The global monsoon across time scales: Mechanisms and outstanding issues


1. Introduction

In recent decades there has been a dramatic increase in the number of publications devoted to monsoon variability, both from the modern-era and paleo-climatological perspectives. The satellite record has extended monsoon observations beyond continental records and into data-scarce oceanic regions. The appearance of speleothem and ice-core records, together with deep-sea and terrestrial sediments, has enhanced the resolution of proxy records to an unprecedented level. The rapid growth of research interest has also led to a serious divergence in opinions on some issues. Is there coherent variability of the regional monsoons? Do the monsoons respond directly to orbital forcing, or indirectly through ice-sheet changes? Was the rise of the Tibetan Plateau responsible for the establishment of the modern Asian monsoon system? These questions now motivate a joint effort of the modern and paleo-monsoon communities to review progress made, identify instances of incomplete or inconsistent evidence, and address outstanding issues.

In order to better understand the dynamics of monsoon variability, the PAGES Working Group “Global Monsoon and Low-Latitude Processes: Evolution and Variability” was established in 2007. Two successive Global Monsoon (GM) symposia (2008 and 2010), and two writing workshops (2012, 2014) were held in Shanghai, where paleo-and modern climatologists, data producers and modelers met together to help dissect the mechanisms and controlling factors of monsoon variability at various temporal-spatial scales.
to compare regional monsoon systems across a broad range of timescales, and to unravel the mechanisms causing variations in the global monsoon system and regional deviations from the global mean. As products of this effort, a GM special issue was published in “Climate Dynamics” (Wang et al., 2012a), and the first synthesis papers published in “Climate of the Past”, under a title “The global monsoon across timescales: coherent variability of regional monsoons” (Wang et al., 2014a). The Working Group’s efforts have proven to be rewarding. As the concept of a “global monsoon” is relatively new, its application in paleoclimatology was a contentious issue as recently as a few years ago, but in recent years a flurry of publications approaching the monsoon as a global system has emerged.

During this time, a remarkable development has been the growing body of high-resolution paleo-monsoon reconstructions extending to a deeper geological past. New speleothem records have been published from various continents (Fig. 1), high-resolution terrestrial and marine paleo-monsoon sequences have been generated from the Southern Hemisphere, where paleo-monsoon studies were previously sparse, and numerous coring or drilling expeditions to the monsoonal oceans have been implemented. In recent years (2013–2015), six deep-ocean drilling expeditions were completed by IODP (International Ocean Discovery Program) to explore the Cenozoic history of the Indian, East Asian and Australian monsoons (see Fig. 1, Table 1). All of these efforts have provided an unprecedented opportunity to gain new sets of high-quality data on a global scale and to reveal a profound insight into the mechanisms of global monsoon variability.

Presented here is the second synthesis paper of the PAGES GM Working Group. In the first paper, we provided observation and proxy data to show that the regional monsoons can vary coherently in various contexts, in support of the GM concept across timescales. As a follow-up, this second paper gives an in-depth discussion of driving mechanisms and outstanding issues in GM studies. We start with an introduction of the GM as a planetary scale circulation system and a brief accounting of why it exhibits regional structure. We then analyze the monsoon formation mechanisms, together with major processes that drive monsoon variations, including external forcings and internal feedbacks. Afterwards, we move to the geological records, reviewing the factors affecting solar forcing of GM variability, and the underlying features that generate regional monsoon structure. Lastly, special attention is paid to persisting scientific debates, with an attempt to reveal the key issues and to provide suggestions for future research.

2. Monsoon climate and its formation

2.1. Modern regional monsoons and global monsoon systems

Despite hundreds of years of research history and thousands of resultant publications, fundamental monsoon characteristics such as their basic definition and geographic range remain debated. It is of primary importance thus to clarify the definition of a monsoon and to briefly introduce the existing regional monsoons before moving to a synthesis discussion of the GM.

2.1.1. Regional monsoon systems

2.1.1.1. Definition of monsoon climate. The modern definition of monsoon climate is based on both the annual reversal of surface winds and the contrast between rainy summer and dry winter seasons (Webster, 1987. Part 1), in contrast to the classical delineation of monsoon regimes since the early 20th century, which was based solely on the annual reversal of surface winds (Hann, 1908; Khromov, 1957; Ramage, 1971; Li and Zeng, 2000). The classical definition confines monsoons to the eastern hemisphere and includes the Asian, Australian, tropical African, and Indian Ocean monsoon systems. Defined in this manner, the identified regions also include some mid-latitude areas.

Table 1
Recent IODP expeditions exploring paleo-monsoon.

<table>
<thead>
<tr>
<th>IODP</th>
<th>Sea area and primary scientific theme</th>
<th>Dates</th>
<th>IODP sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>346</td>
<td>Sea of Japan and East China Sea, Asian monsoon</td>
<td>Jul-Sept 2013</td>
<td>U1422-U1430</td>
</tr>
<tr>
<td>353</td>
<td>Bay of Bengal, Indian monsoon precipitation</td>
<td>Nov 2014-Jan.2015</td>
<td>U1443-U1448</td>
</tr>
<tr>
<td>354</td>
<td>Bengal Fan, Himalaya and climate</td>
<td>Jan-March 2015</td>
<td>U1449-U1455</td>
</tr>
<tr>
<td>355</td>
<td>Arabian Sea, South Asian monsoon and tectonics</td>
<td>March-May 2015</td>
<td>U1456-U1457</td>
</tr>
<tr>
<td>356</td>
<td>NW shelf of Australia, Indonesian throughflow and Australian monsoon</td>
<td>Jul-Sept 2015</td>
<td>U1458-U1464</td>
</tr>
<tr>
<td>359</td>
<td>Maldives, Indian Ocean sea-level and monsoon</td>
<td>Sept-Nov 2015</td>
<td>U1465-U1472</td>
</tr>
</tbody>
</table>
which exhibit winter precipitation maxima and belong to Mediterranean climate regimes. The modern definition is motivated in part by the considerable socio-economic and scientific importance of monsoon rainfall, and thus delineating monsoon domains based on precipitation is imperative and advantageous. When determined by rainfall characteristics, monsoon regions extend across both the eastern and western hemispheres and include the North and South American, and southern African monsoons (Wang, 1994).

Fig. 2 shows the monsoon domains defined by two simple criteria in rainfall characteristics: (a) a summer-minus-winter precipitation differential exceeding 300 mm and (b) an proportional summer precipitation threshold exceeding 55% of the annual total (Wang and Ding, 2008). Here the local summer denotes the extended season, including May through September for the northern hemisphere (NH) and November through March for the southern hemisphere (SH). The defined monsoon domains include all eight regional monsoons: Indian (IN), western North Pacific (WNP), East Asian (EA), Australian (AUS), North American (NAM), South American (SAM), North African (NAF), and South African (SAF) (Fig. 2) (Yim et al., 2014a). Note that the monsoons entail substantial oceanic regions, including the marginal seas in South and East Asia, and the tropical western North Pacific (WNP), southwest Indian Ocean, and tropical eastern North Pacific Ocean. The oceanic monsoons are integral parts of the regional monsoon systems, but overlooked in many paleo-monsoon studies as discussed later.

2.1.1.2. Tropical and sub-tropical monsoons. A majority of the above listed regional monsoons belong to tropical monsoons in nature except for the EA monsoon and southern part of the South American monsoon. The tropical monsoons originate primarily from the meridional thermal contrast between land and ocean or between hemispheres. The tropical monsoons are characterized generally by an annual reversal of both zonal and cross equatorial winds; and thus can be identified by regions where the local summer minus winter westerly zonal wind speed at 850 hPa exceeds 50% of the annual mean (Wang and Ding, 2008). Tropical monsoon wind domains identified in this manner are outlined by the black lines in Fig. 2 and are in an excellent agreement with those defined by Ramage (1971) except that the North and South American monsoons were missed in Ramage (1971). The monsoon westerly wind and precipitation domains generally do not coincide but they are dynamically consistent: The westerly monsoon domains are generally situated on the equatorward and westward sides of monsoonals clouds and rainfall (Fig. 2), arising from the westward propagation of Rossby waves generated by these precipitation heat sources (Gill, 1980).

The strongest regional monsoon, the Asian monsoon, consists of three sub-monsoon systems, namely the South Asian (or Indian), western North Pacific (WNP), and East Asian (EA) monsoons, which are separated roughly by 105°E and 20°N (Fig. 3). The East Asian monsoon (EAM) is located to the east of the Tibetan Plateau while the Indian monsoon is located to the south of the Tibetan Plateau. Both modeling experiments and the geological record suggest that the East Asian monsoon is more sensitive than the Indian monsoon to the uplift of Tibet (Wang et al., 2003a; Yanai and Wu, 2006; Chen et al., 2014). Without the Tibetan Plateau, the Indian summer monsoon (ISM) would be much weaker than the present day, whereas East Asia would be dominated by subtropical dry zones and its monsoon would disappear (Kitoh, 2004). The summer/winter precipitation ratio is much higher in the Indian sector than in the East Asia region (Clemens et al., 2010;...
While the three sub-systems are dynamically interconnected, they respond differently to external forcing and internal feedbacks (Wang et al., 2001a), and ENSO-related variability is much stronger in the East Asian monsoon than the Indian monsoon region (Wang et al., 2003a).

Of particular interest is the subtropical EAM, which uniquely extends poleward of ~50th latitude and thus deserves special consideration. The EAM is a subtropical-extratropical monsoon and is characterized by an annual reversal of the meridional winds (Guo, 1983), because it is largely controlled by the zonal thermal contrast between the Asian continent (including the Tibetan Plateau) and the Pacific Ocean, which together generate a pressure gradient between the Asian Low and the western Pacific subtropical high during summer. The summer southerly and winter northerly flows are rotational winds that are driven by this pressure gradient and Earth’s rotation (Fig. 3). Summer rainfall over EA is supplied by the northward moisture transport on the western and poleward side of the western Pacific subtropical high originated from Indian and Pacific Oceans. The EA monsoon wind domain can be defined based on monsoon southerly flow in an analogous manner to defining the tropical monsoons (Wang and Ding, 2008). The existence of the EA subtropical monsoon depends critically on the presence of a subtropical high to the east of a continental low. The western part of the ridge lines of the subtropical high divide the monsoon region into the tropical monsoon to its equatorward side and the subtropical monsoon to its poleward side. The southeast part of the South American monsoon is also a subtropical monsoon as it is located on the westward and poleward side of the South Atlantic subtropical high. Since the subtropical highs in the western Pacific and South Atlantic are coupled with the corresponding tropical monsoon trough in the WNP and South America, respectively, the subtropical monsoons are also linked to the ITCZ via the associated meridional flow.

Note that during northern winter the massive Eurasian continental cooling produces the world strongest continental cold air mass (the Siberian High) and the globe’s most powerful winter monsoon over East Asia and the WNP, which can affect a quarter of the northern hemisphere and extend from high latitudes to the equator, where it influences the Inter-Tropical Convergence Zone (ITCZ), a region of seasonally persistent deep convection and convergence in the deep tropics. In contrast to the northerly and northwesterly winter monsoon winds in the subtropical-midlatitude EA, the winter monsoons in the tropical regions such as IN, WNP, NAF and NA are characterized by near-surface northeasterlies. The intensity of the winter monsoon impacts both terrestrial and marine records in East Asia, such as in loess sequences in North China, and in deep-sea sediments in the South China Sea (see Wang et al., 2005, for a review). Remarkably, the wind-driven primary productivity in the South China Sea reaches its maximum in winter, rather than summer (Zhao et al., 2009). All these features make the East Asian monsoon unique amongst the regional monsoons.

2.1.1.3. Causes of monsoon regionality. Each regional monsoon has its unique features due to its specific land-ocean configuration and topography, in addition to different atmosphere-ocean-land interactions. The third pole-Tibetan Plateau acts as a heat source during summer, which greatly enhances the north-south temperature gradients and, consequently, the ISM. The existence of the east African highlands strengthens the cross-equatorial flow and Somalia Jet, inducing upwelling in the west Arabian Sea and a large associated zonal sea surface temperature (SST) gradient that accelerates southwest Indian monsoon winds. As a result of this and other influences, the ISM is the strongest on the Earth. The North African summer monsoon is a more typical monsoon region because it is driven by the strong land-ocean thermal contrast between the cold tongue of the equatorial Atlantic and extremely hot Sahara Desert. In the western hemisphere, the North American low-latitude land area is relatively small compared to that of the neighboring ocean and strong trade winds prevail over the eastern Pacific and Atlantic in both hemispheres. In addition, the oceanic ITCZ remains in the NH year-round due to the existence of the vast equatorial Pacific and Atlantic cold tongues of SST. Therefore, the annual variation of the surface winds is not strong enough to achieve zonal reversal. As such the American monsoons are characterized primarily by their strongly seasonal rainfall and are generally much weaker than the eastern hemisphere’s monsoon regions. Regardless of all of their individual distinctions, however, these regional monsoons can often be viewed as comprising parts of a global system, the GM.

2.1.2. Global monsoon system

Physically, the GM may be viewed as a forced response of the coupled climate system to the annual cycle of solar insolation. All regional monsoons are bounded and synchronized by the solar forcing variation and therefore form a unified planetary-scale circulation system regardless of regional distinctions (Trenberth et al., 2000). This unified GM system represents the leading empirical orthogonal function (EOF) mode of the annual variations of precipitation and circulation in the global tropics and subtropics (Wang and Ding, 2008).

2.1.2.1. Global monsoon circulation. The leading mode of the global monsoon system is characterized by a gross equatorially antisymmetric (or hemispheric anti-phase) pattern of precipitation and circulation. Fig. 4 shows the total low-level wind fields toward the end of the two solstice seasons along with seasonal migration of monsoon precipitation. The total wind field can be decomposed into its divergent and rotational components. Due to Earth’s rotational effect, the global monsoon circulation is dominated by the rotational wind component even in the deep tropics. However, the divergent circulation is directly linked to precipitation and zonal mean meridional circulation (Hadley circulation). The black lines outline the regions where zonal winds reverse direction between February and August, which mainly occurs in tropical Asia, Africa and Australia, and the Indo-Pacific Warm Pool regions. The western hemisphere oceans are dominated by subtropical highs and associated trade winds in both the North and South Pacific, and North and South Atlantic.

A global view of the upper-level (~12 km high from the surface) divergent circulation associated with GM system is shown in Fig. 5. The low-level divergent winds, at ~1.5 km, show roughly an opposite circulation pattern (figure not shown). The strengths of the divergent winds are generally one third of the rotational winds in the deep tropics and an order of magnitude smaller than the rotational winds elsewhere. But the divergent winds are extremely important as they directly link to vertical motions that produce clouds and precipitation. The divergent circulation is largely driven by local summer monsoon precipitation heating. As shown in Fig. 5, the rainy seasons in the summer hemisphere coincide with the upper-level divergence (and low-level convergence) and the dry monsoon seasons and deserts in the winter hemisphere correspond to upper-level convergence and sinking motion. The corresponding upper-level meridional winds blow across the equator from summer to winter hemispheres and the low-level meridional winds blow in an opposite direction to approximately balance the mass budget of each hemisphere.

In this context, the GM can be defined as a planetary scale circulation system with a seasonal reversal of the three-dimensional monsoon circulation that is accompanied by meridional migration of monsoon rainfall. All regional monsoons are driven and synchronized by the annual cycle of incoming solar radiation and they are interconnected by the global divergent circulation (Trenberth et al., 2000). Due to their existence in a seasonally variable, deep convective environment they are also governed by common forcing and shared feedback processes. Therefore, the regional monsoons should not be studied purely in isolation. The combined view of the GM sheds light on the monsoon dynamics from a global perspective and helps to better understand some major features of the monsoon responses. In a geographical aspect, the GM
concept is useful also to infer past variability in data sparse regions based on spatial coherence.

2.1.2.2. Global monsoon and ITCZ. The intertropical convergence zone (ITCZ) shown in Fig. 4 is defined by the locations where the monthly mean precipitation is a latitudinal maximum. The ITCZ may be classified into two types: Monsoon trough and trade wind convergence zone. The portion of ITCZ embedded in the monsoon regions experiences a large seasonal meridional migration and the low-level winds on the ITCZ’s equatorward side reverse their direction annually (Fig. 4). This portion of the ITCZ is called a monsoon trough (Ramage, 1971). Conversely, in the central Pacific and Atlantic Oceans, where the ITCZ’s annual displacement is less than a few degrees of latitude and the zonal winds (trade winds) on both sides of the ITCZ do not change directions between summer and winter, a monsoon is absent. These portions of the ITCZ can be referred to as ‘trade wind convergence zones’ (Wang, 1994).

In the trade wind convergence zones, the monsoon is absent and the ITCZ stays in the northern hemisphere throughout the year due to atmosphere-ocean interactions under this unique land-ocean configuration. For instance, in the eastern Pacific the northwest-southeast tilted coast line anchors the cold tongue in the equatorial Pacific, preventing

Fig. 4. Climatological low-level (850 hPa) winds (arrows) and precipitation rate (shading, mm/day) during (a) August and (b) February. The green lines outline the monsoon domains defined by precipitation characteristics and the black lines outline the tropical monsoon domains defined by the seasonal reversal of zonal wind. The red lines indicate ITCZ positions (solid for monsoon trough and dashed for trade wind convergence).

Fig. 5. Global monsoon divergent circulation. Climatological mean (1979–2012) (a) JJA and (b) DJF precipitation (shading in units of mm/day), and upper-tropospheric velocity potential (contours in units of 10^6 m^2/s) and divergent component of vector winds (unit m/s) at 200 hPa (~12 km). Local summer (winter) monsoon regions are outlined by red (blue) curves. The dry regions, where the local summer precipitation is <1 mm/day are outlined by yellow curves. The upper-level divergence (convergence) centers are indicated by green (brown) dashed circles. The rainfall and wind data are derived from GPCP and NCEP-DOE, respectively.
the ITCZ from moving into the southern Hemisphere (Philander et al., 1996). Energetic constraints may also play a role in determining the mean ITCZ location (e.g., Frierson et al., 2013).

Since the tropical monsoon trough (from Africa to Australia and over America) occupies more than three quarters of the ITCZ (Fig. 2), the annual migration of the ITCZ is largely driven by the GM system.

Notably, the two kinds of ITCZ differ in their response to external forcing as the tropical climate changes. For example, the trade winds respond to precession cycles and exhibit a much more linear variation at the orbital time scale (e.g., Beaufort et al., 1997, 2001; Tachikawa et al., 2011) than do the regional monsoons given the complicity of interactions between the land and ocean, as discussed later in Chapter 4.

2.1.2.3. Global monsoon and Hadley circulation. The GM system plays a key role in driving the Hadley circulation, i.e., the zonal mean meridional circulation, which has considerable longitudinal variation (Fig. 5). The zonal mean circulation provides a rather simplistic view of the three dimensional divergent circulation. As shown by Fig. 5, during northern summer, the latent heat released in the heavy summer monsoon rains lift air upward and forms a local maximum in upper-level divergence. The strongest divergence centers are seen over the South Asian-Western North Pacific monsoon regions with other weaker centers being located in the North American and Northern African summer monsoon regions. The NH summer monsoons drive southward cross-equatorial flows aloft that sink in the SH winter monsoon regions in addition to the South Pacific and South Atlantic subtropical highs (Fig. 5a). During NH winter (DJF), the upper-level divergence centers are coupled to the SH summer monsoon regions and the upper-level divergent winds flow from SH to NH in both the winter monsoon regions and the North Pacific and North Atlantic subtropical High regions (Fig. 5b). These meridional branches of divergent monsoon flows are called “lateral” monsoons by Webster et al. (1998) and form the backbone of the Hadley circulation.

2.1.2.4. Global monsoon, hydrological cycle and desert. The annual variation of precipitation and associated hydrological cycle is driven to a large degree by the GM system. During JJA, about 70% of the tropical-subtropical precipitation falls in the NH summer monsoon region (Fig. 5a). The latent heat released in monsoon rains provides a major source of energy to maintain the atmospheric engine. A similar dominant contribution of the monsoons to the hydrological cycle exists for DJF. Thus, the GM is a key component of the global hydrological cycle. In an annual mean sense, the total monsoon area occupies about 19% of the total surface area of the earth, while the total monsoon rainfall accounts for about 31% of global precipitation (Wang and Ding, 2008).

As shown in Fig. 4, the subtropical highs and associated giant anticyclonic circulations dominate the global sub-tropics except over the Eurasian continent. Much of the desert areas are closely associated with the subsidence of the subtropical highs. Conventionally, the subtropical dry climate has been interpreted as resulting from the descending branch of the Hadley circulation; but this is true only for the local winter season. For instance, upper-level convergence (descending motion) occurs over the SH subtropical highs during southern winter (JJA, Fig. 5a) and over the NH subtropical highs during northern winter (DJF, Fig. 5b). However, during the local summer season, the sinking motion is not driven by the Hadley circulation but rather, for the NH, by the upper-level convergence (sinking motion) induced by the NH summer monsoons (Fig. 5a, the “transverse monsoons”). Similarly, sinking motion in the SH summer subtropical Highs is also linked to the SH summer monsoons (Fig. 5b). The monsoon-subtropical High linkage in the summer hemisphere is thus generally strong. Such couplings are key to explaining why the North Pacific Subtropical High is much stronger during northern summer (when the Asian summer monsoon induced sinking motion is strong) rather than during northern winter (when the sinking branch of the Hadley cell is strong).

Note that dry regions are generally located to the western and poleward side of the summer monsoons due to the descent resulting from the interaction between monsoon heating-induced Rossby waves and the mean westerly flow on their poleward side (Hoskins, 1996; Rodwell and Hoskins, 1996; Hoskins and Wang, 2006). The resultant coupled monsoon - desert interactions govern a majority of the global tropics and sub-tropics, including the Australian and South American deserts in the SH and the Sahara-Arabian deserts in the NH (Figs. 2, 5). In contrast, the deserts in the Asian interiors are overall at much higher latitudes than other arid regions, and are essentially independent of the subtropical highs. The deserts in Central Asia are mainly related to the moisture barrier and the related thermo-dynamic effects of the Himalayan-Tibetan complex, the Siberian high-pressure cell, and the remote distance from the oceans (Chen et al., 1991; Kutzbach et al., 1989; Wang et al., 2005; Wu et al., 2012b). Similarly, the North American deserts are also linked in part to the rain-shadow development on the leeward side of mountains (Kutzbach et al., 1989).

From the geological perspective of the monsoon-desert coupling, desert lands could also be discriminated into planetary deserts, such as the Sahara related to the tropical African monsoon, and inland deserts, such as those in the Asian interior associated with the sub-tropical Asian monsoon. We interpret that the Paleogene circulation patterns in Asia are quite similar to today's African monsoon-desert coupling. The zonally extended arid belt is similar to the Sahara Desert and both are linked to the subtropical high-pressure zone. Today, the regions south of the Sahara are mainly under the influence of the African monsoon. This is also similar to southern-most China during the Paleogene when a tropical monsoon, primarily forced by low-latitude insolation, was present, as shown by both geological data (Guo et al., 2008; Sun and Wang, 2005) and climate simulations (Zhang et al., 2012). In the region north to the mid-latitude dry belt, a Mediterranean-like climate would have persisted (Guo, 2010; Zhang et al., 2012), as is currently the case for the North African continent. It is hypothesized that both tropical monsoons and subtropical aridity could be traced back to the very early history of Earth (Guo et al., 2008), and formation of the subtropical drylands, such as the Sahara region, should be primarily dependent on the timing of continental drift to subtropical latitudes, although uplift of eastern African topography would have enhanced aridity (Sepulchre et al., 2006). These hypotheses are consistent with climate model results showing that both African monsoon and aridity in the continental interior may have co-existed at least by the Oligocene (Fluteau et al., 1999) and a weaker-than-average Australian monsoon would be present since the Miocene (Herdal et al., 2011). However, we should be aware that relatively arid conditions do not necessarily lead to the formation of deserts, which only occurred when aridity reaches a certain threshold and which in turn is likely modulated by other factors such as the ice-building at northern high-latitudes in the mid-Pliocene that led to enhanced aridity in the Asian interior (Guo et al., 2004).

2.2. Formation and variability mechanisms of monsoon climate

2.2.1. Formation of monsoon

In the pioneering theory of Halley (1686), the monsoon circulation was perceived as a land-sea breeze circulation caused by land-sea thermal contrast. This theory assumes an annual variation of solar radiation as an imposed forcing. As such, the land-ocean configuration (tectonic forcing) and related thermal contrast determine fundamental patterns and three-dimensional structures of the monsoons, and are important for understanding their individual regional characteristics (Webster, 1987).

From the GM point of view, the hemispheric thermal contrast driven by insolation’s annual cycle is fundamental to monsoon formation, whereas land-sea thermal contrast and topography are generally critical for the formation of regional monsoon characteristics. The land-sea
The monsoon actively interacts with the hydrological cycle. On one hand, the monsoon circulations transport moisture from trade-wind ocean regimes to monsoon regions, and from winter hemisphere to summer hemisphere, which feeds monsoon rains. The latent heat released in the monsoon rains, combined with the associated radiative heating of clouds, thereby strengthen the monsoon circulation and re-inforce this large-scale exchange of moisture and energy, forming a positive feedback between the monsoon circulation and monsoon rainfall. Therefore, the interaction between GM and global hydrological cycle plays a vital role in global climate change.

Mountains, by lifting the impinging monsoon flows, lock monsoon rains in the windward side and reduce monsoon rain in the leeside, significantly changing the geographic distribution of monsoon rainfall. High land plateaus also provide an elevated heat source or sink, through surface heat flux exchange, and thus significantly affect monsoon circulations. In sum, a broad range of atmosphere-ocean-land interactions is fundamental in determining regional monsoon characteristics.

2.2.2. Drivers of monsoon variability

As stated in our first synthesis paper, the regional monsoons can vary coherently across timescales, from interannual, interdecadal, centennial, millennial, and up to orbital and tectonic timescales. Meanwhile, each subsystem within the global monsoon system has its own features, depending on its unique geographic and topographic conditions (Wang et al., 2014a). The present paper focuses on the drivers of monsoon variability both in space and time.

If the Earth’s monsoon system is considered across time scales, the drivers of its variability include energetic influences, interactions with the Earth’s surface including ice sheets and land-sea distributions, and air-sea-land interactions within the coupled climate system (Table 2). The ultimate drivers, i.e. the energy sources, of the Earth system are twofold: the external forcing of the system as manifested in solar radiation, and the internal forcing by the Earth’s interior. The distribution of the solar flux at the top of atmosphere is subject to spatio-temporal variations caused by cyclic geometric changes of the Earth’s orbits in response to gravitational forcing within the solar system, largely due to the Sun and Moon. The Earth’s interior processes, mainly mantle cycling, drive tectonic deformations responsible for changes in land-sea distribution and topography, and magmatic processes resulting in volcanic degassing. All of these can influence the monsoon system.

On the basis of incoming energy, monsoon variability responds directly to the complex interactions within the water cycle and carbon cycle (Table 2). The water cycle is the main carrier of climate processes, and comes into the GM system through both the physical movement of air and water masses, and phase transitions of water. The effects of the carbon cycle on the monsoon include both inorganic carbon such as greenhouse gases and organic carbon such as land vegetation cover.

Meanwhile, monsoon variability may also arise from dynamic and thermodynamic feedback processes within the climate system. Examples are the El Niño-Southern Oscillation (ENSO), the Atlantic multi-decadal oscillation (AMO), and interdecadal Pacific Oscillation (IPO), which cause monsoon variability mainly on interannual to multidecadal time scales (Table 2). The three groups of factors listed in Table 2 range from external forcings to internal feedbacks: “Energy sources” belong to the typical external forcings, while “air-sea interactions” are purely internal feedbacks. However, for the factors grouped under the “Water cycle” and “Carbon cycle”, the separation of “external forcing” and “internal feedback” is conditional and heavily dependent on timescale. The ice sheets, for example, vary in size slowly and can be considered as an external forcing in the context of modern climatology, yet they comprise a crucial internal feedback in the context of glacial cycles on orbital timescales ($\approx 10^4$ ka).

Thus, the GM system varies in space and time in response to both external forcings and internal feedbacks. As already discussed, the ultimate driver of monsoon climate is the annual cycle of solar insolation,
in a close connection with the hydrological cycle and surface interactions. The water and energy cycles are fundamental components of the climate system and describe the flow of energy entering the climate system at the top-of-atmosphere, the movement and redistribution of water and energy within the atmosphere, and exchanges of heat and moisture at the air-surface interface. The GM is fundamental to these flows and exchanges as it comprises a significant portion of the globe’s rainfall and is fundamental to the atmospheric and ocean circulations that not only govern the flow of water and energy within the climate system but also play a key role in determining the climate system’s albedo and emission of longwave radiation to space. Given the intertwined nature of the GM with the water and energy cycles, it is therefore also susceptible to changes in external forcing.

The ties between the GM and the water and energy cycles mentioned above are rooted in the constraint imposed by the surface energy balance. When the solar flux that reaches the surface is altered, so is the energy available for evaporation. In turn, both the net flux of moisture into the atmosphere and rainfall are altered. Changes in the surface solar flux typically arise from changes in the Sun’s radiative output or from changes in atmospheric opacity arising from variability in aerosols from volcanic or other natural sources, or anthropogenic emissions. Though both solar and aerosol variability share a common mechanism in influencing GM rainfall, principally via the surface energy budget, they are often distinguished by their cyclic versus episodic nature. Any changes in the solar flux or atmospheric composition and opacity may result in monsoon variability at various time scales.

On the other hand, the features of the underlying surface are largely responsible for the different response to the external forcing among regional monsoons. These include geographic features such as the distribution of coastlines and mountains, oceanographic features such as SST and associated upwelling zones, and cryospheric features such as the size of continental ice sheet. All these can influence the atmospheric circulations and hydrological cycle through their mechanical and thermal effects and albedo. For example, the asymmetric pattern of continents on the modern Earth explains the NH position of the ITCZ and the dominant role of the NH monsoons in the GM system. The unique size and structure of the Asian continent are responsible for the exceptional strength of the Asian monsoon.

Finally, monsoon variability may arise from dynamic and thermodynamic feedback processes within the climate system. The external forcings and internal feedbacks driving monsoon variability at different time scales can be summarized as in the following diagram (Fig. 6). In the following chapters, the drivers of monsoon variability will be examined in the order shown in Fig. 6: air-sea coupled variability (Chapter 3), solar flux (Chapter 4) and underlying surface (Chapter 5).

3. Internal interaction mechanisms of monsoon variability

Climate variability on interannual to multidecadal time scales arises largely from dynamic and thermodynamic feedback processes within the Earth climate system (e.g., Liu, 2012). These feedback processes are thus referred to as internal interactions (as opposed to variability arising from volcanic, solar, or anthropogenic influences which are typically regarded as being “forced”). As various components within the climate system vary at different rates, these internal interactions span a wide range of time scales. An example is the continental ice sheet. The ice sheet size varies on the orbital time scale of $10^4$ years due to internal interactions within the climate system, however, on shorter time scales including the human dimension, it can be regarded as “external forcing” to monsoon variation. Another example is millennial-scale monsoon variability causally linked to internal reorganizations among atmosphere and ocean circulations, such as the AMOC (Atlantic meridional overturning circulation), yet at shorter time scales, AMOC-related SST patterns may be considered as boundary conditions of external forcing. In the extreme, SST patterns generally can be considered as a boundary forcing, such as commonly assumed for example in weather forecasting.

Although the internal interaction mechanisms of monsoon variability cover a broad range of time scales, our discussion in this chapter mainly focuses on interannual to multidecadal time scales, but extending to millennial time scale and beyond. Extensively discussed are monsoon interactions with ENSO at low latitudes, with westerlies dominating in middle latitudes, and with the AMO (Atlantic Multidecadal Oscillation) from high to low latitudes. Of those, the most important mode of climate variability, ENSO, is perhaps the most robust and canonical example of atmosphere-ocean interaction.

3.1. Low-latitude processes determining GM variability

At the global scale, monsoon variability is influenced by the two extremes of the climate system: the tropical SST and the polar ice-sheets. The atmospheric deep convection in the tropics depends on the SST gradients of the warm pool and cold tongue. The oceanic deep convection in high-latitudes gives rise to deep-water formation. The convective instability in the two locations largely determines the global climate including variability of the global monsoon (Webster, 1994; Pierrehumbert, 2000).

3.1.1. ENSO - monsoon linkage

Monsoon interannual variations have been studied primarily on regional scales due to their unique characteristics associated with
specific land-ocean configurations and differing feedback processes. Key regional monsoon responses include those in South Asia (e.g., Webster et al., 1998), East Asia (e.g., Tao and Chen, 1987), Australia (e.g., McBride, 1987), Africa (e.g., Nicholson and Kim, 1997), North America (e.g., Higgins et al., 2003), and South America (e.g., Zhou and Lau, 1998). The regional approach, however, does not seek to identify coherent variability over the globe or its governing influences on planetary scales.

Is there any coherent variability across regional monsoons on the interannual time scale? To what extent are the interannual variations of the GMP driven by internal feedback processes? Wang et al. (2012b) showed that from one monsoon year (May to the next April) to the next, most continental monsoon regions, separated by vast areas of arid trade winds and deserts vary in a coherent manner driven by ENSO. In general, during El Niño, monsoonal rainfall over land tends to decrease over most regions, so the total amount of global-land monsoon rainfall tends to be anti-correlated with ENSO. However, ENSO has a stronger influence on the Northern Hemisphere summer monsoon (NHSM) than on the Southern Hemisphere summer monsoon (SHSM) and it also tends to more strongly influence continental monsoon rainfall than oceanic. As a result, the total amount of the NHSM land rainfall is highly related to ENSO with correlation coefficient $r = 0.85$ for 1979–2008 (Wang et al., 2013b). Thus, the regional monsoons are coordinated not only by external (e.g., orbital) forcing but also by internal feedback processes such as ENSO.

How does ENSO affect global and regional monsoons? ENSO affects the GM directly through its alteration of the Walker circulation and associated atmospheric teleconnections (Kumar et al., 1999; Turner et al., 2005; Wang et al., 2000; Wen et al., 2016). Specifically, the impacts of the ENSO on global and regional monsoons are season-dependent and the mechanisms by which ENSO affects monsoon are also season- and region-dependent (Wang et al., 2003b). During the northern summer of a developing El Niño, the SST anomaly in the central-eastern equatorial Pacific exhibits an amplitude about half of the mature phase. This SST anomaly can weaken the Walker circulation, which suppresses convective heating over the maritime continent. The suppressed heating over the maritime continent in turn generates an off-equatorial low-level anticyclonic circulation (as a Rossby wave response) which decreases South Asia monsoon rainfall (Wang et al., 2003b). The response of the zonal circulation, not only in the Pacific but also over the Indian and Atlantic Oceans, decreases West African rainfall as well.

During the mature phase of El Niño in boreal winter when SHSM reach their peak, the weakened Walker circulation and associated teleconnections directly suppress the Indonesian-Australian summer monsoon, the northern South American Monsoon, and the Southern African summer monsoon over land (Ropelewski and Halpert, 1987). La Niña events tend to have the opposite impact (i.e. enhanced monsoon rainfall over land during monsoon years) in many regional monsoon regions.

### 3.1.2. Coupled monsoon-ocean system

The monsoon and ocean are coupled (Webster et al., 1998). It is important to recognize that atmosphere-ocean interaction in the monsoon regions is a critical source of monsoon variability (Wang et al., 2003b). ENSO effects on monsoon can be considerably modified by local monsoon-ocean interaction over Indo-Pacific Warm Pool. The monsoon-ocean interaction over Indian Ocean can offset the remote ENSO impacts by inducing Indian Ocean Dipole (or zonal) mode (Webster et al., 1999; Saji et al., 1999; Li et al., 2003) or by directly warming northern Indian Ocean (Lau and Nath, 2000).

During the boreal summer of the decaying stage of a strong El Niño, the atmosphere-ocean interaction plays a major role in prolonging ENSO impacts in the EAM region (Wang et al., 2000). Over the Indo-Pacific warm ocean, a prominent Western Pacific anticyclonic anomaly (WPAC) is coupled with pronounced SST dipole anomalies: a cooling to its east in the Western Pacific and a warming to its west over the northern Indian Ocean and EA marginal seas from the El Nino developing fall to the decaying summer (Wang et al., 2003b). This coupling provides a positive thermodynamic feedback between the WPAC and underlying SST dipole anomaly, which amplifies the WPAC from the fall to spring and maintains it through the El Niño decay summer (Wang et al., 2000; Lau and Nath, 2003; Lau et al., 2004; Lau and Wang, 2006; Chowdary et al., 2010; Xiang et al., 2013), which enhances EASM rainfall over the Yangtze River Valley and causes northern Indian Ocean warming (Du et al., 2009). In turn, the northern Indian Ocean warming tends to increase precipitation heating that excites equatorial Kelvin wave easterlies over the far western Pacific, and the associated anticyclonic shear vorticity can enhance the anomalous WPAC (Yang et al., 2007, 2009; Xie et al., 2009; Wu et al., 2010). Note that the forcing of the warming Indian Ocean to the anticyclone in the western Pacific is only evident in late boreal summer when the monsoon trough is established. It is the SST cooling in the western Pacific that drives the anticyclone in early summer before the establishment of the monsoon trough in the western Pacific (Wu et al., 2009).

The delayed occurrence of the basin-wide warming over the Indian Ocean is a result of the atmosphere-ocean interaction induced by the remote forcing from the eastern-central Pacific through atmospheric teleconnection. This delayed Indian Ocean warming (capacitor) effect should be understood in terms of a coupled atmosphere-ocean system, rather than treating the Indian Ocean warming as a forcing to the atmosphere as has been done in some AMIP (Atmospheric Model Intercomparison Program) experiments, where SST is prescribed. The notion of monsoon-warm ocean as a coupled system has been elaborated in many previous studies (e.g., Webster et al., 1998; Lau and Nath, 2000; Wang et al., 2003b). It was recognized that AMIP experiments cannot simulate the monsoon anomalies even given the strongest 1997–98 El Niño forcing (Wang et al., 2003b), and over the summer monsoon regions the SST is generally a passive response to atmospheric forcing (Wang et al., 2005b). Therefore, the effect of the Indian Ocean warming should be demonstrated by numerical experiments with a coupled model as properly done by Lau et al. (2005), Chowdary et al. (2010), Xiang et al. (2013).

#### 3.1.3. Warm pool SST-monsoon connections

The Indo-Pacific Warm Pool (IPWP) or Western Pacific Warm Pool (WPWP), is the heat engine for the globe’s climate system, and its vast moisture and heat exchange profoundly impact climate worldwide. The oceanic warm pool is defined by the 28 °C SST threshold, which in the present era covers nearly 30% of the global ocean surface (Hoyos and Webster, 2012). The dominant latent heat release in the tropics is associated with deep convective activity, which in turn is dictated by the distribution of SST. Small variations in SST of the IPWP influence the location and strength of convection in the rising limb of the Hadley and Walker circulations, perturb planetary-scale atmospheric circulations, and influence tropical hydrology (Qu et al., 2005). The IPWP variations, consequently, have a twofold impact on the global monsoon, through the intensity of hydrological cycling related to deep convection, and though ENSO activity related to zonal asymmetry of the tropical ocean.

Historically, the warm pool varies not only in its area, but also in the intensity of its deep convection. An enhanced IPWP causes an increase in the average intensity of tropical cyclones and the monsoon circulation (Hoyos and Webster, 2012). This positive correlation is well supported by geological records. This is well supported by geological records. The LGM SST in the IPWP was nearly 3 °C lower than the present (Lea et al., 2000), and the warm pool defined by 28 °C isotherm disappeared (Chen et al., 2005). There is a correlation between cooler SST in the IPWP and reduced monsoon rainfall in neighboring regions, sea surface salinity anomalies on multi-decadal through to millennial timescales there are remotely forced by the Asian monsoon/ITCZ (Oppo et al., 2009).
3.2. Interdecadal variability of global monsoon

On interdecadal timescales, numerous studies have investigated the linkage between regional monsoons and other major modes of climate variability. For instance, Indian summer monsoon precipitation has been shown to exhibit a correlation with both the North Atlantic Oscillation (Goswami et al., 2006) and multidecadal variations of Pacific SST (Meehl and Hu, 2006). Northern China’s rainfall has been found to be correlated with the Pacific Decadal Oscillation (PDO; Qian and Zhou, 2013). There are linkages between AMO and Asian summer monsoon (Zhang and Delworth, 2006; Lu et al., 2006), the North African monsoon (Gaetani and Mohino, 2013) and North American monsoon (Sutton and Hodson, 2005). The Australian summer monsoon has been linked to PDO (Mantua and Hare, 2002; Power et al., 1999a). A variety of decadal and interdecadal variations of regional monsoons has been identified, with differing periodicity and phase change points (Yim et al., 2014b). Variability in the global monsoon during the 20th century has been documented. Wang and Ding (2006) found a decreasing trend of global land monsoon precipitation in the last 50 years. Of the 20th century, mainly occurring before the late 1970s. An extended analysis to the whole 20th century by Zhang and Zhou (2011) further demonstrated that the decreasing trend is mainly dominated by the monsoons in the northern hemisphere and is part of decadal variability rather than a purely linear trend. The model simulations by Zhou et al. (2008) demonstrated that the decadal variability of global land monsoon precipitation is modulated by the tropical SST anomalies associated with the phase transition of IPO/PDO.

Is there a coherent global structure of GM interdecadal variability? What gives rise to the GM interdecadal variability? Global precipitation observations (over both land and ocean) are available only after the late 1970s. Wang et al. (2013a) showed that during this period, a coherent decadal change of precipitation and circulation emerges in the entirety of NHSM system. The intensity of the NHSM can be measured by a NHSM circulation index which is defined by the vertical shear of zonal winds between 850 hPa and 200 hPa averaged in (0°–20°N, 120°W–120°E). The NHSM circulation index is highly correlated with the NHSM rainfall intensity over the modern record (r = 0.85 for 1979–2011). The NHSM circulation index was constructed for the period of 1871–2010 by using monthly circulation data taken from 20th Century Reanalysis for 1871–2010 (Compo et al., 2011).

It was found that on the decadal time scale, the NHSM circulation index (vertical wind shear, VWS) is significantly correlated to the so-called Mega-ENSO index (r = 0.60 for the period 1871–2010) (Fig. 7a). This correlation is higher (r = 0.77) for the period of 1958–2010 by using ERA-40 reanalysis data (Uppala et al., 2005). The Mega-ENSO index is defined by using the SST anomaly averaged over the western Pacific K-shape region minus that over the eastern Pacific triangle region between 40°S and 40°N (Wang et al., 2013b). The mega-ENSO index is an integrated measure of the interannual-to-multidecadal variability of Pacific basin-wide SST: It represents, on interannual time scale, the ENSO (NINO3.4 index) (r = 0.84 for 1901–2012) and, on decadal-multidecadal scale, the Pacific Decadal Oscillation (Mantua and Hare, 2002, r = 0.81) and the Interdecadal Pacific Oscillation (Power et al., 1999b). In addition, the NHSM circulation index is also significantly correlated with the AMO, especially on the decadal time scale (Fig. 7b). Here the AMO index is defined by detrended SST anomalies averaged over the North Atlantic between the equator and 60°N. About half of the observed decadal variance of the NHSM can be captured by a combination of the Mega-ENSO and AMO (r = 0.70) (Fig. 7c).

Though strong, the correlation does not mandate causality. Physically, the eastern Pacific cooling and the western Pacific warming (Mega-LaNina) are coupled with the strengthening of the Pacific subtropical high in both the North and South Pacific and their associated trade winds, increasing the Walker circulation and causing moisture to converge into the Asian and African monsoon regions. The Mega-La Nina also tends to increase precipitation over the central and South American monsoon regions. Through tropical teleconnection, Mega-La Nina induces enhanced westerlies over the tropical Atlantic, thus strengthening Northern African monsoon. The cumulative effect of these changes is that the NHSM would increase. The ways by which Mega-El Nino affects GM are primarily opposite to Mega-La Nina. The American monsoon strength and Northern African monsoon strength are significantly linked to the AMO (Enfield et al., 2001; Zhang and Delworth, 2006). In addition, a cold phase of the AMO (North Atlantic cooling) can shift ITCZ southward and weakens Indian and East Asian summer monsoon as well (Zhang and Delworth, 2005; Goswami et al., 2006; Lewis et al., 2010).

In sum, although there are issues regarding whether the GM can itself be viewed as a major mode of climate variability, the evidence provided here suggests that on interannual-to-multidecadal time scale, there is coherent GM variability that is driven by internal feedback modes such as ENSO, mega-ENSO, the PDO/IPO, and the AMO. On centennial or longer time scales, changes in incoming solar radiation and volcanic forcing can generate a significant and coherent GM response. Comparison of the differences and commonality among different regional monsoons will further our understanding of the fundamental processes that drive monsoon variability.

3.3. High-latitude processes determining GM variability

The fact that the interdecadal variability of the GM is linked to AMO suggests that on decadal or longer time scale, the high-latitude
processes may play a very important role. According to instrumental records, North Atlantic variability is characterized by an alternation between warm and cold SST anomalies on a timescale of 60–80 years, referred to as the AMO (Schlesinger and Ramankutty, 1994; Kerr, 2000). The AMO is linked with various decadal climate fluctuations, ranging from variations in Atlantic hurricanes to global temperatures. The forcing mechanism pacing the AMO remains subject to considerable debate (see the review of, e.g., Liu, 2012). On one hand, the AMO may be driven in part by internal ocean variability and is related to fluctuations in the AMOC (Knudsen et al., 2014; McCarthy et al., 2015). As discussed later, North Atlantic SSTs in the geological records are likely influenced by variations in the ocean circulation, and particularly the AMOC, which in turn is driven by high-latitude cryospheric processes (Wolfe and Cessi, 2011). On the other hand, it was shown that the influence of solar and volcanic forcing on the AMO played a dominant role in pacing the AMO after termination of the Little Ice Age (Knudsen et al., 2014).

### 3.3.1. Atlantic meridional overturning circulation and its impact on GM

The surface conditions of the North Atlantic have been the center of attention in modern and paleo-climatology because of their recurrent abrupt changes and the role as a significant driver of Northern Hemisphere climate. The AMOC is the major vehicle driving hemispheric heat transport in the ocean, and its intensity heavily depends on the density of the upper ocean in the North Atlantic. Drastic enhancement of the sea ice export from the Arctic, or melting water from the continental ice sheet, can cause a freshering of the North Atlantic, known as “fresh water hosing”, and lead to an altered overturning state (Häkkinen, 1999), giving rise to rapid SST changes in the region. A rapid drop in Northen Hemisphere SST of about 0.3°C around 1970 can be attributed to the widespread freshening of the upper 500–800 m layer of the northern North Atlantic, known as the “Great Salinity Anomaly” (Dickson et al., 1988; Thompson et al., 2010), supporting the notion that the observed AMO change may be resulted from internal variability.

A large number of observations and modeling works confirmed the notion that the cold North Atlantic SST leads to week monsoons in Asia and Africa (e.g., Gupta et al., 2003; Mulitza et al., 2008; Marzin et al., 2013; Chang et al., 2008). Model simulations (e.g. Zhang and Delworth, 2005; Lewis et al., 2010) demonstrate that the global response to an imposed addition of freshwater to the North Atlantic may substantially weaken the AMOC, and thus significantly change interhemispheric and meridional temperature contrasts, resulting in the southward displacement of the ITCZ, an El Niño-like pattern in the southeastern tropical Pacific, an overall weakening of the Asian summer monsoons, and an intensification of the South American and Indonesian-Australian monsoons (Fig. 8).

How does the AMOC affect GM? A weakened AMOC result in decreasing temperature in the North Atlantic and thus increasing meridional gradient of Atlantic SST. This, in turn, results in a southward ITCZ shift and subsequently an alteration of the Hadley circulation, which results in the low-latitude interhemispheric precipitation seesaw pattern (e.g. Lindzen and Hou, 1988; Clement et al., 2004; Wang et al., 2004; Wang et al., 2007b; Kang et al., 2008). Such modification produces intensified uplift of moist air in the southern low latitudes but strong subsidence in the north (Lindzen and Hou, 1988; Chiang and Bitz, 2005; Clement and Peterson, 2008). As a result, the Asian, North African and North American summer monsoons are weakened (e.g. Stuger et al., 2011; Chiang and Friedman, 2012; Otto-Bliesner et al., 2014; Wen et al., 2016a), while the South American and South African summer monsoons intensify (Nace et al., 2014), and vice versa.

Climate models further suggest that a warm phase of the AMO strengthens summer rainfall over India and the Sahel. Paleoclimatological studies have confirmed this pattern—increased rainfall in the AMO warm phase, decreased in the cold phase—for the Sahel over the past 3000 years (Zhang and Delworth, 2006). The same link was found not only in East Asia, but also in the American tropics. The simulations show that a small, 0.1 Sv freshwater addition in the northern North Atlantic can drive drying in the South American monsoon region. By contrast, a 1.0 Sv freshwater addition can lead to a complete shutdown of the AMOC, a considerable southward shift of the ITCZ, and a change in the seasonal cycle of precipitation over Amazonia (Parsons et al., 2014).

In sum, the THC variations driven by high-latitude processes have recently been shown to exert a strong influence on regional monsoon climates in a globally coordinated way.

### 3.3.2. Millennial-scale monsoon oscillations driven by high latitude processes

While the GM oscillation pattern on millennial timescales is well characterized (Cheng et al., 2016), mechanisms driving such an oscillation remain under investigation. Among several hypotheses (e.g., solar variability and natural instability of ice sheet), changes in the AMOC at high N. Atlantic latitudes appear to be the most promising candidate (e.g. Broecker et al., 1985; Clark et al., 2002; Broecker, 2003; Alley, 2007; Clement and Peterson, 2008). It has been shown that the AMOC-triggered climate events can propagate worldwide though the ‘bipolar seesaw’ mechanism (Stocker et al., 1992; Broecker, 1998; Stocker and Johnsen, 2003). Although the climate oscillations with the largest amplitude and rate are indeed centered near Greenland/the North Atlantic Ocean, the smaller, a more gradual climate oscillation of nearly opposite sign in Antarctica suggests a scenario that can be reasonably well explained by the ‘bipolar seesaw’ mechanism. A recent study further demonstrated a close anti-correlation between millennial Asian monsoon and Antarctic temperature events throughout the past 640 ka (Cheng et al., 2016), consistent with the ‘bipolar seasaw’ mechanism. The authors also showed that boreal summer insolation influences the pacing of major millennial-scale Asian monsoon events, including weak monsoon intervals during glacial terminations. In addition, a precise phase comparison of high resolution ice core records further reveals a ~210 ± 95 (2σ) years mlead of abrupt Greenland warming (cooling) to corresponding Antarctic cooling (warming) onset, in line with the aforementioned mechanism (WAIS Divide Project Members, 2015).

This mechanism linking NH and SH monsoons and AMOC variability seen for interdecadal climate variability (see Fig. 8) may equally well validate on millennial time scale. This AMOC impact on Afro-Asian monsoon is consistent with recent modeling results and observational data during the last glaciation (Otto-Bliesner et al., 2014; Wen et al., 2016a).

As an alternative to the hypothesis of freshwater perturbations, ice-sheet dynamics provide another apparent explanation of millennial events that have occurred in the North Atlantic (Petersen et al., 2013; Dokken et al., 2013), although the hypothesis needs further validation to explain the global pattern of the climate events as aforementioned.

Notably, certain features in GM proxy records from both hemispheres bear some similarities to Antarctic temperature changes. For example, some portion of the millennial-events and orbital scale responses with a more gradual trend are more similar to Antarctic records than to Greenland records. Accordingly, it has been argued that monsoon oscillations on millennial-scale during the last glacial period in equatorial Africa (Mulitza and Rühlemann, 2000; Brown et al., 2007; Weldeab, 2012), Indian monsoon (Cai et al., 2006; Caley et al., 2013), and East Asian monsoon (Cai et al., 2006; Rohling et al., 2009) regions might be driven or influenced by temperature change in high SH latitudes. However, there appears to be a paradox with the aforementioned scenario. Similarity of millennial event features in the GM record with Antarctic temperature records could also result from the mechanism that modulates Antarctic temperatures such as the SST adjustment (Deplazes et al., 2013) and a possible thermal reservoir effect (Stocker and Johnsen, 2003). On the other hand, if the millennial monsoon events in low latitudes of the NH were driven by temperature changes...
at high latitudes in the SH, it would be crucial for the same mechanism to be able to explain the observed broad similarities but nearly anti-phased patterns between monsoon events in two hemispheres (Wang et al., 2014a). Alternatively, it is plausible that low latitude GM changes may be one of the important links in the ‘bipolar seesaw’ mechanism, and thus the millennial pattern of Antarctic temperature change could also be viewed partially as a result of low-latitude GM changes, rather than as a direct cause of the GM variability. A recent study suggests that the similarity observed in some cases between Antarctic temperature and Asian monsoon δ18O records could be a manifestation of their responses to a common factor such as SST changes resulting from perturbations in the ocean circulation triggered by changes in the AMOC modes, rather than a direct causal link of one to the other (Kathayat et al., 2016).

The third hypothesis suggests that changes in ENSO can trigger abrupt climate change and affect oceanic salinity and temperature for a few centuries. The more frequent and stronger El Niño and La Niña modes (the ‘super ENSO’) may correspond respectively to the high-latitude stadial and interstadial conditions (e.g. Cane, 1998; Clement et al., 2001; Stott et al., 2002). This hypothesis lacks critical evidence to support the existence of two distinct operating modes of the tropical ocean-atmosphere system, where each mode can be locked in for many centuries or millennia (Broecker, 2003). Nevertheless, more studies are clearly necessary to further assess the role that the GM climate plays in triggering, propagating, and amplifying millennial events to global scale.

In sum, the largest amplitude and more abrupt millennial oscillations appear to be centered near Greenland and the North Atlantic, while small and gradual changes of nearly opposite sign are typical in Antarctica, a connection that can be reasonably explained by the ‘bipolar seesaw’ mechanism. In the climate context of the bipolar seesaw, low-latitude monsoon changes on millennial-scale manifest a broadly anti-phased relationship between the northern and southern hemispheres. It is also plausible that some centennial-multidecadal monsoon events have coincided with a similar anti-phased relation between the hemispheres as well. Changes in the rate of the AMOC have been suggested to be a key process that may explain most aspects of such millennial-scale changes (e.g. Broecker and Denton, 1989; Clark et al., 2002) and centennial- multidecadal scales as well (Zhang and Delworth, 2006). A surprising observation in the Mesoamerican monsoon was the presence of strong monsoon precipitation in the LGM, which is attributed to an active but shallow AMOC and proximity to the ITCZ (Lachniet et al., 2013). However, our understanding of AMOC dynamics and its climate impact is still limited and further study is necessary to ascertain the link between AMOC and distinct monsoon responses on wide range of sub-orbital timescales.

4. Radiative forcing mechanism

This section discusses effects of one type of external forcing: radiative forcing induced by changes in solar radiation and atmospheric composition. Earth’s climate system is driven primarily by heat energy arriving from the Sun in the form of electromagnetic radiation. Solar radiation arriving at the top of Earth’s atmosphere is subject to a set of changes before entering into the climate system: the geometric changes of the Earth’s orbit modify its seasonal and latitudinal distribution, the composition and opacity of the atmosphere affects the reflection and scattering of the incoming radiation back to space, and the underlying surface controls the absorbing capacity and geographic distribution of heating. This chapter considers all aspects except for the underlying surface, which is the subject of the next chapter.

4.1. Variations of solar intensity

To the general public, solar insolation may seem to be a constant, but it is not. More than forty years ago, Sagan and Mullen (1972) found from a solar model that our Sun was likely about 25% less luminous 2.3 billion years ago than it is today, and that the Earth surface temperature would have been below the freezing point of seawater if all climate feedback processes and other forcings were the same as today. This freezing Earth scenario, however, conflicts with geological evidence of the Archaean world. A variety of hypotheses have been proposed to solve this “early faint Sun paradox”, including a high atmospheric concentration of greenhouse gases (see Feulner, 2012 for a review) or the presence of NH₃ in the atmosphere protected by a haze layer (Wolf and Toon, 2010). So far, this paradox remains unsolved. To our knowledge, no paper has been published to address the Precambrian monsoon, nevertheless there is no reason to suspect the presence of a monsoon in the Archaean world. Yet, it remains a challenge to characterize and understand the response of the monsoon circulation and hydrological cycle to the radiation forcing from the ancient faint young Sun.

To a much lesser extent, solar intensity varies with solar cycles, which were discovered in the 19th century. On decadal timescales, variability in solar intensity is governed primarily by the 11-year cycle of sunspot activity and is on the order of 0.2 W m⁻² in the global average. It remains unclear precisely how such variability influences the GM, and two mechanisms have been proposed including a “top-down” stratospheric response of ozone to fluctuations of solar intensity and the “bottom-up” coupled ocean-atmosphere surface response. Meehl et al. (2011) suggested that the two mechanisms acting in concert exert a considerable influence on Indo-Pacific climate and thus a large fraction of the GM domain. Quantifying the magnitude of such effects remains a challenge however due to the difficulty of disentangling such low frequency variability from an observational record.
containing the multiple, simultaneous, and uncertain influences of aerosols, changes in land use and irrigation, solar and internal variability, and increases in greenhouse gases.

There is also considerable evidence that, on the timescale of centuries, quasi-cyclic variability in solar ebbility, and increases in greenhouse gases, aero-
sols, changes in land use and irrigation, solar and internal varia-
tory (annual mean, annual cycle, and GMP domain etc.) was found to
to be comparable to the assimilated climatology in
understand how the GM evolves under estimated forcings of the last
coupled atmosphere-ocean climate model (Legutke and Voss, 1999), to
to the top three climate models that participated in the
therefore, that variations in solar irradiance over the last centuries have
been much smaller than previously thought, and modeling with a weak
or e
instance, that variations in solar irradiance over the last centuries have
been much smaller than previously thought, and modeling with a weak
solar forcing could yield quite different results (i.e., Jungclaus et al.,
4.2. Earth’s orbital forcing

Orbital forcing is perhaps the most rapidly developed part in paleo-
climate research over the last half century. The seasonal and latitudinal
distribution of solar radiation reaching the Earth’s surface is subject to
to their forcing, then discuss the related hemispheric contrast
and the differences in the Hot- and Ice-House worlds.

4.2.1. Monsoon response to orbital forcing

Over the Phanerozoic eon, solar output is generally stable except for
sunspot cycles. Yet, the change of Earth’s orbit can induce significant
change in insolation and therefore provides a major forcing on the GM.
The impact of orbital forcing on the monsoons was first studied for the
Holocene Afro-Asian monsoon in an atmospheric general circulation model by Kutzbach (1981). Enhanced summer insolation in the NH in
the early to mid-Holocene (Fig. 9a) was reasoned to strengthen the NH
summer monsoon over land, as proposed in both models and reconstructions (e.g. Kutzbach and Otto-Bliesner, 1982; Kutzbach and Street-Perrott, 1985; Liu et al., 2003). Monsoon response to orbital forcing, therefore, can be
as shown in Fig. 9c where the obliquity is
Fig. 9 shows an example of insolation forcing in the mid-Holocene (6 ka)
relative to the present, when the perihelion occurs in September. The
insolation is enhanced by up to ~30 W/m² in high northern latitudes in
boreal summer (JJA) (Fig. 9a). This insolation change is mainly caused by
precession, as shown in Fig. 9c where the obliquity is fixed as the present-day value. Increased obliquity in the mid-Holocene (24.1° at
6 ka vs. 23.5° at 0 ka) also induces a seasonal change in solar insolation.
In local summer, an increased tilt increases the insolation at mid-
to high latitudes (Fig. 9d), because these regions face more directly toward

Fig. 9. Latitudinal distribution of the change of seasonal insolation between 6 ka and 0 ka. (a) total, (c) precessional component, (d) obliquity component. The annual mean change is shown in (b), which is due entirely to obliquity change.

In the model run without variability in the external forcing, no
significant centennial variations were identified in monsoonal precipi-
tation in the northern Hemisphere (NHMP), southern hemisphere (SHMP), or the globe (GMP). Further the NHMP and SHMP were found to
be uncorrelated, indicating that coherent interhemispheric monsoon
variability was largely nonexistent in the absence of external forcing.
In the run with variability in the imposed external forcing, on the other
hand, the NHMP and SHMP were strongly correlated on centennial to
sub-millennial time scales. Strong GMP is simulated from 1030 to 1240,
Medieval Climate Anomaly, while weak GMP is simulated
sub-millennial time scales. Strong GMP is simulated from 1030 to 1240,
the Little Ice Age from 1450 to 1850 with three minima occurring,
respectively, around 1460, 1685, and 1800. These three
minima coincided with the Spörer Minimum (1420–1570), the
Maunder Minimum (1645–1715), and the Dalton Minimum (1790–1820), periods of low sunspot activity and, in the two latter
cases, increased volcanic activity (Soon and Yaskell, 2003; Haltia-Hovi
et al., 2007). This strongly suggests a connection between GMP and
centennial-scale modulation of the solar and/or volcanic radiative
 forcings (or effective solar forcing). Liu et al. (2009) further note that
GM strength is strongly correlated to the inter-hemispheric temperature
contrast and that, on centennial timescales, the GM strength responds
more directly to the effective solar forcing than the concurrent forced
response in global mean surface temperature.

Despite of the remarkable progress, it is not easy to separate climate
variability into components attributable to external forcing and internal
climate variability (Schurer et al., 2013). There is an opinion, for

In local summer, an increased tilt increases the insolation at mid-
to high latitudes (Fig. 9d), because these regions face more directly toward

In local summer, an increased tilt increases the insolation at mid-
to high latitudes (Fig. 9d), because these regions face more directly toward

In local summer, an increased tilt increases the insolation at mid-
to high latitudes (Fig. 9d), because these regions face more directly toward
the Sun. In winter, however, there is little change of insolation, because there is little insolation during the winter half year, especially in the polar regions where the insolation is zero in winter and thus an increased tilt doesn't reduce insolation. The increased insolation in summer and lack of insolation change in winter also enhances the seasonal cycle of insolation at mid- and high latitudes. Furthermore, the insolation shows a small change in the annual mean, with a net cooling in the tropics of \(-1\) W/m\(^2\) and a warming at high latitudes of \(+4\) W/m\(^2\) (Fig. 9b). These opposing changes of annual mean insolation at low and high latitudes can be understood simply as a direct response to increased obliquity. At 6 ka, the seasonal cycle of insolation is enhanced in the NH, but decreased in the SH, by about 10\%, mainly due to the change of perihelion.

The Mid-Holocene response of the GM to orbital forcing is dominated by the enhanced summer and winter monsoons in the NH and reduced summer and winter monsoons in the SH, as a direct response to the changing seasonal cycle of local insolation. Fig. 10a–c show the boreal summer climate response to insolation forcing at 6 ka averaged across an ensemble of coupled ocean-atmospheric general circulation models (Braconnot et al., 2007a, 2007b). Surface temperature increases over all the continents, reaching over 2 °C in Eurasia (Fig. 10a), in direct response to increased insolation. This leads to a low surface pressure, an intensified monsoon southerly wind and moisture flux convergence and, in turn enhanced rainfall across the global monsoons, from North Africa to South and East Asia, and to some extent, in North America (Fig. 10b,c). Over the subtropics of the East Asia and North America, the increased summer monsoon rainfall is accompanied by drying in some neighbouring regions due to compensating descending flows (Harrison et al., 2003; Liu et al., 2003, 2004). The dry regions can't be explained simply as the westward propagation of Rossby wave in the subtropics (Rodwell and Hoskins, 1996), because the local monsoon response consists of the response to local insolation changes as well as the remote response. In addition, some regions, such as the subtropical East Asia and North America, are also strongly influenced by the westerly jet (Molnar et al., 2010; Chiang et al., 2015). In austral summer, reduced summer insolation in the SH cools the surface over all continents (Fig. 10d). In the SH, the cold land generates high pressure, reduces rainfall and northerly surface wind and in turn moisture flux divergence over the SH summer monsoon regions in South America, southern Africa and Australia (Fig. 10e, f). At the same time, reduced insolation in the NH also enhances the NH winter monsoon, with a strong southerly wind that extends from North Africa to Eastern Asia, and eventually across the equator (Fig. 10f), increasing precipitation over the ocean. This enhanced winter monsoon is clearest in the western Indian Ocean and is associated with a significant increase in the Siberian High and Asian winter monsoon. Overall, in the early to mid-Holocene, the seasonal cycle of insolation is increased in the NH, but decreased in the SH. This enhances the GM in both summer and winter in the NH, but weakens the GM in both the summer and winter in the SH. These overall features of the GM have been confirmed in both

Fig. 10. Seasonal climate response of the coupled ocean-atmosphere system to the insolation forcing at 6 ka relative to 0 ka, for boreal summer (JA) (a) surface temperature (°C), (b) precipitation and (c) surface wind. The response is derived as the difference between the CGCM simulations forced by insolation at 6 ka and 0 ka in PMIP2. (d), (e) and (f) are the same as (a), (b) and (c), respectively, except for being boreal winter (JF). Similar to Fig. 9, contour intervals ± 0.2, ± 0.5, ± 1.0, ± 3.0 °C for temperature, and ± 0.2, ± 0.5, ± 1.0, ± 3.0 mmd\(^{-1}\) for precipitation, dark (light) grey shading for temperature warming (cooling) > 0.5 °C, and for precipitation increases (decreases) > 0.5 mmd\(^{-1}\), hatched lines in (a) and (d) for regions where the mean intermodel difference exceeds 1.5 °C. Nine models are considered in the ensemble mean for each period. (adapted from the model output of Braconnot et al., 2007a).
models and observations (e.g. COHMAP members, 1988; Liu et al., 2003 and references therein).

The Mid-Holocene example illustrates the sensitivity of climate change to orbital forcing, and in particular to precession, and the complicated spatio-temporal variations in the insolation response to the orbital forcing, cautioning against simplistic approaches to orbital forcing in paleo-climate interpretations. Recent studies suggest that obliquity has a direct effect on monsoon variability, and we will discuss it later (“Section 6.3.2”).

As the temporal resolution of geological records declines with age, the ~20 ka precessional cycles are not easily recognizable in deep-time geological archives. Fortunately, the climate amplitudes of precession cycles are modulated by short and long eccentricity of 100 ka and 400 ka cycles, respectively. Of particular value is the 400 ka long eccentricity which is the most stable orbital parameter throughout geologic history (Berger et al., 1992; Matthews and Frohlich, 2002) and has been extensively reported from marine sequences since the Paleozoic (e.g., Horton et al., 2012; De Vleeschouwer et al., 2013). The 400 ka signal is most prominent in inorganic $\delta^{13}C$ records of marine sediments (e.g., Wang et al., 2010; Tian et al., 2011). Since the long eccentricity cycle pervades the $\delta^{13}C$ and other biogeochemical records in geologic history, and 400 ka rhythms in the oceanic $\delta^{13}C$ sequence can be hypothetically explained as arising from changes in the ratio between particulate and dissolved organic carbon (POC/DOC) in the ocean, which depend on the monsoon-controlled nutrient supply (Wang et al., 2014b). The 400 ka cycles are likened to the Earth’s “heartbeat” (Pâlâke et al., 2006) and represent a potential archive of GM variability for the pre-Quaternary world (see our first synthesis, “6.3 Eccentricity modulation” in Wang et al., 2014a). The global impact of long eccentricity cycles extends beyond climate changes. For example, the 400 ka rhythms are found to correspond to third-order sequences in stratigraphy (Strasser et al., 2000) and to pace the pulses of species turnover in small mammals (van Dam et al., 2006). Furthermore, the amplitude of the long eccentricity itself is modulated by still longer cycles of 2.44 ma and 9 ma (Boulila et al., 2012; Sprovieri et al., 2013), opening up new opportunities for GM research in deep-time geology.

4.2.2. Hemispheric contrast

On orbital timescales, the GM is driven predominantly by preces- sion-induced changes in solar radiation because unlike eccentricity, precession does not change total annual insolation but rather affects the seasonal distribution of solar radiation. As a result, any increase in northern tropical-subtropical summer insolation is balanced by a corresponding decrease in southern tropical-subtropical summer insolation and vice versa. Indeed, interhemispheric contrasts in the GM on precessional timescales are well manifested in proxy records, which exhibit grossly anti-phased relationship between hemispheres (Wang et al., 2014a, “6.2 Precession and inter-hemispheric contrasts”). The climate model simulations ranging from the low-resolution general circulation model of global atmosphere (Kutzbach (1981) to the current state-of-the-art atmospheric general circulation model (AGCM) with an embedded scheme for water isotopes (Battisti et al., 2014), also demonstrate a clear GM interhemispheric contrast (Schmidt et al., 2007; Braconnot et al., 2008; Zhao and Harrison, 2012; Li et al., 2013; Liu et al., 2014; Battisti et al., 2014). The modelling results indicate that at times of high NH insolation, the summer monsoon (June to August) rain belt in the NH, particularly over the Asian, Northern South American and North African monsoon domains, penetrates northward further inland due to increased land-ocean thermal contrast, whereas summer monsoon rainfall (December to February) in the SH, particularly over the South American, South African and Indo-Australian monsoon domains, is significantly reduced (Fig. 10b,e). This hemispheric contrast is caused by the latitudinal migration of the ITCZ. Paleo-records indicate that the ITCZ migrates towards a differentially warming hemisphere together with a shift of the rain belt (Schneider et al., 2014). Essentially, precession-induced insolation variations are strongly expressed in rainfall at low latitudes, resulting in an out-of-phase pattern between hemispheres.

Potentially, other climate forcings such as the global ice volume and atmospheric CO$_2$, can modify the inter-hemispheric contrast. For example, a recent climate model experiment invoking the ice volume condition and atmospheric CO$_2$ concentrations, found a smaller contribution from both eccentricity and obliquity to the monsoon variability (Weber and Tuenter, 2011). In addition, observations show that equatorially asymmetric thermal forcing leads to a cross-equatorial pressure gradient, which in turn drives moisture flow across the equator. This mechanism may be particularly important in driving orbital-scale variability in the Asian-Australian monsoon domain (e.g. An et al., 2011). On the basis of this consideration, the insolation differential between two hemispheres has been proposed as a major monsoon forcing (e.g. Zhuo et al., 2014). However, the insolation differential effectively does not change the phase at precessional bands nor its major spatial patterns and therefore, it remains an open question as to what degree the cross-equatorial forcing, in comparison to NH or SH insolation, contributes to monsoon variability on orbital scales.

The climate variance in the equatorial-tropics (low latitudes, e.g. < 10°), however, tends to exhibit complex behavior at precessional timescales in both proxy reconstructions (e.g. Cheng et al., 2012a, 2012b, 2013) and climate simulations (e.g. Zhao and Harrison, 2012; Liu et al., 2014; Battisti et al., 2014). For instance, on the precessional timescale, the South American monsoon exhibits a nearly anti-phased relationship between eastern and western Amazonia at ~6°S (Cruz et al., 2009; Cheng et al., 2013), which is remarkably consistent with results from a recent climate model study (Battisti et al., 2014). The aforementioned disparities at lower latitudes may result from changes in the Walker circulation associated with monsoonal deep convection over equatorial-tropical regions, tied to specific insolation and/or glacial boundary conditions and complex atmosphere-ocean interactions (Clement et al., 2004; Cruz et al., 2009; Cheng et al., 2013).

Proxy records show that orbital-scale GM variations are punctuated by a series of millennial-length events at least over the past several glacial-interglacial cycles (Cheng et al., 2016). This pattern of millennial climate variability also exhibits a broad interhemispheric contrast in that weak monsoon events in the NH correlate with strong monsoon events in the SH and vice versa (Wang et al., 2014a, “5.1”). While the interhemispheric contrast on the precessional timescale is ultimately linked to changes in external forcing (i.e. annual variation of insolation and resulted land-sea thermal contrast), the millennial GM variability may be causally linked to internal reorganizations among atmospheric and oceanic circulations, such as the AMOC, which can change the interhemispheric thermal gradient, and in turn, can shift the mean latitudinal location of the ITCZ, producing hemispherically asymmetric monsoon events. Due to the different forcing mechanisms at orbital (by precessional forcing) and millennial (by AMOC) timescales, the GM response can exhibit different features. For example, the intensity of the summer and winter monsoons tends to be in phase at the orbital time scale, because of the dominance of the precessional forcing, but out of phase for millennial variability, because of the interhemispheric bipolar temperature forcing associated with the AMOC, as suggested by model simulations for the East Asia monsoons (Wen et al., 2016a).

4.2.3. Regional differences

Insolation is a direct forcing to monsoons, yet over land the monsoon response can be affected by feedbacks from nearby oceans and land surface processes. These feedbacks can differ substantially among regional monsoons and therefore may contribute to the complex regionality of the GM. The oceanic feedback on the GM can be studied by a comparison of coupled and atmosphere-only model simulations with prescribed SST during, for example, the Holocene. With an active ocean, the North African summer monsoon is enhanced significantly, with the rain belt moving further north in better agreement with paleo-reconstructions, reflecting a positive oceanic feedback on the North.
The Holocene, similar monsoon responses to orbital forcing have also been simulated in coupled climate models spanning the last 300 ka with the model orbital forcing accelerated (Liu et al., 2006c; Kutzbach et al., 2008; Chen et al., 2010b, also see Section 5.3, for phase relations between forcing and responses). At even longer timescales, among the past 5 ma records of oceanic δ13C from different monsoon regions, the precession and long-eccentricity forcing is best manifested in the Mediterranean sequences of the African monsoon with intercalated sapropels generated by Nile flooding (Wang et al., 2010). This is consistent with simple energy balance model simulations 30 years ago that show that the largest thermal response to the precession cycles occurs over Africa (Short and Mengel, 1986), and that the interaction between the biannual passage of the Sun across the equator and the seasonal timing of perihelion increase the tropical climate response to eccentricity cycles, suggesting a strong response of the African monsoon to the 400 ka and 100 ka cycles (Short et al., 1991; Crowley et al., 1992). A recent comparison of late Pliocene marine sequence between the Mediterranean and the South China Sea arrived at a similar conclusion that the former has a much cleaner precession and eccentricity in its δ13C and other monsoon-related proxy records (Li and Wang, 2015).

In general, the monsoons are driven by solar radiation globally, whereas the specific features of the underlying surface generate differences between regional monsoon systems. With its superior size, Asia hosts the largest among all the regional monsoon systems and includes three different sub-systems, of which the East Asian monsoon reaches the highest latitude (45°N) (Fig. 3). In the SH, the Australian monsoon is most susceptible to the NH influence because of its close Asian connection and oceanic surroundings. In terms of orbital forcing, the African monsoon exerts the most robust precession and eccentricity signals, apparently because of its equator-straddling position.

All regional monsoons respond to the cyclic changes of insolation induced by Earth’s orbits, primarily the precession cycle of ~20 ka. As shown by numerical modeling, precession-related peaks in the

African monsoon in response to orbital forcing (Kutzbach and Liu, 1997; Braconnot et al., 2000, 2007b; Liu et al., 2003; Zhao et al., 2005). The northward migration of the ITCZ over West Africa ranges from 2 to 10° in PMIP2 models, accompanied by a stronger northward SST gradient in the tropical Atlantic and an enhancement of Sahel precipitation (Braconnot et al., 2007b). The ocean also has feedbacks on other regional monsoons across the globe, some positive and some negative. As revealed by climate simulations of the Holocene, the ocean feedback is positive for the early stage of the African monsoon, but strongly negative for the Indian and East Asian monsoons (Marzin and Braconnot, 2009a). One interesting feature of the oceanic feedback is on the Australian summer monsoon. Countering the direct effect of a reduction in insolation in austral summer, which tends to weaken the summer monsoon in the SH, oceanic feedbacks induce a significant increase of monsoon wind and rainfall over northwestern Australia, because of the feedback between monsoon winds, coastal oceanic upwelling, and SST. This SST effect in enhancing the Australia monsoon could overwhelm rainfall reductions caused by the direct insolation effect and lead to a modest increase of total rainfall over northwestern Australia in some coupled models (Liu et al., 2003, 2004; Braconnot et al., 2007b). This may offer an explanation of some proxy evidence suggesting a possible intensification of Australia summer monsoon in the Holocene (e.g. Liu et al., 2004).

The GM can also interact with land ecosystems through the albedo feedback on the energy cycle (Charney, 1975) and evapotranspiration feedback on the hydrological cycle (Shukla and Mintz, 1982). This will be demonstrated on the vegetation-monsoon feedback in the North Africa later in Section 5.3.2.

The Mid-Holocene case convincingly demonstrates that the geographic differences in response to orbital forcing are inherent to the GM. One should not expect an equal change in regional monsoon systems driven by the same orbital forcing, because of the different geographic conditions and feedback processes for each. Before the
amplitude spectra of the maximum summer temperature at the equator are much stronger in Africa than in other regions, such as for example the open Pacific Ocean (Fig. 11A, B). Consequently, precession forcing is clearly manifested in African monsoon proxies compared to proxies of other monsoon systems. The conclusion is well supported by geological data, and can be illustrated by a comparison of the Pliocene records between the African and East Asian monsoons, represented by the Sicily section off North Africa and ODP Site 1143 in the South China Sea, respectively. The 400 ka long sequence of δ18O and δ13C from 3.2 Ma to 2.8 Ma off Africa displays distinct precession cycles of low-latitude climate processes, while the precession signal in the SCS shows much lower power and is juxtaposed with many other frequencies (Fig. 11C–F) (Li and Wang, 2015).

Additionally, the 400 ka and 100 ka peaks of long and short eccentricity at the African site are estimated in modeling experiments to be a factor of 3 stronger than in the Pacific in modeling experiments (Fig. 11A, B), because the effects of eccentricity in modulating the seasonal cycle through the precession cycle are more pronounced over land than over ocean (Short et al., 1991). The modeling results agree well with proxy data. From a global comparison of marine δ18O and δ13C time series over the last 5 ma, the highest coherence and an almost in-phase relationship between isotope records and long eccentricity cycles is found for the Mediterranean planktonic record, that contains sapropel deposits resulting from extensive Nile flooding caused by an intensified African monsoon (Wang et al., 2010).

4.2.4. Hot-house and ice-house world

The geological history, at least in the Phanerozoic, is characterized by alternation of warm “Hot-House” (or “Green-House”) and cold “Ice-House” climate regimes (Miller et al., 1991). For the past 34 million years, Earth has had ice on both poles and experiences its Ice-House stage, while a warm equable “Hot-House” climate existed during the mostly ice-free Cretaceous period (65–140 ma ago). The two regimes could differ significantly in their respective climate responses to the three orbital components.

In the Ice-House world, like the Quaternary, ice sheets and sea ice generate high atmospheric pressure in the polar regions through the ice albedo effect, providing a relatively stable environment in the global climate system. This has been speculated to be an explanation for the presence of two signals in most Quaternary monsoon records: a 20 ka precession signal of low-latitude insolation, and a 40 ka obliquity (or 100 ka in late Quaternary) signal of high-latitude ice-sheet changes. Under the Hot-House regime, by contrast, atmospheric pressure in the ice-free polar regions is thought to be subject to seasonal reversal that leads to instability in the polar and subtropical fronts and consequently in the midlatitude westerlies. In this state, the high- and mid-latitude ocean becomes more turbulent, but the equatorial system remains stable with steady easterly winds, as speculated by Hay (2008, 2011), resulting in an enhancement of precession forcing effects in climate variations.

If the above inference holds true, climate variations in the Hot-House Cretaceous should be dominated by low-latitude processes with periodicities associated with precession and eccentricity, as has indeed been observed in the proxy record. Herbert (1997) compared the orbital pacing of marine carbon cycle in deep sea sediments over the past 125 ma, and found clear differences between the Ice- vs Hot-House regimes. Analyzing the variations in carbonate accumulation, he found much of the variance carries the “polar” signal of 40 ka obliquity forcing for the past 35 ma of the Late Cenozoic Ice-House. By contrast, the carbonate cycle is dominated by precessional forcing modulated by 100, 400, and 2400 ka eccentricity beforehand, in the “Hot-House” world. Because the same feature is observed in both Hemispheres, and the carbonate accumulation varies together with biogenic silica, the signal of orbital forcing is believed to be indicative of the global biological production rather than carbonate dissolution or other kinds of local effects. Similar eccentricity and precession signals were also revealed in the early Cretaceous deposition of eolian dust transported from Africa to Mexico by easterlies (Latta et al., 2006), again confirming the existence of strong low-latitude variations in Hot-House climates.

As discussed previously, the 400 ka long eccentricity is best recorded in marine δ13C and carbonate records because of the long residence time of carbon in the oceanic reservoir (105 years), and the long-term cycle in oceanic carbon reservoir can be hypothetically attributed to changes in the ratio between particulate and dissolved organic carbon (POC/DOC) in the ocean, depending on the monsoon-controlled nutrient supply (‘DOC hypothsis’; see Wang et al., 2014a, 2014b). Therefore, low-latitude processes including the monsoons probably played a dominant role in Cretaceous climate variability, but the climate regime changed at the Eocene/Oligocene boundary around 35 ma ago when the polar ice cap started to develop. The Oligocene deep-sea record reveals both 400 ka and 100 ka eccentricity and 1.2 ma obliquity cycles in glacial and carbon cycle events (Pälike et al., 2006), suggesting that the oceanic carbon reservoir responds both to high-latitude ice-sheet and low-latitude monsoon variations.

The role of high-latitude processes in the global climate becomes stronger with the growth of ice sheets. Remarkable are two events of rapid ice-sheet growth at 13.9 Ma and 1.6 Ma, the former in Antarctic and the latter in Arctic. Ocean restructuring at 1.6 Ma marked by the isolation of a sluggish abyss under the Southern Ocean has obscured the long-eccentricity 400 ka signal in oceanic δ13C (Wang et al., 2010, 2014b). A similar change occurred in the Miocene at 13.9 Ma when the 400 ka cyclicity in δ13C records flattened out together with a drastic cooling and Antarctic ice-sheet expansion (Holbourn et al., 2007; Tian et al., 2009). The above cases illustrate how ice sheet development modifies the climate response to the orbital forcing. It will be interesting to compare which climate regime strengthens the monsoon circulation: Hot-house or Ice-house. Obviously the answer has to wait until more Pre-Quaternary monsoon records become available as the hypotheses on the monsoon responses in the two different worlds remain speculative. So far, there are no studies in state-of-the-art climate models comparing GM responses between the hot-house and ice-house base states.

4.3. Volcanic activity and aerosol forcing

4.3.1. Volcanism and monsoon

Volcanism enters into the climate system through its erupted gas and aerosols. While the aerosols associated with small volcanic events are rained out of the atmosphere on a timescale of weeks, large explosive eruptions inject significant amounts of sulfur into the stratosphere, where it can reside for years, cool Earth by reflecting and scattering incoming sunlight, and warm the stratosphere by absorbing solar and terrestrial radiation. Furthermore, although each volcanic eruption only affects solar forcing for a few years, a series of explosive volcanic events can have an accumulated effect on global climate of up to centennial time scales through their impact on the ocean and sea ice (e.g. Zhong et al., 2011). In response to a volcanic eruption, the immediate cooling effect tends to be stronger over land than over the ocean due to the smaller heat capacity of the former, which reduces the temperature difference between the two and hence decreases monsoon precipitation. It also reduces the energy available at the ocean surface to support evaporation and thereby decreases rainfall globally.

Recently, the volcanism-monsoon relation has become a hot topic in climatology, and a number of works have demonstrated reductions in monsoon precipitation after major volcanic eruption. A comparison between the simulations and observation datasets not only shows the significant decrease of global precipitation following eruptions (e.g. Trenberth and Dai, 2007), but also the largest decrease of rainfall occurring in rainy tropical, monsoonal regions (Iles et al., 2013; Iles and Hegerl, 2014). The drought history in China offers an ideal test of the hypothetical relationship between volcanism and the monsoon. Using proxy and historical data over the past 700 years, Zhuo et al. (2014)
were able to show that volcanic eruptions can cause profound drought arising from weakened monsoon precipitation. According to the simulation results, the reduction in summer precipitation over eastern China can be attributed to a decrease of moisture vapor over the tropical ocean, and the weakening of the summer monsoon due to the reduced land-sea thermal contrast after large volcanic eruptions (Man et al., 2014).

A recent study revealed a latitudinal dependence of volcanism in its climate impact. Using a 1500-year volcanic sensitivity simulation by the Community Earth System Model version 1.0 (CESM1), Liu et al. (2016) investigated whether large volcanoes erupted at different latitudes have distinctive effects on the monsoon in the Northern Hemisphere (NH) and the Southern Hemisphere (SH). They found that the GM precipitation in one hemisphere is enhanced significantly by the remote volcanic forcing occurring in the other hemisphere. This remote volcanic forcing-induced intensification is mainly through circulation change rather than moisture content change. In addition, the NH volcanic eruptions are more efficient in reducing the NH monsoon precipitation than the equatorial ones, and so do the SH eruptions in weakening the SH monsoon, because the equatorial eruptions, despite reducing moisture content, have weaker effects in weakening the off-equatorial monsoon circulation than the subtropical-extratropical volcanos do. The 1783–1784 Laki eruption in Iceland, for example, caused significant NH cooling (by 1–3 °C) and weakening the African and Indian monsoon circulations, with precipitation anomalies of 1–3 mm/day over the Sahel of Africa (Oman et al., 2006), well in line with the recent finding that large summer high-latitude eruptions in the NH cause strong hemispheric cooling and induce an El Niño-like anomaly in the equatorial Pacific (Pausata et al., 2015).

4.3.2. Aerosol and monsoon

Volcanic eruptions, however, are only one of multiple sources of atmospheric aerosols, and the role of desert and anthropogenic aerosols in climate has attracted scientific interest for decades. Two groups of aerosols can be discriminated: absorbing aerosols such as dust and black carbon, which can heat the atmosphere by absorbing solar radiation, and non-absorbing aerosols such as sulfate, which scatter solar radiation and have a relatively weak atmospheric heating effect. Both groups of aerosols cause surface cooling by blocking solar radiation from Earth’s surface and thereby influencing monsoon climate: a process referred to as “global dimming effect” (Stanhill and Cohen, 2001). Now the direct climate impact of aerosols though solar absorption is broadly recognized, and the aerosol forcing of tropical rainfall and monsoon variations is generally accepted (Lau et al., 2006).

Geographically, the coincidence of natural and anthropogenic aerosol sources in Asia leads to high aerosol concentrations throughout the southern part of the continent where the Asian monsoon prevails (Chung et al., 2005). The aerosol forcing of the Asian monsoon has subsequently become a subject of extensive research. For example, it was found that dust-induced heating of the atmosphere over North Africa and West Asia rapidly modulates monsoon rainfall over central India, and dust and precipitation vary in concert over timescales of about a week (Vinoj et al., 2014). It was also shown that the increase of black carbon and other aerosols in South Asia since 1930 has been accompanied by strong negative trends in surface solar radiation and summer monsoon rainfall, suggesting a weakening effect of aerosols on monsoon precipitation (Ramanathan et al., 2005). The Indian monsoon rainfall was found to be closely associated with the Middle East dust aerosols on seasonal timescales (Jin et al., 2015), and the rainfall response to dust was attributed to its heating effect in the mid-to-upper troposphere, which enhances monsoon circulation (Jin et al., 2016). Modeling results also revealed that the aerosol forcing plays a primary role in driving the weakened low-level monsoon circulation in East Asia (Song et al., 2014). Great attention has been paid to mutual interactions between aerosols and the monsoon in China when weakening monsoon circulation in recent decades has likely helped to increase regional aerosol concentrations, the substantial increase in Chinese air pollutants has also likely decreased the temperature difference between land and sea, further weakening the monsoon circulation (Wu et al., 2016).

On a global scale, the monsoon-aerosol interaction can be best exemplified by the Asian monsoon-aerosol interaction as Asia is a primary source of emissions of diverse species of aerosols from both anthropogenic and natural origins and the Asian monsoon system is the strongest monsoon system. Li et al. (2016) provided a comprehensive review of studies on Asian aerosols, monsoons, and their interactions. They pointed out that distributions of aerosol loading are strongly influenced by distinct weather and climatic regimes, which are, in turn, modulated by aerosol effects. On a continental scale, aerosols reduce surface insulation and weaken the land-ocean thermal contrast, thus inhibiting the development of monsoons. Locally, aerosol radiative effects alter the thermodynamic stability and convective potential of the lower atmosphere leading to reduced temperatures, increased atmospheric stability, and weakened atmospheric circulations. The atmospheric thermodynamic state, which determines the formation of clouds, convection, and precipitation, may also be altered by aerosols serving as cloud condensation nuclei or ice nuclei. Absorbing aerosols such as black carbon and desert dust in Asian monsoon regions may also induce dynamical feedback processes, leading to a strengthening of the early monsoon and affecting the subsequent evolution of the monsoon. Many mechanisms have been put forth regarding how aerosols modulate the amplitude, frequency, intensity, and phase of different monsoon climate variables.

There is a wide range of theoretical, observational, and modeling findings on the Asian monsoon, aerosols, and their interactions. A new paradigm proposed by Li et al. (2016) hypothesizes that natural aerosols such as desert dust, black carbon from biomass burning, and biogenic aerosols from vegetation are considered integral components of an intrinsic aerosol-monsoon climate system, subject to external forcing of global warming, anthropogenic aerosols, and land use and change.

Despite this progress, aerosols remains a major source of uncertainty in both the interpretation of paleoclimate records and future climate projections, and our understanding of the aerosol forcing in monsoon variations is far from adequate. Noteworthy is the indirect climate impact of aerosol, such as the possible contribution of anthropogenic sulfate to the Sahelian drying trend in the 20th century (Rotstayn and Lohmann, 2002). From a global perspective, the combined effects of aerosols from different parts of the world are likely to alter the large-scale heating and pressure gradients, induce changes in the atmospheric general circulation, and influence clouds and rainfall (Lau et al., 2008), yet further work is required to understand the multiple connections between aerosol and rainfall variability in the monsoon regions (e.g., Gautam et al., 2009; Kuhlmann and Quaas, 2010).

4.4. Greenhouse gas forcing

Greenhouse gases (GHG) in the atmosphere absorb and emit radiation, and the increase of GHG concentration changes the flows of energy and water through the climate system. It is known that an intensification of the hydrologic cycle is a robustly simulated response to increased greenhouse gas concentrations as a consequence of increases in the surface evaporative and net downward longwave fluxes. Typically, these changes result in intensification of global monsoon rainfall. Future changes of GM under GHG-induced global warming was assessed by using 20 coupled models that participated in the phase five of Coupled Model Intercomparison Project (CMIP5) (Lee and Wang, 2013). They find that the land monsoon domain over Asia tends to expand westward by 10.6%; the annual mean, annual range of GM precipitation and the percentage of local summer rainfall will all amplify at a significant level over most of the global monsoon region; changes in monsoon precipitation exhibit huge differences between the NH and the SH. The NH monsoon precipitation will increase.
significantly due to increase in temperature difference between the NH and SH, significant enhancement of the Hadley circulation, and atmospheric moistening, against stabilization of troposphere. Wang et al., 2014c explained that the “warm land-cool ocean” favors the entire Asian-Australian monsoon rainfall increase by generation of an east-west asymmetry in the sea level pressure field. On the other hand, the warm NH-cool SH induced hemispheric SLP difference favors the ISM, but reduces the Australian summer monsoon rainfall. While the Indian monsoon precipitation will increase, the monsoon circulation will weaken (May 2002; Ueda et al., 2006). While modeling results show an increase in rainfall and its variance over the Sahel with enhanced GHG forcing (Caminade et al., 2006), there are significant uncertainties in the projections among the state-of-the-art climate models (Giannini et al., 2008; Collins et al., 2013; Christensen et al., 2013). The CMIP5 models also show large spreads in the projection of monsoon rainfall changes over East Asia due to the uncertainty in the projection of the western Pacific tropical high which controls the water vapor supply to monsoon rainfall (He et al., 2015). The uncertainty in Indian monsoon rainfall projection can be largely explained by the climate sensitivities of models to GHG forcing (Chen and Zhou, 2015). The response of North America changes in total monsoon season rainfall are small and insignificant (Cook and Seager, 2013).

Above we discussed solar-volcanic and greenhouse gas forcing separately, so now let’s compare the two effects on global monsoon and precipitation. According to transient simulations, solar-volcanic effects dominate most of the slow climate variations within the past millennium, while the impacts of greenhouse gases have dominated since the second half of the 20th century (Ammann et al., 2007). Recent examination of global precipitation changes over the last millennium and future projections to the end of the 21st century has revealed that the solar, volcanic, and GHG forcings induce contrasting changes in global precipitation and SST (Liu et al., 2013). Under global warming, precipitation is likely to increase at high latitudes and the tropics and to decrease in already dry subtropical regions (Durack et al., 2012). The absolute magnitude and regional details of such changes, however, remain intensely debated (Held and Soden, 2006; Wentz et al., 2007). Liu et al. (2013) found that for every degree rise in global temperature, the global rainfall rate since the Industrial Revolution has increased at a smaller rate (less by about 40%) than during the medieval warming period in simulations using the ECHO-G model. This difference results from different sea surface temperature patterns. When warming is due to increased greenhouse gases, the gradient of SST across the tropical Pacific weakens, but when it is due to increased solar radiation, the gradient increases. For the same global surface temperature increase the latter pattern produces less rainfall, notably over tropical land, which explains why in the model the late 20th century is warmer than the Medieval Warm Period but precipitation is less. The role of SST gradient in global precipitation is well known from El Niño studies (Ropelewski and Halpert, 1996). Proxy reconstructions also indicate that the gradient strengthened in past millennial periods when the Earth warmed due to increased solar radiation (Adams et al., 2003; Cobb et al., 2003; Mann et al., 2005), although the response of the zonal SST gradient to global cooling at the Last Glacial Maximum (LGM) remains uncertain in both the observation and models (e.g. Liu et al., 2005). In contrast, in most model projections of future greenhouse warming this gradient weakens (Vecchi et al., 2006).

Why do the greenhouse and solar radiation warming have different effects on the Pacific SST gradient? If we ignore the complex effects related to clouds, adding long-wave absorbers, that is heat-trapping greenhouse gases, to the atmosphere decreases the usual potential temperature difference between the surface and the top of the atmosphere, making the atmosphere more stable. The increased atmospheric stability weakens the trade winds, resulting in stronger warming in the eastern than the western Pacific, thus reducing the usual SST gradient—a situation similar to El Niño. Solar radiation, on the other hand, heats the Earth’s surface, increasing the usual temperature difference between the surface and the top of the atmosphere without weakening the trade winds. The result is that heating warms the western Pacific, while the eastern Pacific remains cool from the usual ocean upwelling.

The different effect on global precipitation by greenhouse and solar radiative warming is also consistent with the global tropospheric energy budget (Allen and Ingram, 2002), which requires an approximate balance between the latent heat released in precipitation and radiative cooling. The tropospheric cooling is less for increased greenhouse gases, which add radiative absorbers to the troposphere, than for increased solar heating, which is concentrated at the Earth’s surface. Thus warming due to increased GHG has a different climate fingerprint than warming due to solar radiation changes.

By analysis of the ECHO-G millennial simulations (Legutke and Voss, 1999), Liu et al. (2012) further found that the NHSM precipitation responds to GHG forcing more sensitively, while the SHSM precipitation responds to the solar-volcanic radiative forcing more sensitively. The NHSM is enhanced by increased NH land-ocean thermal contrast and NH-minus-SH thermal contrast. On the other hand, the SHSM is strengthened by enhanced SH subtropical highs and the east-west mass contrast between Southeast Pacific and tropical Indian Ocean. Intensification of the summer GM is associated with (a) increased global land-ocean thermal contrast, (b) reinforced zonal mass contrast between Southeast Pacific and tropical Indian Ocean, and (c) enhanced circum-global SH subtropical highs. The physical mechanisms revealed here add to the understanding of future change of the GM, and also are valuable in the interpretation of paleo-records and attribution of forcing mechanisms.

5. Underlying surface

Solar forcing is the basis for the global nature of the monsoons as it drives variations in a globally coherent way. Contrasts among the various regional monsoons are largely driven by the underlying surface. Each of the regional monsoons is marked by the specific surface features that shape its unique response to forcing. These features include the land-sea distribution, ocean circulations and land topography, which all can vary basically on geological timescales. Tectonic deformations, glacial cycles and sea-level changes are among the major processes modifying the underlying surface and hence the monsoon themselves.

5.1. Land-sea distribution

The geographic distribution of continents is the primary feature of the underlying surface that determines the monsoon climate pattern. Although monsoons as an atmospheric circulation system can exist even on an aqua-planet without land-sea contrast (Yano and McBride, 1998; Chao and Chen, 2001; Schneider and Bordoni, 2008), in geological reality, however, the continents on Earth came into existence together with the ocean (Valley et al., 2014). There are two ways for the Earth’s surface to force the monsoon variations, either through the Earth’s albedo, or the atmospheric circulation; and both are primarily determined by the land-sea distribution.

5.1.1. Areal proportion of land and sea

As a system of coupled interactions, the global monsoon is sensitive to the proportion between the total areas of land and sea. This proportion is subject to constant variations with sea-level changes that can be traced back to the early Archeozoic, nearly 4 billion years ago (Eriksson, 1999). Fluctuations in global sea level result from changes in the volume of water in the ocean or the volume of the ocean basins.
Water-volume changes are dominated by the growth and decay of continental ice sheets, and the changes in ocean basin volume are dominated by slow variations in sea-floor spreading rates or ocean ridge lengths (Miller et al., 2005). Birth of new oceans, for example, decreases the mean age of the total seafloor and hence reduces the global volume of ocean basins, as the oceanic crust deepens with age due to thermal subsidence (Cogné et al., 2006). However, it is only the changes in the glacial cycles that have been studied for their impact on the monsoon climate.

The growth of ice sheets during glaciations causes sea level to drop and reduces ocean area. Of particular significance are the broad shelf seas between Asia and Australia, the so-called “maritime continent”. Climatically this is the centre of the Indo-Pacific Warm Pool (Fig. 12), a source of latent heat and moisture for the global atmospheric circulation, and a key area in energy transfer between the Pacific and Indian oceans, particularly during ENSO events. During the LGM, with the 120 m of global sea level drop, the Sunda Shelf (“Great Asian Bank”) and the Sahul Shelf (“Great Australian Bank”) emerged (Fig. 12B,C). Together with the exposed East China Sea Shelf, a total of ~3.900,000 km² sea area emerged as low land, greatly changing the sea/land ratio in the East Asian and Australian monsoon regions (Wang, 1999).

The geographic changes had serious climate consequences. With the decline of sea/land areal ratio and SST in the last glaciation, monsoon precipitation was reduced considerably in the warm pool region (De Deckker et al., 2002), from the Sundaland (Bird et al., 2005) to the Great Australian Bank (Fig. 12), a Warm Pool (Fig. 12), a source of latent heat and moisture for the global atmospheric circulation, and a key area in energy transfer between the Pacific and Indian oceans, particularly during ENSO events. During the LGM, with the 120 m of global sea level drop, the Sunda Shelf (“Great Asian Bank”) and the Sahul Shelf (“Great Australian Bank”) emerged (Fig. 12B,C). Together with the exposed East China Sea Shelf, a total of ~3.900,000 km² sea area emerged as low land, greatly changing the sea/land ratio in the East Asian and Australian monsoon regions (Wang, 1999).

The geographic changes had serious climate consequences. With the decline of sea/land areal ratio and SST in the last glaciation, monsoon precipitation was reduced considerably in the warm pool region (De Deckker et al., 2002), from the Sundaland (Bird et al., 2005) to the Sahul Shelf (Kawamura et al., 2006). The results from the multi-model ensemble of seven models participated in the CMIP5 show during LGM the GM precipitation is reduced by about 10% and the monsoon intensity is generally weakened over most monsoon regions except the Indonesia-Australian monsoon which is strengthened likely due to the changed land-ocean configuration in the vicinity of the Maritime Continent (Yan et al., 2016). Meanwhile, the paleynological proxy data reveal diverse vegetation patterns in this glacial interval. Thus, a savanna corridor in the southern Sundaland hindered biogeographic exchange (Bird et al., 2005; Wurster et al., 2010), whereas tropical rainforest persisted in its northern part (Sun et al., 2000; Wang et al., 2009). Apparently, the regional monsoon circulation is very sensitive to changes in sea/land areal proportion in the tropical region.

Another aspect is the water exchange between the Earth’s surface and interior. The ocean is losing water to the mantle because the subduction of water surpasses the water emission from the mantle. Presumably, the total water mass decreased by 6 to 10% over the entire Phanerozoic since ca. 540 ma ago due to the subduction into the mantle (Wallmann, 2001). However, there is no record supporting large-scale sea-level drop in the geological history. Within the Phanerozoic, for example, the amplitude of sea-level fluctuations are limited to within 200–300 m (Haq et al., 1988; Haq and Schutter, 2008), a small change compared to an average ocean depth of nearly 4 km. Evidently, the proportions of H₂O stored in Earth’s interior and present in the exosphere were established relatively early in Earth history and have changed little since (Hirschmann, 2006). Nevertheless, the question of the water and ocean volume changes on the Earth, and their impact on the monsoons remains open.

5.1.2. Position of land and sea

Around twenty years ago, the effect of land-sea geometry and topography on monsoon circulations began to be studied in climate models (Dirmeyer, 1998; Meehl, 1992, 1994). The geographic position of the continents was found to be crucial to monsoon development. As a general trend of the Phanerozoic, the continents have been moving from South to North. Phanerozoic paleogeography started with the amalgamation of Eastern and Western Gondwana into a super-continent in the SH (Buchan, 2004), and the northward shift since then has resulted in the modern interhemispherically asymmetric patterns with continents disproportionately located in the NH (e.g., Stampfli and Borel, 2002), which explains the NH predominance in the GM response to orbital forcing in the Quaternary records.

The latitude of the continents is critical to the establishment of a distinct seasonality in precipitation and a summer monsoon circulation. As shown by modeling experiments, low-latitude positioning of the continents (south of the westerly storm track) tends to lead to landlocked precipitation maxima, and poleward positioning (near the westerly storm track) tends to lead to a Mediterranean climate. Variations in the meridional extent of land affect the strength of the Hadley cell and the degree of seasonality in the climate over land (Dirmeyer, 1998).

Another significant aspect is the orientation of the ocean. A typical example of a zonally oriented ocean was the Tethys. The Mesozoic Tethys Ocean existed between the continents of Gondwana and Laurasia, extending along the equator. In the early Mesozoic, the Tethys interacts with the Pangea supercontinent, generating a “megamonsoon” as evidenced by the Triassic and Jurassic record in the US (see Wang and Li, 2009, for a brief review). In the Cretaceous, a hypothetical westward-flowing Tethys circums-global current was formed by tectonic deformation (Barron and Peterson, 1989). Although the monsoon was probably weak, local surface currents in the Tethys Seaway might display seasonal reversal during the south Eurasian monsoon months (Bush, 1997).

In contrast, meridionally oriented seaways have repeatedly occurred over geological history, such as the Cretaceous Western Interior Seaway in the United States and the Paleogene Turgai or West Siberian Sea in Russia, which formed a connection with the Arctic basin. The humid climate in West Siberia at 50–55 Ma was speculatively ascribed to a monsoon circulation related to the sea transgression in West Siberia (Akhmetiev et al., 2012). On the other hand, the Paratethys, an epicontinental sea stretching over Eurasia 30 Ma ago, progressively receded during the Miocene, resulting in continentalization of the Asian interior and the enhancement of monsoon circulation (Ramstein et al., 1997). Although a straightforward connection between sea transgressions and monsoons is debatable, it will be of great interest to further explore the climate impacts of meridionally-oriented seaways.

5.1.3. Continent assembly and fragmentation

It is well established that the continents are assembling and dispersing within the Wilson cycle. A Wilson cycle lasts hundreds of
million years and consists of alternation of sea-floor spreading and continent collision, resulting in formation and collapse of supercontinents. The recent two supercontinents, Rodinia and Pangaea, formed ~1000 ma and ~300 ma ago, respectively (Bradley, 2011). As suggested by numerical modeling and geological evidence, the Pangaea supercontinent straddling the Equator gave rise to a global-scale monsoon system, the “Megamonsoon”, with a seasonal reversal of circulation and large-scale migration of the ITCZ over Pangaea (Kutzbach and Gallimore, 1989). Meanwhile, the enhanced global monsoon means also enhanced aridity in non-monsoon regions deep inland, thus a “mega-drought” often co-occurs with a “mega-monsoon”. Regardless of the cyclic nature of continent breakup and assembly, the scatter of continental fragments varies with various Wilson cycles. It was noticed that after the collapse of the last supercontinent, Pangaea, there are presently higher amounts of continental land mass units than after the breakup of previous supercontinents Rodinia or Pannotia (Cogné and Humler, 2008), and its climatic consequences have yet to be explored.

The breakup of Pangaea began at about 180 Ma ago (Early Jurassic) and continues today. The fragmentation of Pangaea and establishment of the modern monsoon sub-systems were discussed in our companion paper (Wang et al., 2014a, “7. Global monsoon at a tectonic timescale”). Since the Hot House/Ice House transition occurred in the late stage of the super-continent breakup in the middle Cenozoic (see “Section 4.2.4”), it’s intriguing to know whether the climate regime change is related to the geographical reorganization. Fig. 13 compares two patterns of land-sea distributions for the late-Cenozoic Ice-House stage, and the Cretaceous Hot-House stage, represented by the modern (upper panel) and the mid-Cretaceous geography (lower panel). The Cretaceous Tethys Seaway provided a low-latitude communication between all of the major ocean basins, and there were several meridional seaways connecting the polar regions with the tropics. Such a land-sea distribution pattern was speculated to be favorable for the Cretaceous warm equable climate, with much lower meridional and ocean-continent temperature gradients than those of today (Hay, 2011). By contrast, the throughflow of the circum-Antarctic seaway in the Late Cenozoic benefits the development of a polar ice-sheet, and the cut-off of the circum-global tropical seaway may have contributed to the establishment of the regional monsoons, such as the East Asian monsoon around the Oligo/Miocene boundary (see Wang et al., 2014a, “7.3.1”).

The scarcity of deep-time paleoclimate data precludes the possibility of dating the establishment of most regional monsoon systems. The recent IODP expeditions to the monsoon regions are promising in providing a deep-sea sediment record to trace back the regional monsoon history (Fig. 1; Table 1). An example is the IODP 354 expedition to the Bay of Bengal that found the transition from carbonates to turbidites around the end of Oligocene, implying a radical change in hydrology of the Indian monsoon region (France-Lanord et al., 2015), probably related to the establishment of the Asian monsoon system. Similar results were obtained from IODP expedition to the Indian Ocean that found the onset of “proto”-South Asian monsoon at 25 Ma which suddenly strengthened at 12.9 Ma (Betzler et al., 2016).

5.1.4. Seaway and monsoon climate

Monsoons are thermally driven phenomena; hence their intensity inherently depends upon temperature fields of the land and ocean. The ocean has two ways to influence the monsoons: either via SST of nearby seas at the regional scale, or via climatic teleconnections from the remote oceans at the global scale. In both cases, the ocean circulation plays a significant role, which, in turn, is controlled by the distribution of connecting seaways. As the SST-monsoon connection was discussed in Chapter 3 on internal interactions, here we examine the role of seaways that control the ocean circulations as one of specific features of the underlying surface.

The opening and closure of a seaway can reorganize an ocean circulation and alter ocean heat transport and moisture fluxes, with a climate significance that can hardly be overstated. As seen in Fig. 13, a circum-polar seaway enhances the role of high-latitude processes in global climate, whereas a low-latitude communication of the global ocean may lead to a warm equable climate. Geological history knows many cases when tectonic-induced opening and closure of oceanic passageways led to the reorganization of ocean circulations. The early Palaeocene formation of the Panama Isthmus and the final closure of the Central American Seaway at about 2.7 Ma cut off the Atlantic-Pacific connection and triggered the intensification of Northern Hemisphere glaciation, resulting in restructuring of the global atmospheric and oceanic circulations (Driscol and Haug, 1998). Meanwhile, the progressive closure of the Seaway led to the Pacific-Caribbean salinity contrast (Steph et al., 2006), exerting a direct impact on the American monsoons.

Another example is the closure of the Indonesian gateway between the Indian and Pacific Ocean. The Indonesian Throughflow transports excess heat and freshwater from the Pacific to the Indian Ocean through the gateway, and its variability is related to the global climate, as well as to the intensity and position of the WPWP (Meyers, 1996). Intensity changes in the Indonesian Throughflow influence ENSO and the Asian and Australian monsoons (Webster et al., 1998; Wajsockic, 2002; Sprintall et al., 2003). The closure of the Indonesian gateway and the formation of the restricted Indonesian Throughflow resulted from the northern movement of the Australian Plate, and its history has been a subject of research for decades (e.g., Kennett et al., 1985). Although deep water exchange between the two oceans has been restricted as early as the Early Miocene (Kuhnt et al., 2004), the closure of the Indonesian Throughflow (ITF) did not occur until ~10 ma ago, giving rise to the formation of the WPWP (Li et al., 2006; Gallagher et al., 2009). As suggested, the northward displacement of New Guinea about 5 Ma ago further restricted the gateway and switched the source of throughflow from warm South Pacific to relatively cold North Pacific waters, which has been suggested to have reduced monsoon rainfall over eastern Africa and influenced the evolution of hominids there 3–4 ma ago (Cane and Molnar, 2001).

The Indonesian gateway also has a profound impact on the Southern Hemisphere. A typical case is the Leeuwin Current off Western Australia, a warm current flowing southwards near the continent’s western coast and contributing greatly to the aridity of the region. It is the only south-flowing eastern boundary current in the SH, and its...

Fig. 13. Paleogeographic comparison of the modern and mid-Cretaceous worlds. Note the different latitudinal positions of the circum-global seaway (Hay, 2011).
volume and intensity are reduced in glacial periods because of the reductions in Indonesian Throughflow and an increase of the opposing West Australian Current (Spooner et al., 2011). Actually, the present Leeuwin Current was established in the Pliocene when the Indonesian Gateway took on an expression resembling its present form and enabling a full development of Indonesian throughflow (Wykoff et al., 2009). As a result, the tropical area of West Australia around 20°S is characterized by relatively dry glacialal and humid interglacials, contrasting with the typical southern hemisphere climate variability of humid glacial and dry interglacials (Stuut et al., 2014). Obviously, this observation can be explained by the influence of the unusual Leeuwin Current and the close connection between the Australian and Asian monsoons. As simulated in some climate models (see Section 4.2.2), the SST feedback in the Leeuwin Current region can even overwhelm the direct local insolation effect to exert a dominant influence on the Australia monsoon (Liu et al., 2003, 2004).

Another aspect is the zonal SST contrast. The WPWP, the East Pacific cold tongue (EPCT), and the atmospheric Walker circulations form a tightly coupled system. The SST evolution of both the WPWP and EPCT may influence ENSO frequency and intensity. However, the coupled system was formed in response to the closure of gateways, the Indonesian seaway on the West and the Panama seaway on the East. On the basis of paleo-SST records in deep-sea sediments, the WPWP came into existence about 10 Ma ago, as a result of the oceanographic barriers between the Indian and Pacific Oceans (Li et al., 2006; Gallagher et al., 2009), and intensified around 4 Ma (Sato et al., 2008). The formation of the EPCT was dated to ~3 Ma (Ravelo et al., 2004; Wara et al., 2005), with the meridional SST gradient between the equator and subtropics remarkably increased from 4 Ma to 2 Ma (Brierley et al., 2009). The findings of warm climate in the EPCT before 3 Ma has led to a hypothesis of a “permanent El Niño-like state” in the Pliocene before 3 Ma (Philander and Fedorov, 2003; Ravelo et al., 2004), which is challenged by later works (e.g., Lawrence et al., 2006). Unfortunately, there is little work on the historical connection between the zonal SST gradients in the tropical ocean and the monsoon circulation.

5.2. Orography

It is now widely accepted that orography has significant impacts on monsoon circulations. Intensive studies have been carried out to understand the links between Asian monsoon climate and the formation of the Himalayan-Tibetan complexes. As reviewed in our first synthesis paper (Wang et al., 2014a, 7.3.1.), the transition from the zonal to monsoonal climate pattern in Asia occurred around the Oligocene/Miocene boundary, but the causal link between climate and tectonics remains unclear.

Yeh et al. (1957) and Flohn (1957) were the first to find that the Tibetan Plateau acts as an elevated heat source for the atmosphere in summer to drive the Asian monsoon. Later, general circulation models (GCMs) were employed to study the effect of Tibetan Plateau on the Asian monsoon with different levels of uplift (e.g., Kutzbach et al., 1993; Prehl and Kutzbach, 1997). It has become clear that the Tibetan Plateau is a heat sink in winter but a heat source in summer (Fig. 14A,B). Thus, the Tibetan Plateau represents a huge air pump and influences the annual cycle of the global atmospheric circulation and monsoon climate (see Yanai and Wu, 2006, for a review).

This thermal mechanism, however, was recently challenged by an alternative hypothesis of mechanical barrier blocking mechanism. In a climate model the South Asian summer monsoon was found largely unaffected by the removal of the plateau, provided the narrow orography of adjacent mountain ranges was preserved (Boo and Kuang, 2010, 2013; Molnar et al., 2010). The authors hypothesize that orography creates a strong monsoon by serving mainly as a thermal insulator, blocking the cold and dry air from the inland northern Asia from penetrating into South Asia. According to this interpretation, therefore, it is the Himalayas, not the Tibetan Plateau, that represents the essential topographic impact on the Asian monsoon (Fig. 14C; Cane, 2010).

This mechanical barrier mechanism, however, is inconsistent with later studies in other climate models. Using a different AGCM, Wu et al. (2012a) showed that the Asian summer monsoon system is mainly controlled by the thermal forcing induced by the southern slope of the Tibetan Plateau, whereas the orographic mechanical blocking forcing is not essential. Recently, a new series of experiments in another state-of-the-art CGCM are also inconsistent with the barrier blocking mechanism. Instead, a candle-like latent heating in the middle troposphere was found along the southern edge of the Tibetan Plateau in the experiments (Fig. 14D), whereas the candle heating disappears in the no-Tibetan Plateau experiment (Chen et al., 2014). This candle heating is mainly latent heating released by the deep convection along the southern slope of the Tibetan Plateau, and is triggered by sensible heating induced over the southern slope, largely consistent with Wu et al. (2012a).

Fig. 14. Diagrams showing different views on climate effect of the Himalayan-Tibetan complexes. A-B. Thermal mechanism (Kutzbach et al., 1993), showing monsoon impact of summer heating of high plateau on large-scale atmospheric circulation and precipitation (A. side view, B. top view). B. Barrier blocking mechanism, showing the role of moisture convection as the main driver of the South Asian monsoon (Cane, 2010). C. Candle heating mechanism (Chen et al., 2014), showing meridional cross section of total diabatic heating (K day^-1) averaged between 85° and 95°E for full Tibetan Plateau experiment.
Furthermore, linking the Iranian Plateau to the Tibetan Plateau in modeling substantially reduces the precipitation over Africa and increases the precipitation over the Arabian Sea and the northern Indian subcontinent, effectively contributing to the development of the South Asian summer monsoon (Wu et al., 2012b).

Despite numerous efforts over decades, the mechanisms underlying the Tibetan Plateau-monsoon interaction still remain to be fully understood. There must be, for example, a barrier effect of the Tibetan Plateau on the Asian monsoon, especially on the South Asia summer monsoon, but its mechanism requires a better understanding. A major question is the different effect of the plateau uplift on different regional monsoon systems. Indeed, modeling studies show that the effect of Himalayas is confined mainly to the South Asia summer monsoon over India, while the orographic effect of the Tibetan Plateau is dominated to the East Asian summer monsoon over eastern China (Liu and Yin, 2002; Chen et al., 2014). As shown by a new modeling study, tropical monsoon climates should exist south of 20°N in South and Southeast Asia before the uplift of the Tibetan Plateau, due to the seasonal shifts of the ITCZ, while the East Asian monsoon was entirely absent in the extratropics (Liu et al., 2015).

The plateau uplift and monsoon evolution have long been regarded as a canonical example of a tectonic-climate linkage, yet this linkage remains unproven because of data scarcity. The Indian-Eurasian collision began approximately 55 Ma ago, but little is known about what happened since then due to uplift of the Himalayan—Tibetan complex, with particular uncertainty in its effect on monsoon location and intensity. Among the major pulses of the Himalayan—Tibetan complex uplift, convincing evidence exists only around the Oligocene-Miocene boundary at 25–20 Ma when the EAM intensified and desertification started in inland Asia in association with the uplift of the northern Tibet (Guo, 2015; Tada et al., 2016). Therefore, the effect of Tibetan Plateau uplift might differ significantly among the different regional monsoons. It remains a challenge to verify these modeling results with geological records.

The Cenozoic uplift-climate connection has been a research focus in Earth science for at least 30 years (see Ruddiman et al., 1997 for a review). The role of the global mountain uplift (MU), which occurred during the middle and late Cenozoic, in modulating GM precipitation was recently investigated by Lee et al. (2015) using the Meteorological Research Institute atmosphere-ocean coupled model experiments. The annual mean precipitation over the entire globe remains about the same from the no-mountain experiment (MU0) to the realistic MU (MU1). However, over the Asian-Australian monsoon region and Americas precipitation increases by about 16% and 9%, respectively. In the Northern Hemisphere, the rainy seasons are lengthened as a result of the earlier onset of the summer monsoon. The East Asian monsoon is a unique consequence of the MU, while other monsoons are attributed primarily to the land-sea distribution. The strength of the global monsoon is shown to be substantially affected by MU. In particular, the MU has a greater impact during transition seasons than solstice seasons.

In general, orogeny is inherent to all of the geological times, but significantly intensifies with the collision of continental plates during the assembling stage in the Wilson cycle. Probably the most remarkable case was the Central Pangaean Mountain chain in the Late Paleozoic around 300 Ma ago, of possibly ca. 1000 km wide and 5000–7000 km long (Ziegler et al., 1997). Mid-Carboniferous collision of Gondwana-land and Laurasia resulted in construction of this E-W oriented equatorial mountain chain, and the early modeling results revealed its amplification effect on monsoon precipitation (Otto-Blesner, 1998), causing a transition from humid to arid climate (Parrish, 1993). However, later studies show that the rates of mountain uplift and erosion (~10^7 years) were much too slow to explain most of the relatively rapid (10^5-10^7 years) climate changes, and the influence of orogeny was smaller than that of other factors, such as atmospheric pCO2 and continental glaciation (Tabor and Poulsen, 2008). Consequently, regional uplift/erosion of the mountains has a secondary effect on tropical precipitation (Pryor and Poulsen, 2008). Given the limited paleo-climate datasets available, the hypothesis of a tectonic-climate connection will remain an open question for a long time, especially in the deep time geology.

5.3. Earth’s albedo

5.3.1. Albedo and monsoon

Albedo affects earth’s climate and has naturally strong impacts on the global monsoon. From a modern perspective, the effects of albedo can be addressed through combined analyses of observation data and climate models. However, it remains difficult to explore the roles of albedo in a paleo-perspective, mainly because of the uncertainty in reconstructing past variability in the relevant fields (snow cover, sea ice, ice sheets, vegetation and soil types etc.). Moreover, the effects of albedo on paleoclimatic timescales are mostly considered in an integrated way, for example, in exploring the ice-sheets impacts on the monsoon circulations (Cheng et al., 2009).

The influence of albedo on the Asian monsoon is among the most explored. A comprehensive review with regards to the underlying land surface on the modern Asian monsoon was given in Wang (2006). More than a century ago, Blanford (1884) and Walker (1910) documented an inverse relation between the winter snow cover over the Tibetan Plateau and the subsequent Indian summer monsoon rainfall, that is, an anomalously higher snow cover corresponds to a shortfall of monsoon rainfall in the following summer. Physically, it is hypothesized that the higher snow cover reduces the Tibetan Plateau heat source in the following summer, due to reduced absorption of solar insolation, increased snow melting, and in turn, a cooler land surface and weaker Indian monsoon. An inverse relation between the winter snow cover and the subsequent Indian summer monsoon rainfall was generally supported by satellite data (e.g. Parthasarathy and Yang, 1995; Yang, 1996) and climate models (e.g. Shen et al., 1998; Barai and Marx, 2000; Becker et al., 2001). Barai and Shukla (1999) did not find a significant relation between the Himalayan snow cover and subsequent monsoon rainfall. Some studies (e.g. Ueda et al., 2003) suggested a rather weak dynamic link between the snow cover and Asian summer monsoon.

The East Asian monsoon is somewhat different from the south Asian monsoon (Tao and Chen, 1987), as the snow-rainfall connection in China is rather complex. A positive correlation between Tibetan winter snow and subsequent summer precipitation over the middle and lower reaches of the Yangtze River may exist (Chen et al., 2000; Wu and Qian, 2003). A close relationship exists between the interdecadal increase of snow depth over the Tibetan Plateau during March–April, enhanced summer rainfall over the Yangtze River valley, and reduced rainfall in the southeast coast of China and the Indochina peninsula (Zhang et al., 2004). It has been proposed that the increased albedo leads to a reduction of the heat source over the Tibetan Plateau in the following summer, through increased albedo and snow melting, thus reducing the zonal land-sea thermal contrast, and in turn the EASM (Ding and Chan, 2005). While albedo variability over the Tibetan Plateau on geological time scales likely has a long-term influence on both global climate and local atmospheric circulation, relatively little has been documented. Numerical modeling, however, shows that the changes in snow cover over the Tibetan Plateau, for example, play a critical role in reinforcing the 9.5 kyr BP monsoon in India during spring (Marzin and Braconnot, 2009b).

The impact of sea ice on the monsoon systems are mostly focused on modern climate processes, as has been thoroughly addressed (e.g. Wang, 2006). The sea ice also plays an important role in changing global ITCZ position and Hadley circulation, and thus the GM (Chiang and Bitz, 2005). From a geological perspective, Guo et al. (2004) demonstrated that increased sea-ice extent at the northern high-latitudes during the late Pliocene and early Quaternary enhanced the inland aridity and winter monsoon in Asia. In contrast, the summer monsoon
was largely weakened during the same interval. During the marine oxygen isotope stage 13 (~500 ka ago), increased sea-ice extent over the Southern Ocean and decreased sea-ice extent in the circum-Arctic region was associated with strengthened Asian, Indian and African summer monsoon (Guo et al., 2009).

The impact of ice sheets on the monsoon circulations is likely to arise from their influences on both albedo and topography. Yin et al. (2009) explored these influences individually, suggesting that albedo and topography of ice-sheets may have distinct effects on the Asian monsoon. In particular, the orographic effect was found to generate a wave train that reinforces atmospheric vertical motion and precipitation over East China.

5.3.2. Vegetation feedback

As mentioned above (Section 4.2.3), the GM interacts with the land ecosystem through albedo and evapo-transpiration feedbacks. The vegetation-monsoon feedback has been studied most intensively in North Africa. Early modelling studies suggest that a decrease in vegetation increases surface albedo and reduces energy absorption over land, reducing convection activity and monsoon rainfall. The reduced rainfall further reduces vegetation, forming a positive feedback (Clausen, 1997; Clausen and Gaylor, 1997). A decrease in vegetation also reduces evapotranspiration, which reduces moisture supply and, in turn, monsoon rainfall, enhancing the positive feedback. This positive feedback may generate abrupt changes of the Africa monsoon (Clausen et al., 1999). Vegetation may also impact the monsoon remotely, for example, with cooling in Siberia (tundra replacing trees) weakening the Asian summer monsoon (Crucifix and Hewitt, 2005). In general, the forcing of the atmospheric circulation is primarily due to the change of vegetation cover in Siberia and Tibet and a strong feedback loop involved; reduced vegetation increases snow cover, which increases the surface albedo, and therefore leads to a surface cooling. These feedbacks and the resultant change in the large-scale temperature gradient cause a southward shift of the tropical front, which reduces the tropical easterly jet and, because of momentum conservation, reduces the low-level monsoon flow.

It should be noted, however, that vegetation feedbacks with the monsoons are complex (Liu et al., 2010) and the quantification of their effects in the real world remains challenging (Liu et al., 2006a). Some models exhibit positive regional vegetation feedbacks, while in others, negative feedbacks dominate (Liu et al., 2006b; Braconnot et al., 2007b). Furthermore, how these feedbacks affect the GM in different regions has not been studied extensively. Yet, these feedbacks can be important not only for understanding past, but also present and future, GM changes. An outstanding case is the Sahel drought in the last three decades of the 20th century, highlighting the need for a better understanding of the African monsoon dynamics. During the “urging the thirtie about 5000 years ago, the dust reduction strengthens the vegetation, highlighting the need for a better understanding of the African monsoon dynamics.r than the vegetation-change case only. Thus, changes in Saharan dust loadings are not less important than those of vegetation (Pausata et al., 2016). Therefore, much study is still needed to understand monsoon feedbacks with the ocean and land.

6. Outstanding issues

The rapid growth in the number of publications on global and regional monsoons in recent years has been accompanied by an increased debate on several issues. In many cases, the traditional wisdom about monsoon drivers and effects has come into question with new observational and modeling results. In this chapter, we discuss several such key issues, with an emphasis on time scales beyond the instrumental record.

6.1. What is a monsoon: wind or rain?

With its extremely broad use in climate science, the term monsoon has become somewhat ambiguous across the modern and paleoclimate communities. The definition varies from being a strict measure of seasonal wind reversal at one extreme, to a sign of enhanced precipitation at the other. However, if monsoons are defined by seasonal reversals of the zonal wind field, the East Asian monsoons and some portion of American monsoons would not be defined as monsoonal, while the Mediterranean and Arctic Ocean regions would be monsoonal if just annual reversal of prevailing surface winds is considered (see Fig. 1a in An et al., 2015). In contrast, if only total precipitation is considered in defining a monsoon index without seasonality, all humid climates would be classified as monsoonal.

A serious question arises from these ambiguities: why is monsoon research needed for addressing ecological and socioeconomic concerns? Until the 19th century, an understanding of monsoon winds was crucial for navigation. Afterwards, when steamboats replaced sailboats, monsoon precipitation became the major social concern because of the impacts associated with floods and droughts. Today, monsoons are recognized as being central to hydrological cycle, with broad relevance for both modern and paleo-climatological research. Consequently, a purely academic approach to monsoon definition has not been generally accepted and goes beyond the scope of the present synthesis. Given this, and in an effort to focus on societal needs, monsoon climate is defined here as being characterized by an annual reversal of surface winds coupled with the contrast between a rainy summer and a dry winter. However, this dual nature of the monsoon can also generate misconceptions, as the two features, wind and rain, are not always satisfied simultaneously. The modern monsoon wind and precipitation domains are dynamically consistent on a regional scale across a monsoon domain, but they do not have to mutually coincide within the area, as illustrated in Fig. 2. The potential discrepancy is further amplified in paleo-monsoon studies when proxy records are used (see Wang et al., 2014a, “3.2 Proxies for local and global monsoons”). This is simply because a thermal contrast alone can produce pressure gradients that drive winds, whereas precipitation depends also on moisture availability as well as other conditions such as the convergence of low level wind fields. The contrast becomes important and cannot be ignored in some contexts, such as if the monsoon response to orbital forcing is discussed, as wind-based paleo-records, such as those influenced by upwelling-induced productivity, might be inconsistent with those based on rain-related δ18O. In view of this possible discrepancy, the scientific community has tried to collect and compare geological sequences of both rain and wind proxies for reconstruction, for example, of the Indian monsoon history. Nearly 30 years ago, deep-sea sediment records over the past 14 ma were recovered from the Western Arabian Sea by ODP Leg 117, and a monsoon-driven upwelling history of the Plio-Pleistocene was reconstructed using basically foraminifers and eolian dust (Kroon, 1991; Clemens et al., 1996), both being indicative of Indian monsoon winds. Recently, IODP Expedition 353 in the Bay of Bengal was targeted at revealing the history of the “Indian monsoon rainfall”, on the basis of surface salinity changes. The new data will yield opportunities to explore the extent to which basin-scale monsoon winds and continental precipitation are coupled over a range of time and space scales (Clemens et al., 2015). They will also provide a critical test of speleothem records from the region that show the growth phase (or pluvial periods) concurred broadly with high NH summer insolation (e.g., Fleitmann et al., 2011).

At this point we caution against the abuse of the term “monsoon”. It is quite often that any climate change in low-latitude regions is attributed to “monsoon variations” without sufficient reasoning. Neither all winds nor all rains are monsoonal, even in the modern monsoon regions. In the modern Sea of Japan, for example, the eolian dust flux is estimated at about 0.2 to 4.5 g/cm²/ka, carried by the westerly jet, but the winter monsoon wind can also contribute to this dust transport. Thus it is essential to distinguish between the two winds on the basis of
analysis of the dust source regions (Nagashima et al., 2007). On the other hand, the precipitation pattern in the Asian hinterland may very much depend on the relationship between the midlatitude westerly and the monsoon, and it is thus crucial to discriminate between the two regimes in the paleo-record interpretation (Chen et al., 2008, 2012a). Indeed, a recent study also shows that a seemingly consistent “summer monsoon” signal in northwestern China is caused by changes in the westerly storm track in winter (Kutzbach et al., 2013). Given the fact that changes in modern precipitation within a given monsoon domain might differ depending on location (e.g., Liu et al., 2014), it should also be cautioned against using a single local rainfall record in reconstructing the history of a regional monsoon. It would be misleading to attribute enhanced humidity to the summer monsoon without convincing physical arguments for certain geological times such as the Eocene when the entire world enjoyed relatively high humidity.

The same problem applies to the GM. An intriguing question is which climate regime benefits the monsoon circulation: Hot-house or Ice-house? The answer apparently depends on the monsoon definition. In an ice-free Hot-house world, such as in the Late Cretaceous, the reduced thermal contrast and the sluggish circulation does not drive intense winds, whereas high temperatures increase water vapor content according to the Clausius–Clapeyron relation, and hence increase precipitation (Hay, 2008). By contrast, development of ice caps in the polar regions drive migrations of the ITCZ in glacial cycles, and thereby enhances the monsoon response to orbital forcing (Peyser and Poulsen, 2008), even though water vapor concentrations should be lower than in a Hot-house world. Consequently, in a Hot-house world orbital-scale climate variations are dominated by low-latitude precession and eccentricity forcing periodicities (see “Section 4.2.4 Hothouse and Ice-house world”), whereas in an Ice-house world the variation amplitudes of orbital cycles in climate changes are amplified. This may also explain why the 400 ka periodicity is most prominent in the Oligocene δ18O records (H. Pälike, personal communication), at the transit time when the world climate entered into the Ice-house regime.

6.2. Stable isotopes and hydrological cycle

Among the numerous proxies used in paleo-monsoon reconstructions, the oxygen isotope is by far the most effective, as it has been extensively exploited as both a marine and terrestrial climate archive. When C. Emiliani (1955) pioneered the use of δ18O in foraminiferal tests to study the late Quaternary climate in the 1950s, δ18O was interpreted mainly as a record of seawater temperature. Later, in the later 1960s, N. Shackleton found that δ18O signals over that interval are instead primarily an indicator of changing ice sheet volume, resulting from isotopic fractionation with the transfer between liquid and solid phase of H2O. In the past two decades, δ18O analysis has become widely applied to speleothem carbonate and ice-core air bubbles as a proxy for hydrological cycle intensity at low latitudes, and particularly, monsoon precipitation. This is a logical extension of the use of δ18O as a proxy, because isotopic fractionation occurs also with phase transitions between liquid and gaseous H2O, which is, however, much more complicated than the transition between ice and water.

6.2.1. Speleothem δ18O as monsoon proxy

One of the most exciting advances in Quaternary paleoclimatology over recent years is the synthesis of high-resolution speleothem δ18O records from the Asian and the South American monsoon regions, which reveal the predominance of precessional forcing and an anti-phased relationship between the two hemispheres (Cheng et al., 2012a). A striking finding is the correlation of the δ18O amplitude variations between speleothem records from China and atmospheric O2 records from the Vostok ice core (Petit et al., 1999). The likely mechanism behind the covariability of the two δ18O records is the change in the average isotopic composition of meteoric precipitation and relative humidity driven by the hydrological cycle (Wang et al., 2008; Severinghaus et al., 2009). The revelation that the orbital-band δ18O response is dominated by ~20 ka changes arising from precession is in close agreement with Kutzbach’s orbital monsoon hypothesis that tropical monsoons vary in response directly to the intensity of summer (July) solar insolation (Kutzbach, 1981).

The orbital monsoon mechanism, however, is in conflict with the ‘summer monsoon factor’ record in the Indian Ocean based on multiple marine proxies, which show relatively little power at precession (Clemens and Prell, 2003). In addition, the Indian Ocean record suggests a more complicated response tied also substantially to latent heat transfer from the SH and includes a significant phase lag of 8 ka (Clemens and Prell, 2003) or 11 ka (Caley et al., 2011) relative to June insolation at the precessional band, as compared to only ~3 ka or 30° phase lag in the speleothem case. This led Ruddiman (2006) to suggest that the marine proxies may not be tightly coupled to the summer monsoon, but rather may respond to other orbital timescale influences, such as the changes in ocean nutrients related to ice-volume cycles or to the strength of the winter monsoon. This viewpoint has aroused intense debates. At issue are the monsoon proxies: speleothem δ18O vs “summer monsoon factor”, which serves a real proxy for the summer monsoon? This issue requires a better understanding of the nature of speleothem δ18O as well as of monsoon dynamics.

6.2.2. Understanding δ18O in hydrological cycle

The speleothem δ18O signatures can be influenced by multiple factors. Robust replications among various speleothem δ18O records from different caves over a broad Asian region suggest that the records must chiefly represent changes in precipitation δ18O values of regional extent (e.g., Cheng et al., 2006). Nevertheless, the climatic interpretation of δ18O records, especially from East Asia, has become a subject of considerable debate. Variability in the speleothem δ18O record from East Asia has been attributed to various factors, such as winter temperature forcing (Clemens et al., 2010), changes in the vapor source (e.g. Indian vs Pacific Ocean, Maher, 2008), and the response of upstream moisture sources (LeGrande and Schmidt, 2009; Pausata et al., 2011a; Lewis et al., 2010; Liu et al., 2014; Thomas et al., 2016). Accordingly, an urgent question is how to correctly use the δ18O signature to reconstruct the hydrological cycle.

Oxygen isotope records have been frequently used to document regional hydrological changes, such as moisture sources, seasonal precipitation patterns and surface air temperature. Stable isotopes have been recognized as key tracers for tracking moisture sources in air masses (e.g., Noone and Sturm, 2010). Based on modern observations of the seasonal cycle and the spatial distribution of precipitation δ18O, lower precipitation values of δ18O tend to correspond to enhanced precipitation at low latitudes, particularly in certain monsoon regions, while at high latitudes they correspond to cooler temperatures. This difference is conventionally attributed to the “amount” and “temperature” effects, respectively (Dansgaard, 1964; Gat, 1996; Fricke and O’Neill, 1999). In low latitudes, the negative correlation between rainfall amount and its δ18O, or the “amount effect”, appears to be dominant over oceanic ITCZ regions (Dansgaard, 1964), while it becomes rather complicated over monsoonal land regions (e.g. Lachniet, 2009).

In the study of paleomonsoon changes, δ18O of stalagmites is widely used by the community as an indicator of monsoon variability. The δ18O signals were initially interpreted as reflecting the summer versus winter precipitation in the East Asian monsoon region (e.g. Wang et al., 2001b) or rainfall amount in the South American monsoon region (e.g. Cruz et al., 2005), and hence they were interpreted as being indicative of monsoon intensity to first order in both cases. In recent years, many other factors must have recognized as having their imprints on stalagmite δ18O data as well, such as glacial-interglacial seawater δ18O variations, temperature, the distance of moisture transport, sea surface temperature of moisture sources, the proportion of water vapor derived from different sources, and many potential local effects (e.g. Lachniet, 2009). Instead of quantifying all these factors, which is impractical at
the present stage, the replication test among different speleothem $\delta^{18}O$ records from different caves was used as a critical step towards achieving robust precipitation $\delta^{18}O$ signals at large regional scales (Cheng et al., 2012b). Some excellent examples come from replications among the Hulu, Dongge and Sanbao records (Yuan et al., 2004; Wang et al., 2008) from the East Asian monsoon domain, different records from Botuverá cave (Cruz et al., 2005; Wang et al., 2007a, 2007b) in the South American monsoon domain, and the Fort Stanton and Bells records (Asmerom et al., 2010; Wagner et al., 2010) from the southwestern North America domain. In most cases, speleothem records can be broadly interpreted as being indicative of precipitation $\delta^{18}O$ (Cheng et al., 2012b). Therefore, an essential question remains regarding how to link precipitation $\delta^{18}O$ documented by speleothem records to climate variables, such as precipitation amount and monsoon intensity. This has been a persistent challenge and one of central debates for the community during the past decade (e.g. Clemens et al., 2010).

One of possible approaches to interpret the cave data is to compare recent speleothem $\delta^{18}O$ data directly to historical and/or instrumental records to assess their possible links to paleoclimate (e.g. Tan et al., 2011; Sinha et al., submitted for publication). However, it is generally unrealistic to extend such modern relationships to climate variations on longer timescales. On the other hand, Yuan et al. (2004) invoked Rayleigh fractionation to show that changes in the isotope fraction of water vapor rained out between tropical sources and the cave site could account for the observed variability in cave records. In other words, lower $\delta^{18}O$ implies higher spatially integrated monsoon rainfall between the tropical monsoon sources and the cave site. Most modeling studies are consistent with this idea in some fashion, although they generally refer to the process as ‘upstream depletion of oxygen isotope’ (e.g., Pausata et al., 2011a; Liu et al., 2014).

Many other efforts have also been made over the past decade in order to better understand the precipitation $\delta^{18}O$ proxy record. Among them, the most important progress is perhaps the recent development in isotope-enabled model simulations, which shed new light on interpretation of precipitation $\delta^{18}O$ changes across a wide range of timescales as seen in speleothem records, as well as on the accompanying climate variations including wind field and rain amount.

Recently, a set of isotope enabled modeling studies has shown rather consistent spatiotemporal variations in precipitation $\delta^{18}O$ values in comparison with speleothem records, both on both millennial (Pausata et al., 2011a, 2011b; Lewis et al., 2010; Liu et al., 2014) and orbital (e.g. Schmidt et al., 2007; LeGrande and Schmidt, 2009; Liu et al., 2014; Battisti et al., 2014) timescales. For instance, both model and speleothem studies resolve the zonal precipitation dipole pattern on precession timescales in South America (Cheng et al., 2013; Battisti et al., 2014), and the general negative correlation between monsoon (both Indian and East Asian monsoons) intensity (also wind speed) and overall precipitation $\delta^{18}O$ (Liu et al., 2014). However, the relationship between precipitation $\delta^{18}O$ and local rainfall amount is complex in the models. It appears that the precipitation $\delta^{18}O$ exhibits a broadly similar pattern that is coherent on the continental scale, indicating overall precipitation amount on a supra-regional-scale rather than local-level, in a monsoon domain. On the other hand, the amount of local precipitation within a same monsoon system often displays considerable spatial disparities varying dynamically on various temporal-scales. Therefore, the precipitation $\delta^{18}O$ record may correspond to upstream rather than merely the local rainfall amount in many cases, such as in southeastern China (e.g., Li et al., 2013; Liu et al., 2014; Battisti et al., 2014) of the East Asian monsoon domain. As such, lower $\delta^{18}O$ generally implies higher spatially integrated monsoon rainfall between the tropical monsoon sources and the cave site and/or higher summer monsoon rainfall in the cave region and vice versa (Cheng et al., 2016).

6.2.3. Phase relations between forcing and monsoon response

One crucial issue in the speleothem $\delta^{18}O$ and “summer monsoon factor” debate is the phase lag of the monsoon response to precessional forcing. Clemens and Prell (2007) argue that the 8 ka precessional lag is caused by the latent heat export from the SH Indian Ocean as well as glacial boundary conditions. Sensitivity experiments with southern insolation alone do show a significant impact of Southern Ocean insolation on the Indian monsoon (Liu et al., 2006c). However, realistic orbital forcing induces insolation changes in both hemispheres, with the NH forcing favoring an in-phase response and the SH forcing favoring an out-of-phase response. In addition, modeling results do not show a significant impact on the Indian monsoon intensity by the latent heat export from the SH Indian Ocean (Kathayat et al., 2016).

Recent climate model simulations of the evolutionary response of global summer monsoons to orbital forcing over past precessional cycles show a strong and positive response of northern (southern) summer monsoon precipitation to northern (southern) summer insolation forcing, and suggests an approximately in-phase relationship between summer monsoon proxies and summer insolation (Kutzbach et al., 2008), largely consistent with the response in Holocene experiments (Fig. 10). These findings are also supported generally by observations (e.g., Ziegler et al., 2010; Cai et al., 2015).

Fig. 15 shows the amplitudes and phases of the responses of JJA and DJF temperature and precipitation to precessional forcing from a 480,000-year simulation in a coupled climate model (Chen et al., 2010a). Overall, the surface temperature response exhibits a global pattern in both boreal summer (JJA) and winter (DJF), with a maximum magnitude in Siberia (Fig. 15a,c). Furthermore, the phase diagram shows that the JJA temperature response, especially over land, is generally in phase with the precession minimum over the globe, corresponding to a warming of NH summer and SH winter (Fig. 5-1a, opposite for DJF in Fig. 5-1-c). Temperature exhibits a small lead of ~ 40° over the ocean relative to the precession forcing because of the thermal inertia of the surface ocean (Kutzbach et al., 2008).

The GM rainfall response (Fig. 15b,d) exhibits a more complex pattern than does temperature. In JJA, the precipitation response exhibits its largest amplitude in the Africa-Asia monsoon band, with a second maximum in Central America (Fig. 15b). Rainfall enhancement is in phase with the precession minimum in the center of the NH monsoonal rainfall increase band, while being out-of-phase near the rainfall increase center, south of the North African monsoon, in eastern China, and in North America. The in-phase response in the central monsoon region is consistent with the driving of the local summer insolation, consistent with the Kutzbach mechanism, while the out-of-phase region is associated with the northward migration of the Africa monsoon rain belt, and the compensating descending motion around the East Asian and North American monsoons (the opposite occurs in DJF for the SH monsoons in Fig. 15d). Overall, current models have not produced the 8 ka–11 ka lag (or its nearly anti-phase relation) of northern monsoon to summer insolation, as proposed by Clemens and Prell (1990, 2003) and Caley et al. (2011). Further studies are needed on both the model and data sides to understand this discrepancy.

6.3. Role of extra-tropical processes

The debate around speleothem $\delta^{18}O$ partially comes down to the competing views on orbital forcings of monsoon variability: Does the monsoon respond directly to the low-latitude insolation, or indirectly through the ice-sheet dominated high-latitude processes? As we discussed already, the speleothem $\delta^{18}O$ curves follow summer insolation at 30°N or 30°S in general (see Figs. 14–15 in Wang et al., 2014a) and differ substantially from the marine $\delta^{18}O$ curves which represent the ice volume changes. However, speleothem $\delta^{18}O$ does not simply record the signal of low-latitude insolation, but is also influenced by high-latitude processes, especially the ice sheet changes.

6.3.1. Ice-sheet vs monsoon signal in $\delta^{18}O$ record

The dominant precessional forcing apparent in the speleothem $\delta^{18}O$ record should not exclude the ice-sheet/CO₂ influence in Quaternary
Speleothem δ18O records bear the imprint of high-latitude climates, including ice-sheet and sea-ice changes, and especially the AMOC. By correlating the Asian stalagmite δ18O data and northern high-latitude proxies, Cheng et al. (2009, 2016) proposed a chain of events triggered by NH summer insolation from the North Atlantic meltwater to the SH termination. Because stalagmite δ18O reflects changes in the overall Asian monsoon activity, these correlations undoubtedly hint at a strong link between monsoon and ice-sheet changes. The ice-induced North Atlantic circulation changes have also been invoked to explain millennial-scale monsoon events (Wang et al., 2008; Cheng et al., 2016). Both modeling results and geological data show that the response to an AMOC collapse includes a southward shift of the ITCZ in both the Atlantic and the Pacific, and a weakening of the Indian and East Asian summer monsoons (Zhang and Delworth, 2005; Lewis et al., 2010; Otto-Bliesner et al., 2014; Wen et al., 2016).

It remains puzzling how to best diagnose East Asian monsoon variations on orbital-scales given the fact that different proxy records are sensitive to different aspects of the monsoon system. Particularly in the Late Quaternary, the Chinese loess records (e.g., magnetic susceptibility and ratio of citrate-bicarbonate-dithionite extractable Fe2O3 versus total Fe2O3, i.e. Fed/Fet, both are related to summer monsoon) show strong signals of 100 ka periodicity (Fig. 16a), while speleothem records from eastern China exhibit a robust and dominant periodicity of ~20 ka (Fig. 16b) with the 100 ka and 40 ka cycles being nearly absent. As the loess Fed/Fet curve partially matches the benthic foraminiferal δ18O (a global ice volume proxy), whereas the stalagmite δ18O curve is dominated by precession cycles, they were used to claim a dominant role of ice-volume and insolation respectively, in driving orbital variations of the monsoon. This paradox has puzzled our community for the past decade. A recently proposed interpretation relies on geographic difference. On the basis of new loess proxy records and model simulations, Sun et al. (2015) suggest that insolation, ice, and CO2 have distinct impacts on summer precipitation changes in East Asia, whereas their relative impacts exhibit spatial structure, with a relatively stronger insolation effect in south China and a dominant ice/C02 influence in north China. The paradox remains, although the above ‘geographic hypothesis’ can partially explain the different observations. This is because speleothem records show persistent and dominant precession cycles, even
from areas very near the Chinese Loess Plateau, such as those from Gansu, Shanxi and Beijing areas (P.Z. Zhang, Z.G. Rao, and X.L. Li, personal communications, 2015). Alternatively, as demonstrated by Sun et al. (2015) and Li et al. (2015), some proxy records of the East Asian summer and/or winter monsoons, such as the inorganic carbonate $\delta^{13}$C from Gulang/Jingyuan loess sections and grain size respectively, do reveal a coexistence of distinct 100 ka, 40 ka and 20 ka periods. This is in contrast to previous loess records with an almost pure 100 ka cycle (Fig. 17 D and E). On the other hand, new speleothem records of the Indian summer monsoon from southern China (Xiaobailong cave records, Cai et al., 2015) and northeast India (Dutt et al., 2015) also point to a possible correspondence between interglacial-glacial (\~{}100 ka) cycles with precession periodicity (\~{}20 ka) in Indian monsoon variations (Fig. 17C), partially reconciling the above paradox.

Recent modeling results also support the idea that climate conditions are spatially distinct in terms of interglacial-glacial changes in both precipitation $\delta^{18}$O (Cai et al., 2015) and amount (Cai et al., 2015; Sun et al., 2015). These results may also explain to some degree the observed spatial variation in Chinese speleothem records that show different offsets of the $\delta^{18}$O values between the LGM and present. However, it is still difficult for model simulations to fully explain the glacial-interglacial difference observed between Xiaobailong and Dongge $\delta^{18}$O records: \~{}4\% higher for the former and nearly the same for the later, regardless of the fact that the two caves are in close proximity and share a similar hydro-climatic condition in terms of their moisture source/trajecotry and precipitation $\delta^{18}$O and amount across glacial-interglacial periods (Cai et al., 2015). A comprehensive solution of the paradox will require work to further understand the environmental significance of various geological monsoon proxies, feedback processes in the long-term transient simulations, and inter-comparisons between observations and modeling results. In interpreting the ice sheet effect in observations, it should be kept in mind that ice volume tends to be correlated with global temperature and $CO_2$, and therefore it remains a challenge to distinguish the impact of ice sheets from $CO_2$ in paleoclimate proxy records.

6.3.2. Glacial cycles and monsoons

The above discussion leads to the question how the glacial cycles and monsoon are related. The interaction between glacial cycles and monsoons is a crucial topic for Quaternary paleo-monsoon studies. Key points with regards to this interaction would mainly include (1) the temporal and spatial changes of amplitudes; (2) the characteristic frequency, in particular the dominance of the \~{}100 ka signals versus the \~{}20 ka ones; and (3) the phase relationships with the potential drivers.

Our previous synthesis (Wang et al., 2014a) clearly showed that different monsoon proxies might provide conflicting understandings of monsoon histories, and naturally, lead to different views about the nature of orbital-scale monsoon dynamics. In this case, taking into account information from various geological records and proxies is a critical step toward better understanding both monsoon history and dynamics at the orbital scale, as illustrated in Fig. 17.

In terms of orbital signals, our earlier analyses based on various geological records (Guo et al., 2012; Wang et al., 2014a) yielded the following main insights. First, the \~{}20 ka frequency is typically strong in almost all monsoon geological records, with little geographical dependence. This ubiquity supports the role of low-latitude insolation changes on the monsoon circulation. Second, monsoon changes at this precessional band are nearly anti-phased between hemispheres to
within the accuracy of their chronology. Third, most geological records show the co-existence of 100 ka and 40 ka signals, expressed by a pattern linked to the glacial-interglacial changes. These changes are essentially synchronous in the SH and NH following the Milankovitch rules, with stronger monsoons during interglacial versus glacial periods.

Based on these observations, Guo et al. (2012) proposed that orbital-scale monsoon changes consist of two main integrated components, one derived from the low-latitude summer insolation changes (the insolation component), and another from glacial-interglacial cycles (the glacial-interglacial component). The insolation monsoon component is almost anti-phased between the SH and NH while the glacial-interglacial monsoon component is roughly synchronous, due to the Milankovitch rules. Further evaluation would need to reconcile monsoon cyclicity, amplitude, and phase issues in considering these potential drivers. Studies using both geological data and climate models are of particular value. The asymmetrical nature of changes of hemispheric climates and related internal climate processes (Guo et al., 2009), are worthy of receiving more attention.

The two-component concept of climate changes discussed above applies not only to terrestrial proxy records such as stalagmite isotopes but also to marine records such as foraminiferal δ¹⁸O. Over the global ocean, deep-sea δ¹⁸O records show similar features driven by orbital forcing of glacial-interglacial cycles of ~100 ka, and hence planktic and benthic foraminiferal δ¹⁸O records provide a standard means in late Quaternary stratigraphic correlation known as SPECMAP and LR04 stacks, respectively (Fig. 18B). According to a recent study, however, the planktic δ¹⁸O curves from the monsoon dominated South China Sea differ from the SPECMAP/LR04 standard by an enhanced role of 20 ka precession forcing (Fig. 18A), showing some similarity to the δ¹⁸O records in stalagmite calcite. As similar planktic δ¹⁸O records also occur in some other low-latitude oceans under monsoon influence, it is speculated that the planktic δ¹⁸O records of the South China Sea type are related to the hydrological cycle, most probably through the orbit-driven variations of the global monsoon (Wang et al., 2016). Obviously, the orbit-forced climate changes are related to both ice sheet/CO₂ and monsoon variations and exhibited in a combination of ~100 ka and ~20 ka cycles, both in terrestrial and marine records. In sum, monsoon variations not only passively follow the glacial signals from high-latitudes, but can directly respond to the solar forcing. Mid-Pliocene Asian monsoon intensification (4.2–2.7 Ma), for example, preceded the initiation of NH glaciation, and monsoon-enhanced rock weathering and organic carbon burial might have lowered the contemporary atmospheric CO₂ and trigged the onset of NH glaciation (Zhang et al., 2009; Lunt et al., 2008).

6.3.3. Westerlies and monsoons

The interplay among low-latitude monsoonal circulations, mid-latitude westerlies, and high-latitude ice volume is an intricate yet crucial aspect of climate research, and some of the coupling between high- and low-latitude climate changes have been hypothetically attributed to the role of mid-latitude westerlies. At present, the westerly path over the North Atlantic is largely controlled by the sea-ice extent and the meridional gradient of SST, whereas the westerly jet path over East Asia probably determines the position of the EASM rain band (Molnar et al., 2010; Chiang et al., 2015). The key link between high and low latitude processes is the Tibetan Plateau. Because of its topographic barrier, the 500-hPa westerly jet axis passes to the south of the Himalayas in late fall to spring and then jumps to the north of the Tibetan Plateau in early summer. The emergence of the subtropical summer monsoon frontal rain band is associated with the jump of the westerly jet axis to the north of the Tibetan Plateau (Fig. 19). The subtropical summer monsoon rain band migrates northward and then disappears as a result of the northward migration of the westerly jet axis during summer. Across geological history, changes in the North Atlantic, the Tibetan Plateau and the low-latitude monsoon region can all cause N–S oscillations in the westerly jet and summer monsoon limit (Nagashima et al., 2007, 2011).

Although the tie between the Asian monsoon and midlatitude westerlies is generally accepted, their interrelation and the connecting mechanism remains a contentious subject. Chen et al. (2008, 2010a) proposed that changes in effective moisture transport by westerlies dominated variability in Central Asia during the Holocene and were out-of-phase with variations in the Asian summer monsoon, both at millennial and centennial time scales. In contrast to the humid early Holocene and drier late Holocene in the Asian monsoon region, they found dry early and wet late Holocene in Arid Central Asia, with changes mainly determined by North Atlantic SSTs and high-latitude air temperatures that together influenced the availability, amount and transport of water vapor. However, an opposite trend was observed in the same region in other proxies, such as for example in δ¹⁸O records of stalagmites from Kesang Cave, Xinjiang, China, which exhibit a processiona growth with abrupt inceptions of low δ¹⁸O records speleothem growth at times of high NH summer insolation followed by gradual δ¹⁸O increases that track decreases of insolation (Cheng et al., 2012a). The Kesang observations appear to suggest possible incursions of Asian summer monsoon rainfall or related moisture into this and/or adjacent areas during times of high insolation. However, this interpretation invoking the summer monsoon is controversial (e.g., Long et al., 2014). Indeed, most recent simulations in state-of-the-art climate models incorporating water isotopes suggest that the response of the Kesang Cave δ¹⁸O is caused mainly by winter rainfall associated with the storm track as in the Central Asian and Mediterranean climates (Kutzbach et al.,

P.X. Wang et al. 
Earth-Science Reviews 174 (2017) 84–121
2013; Kutzbach, personal communication, 2015). Obviously, further observational data are required to clarify the history of interaction between the westerlies and the monsoons. In particular importance is the comparison of Holocene changes of the Atlantic climate in Western Europe, a typical westerly climate, with central Asian climate.

The Sea of Japan may also be important for understanding monsoon-westerly climate interactions, as it is sensitive to both. The Japan Sea receives eolian dust of about 0.2 to 4.5 g/cm²/ka carried by westerlies, while the input of fresh water and nutrients is provided largely by the EASM. The fluctuating ratio between dust and biogenic deposits has generated an alternation in sediment types that is indicative of the relationship between westerlies and the summer monsoon. The 6000 m long deep-sea sediment cores recovered by the IODP Expedition 346 to the Sea of Japan, are expected to yield the most valuable archives to date in revealing the history of the interplay between westerlies and the monsoon over millions of years (Expedition 346 Scientists, 2014).

The above discussion does not exhaust the role of extra-tropical processes in monsoon variations. For example, a newly proposed factor affecting monsoon variability is obliquity, the tilt of the Earth’s rotational axis. The influence of obliquity on incoming solar radiation at low latitudes is small, much inferior to that of precession. Nevertheless, recent studies suggest that obliquity can change the meridional insolation gradient in the summer hemisphere and the interhemispheric insolation gradient, and thereby exert a direct effect on monsoon climate in the low latitudes (Bosmans et al., 2015; Mohanty et al., 2016). Potentially, obliquity might explain a number of abnormal climate phenomena by its effect on meridional gradient and thus deserves closer attention (Rachmayani et al., 2016).

6.5. Other issues

If the mechanisms of monsoon variability are considered across all the space and time scales, the number of unsolved issues would be too many to be covered in this synthesis. One of the root problems is the hemispheric asymmetry. Today, the global monsoon is dominated by the NH, and global monsoon variations in geological archives reveal the dominance of the NH signal. Since the precession forcing of monsoon climate has opposite phase between hemispheres, the global signal of monsoon variability, such as the Dole effect, follows NH insolation (see Wang et al., 2014a, “3.3 Exploring proxies for global monsoon”). This can be easily explained by the predominantly NH position of the continents in the modern world. The geologic history, however, knows time periods when the continents were located disproportionately in the SH, such as during the Paleozoic. Indeed, Late Paleozoic climate model simulations show well-established monsoon circulations in the SH but less well developed ones in the NH (Peyser and Poulsen, 2008). However, it is much more difficult to find geological evidence to support such SH dominance of the global monsoon, as the phase relationship between monsoon variations and hemispheric insolation requires high-resolution stratigraphy to identify phases at the precession time scale which is, unfortunately, much too difficult for the ancient Paleozoic time.

As mentioned earlier, the recent recognition of the oceanic monsoon has far-reaching implications for paleoclimatology. The monsoon oceans comprise an integral part of the GM, and changes in the ratio between oceanic and continental monsoons may affect the use of δ¹⁸O as a monsoon proxy, particularly in the marginal seas. For example, surface water δ¹⁸O in the South China and Sulu Seas shows heavier values during times of strong summer monsoons, and these unexpected results are tentatively attributed to the redistribution of ¹⁸O depleted rainfall from land to the western Pacific (Oppo and Sun, 2005; Oppo et al., 2007, 2009), i.e. a change in the oceanic/continental monsoon ratio. If confirmed, this will add a new consideration in the interpretation of δ¹⁸O. When the monsoon rain zone shifts between land and sea, the δ¹⁸O value in the terrestrial and surface marine should change, including the Dole effect which is a potential proxy of the global monsoon (see Wang et al., 2014a, “3.3.1”). Until today, our knowledge of isotope fractionation in the entire hydrological cycle is deficient. As a recent survey revealed, soil water that evaporates or is tapped by plants exhibits fractionation that is distinct from that running into streams and recharging groundwater. In other words, the precipitation that supplies groundwater recharge and streamflow contrasts with the water that supplies parts of soil water recharge and plant transpiration (Evaristo et al., 2015; Bowen, 2015). This is just an example of the challenges confronting the use of isotopic monsoon proxies, and its solution awaits an improved understanding of the water cycle and the role of various biological and geological reservoirs.

Another key issue raised by recent findings in the modern monsoon studies is the different impact of insolation vs GHG warming. As explained earlier (see “Section 4.4”), when GHG increases, the...
temperature difference between the surface and the top of the atmosphere declines, making the atmosphere more stable. By contrast, heating by enhanced solar radiation increases the tropospheric lapse rate and strengthens the atmospheric circulation (Liu et al., 2013). It is not easy to distinguish climate changes arising from GHG vs insolation forcing in the modern world. But the geological history knows extreme cases of GHG warming, such as during the Cretaceous when CO₂ concentrations exceeded 2000 ppm (Royer et al., 2007), and there is sufficient evidence to indicate its association with sluggish atmospheric and oceanic circulations (Hay, 2008). A remaining challenge, however, is to understand how GHG warming has affected the paleo-monsoons in a Hot-house world.

7. Concluding remarks

1. The primary driver of the GM is solar insolation, while the specific features in the underlying surface are mainly responsible for the differences among regional monsoon systems. GM variability is forced by cyclic changes of solar activity and Earth's orbits over a broad spectrum of time scales, ranging from the annual cycle to millions of years, as well as to major tectonic deformations of the Earth's surface at geological timescales. GHGs, volcanic activity and aerosols also modify the effective insolation reaching the Earth's surface, and hence contribute to GM variability.

2. The changes of the monsoons in different regions exhibit certain coherent features over the globe such that the concept of a GM provides a useful framework for monsoon study. This global coherence of the regional monsoons can be forced externally by the global forcings, notably the solar forcing and the GHGs at long time scales, and can also be induced internally by climate modes within the coupled climate system, such as ENSO and AMOC, across interannual to millennial time scales (Fig. 5, Table 2).

3. The regional monsoons are shaped by the tectonically determined land-sea distribution and the orography of continents. Together with the oceanic circulations and related SST gradients, these underlying features of geological origin provide the thermal and mechanical control of the atmospheric circulation and the hydrological cycle, giving rise to differences between the regional monsoons. Because of the disproportionately NH position of the continents and the closure of circum-equatorial oceanic circulation in the modern world, the monsoon circulations are more robustly established in the NH and GM intensity responds mainly to NH insolation at orbital time scales.

4. In the modern world, monsoons exist in all continents except Antarctica. As the largest continent, the Asian monsoon is the largest among all the regional systems and includes three different sub-systems, of which the East Asian monsoon reaches the highest latitude (45°N). In the SH, the Australian monsoon is most susceptible to NH influences because of its close Asian connection and oceanic surroundings which suppress the insolation influence. In terms of orbital forcing, the African monsoon exerts the "cleanest" precession signal, apparently because of its equator-straddling position.

5. Monsoons vary either in response to external forcing, or as the result of internal interactions. Generally, and except for the seasons, external forcings work on longer time scales often on a global scale, whereas short-term variability in regional monsoon systems are usually caused by internal feedbacks within the climate system. Meanwhile, the development of polar ice-sheets and the oceanic warm pool can exert a crucial influence on the GM. Debatable is the relative importance of the two factors, as it remains unclear to what extent monsoon variability is attributable to high-latitude cryospheric processes, and how much the GM directly responds to solar insolation in a Hot-house world.

6. Over the past century, the term “monsoon” has been given various, and at times conflicting definitions, in part due to inadequate scientific understanding. For scientific purposes, it is advantageous to clarify the mechanism of the circulation and hydrological cycles rather than get into semantic debates about which climate regimes can be labeled as “monsoonal”. In paleoclimatology, an urgent need is to better recognize the modern concept of monsoon climate and to avoid misuse and abuse of the term.

7. Our understanding of the GM beyond the instrumental time scale depends almost exclusively on proxies. Of paramount importance is the further development of monsoon-relevant proxies, including the further reduction of uncertainty for proxies already in use, and the development of new types of proxies that can help identify certain features of the monsoon more accurately. Many debates in paleo-monsoon studies originate from the use of different proxies, but no proxy is free of ambiguity, as its variations can be caused by multiple factors. "The majority rule" does not fit the use of proxies, and the key is to find out which feature of monsoon variability is represented by a given proxy.

8. Monsoons are central to the global hydrological cycle whose significance remains insufficiently understood, particularly in paleoclimatology. Transitions between the water phases pervade and drive the entire climate system, but the paleo-community in the past has preferably focused on solid/liquid phase transitions, while liquid/gaseous transitions went largely unexplored until very recently, largely due to the difficulties in research approaches. Monsoons were likely key to the entire history of the climate system, and their role in the global climate changes should not be considered secondary to that of the ice-sheets, which exists only in the Ice-House world. Despite its great significance, a reconstruction of the GM history for the deep time geology is hampered by a scarcity of data and the immaturity of proxies. The urgent task in reconstruction is to determine when the modern regional monsoon systems were established, and hopefully the new IODP expeditions to the modern monsoon regions (Fig. 1) will recover proxy records of sufficient and unprecedented length and quality to shed new light on these critical questions.

Both of our syntheses are aimed at setting up the GM concept across time scales, yet the recognition of monsoons as a global system does not rule out their regional heterogeneity, originating mainly from contrasts in their underlying surface. On the basis of land-sea distribution, each continent has at least one subsystem. Within one and the same sub-system, the monsoon response to external forcing may display spatial differences, depending on local conditions. Therefore, the GM as a system comprises a hierarchy of regional and local monsoons with different level of similarity, but all show coherent variability driven by a common solar forcing. The GM concept, therefore, is by no means to replace or depreciate research on the regional monsoons; on the contrary, the global vision of the monsoon system helps to dissect the mechanism and controlling factors of the monsoon variability at various temporal-spatial scales.

Acknowledgements

This work was supported by the PAGES project and written on the basis of two PAGES symposia on global monsoon. We thank all participants of the symposia whose contributions form the basis of the present paper. We also thank anonymous reviewers for their constructive comments that significantly improved the manuscript. Jun Tian is deeply acknowledged for his assistance throughout the Working Group's activities, as well as in the manuscript preparation. P. X. Wang is thankful for the support of NNSFC grant 91128000. B. Wang
acknowledges supports from NNSFC Grant 41371209, 41420104002 and NRF of Korea through Korean Ministry of Education, Science and Technology (MEST, #2011-0021927).

References


