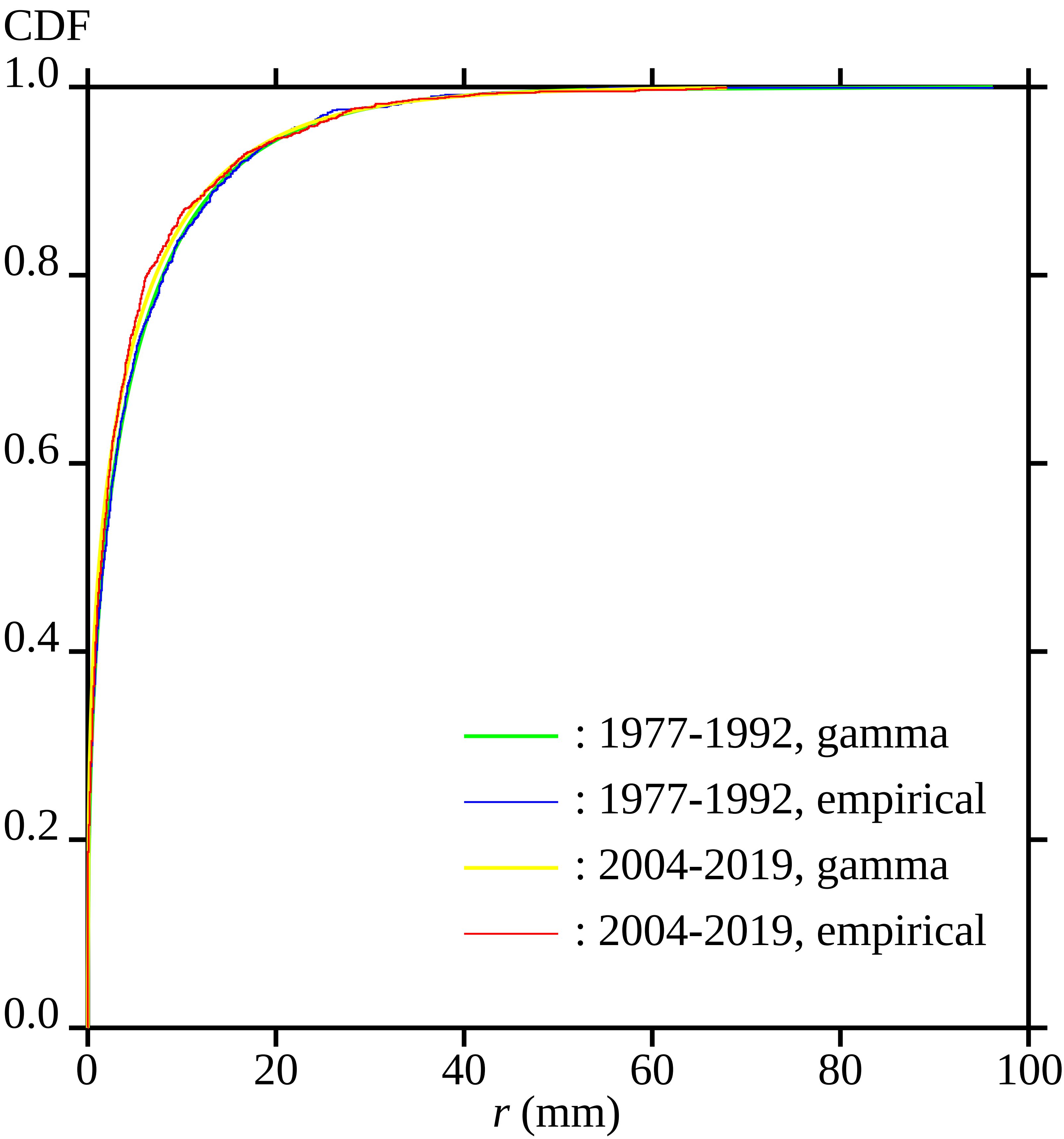
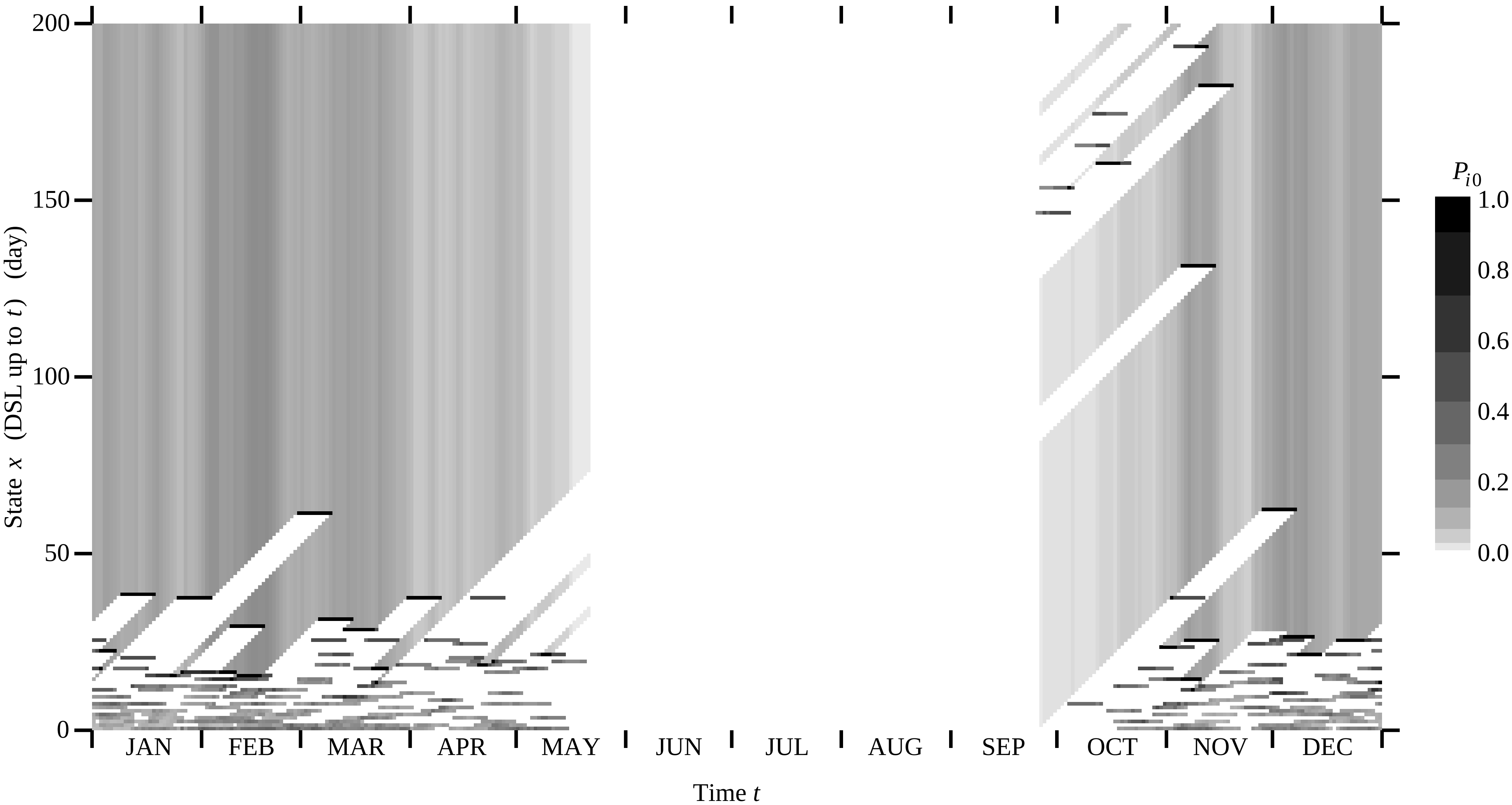


Fig. 6



200
150
100
50
0
State x (DSL up to t) (day)
JAN
FEB
MAR
APR
MAY
JUN
JUL
AUG
SEP
OCT
NOV
DEC
Time t
P_{i0}
1.0
0.8
0.6
0.4
0.2
0.0

Fig. 7

Click here to access/download;colour figure;Fig7-regP5mm7792.eps

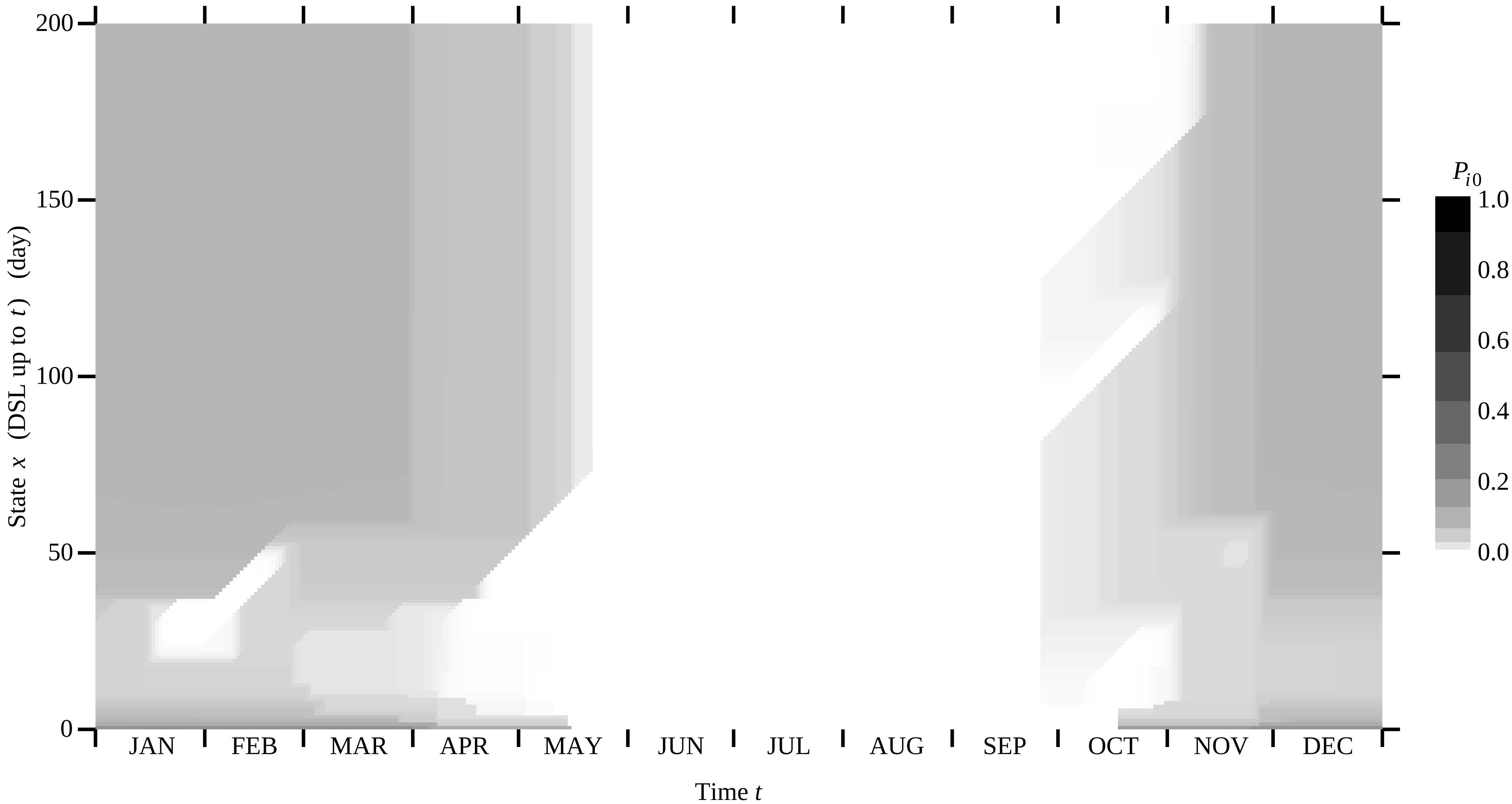
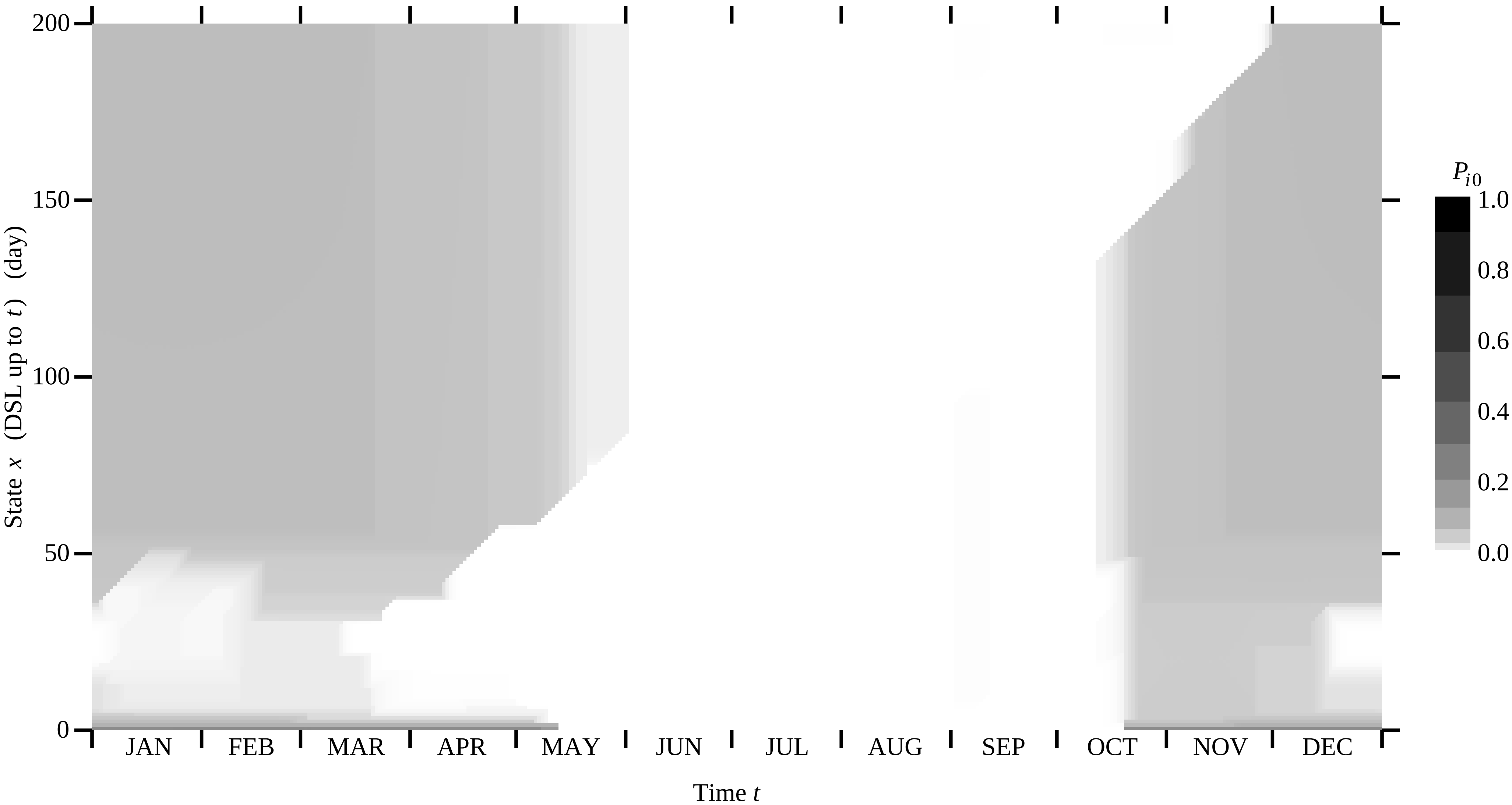
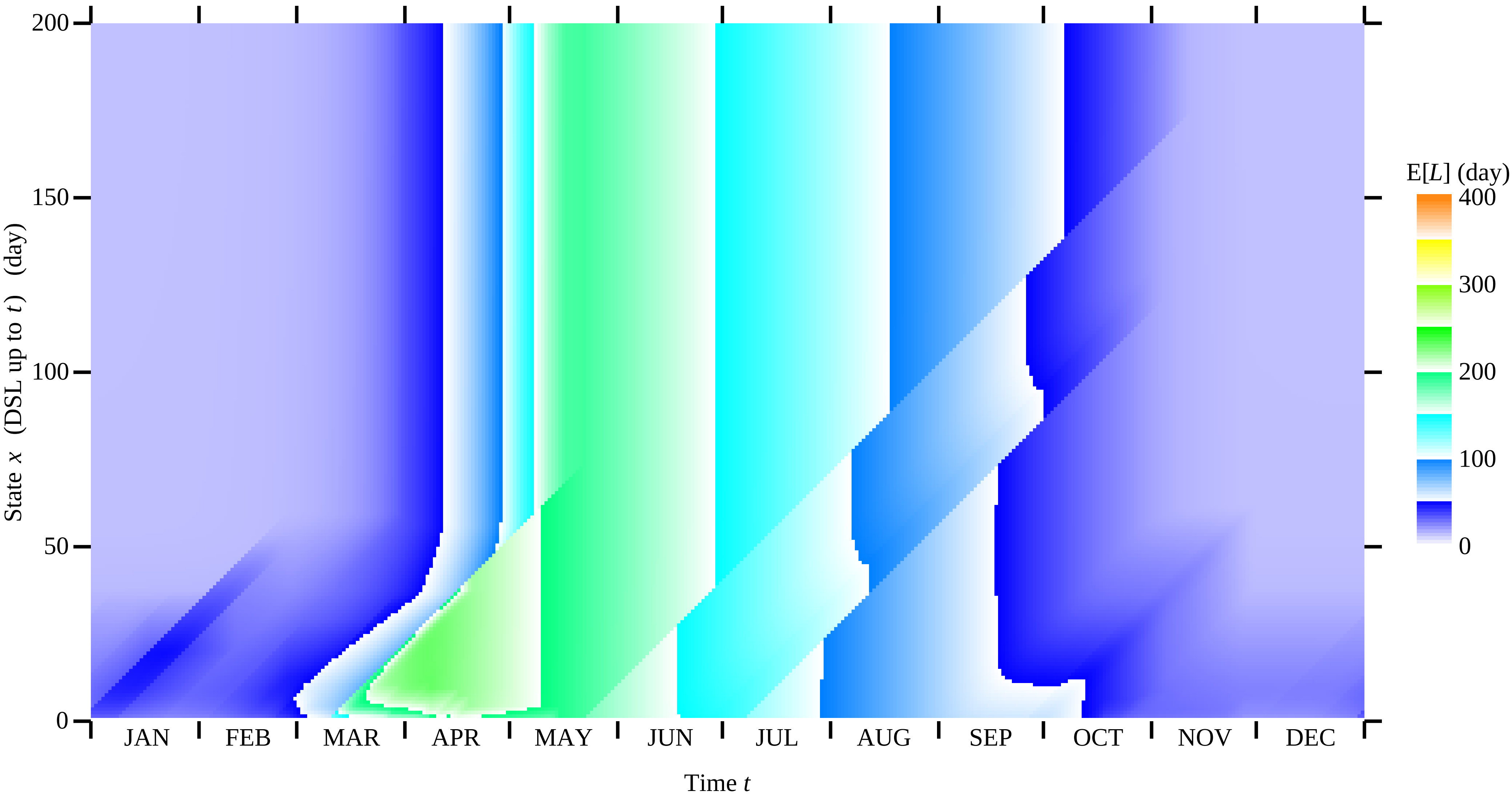


Fig. 8



200
150
100
50
0
State x (DSL up to t) (day)
JAN
FEB
MAR
APR
MAY
JUN
JUL
AUG
SEP
OCT
NOV
DEC
Time t
P_{i0}
1.0
0.8
0.6
0.4
0.2
0.0

Fig. 9



200
150
100
50
0
State x (DSL up to t) (day)
JAN
FEB
MAR
APR
MAY
JUN
JUL
AUG
SEP
OCT
NOV
DEC
Time t
E[L] (day)
400
300
200
100
0

Fig. 10

Click here to access/download;colour figure;Fig10-SD[L]7792.eps

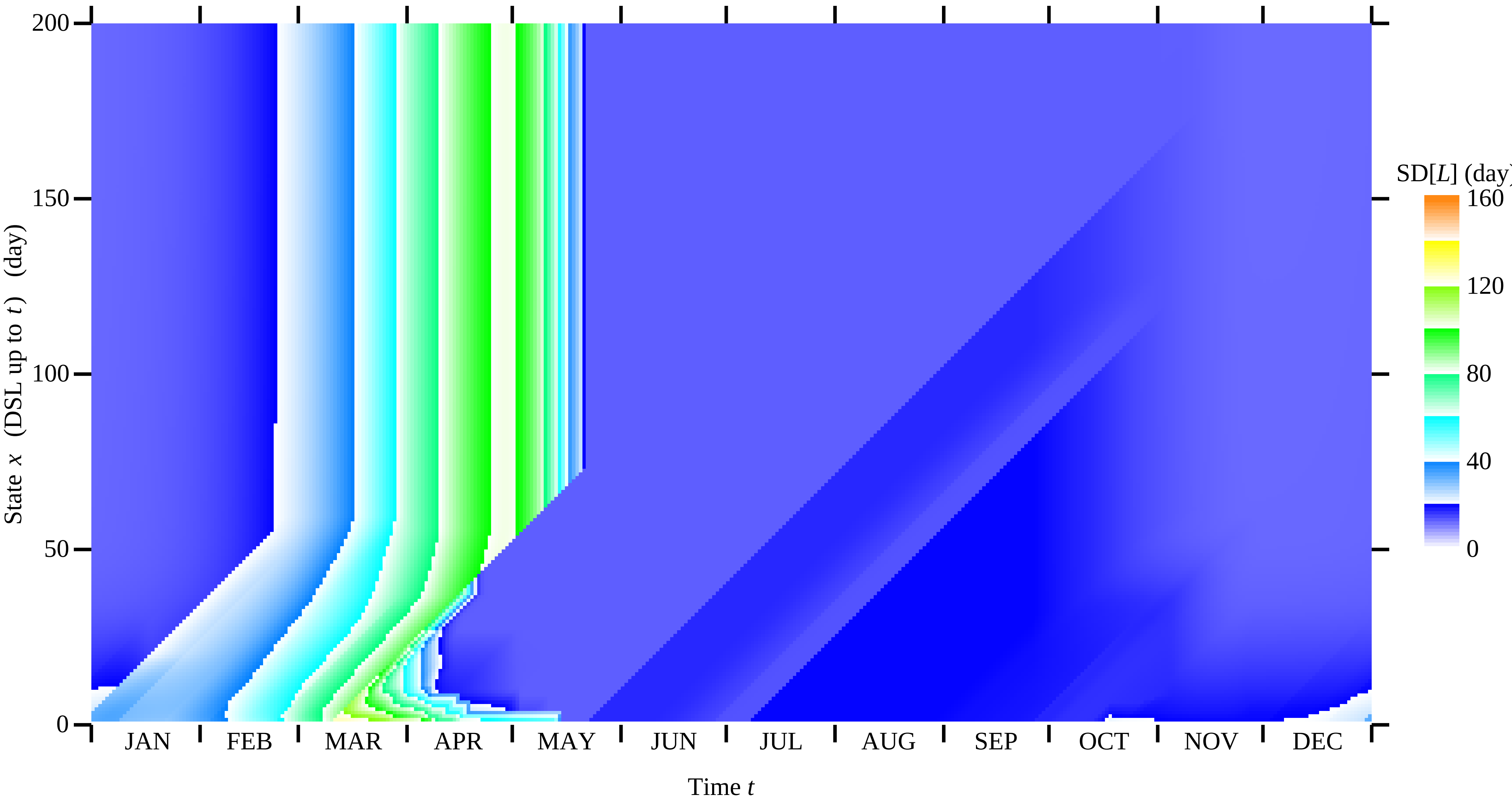


Fig. 11

Click here to access/download;colour figure;Fig11-E[L]0419.eps

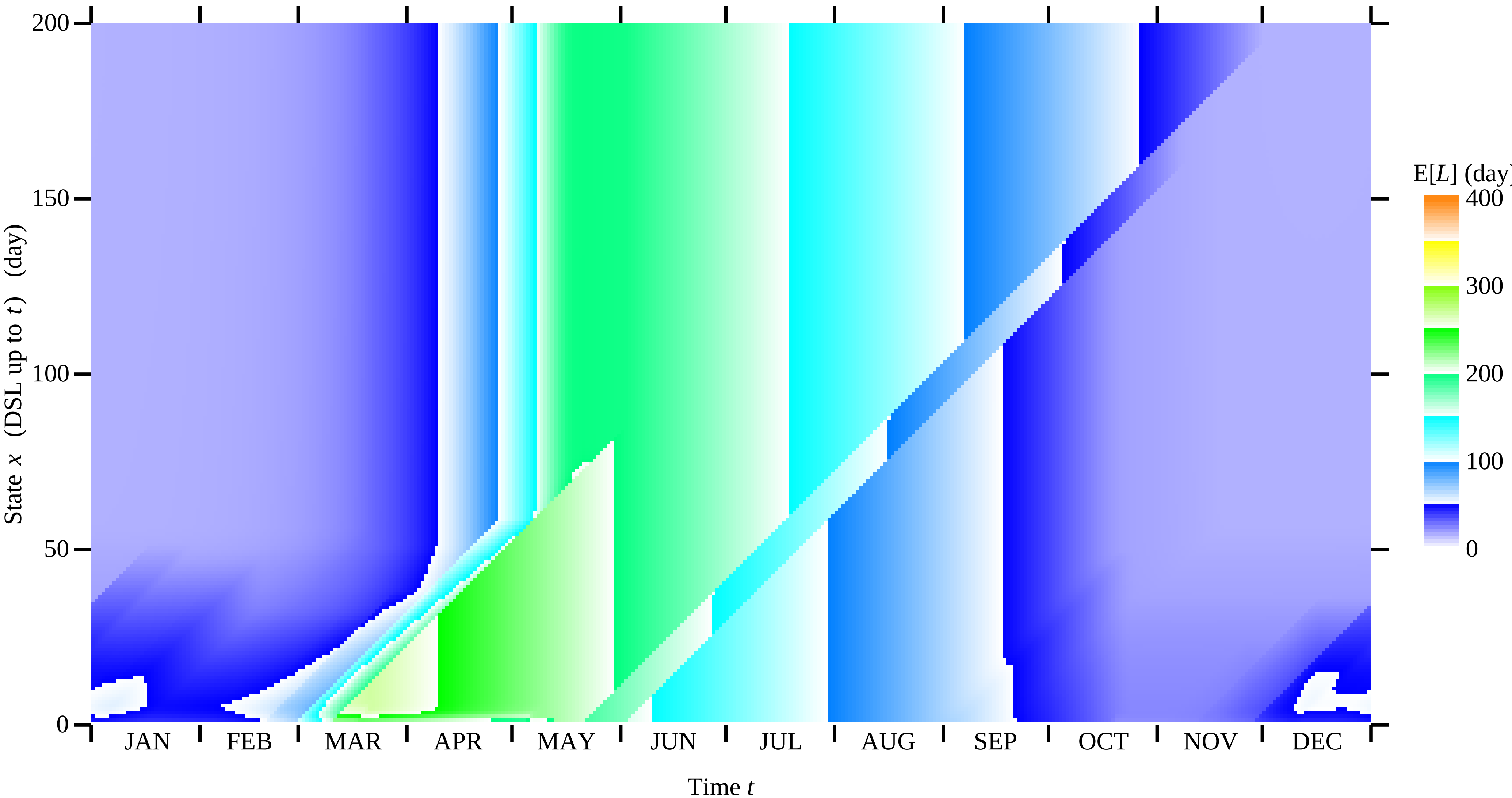


Fig. 12

Click here to access/download;colour figure;Fig12-SD[L]0419.eps

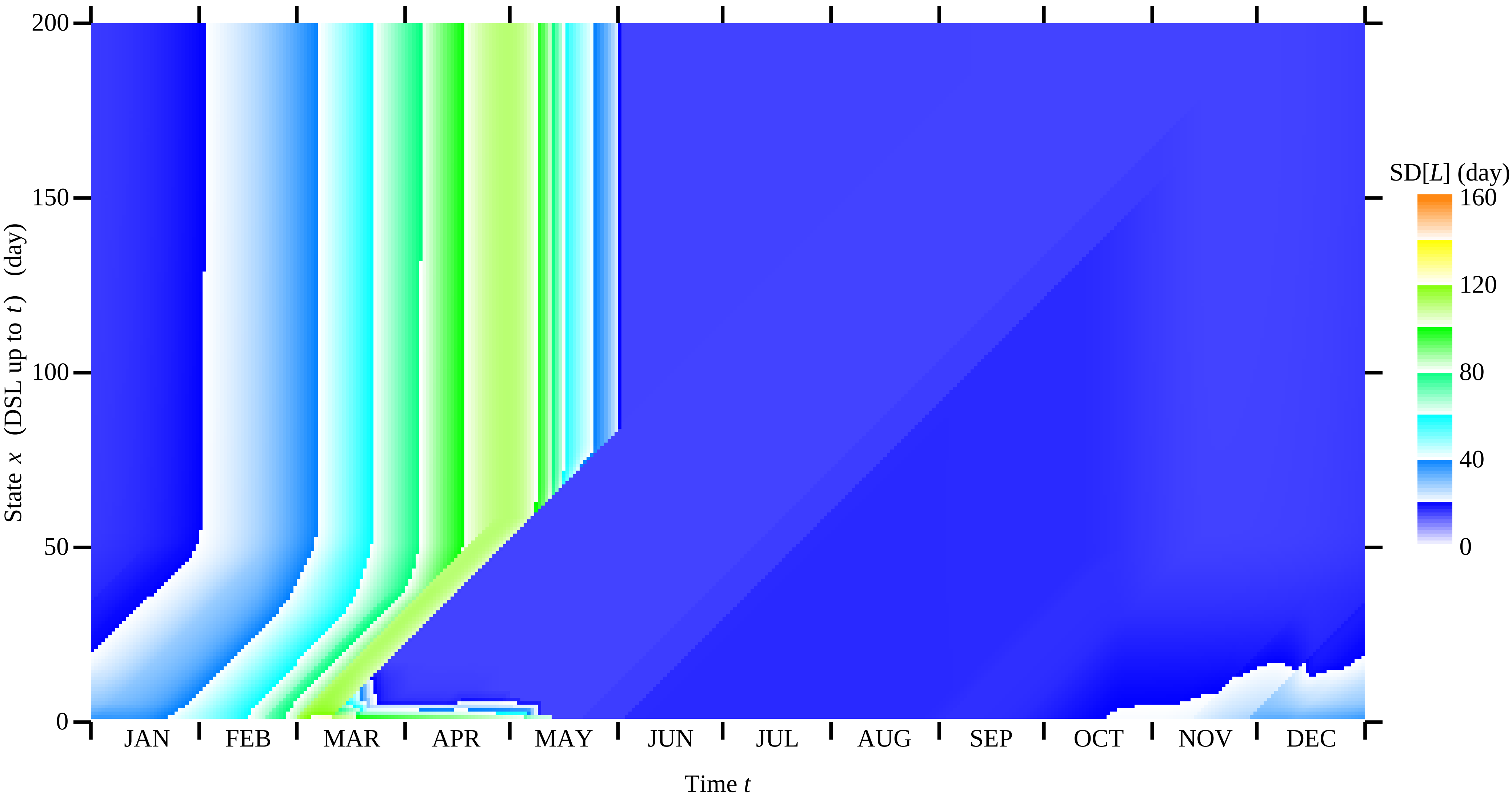


Fig. 13

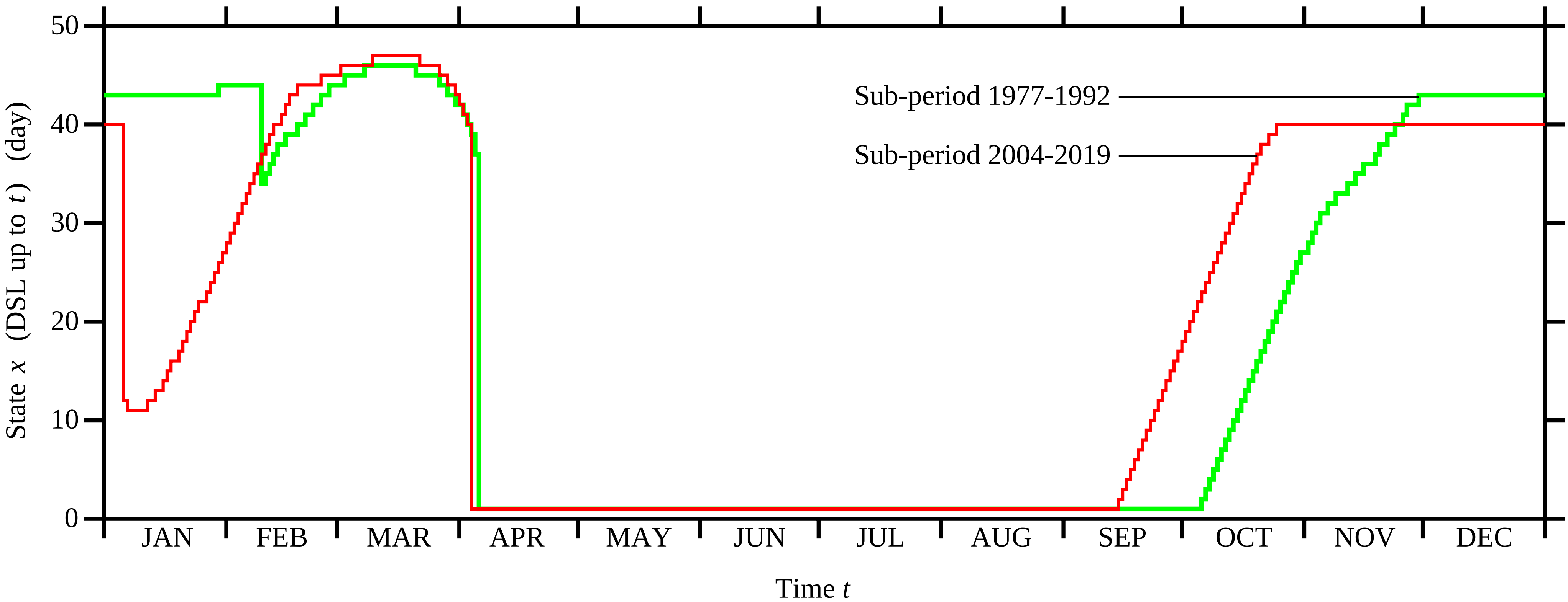
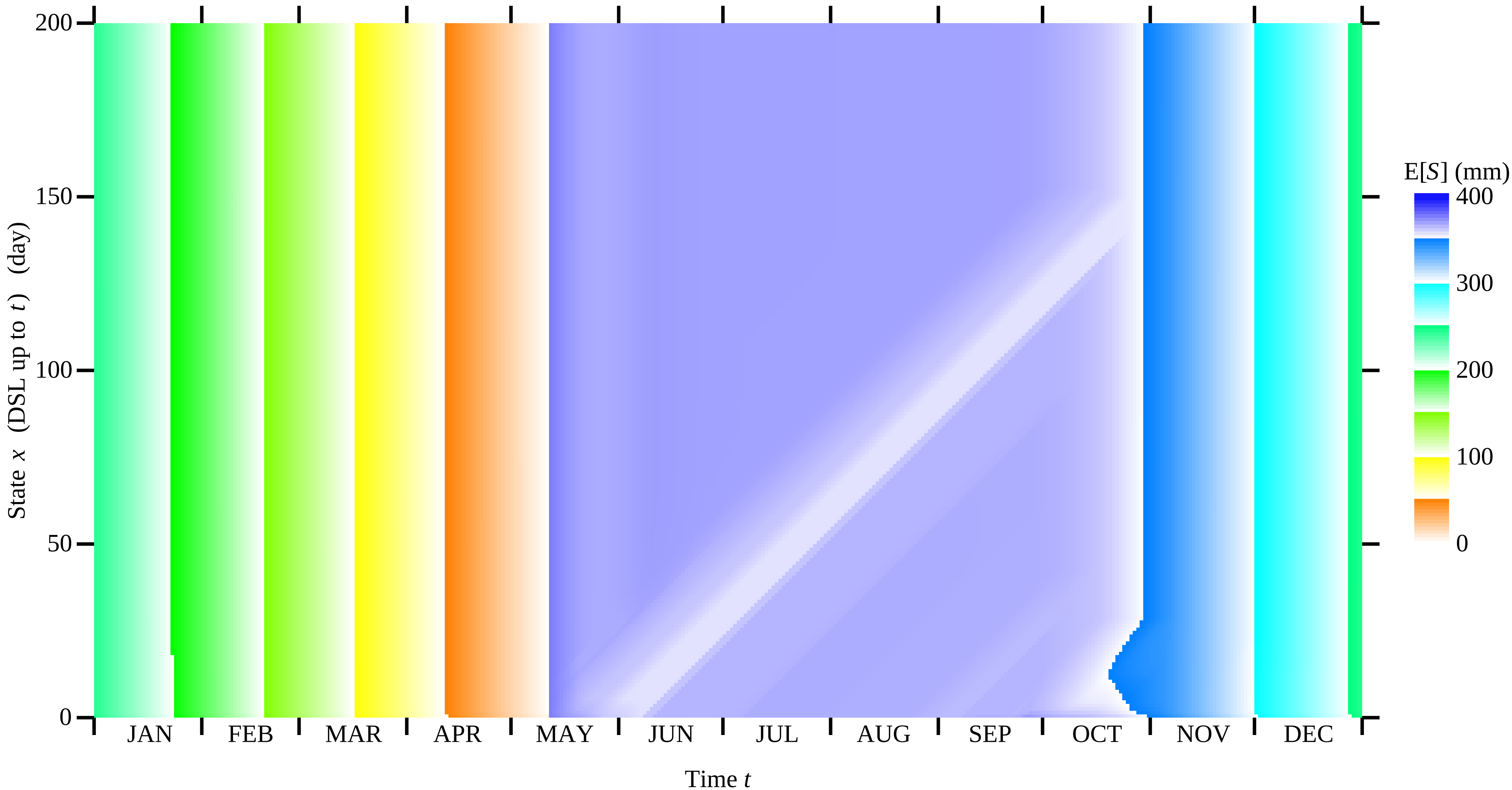

50
40
30
20
10
0
State x (DSL up to t) (day)
Sub-period 1977-1992
Sub-period 2004-2019
JAN
FEB
MAR
APR
MAY
JUN
JUL
AUG
SEP
OCT
NOV
DEC
Time t

Fig. 14



200
150
100
50
0
State x (DSL up to t) (day)
JAN
FEB
MAR
APR
MAY
JUN
JUL
AUG
SEP
OCT
NOV
DEC
Time t
E[S] (mm)
400
300
200
100
0

Fig. 15

Click here to access/download;colour figure;Fig15-SD[S]7792.eps

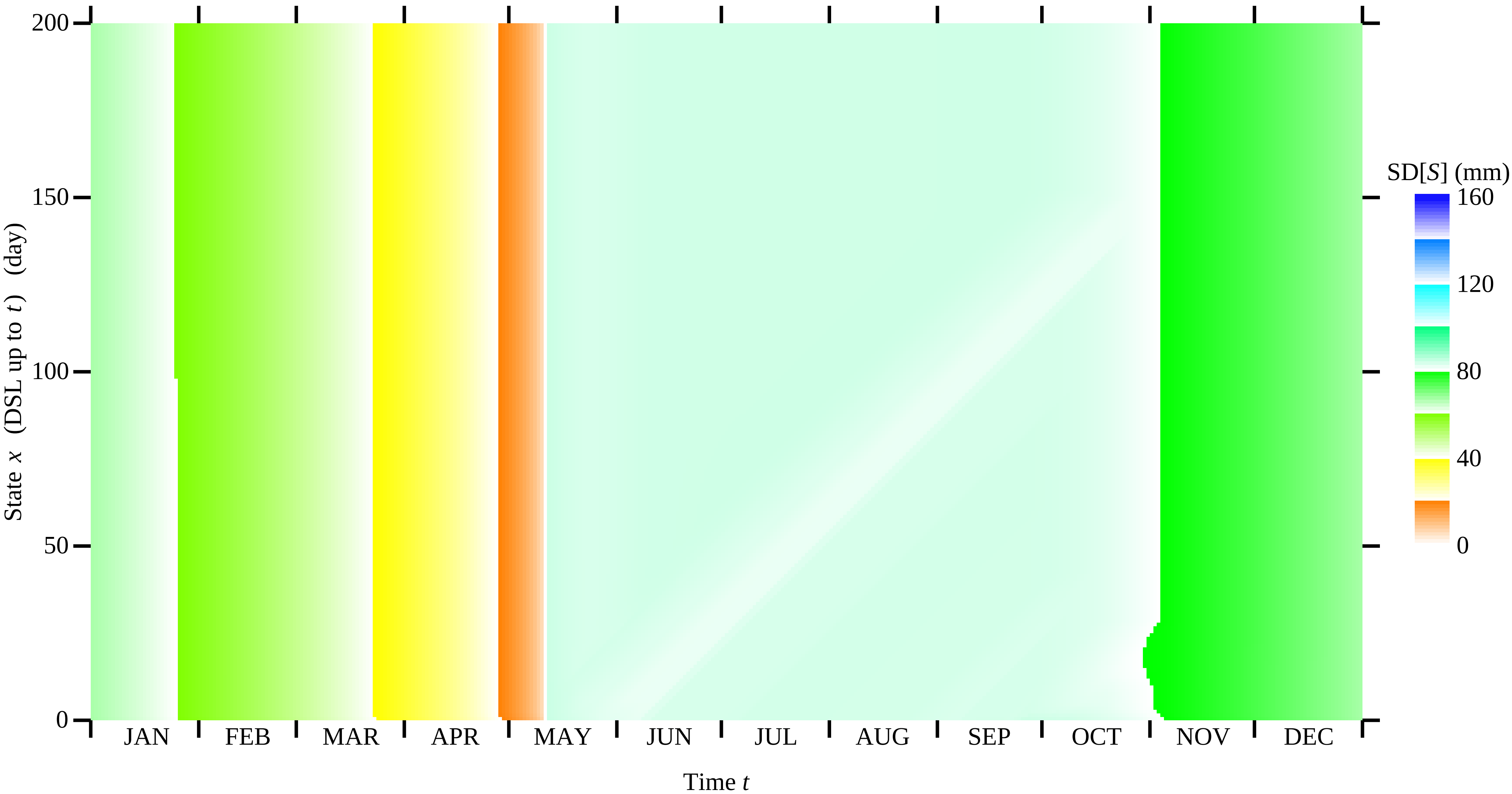


Fig. 16

Click here to access/download;colour figure;Fig16-E[S]0419.eps

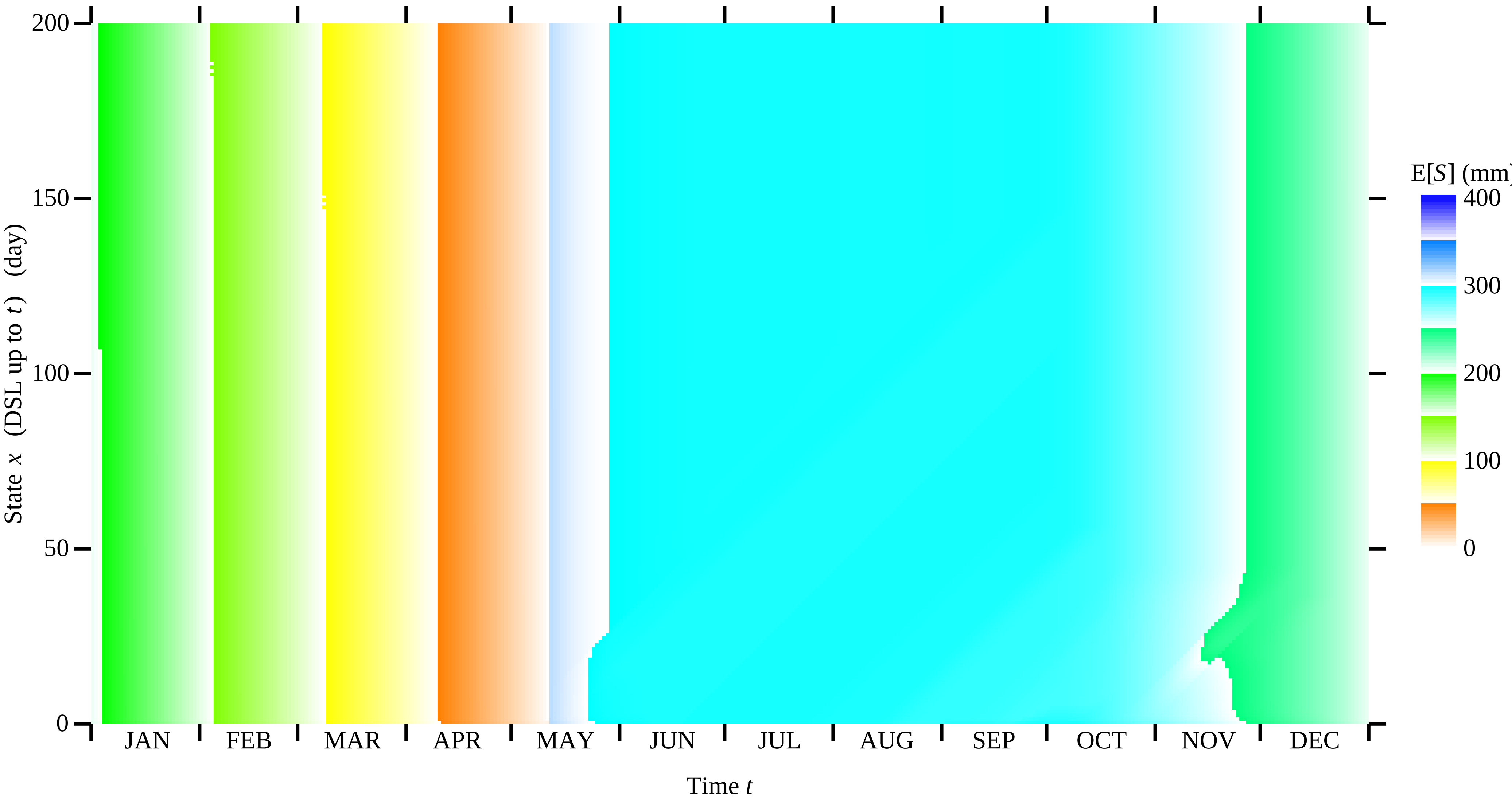
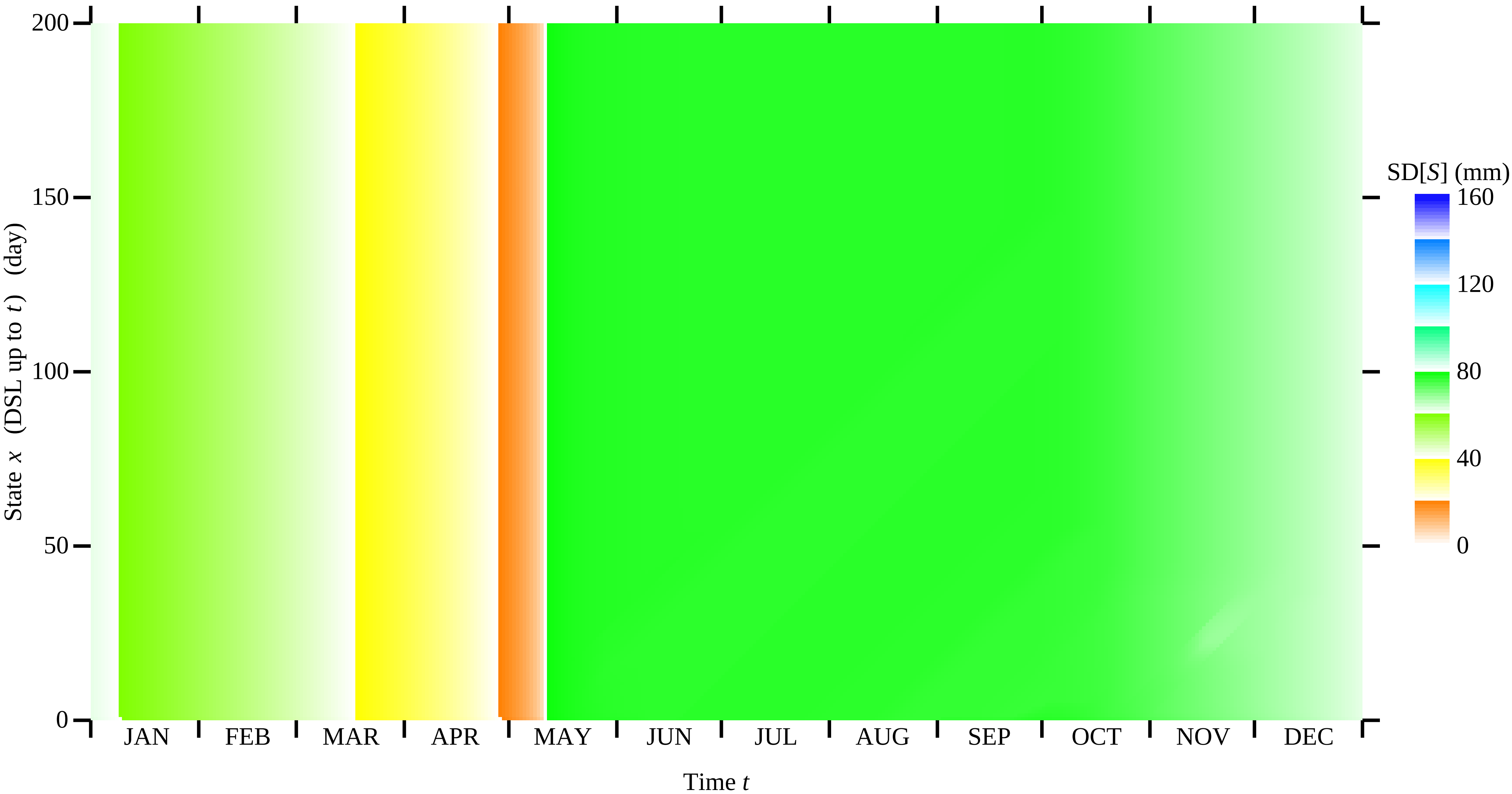


Fig. 17



200
150
100
50
0
State x (DSL up to t) (day)
JAN
FEB
MAR
APR
MAY
JUN
JUL
AUG
SEP
OCT
NOV
DEC
Time t
SD[S] (mm)
160
120
80
40
0

Table 1: Basic statistics of the data sets for the whole period and for the disjoint sub-periods including the selected ones.

Water years	Total number of days in data set	Number of observation days	Number of wet days	Number of dry days	Mean of positive rainfall depths (mm)	Unbiassed sample variance of positive rainfall depths (mm^2)
1974-2019 (The whole period)	16071	15191	2988 (19.7 %)	12203 (80.3 %)	4.809	66.03
1974-1977	943	943	221 (23.4 %)	722 (76.6 %)	4.335	45.41
1977-1992 (The Selected sub-period)	5479	4901	974 (19.9 %)	3927 (80.1 %)	5.018	68.08
1992-2004	4383	4234	827 (19.5 %)	3407 (80.5 %)	4.980	70.73
2004-2019 (The Selected sub-period)	5266	5113	966 (18.9 %)	4147 (81.1 %)	4.560	64.67

Table 2: The depths and numbers of irrigation (Depth (mm) : Number (times)) in a water year for major annual crops in Nineveh Governorate, adapted from Hajim et al. (1996).

Crop	Date of planting	Date of harvesting	AUG	SEP	OCT	NOV	DEC	JAN	FEB	MAR	APR	MAY	JUN	JUL	Total
Wheat	Early NOV	Mid-MAY	-	-	60 : 1	40 : 2	0 : 0	0 : 0	0 : 0	0 : 0	20 : 1	40 : 1	-	-	200
Barley	Early NOV	Early MAY	-	-	60 : 1	40 : 2	0 : 0	0 : 0	0 : 0	0 : 0	0 : 0	0 : 0	-	-	140
Clover	Early OCT	Late MAY	-	40 : 1	40 : 3	20 : 1	0 : 0	0 : 0	0 : 0	0 : 0	20 : 1	40 : 3	-	-	340
Flax	Early NOV	Late MAY	-	-	60 : 1	60 : 1	0 : 0	0 : 0	0 : 0	0 : 0	30 : 1	60 : 1	-	-	210
Sugar beet	Mid-OCT	Mid-JUN	-	60 : 1	40 : 2	20 : 1	0 : 0	0 : 0	0 : 0	0 : 0	40 : 1	60 : 3	60 : 2	-	500
Potato	Mid-AUG	Mid-DEC	30 : 5	30 : 5	30 : 4	30 : 2	0 : 0	-	-	-	-	-	-	-	480
Onion	SEP	MAY	-	20 : 2	20 : 2	20 : 1	0 : 0	0 : 0	0 : 0	0 : 0	20 : 2	20 : 3	-	-	200
Cabbage	SEP	MAY	25 : 1	25 : 6	25 : 3	25 : 1	0 : 0	0 : 0	0 : 0	0 : 0	0 : 0	0 : 0	-	-	275

Table 3: Statistical analysis of each dry season during the sub-period 1977-1992, assuming the Gumbel distribution for the length *L*.

Year	Onset	E[*L*]	SD[*L*]	β_{Gum}	μ_{Gum} = Mode	Observed *L*	CDF	Return period	Significance level
1978	March 15th	150.027	123.183	96.045	94.588	262	0.839	6.229	0.481
1979	March 25th	163.262	114.272	89.098	111.834	217	0.736	3.781	0.652
1980	April 29th	193.766	55.999	43.663	168.563	193	0.565	2.297	0.907
1981	April 29th	193.766	55.999	43.663	168.563	165	0.338	1.510	0.773
1982	May 7th	186.542	53.645	41.827	162.399	146	0.228	1.295	0.590
1983	May 15th	191.327	12.614	9.835	185.650	182	0.235	1.307	0.602
1984	May 10th	183.770	52.874	41.226	159.974	160	0.368	1.583	0.819
1985	April 25th	196.654	58.331	45.480	170.402	200	0.594	2.460	0.873
1986	May 1st	191.918	55.481	43.258	166.949	153	0.251	1.336	0.630
1987	March 28th	171.751	108.753	84.794	122.806	205	0.684	3.168	0.737
1989	July 2nd	139.314	13.514	10.537	133.232	131	0.291	1.410	0.695
1990	April 12th	199.231	73.914	57.630	165.966	211	0.633	2.723	0.818
1991	April 11th	201.770	74.269	57.907	168.345	207	0.599	2.492	0.866
1992	May 11th	182.847	52.616	41.025	159.167	179	0.540	2.173	0.933

* No data was available to determine the dry season in the year 1988 due to the Iran-Iraq War

Table 4: Statistical analysis of each dry season during the sub-period 2004-2019, assuming the Gumbel distribution for the length *L*.

Year	Onset	E[*L*]	SD[*L*]	β_{Gum}	μ_{Gum} = Mode	Observed *L*	CDF	Return period	Significance level
2004	April 20th	202.671	86.211	67.219	163.872	198	0.548	2.211	0.925
2005	May 3rd	192.300	80.997	63.153	155.847	202	0.618	2.617	0.840
2006	April 27th	196.861	83.616	65.195	159.229	181	0.489	1.956	0.956
2007	May 16th	212.723	14.753	11.503	206.083	258	0.989	91.735	0.282
2008	March 14th	225.342	106.851	83.311	177.253	224	0.565	2.300	0.907
2009	April 18th	204.357	86.930	67.779	165.233	194	0.520	2.083	0.950
2010	May 4th	198.859	72.803	56.764	166.094	221	0.684	3.162	0.738
2011	April 23rd	200.190	85.092	66.346	161.894	207	0.602	2.516	0.861
2012	March 29th	221.844	93.526	72.922	179.752	207	0.502	2.010	0.962
2013	May 28th	178.835	15.172	11.830	172.007	164	0.140	1.163	0.450
2014	April 18th	204.357	86.930	67.779	165.233	181	0.453	1.827	0.926
2015	May 11th	198.239	63.570	49.565	169.630	118	0.059	1.062	0.338
2017	April 15th	206.877	88.014	68.624	167.266	207	0.571	2.331	0.900
2018	May 13th	215.723	14.752	11.502	209.084	161	0.000	1.000	0.270

* No data was available to determine the dry season in the year 2016 due to the Iraqi Civil War