

---

## Downscaling projections of Indian monsoon rainfall using a nonhomogeneous hidden Markov model

Arthur M. Greene<sup>1\*</sup>, Andrew W. Robertson<sup>1</sup>, Padhraic Smyth<sup>2</sup>, Scott Triglia<sup>2</sup>

<sup>1</sup> *International Research Institute for Climate and Society, USA*, <sup>2</sup> *University of California, Irvine, USA*

\*Correspondence to: International Research Institute for Climate and Society, Palisades, NY 10964 USA. E-mail:  
amg@iri.columbia.edu

---

Downscaled rainfall projections for the Indian summer monsoon are generated using a nonhomogeneous hidden Markov model (NHMM) and information from both a dense observational dataset and an ensemble of general circulation models (GCMs). The projections are conditioned on two types of GCM information, corresponding approximately to dynamic and thermodynamic components of precipitation change. These have opposing effects, with a weakening circulation compensating not quite half of the regional precipitation increase that might otherwise be expected. GCM information is taken at the largest spatial scales consistent with regional physics and modeling constraints, while the NHMM produces a disaggregation consistent with the observed fine-scale spatiotemporal variability. Projections are generated using multimodel mean predictors, with intermodel dispersion providing a measure of the uncertainty due to GCM differences. The downscaled simulations exhibit small increases in the number of dry days, in the average length of dry spells, in mean daily intensity and in the exceedance frequency of nearly all daily rainfall percentiles.

Copyright © 0000 Royal Meteorological Society

*Key Words:* precipitation; Indian monsoon; climate projections; hidden Markov model; downscaling

*Received . . .*

*Citation: . . .*

## 1. Introduction

Precipitation fields in general circulation models (GCMs) do not capture detail at the fine spatial scales of interest in many climate risk management applications. GCM parameterizations also tend to produce biased rainfall distributions (Dai 2006). These factors preclude the direct imputation of localized, temporally disaggregated precipitation changes from GCM simulations. On the other hand, GCMs are the most comprehensive tools yet devised for the quantitative characterization of climate change.

Observational data do represent the fine spatial detail lacking in GCM simulations, and are less likely to be biased with respect to spatial patterns or rainfall distributions. However, they are records of past climate, fundamentally mute with respect to the future.

Here these complementary data types are combined in the framework of a nonhomogeneous hidden Markov model (NHMM). The NHMM decomposes the observed spatiotemporal rainfall variability over a network of locations via a small, discrete set of “hidden” states (so termed because they are not directly observed). Each state comprises a set of rainfall probabilities and wet-day distribution functions for all locations in the network; the states proceed on a daily time step following a first-order Markov process. These characteristics enable the NHMM to represent both spatial covariance over the network and persistence of large-scale weather patterns, which can be associated with the states. It is thus well-suited for the representation of daily rainfall in climate regimes that can be characterized in terms of variably persistent large-scale weather patterns, as can the Indian monsoon.

In the “nonhomogeneous” construction utilized here, the state transition matrix is conditioned on an exogenous predictor, giving it an implicit time dependence. This allows for realistic simulation of the seasonal cycle and, more importantly, provides a pathway for the introduction of GCM-based climate change information.

To generate projections of future rainfall, two predictive fields from an ensemble of GCMs are utilized: the regional mean rainfall change and a dynamical index of monsoon circulation strength. The former serves as a constraint on the NHMM-simulated rainfall change, while the latter is used to condition transition probabilities among the hidden states. These large-scale predictors are analogous to the GCM-derived boundary conditions used to constrain regional dynamical models. The NHMM is used to generate stochastic daily rainfall sequences for both the 20<sup>th</sup> and 21<sup>st</sup> centuries, from which statistics of interest are extracted. Information from the GCMs is utilized in the form of multimodel means; individual models then serve to delineate the range of uncertainty owing to differences among GCMs.

Markov and hidden Markov models of varying structure have been utilized previously for rainfall modeling (e.g., Richardson and Wright 1984; Hughes and Guttorp 1994a,b; Bates *et al.* 1998; Hughes *et al.* 1999; Robertson *et al.* 2004; Mehrotra and Sharma 2005; Robertson *et al.* 2006, 2009; Ailliot *et al.* 2009; Gelati *et al.* 2010). A particular novelty in the present study is the implied decomposition of monsoon rainfall change into dynamic and thermodynamic components, as manifest in the exogenous variable chosen to drive the NHMM and the concomitant scaling of daily intensity distributions. This decomposition is of particular relevance in the context of long-term climate change (see, e.g., Held and Soden 2006).

Section 2 places the NHMM in the context of other statistical downscaling methods. Section 3 provides a modeling overview, while Sec. 4 describes both the observational and GCM data utilized. Section 5 covers key theoretical underpinnings of the method as well as details of its present application. The NHMM simulations are considered in Sec. 6; this is followed in Sec. 7 by a discussion of rainfall changes on both fine space and time scales. Sections 8 and 9 present a validation by way of method comparison and a discussion of uncertainties,

respectively. A summary and conclusions are provided in Sec. 10.

## 2. NHMM in context

As a downscaling method the NHMM could be classified as a hybrid between weather-typing and regression-based models (Wilby *et al.* 2002). Weather states are inferred from patterns of observed daily precipitation and modeled as first-order Markov dependent in time, while a large-scale circulation index is utilized as predictor in a multinomial logistic regression, the dependent variables being the probabilities in the Markovian transition matrix.

Beyond this general affinity, the NHMM bears some resemblance to the “WGEN” model described by Richardson and Wright (1984) as adapted by Wilks (1992) for the purposes of downscaling climate change scenarios and eventually (Wilks 1998), to model spatially correlated rainfall over a network of sites. These models include two states (wet and dry), with daily rainfall fit by either a gamma distribution or, in Wilks (1998), a two-component exponential mixture. Such models require specification of a large number of location-specific parameters. The NHMM, by contrast, can include an arbitrary number of states and implicitly accounts for spatial covariance, in the form of the rainfall patterns associated with the states. As in Wilks (1998), daily rainfall distributions are modeled as a two-component exponential mixture.

The NHMM’s hidden states, although nominally abstractions, were shown in an earlier study (Greene *et al.* 2008) to be sensibly associated with large-scale anomalies in the monsoonal circulation. The persistence of these anomalies in time is captured by the Markovian dependence of the day-to-day state transitions.

## 3. Modeling overview

The NHMM is first trained on observational data for the period 1951-2004, along with a seasonally-varying index of monsoon circulation strength, taken from the GCM ensemble in the form of a multimodel mean. In this

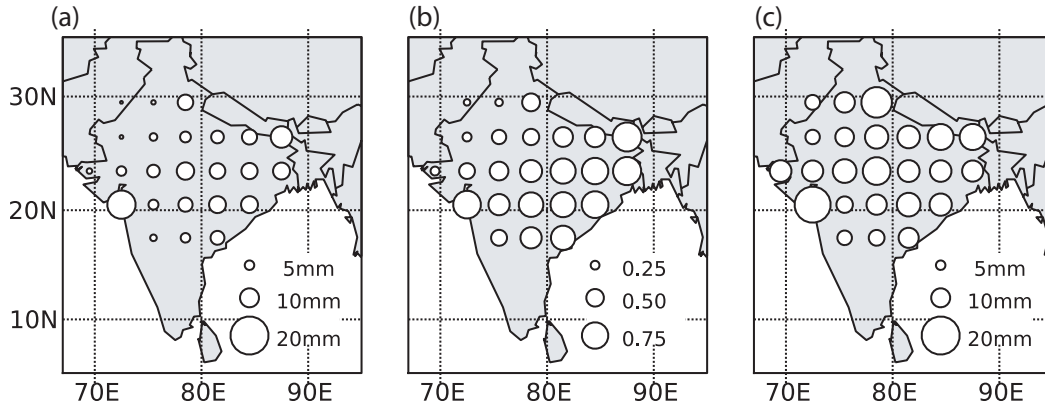
step both the parameters of the hidden states and the relationship between the circulation index and the transition probabilities governing the day-to-day progression of the states are inferred. A bias correction is introduced in order to bring the simulated seasonal cycle into better agreement with observed. The NHMM is then used to generate stochastic sequences of daily rainfall for the periods 1951-2000 and 2070-2099, using in each case the corresponding circulation index. Differences between these two sets of simulations provide a measure of rainfall change owing to changes in the monsoon circulation, and may be thought of as dynamically-induced.

In going to the 21<sup>st</sup> century the circulation weakens. Given the relationships inferred by the NHMM in the learning step, this brings about a reduction in regional mean rainfall. Such a reduction, however, is inconsistent with the multimodel response, in which rainfall *increases*. The discordance is attributed to thermodynamic effects, and resolved by scaling the NHMM daily wet-day amount distributions so that the simulated regional mean rainfall change comes into agreement with the multimodel mean. Attention then focuses on the statistical features of climatically-induced rainfall change, disaggregated in both time and space. It is this downscaling that constitutes the proper domain of the NHMM

## 4. Data

### 4.1. Observations

A daily rainfall dataset for 1951-2004 developed by the India Meteorological Department (Rajeevan *et al.* 2006) constitutes the observational record. This is a 1° gridded product derived from quality-controlled station records and is complete. Only the Jun-Sep (JJAS) values, representing the southwest (summer) monsoon, are utilized. The data are geographically truncated to 15°–29°N, 68°–88°E, which coincides approximately with the main “monsoon zone” (Sikka and Gadgil 1980). This region has been identified in a cluster analysis (Moron, pers. comm) as having



**Figure 1.** Climatological JJAS rainfall for the thinned IMD data: (a) Mean daily rainfall, (b) Probability of rain, (c) Mean daily intensity.

a homogeneous seasonal cycle for precipitation. This facilitates the analysis, since the present NHMM structure is not designed to accommodate seasonal-cycle mixtures. To allow for multiple simulations, given the computer resources at hand, the data were thinned to a resolution of  $3^\circ$ , resulting in a dataset having 24 gridpoints. On a single-processor 2.8 GHz workstation, fitting the NHMM to 54 yr of JJAS daily data on this grid requires ca. 20 min. Note that the effective resolution of the thinned data remains  $1^\circ$ .

Climatological rainfall and its decomposition in terms of probability of rain,  $P(R)$  and mean daily intensity,  $\bar{I}$  (amount on wet days) is shown in Fig. 1, where panels (a) and (b) exhibit a clear gradient, extending from arid Rajasthan and the Thar desert southeast toward the wet coast of the Bay of Bengal. The gradient in  $\bar{I}$ , shown in Fig. 1c, is less clear, owing in part to high intensity values in the central region, but this variable also reflects relative dryness in the northwest.

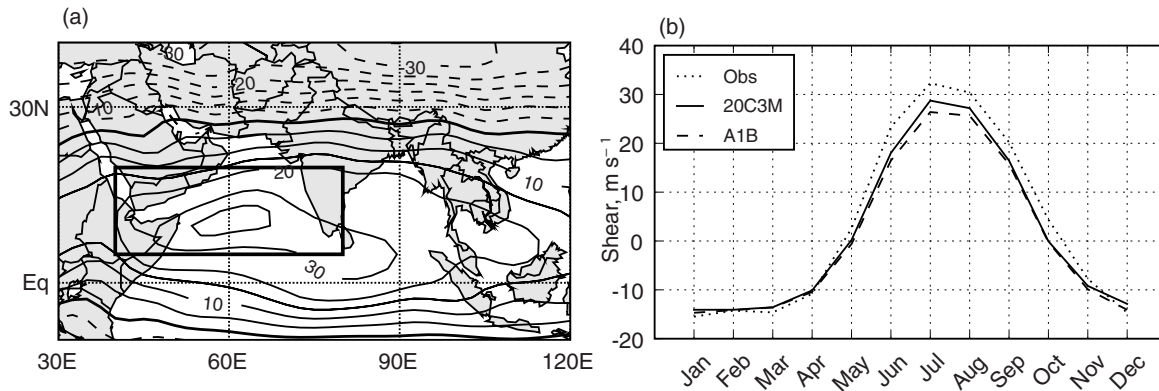
The behavior of the most westerly gridpoint just north of  $20^\circ$  seems somewhat anomalous in the context of this gradient, and signals a moisture source to the west interacting with the escarpment of the Deccan plateau, as opposed to the convective systems that propagate northwestward over the continent from the Bay of Bengal. This is consistent with high rainfall at coastal locations further to the south that are not part of the study dataset.

#### 4.2. Information from GCMs

Precipitation and zonal wind fields from an ensemble of 23 GCMs from the Coupled Model Intercomparison Project version 3 dataset (CMIP3, Meehl *et al.* 2007) are utilized. Simulations from the CMIP3 models underlie much of the research considered in the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC, Solomon *et al.* 2007). The models employed (Table I) comprise all those for which these two fields were available, for both the 20C3M (20<sup>th</sup>-century) and A1B (21<sup>st</sup>-century stabilization) experiments. The data utilized have monthly time resolution. In addition, the 500-mb vertical velocity

**Table I.** GCMs in the CMIP3 archive (Meehl *et al.* 2007) for which both precipitation and zonal wind fields were available for both 20C3M and A1B simulations.

Group	Model
BCCR (Norway)	BCCR-BCM2.0
CCCMA (Canada)	CGCM-3.1
CCCMA	CGCM-3.1(T63)
CNRM (France)	CM3
CSIRO (Australia)	CSIRO-Mk3.0
CSIRO	CSIRO-Mk3.5
GFDL (USA)	GFDL-CM2.0
GFDL	GFDL-CM2.1
GISS (USA)	GISS-AOM
GISS	GISS-EH
GISS	GISS-ER
IAP (China)	FGOALS-g1.0
INGV (Italy)	INGV-ECHAM4
INM (Russia)	INM-CM3.0
IPSL (France)	IPSL-SXG
CCSR/NIES/JAMSTEC (Japan)	MIROC3.2(hires)
CCSR/NIES/JAMSTEC	MIROC3.2(medres)
MPI (Germany)	ECHAM5/MPI-OM
MRI (Japan)	MRI-CGCM2.3.2
NCAR (USA)	CCSM3
NCAR	PCM1
Hadley Centre (UK)	HadCM3
Hadley Centre	HadGEM1



**Figure 2.** Vertical shear of the zonal wind. (a) Climatological JJAS field from the NCEP-NCAR reanalysis for 1951-2000, showing the box defining WSII. (b) Climatological annual cycle, showing the 1951-2000 reanalysis (“Obs”) and multimodel means for 1951-2000 (“20C3M”) and 2070-2099 for the A1B scenario (“A1B”).

field from a subset of these models is employed for a validation exercise.

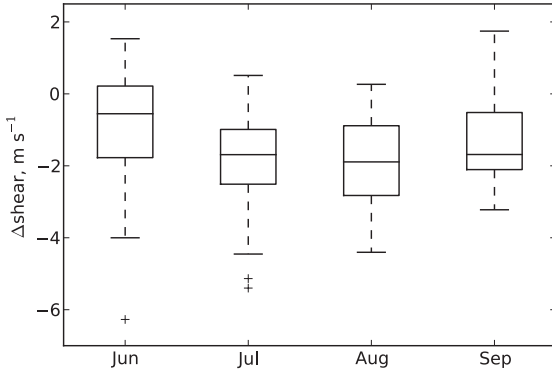
Information regarding the circulation is introduced in the form of the “Westerly Shear Index 1” (WSII), as defined in Wang and Fan (1999). This is the vertical shear of the zonal wind ( $u_{850} - u_{200}$ ) averaged over the box  $5^{\circ} - 20^{\circ}N, 40^{\circ} - 80^{\circ}E$ . Fig. 2a shows that this box encloses a local maximum in the climatological JJAS shear field, as represented in the NCEP-NCAR reanalysis (Kalnay *et al.* 1996).

Wang and Fan, using monthly mean values for JJAS over the 1974-1997 period, found that WSII was both stable in its physical configuration and significantly correlated ( $r = 0.61$ ) with outgoing longwave radiation, indicative of monsoon rainfall, over the Indian subcontinent. Of the climate models utilized (see Table I), a majority also show a local maximum for the shear field in the WSII box. For the exceptions the box invariably sits astride a ridge, the maximum then occurring to the east. The zonal elongation of shear contours is also a robust feature of the GCMs. This pattern concordance extends across centuries, and likely results from the broad land-ocean contrasts and extended high-amplitude topographic relief that characterize this region. On the large scale there is thus relatively robust model agreement, stable across centuries, with respect to monsoon circulation structure.

The WSII index tracks the annual cycle of monsoon winds, as shown in Fig. 2b. During the southwest (summer) monsoon, extending from June through September, values are strongly positive; outside this seasonal window the winds reverse and WSII changes sign. The figure shows the NCEP-NCAR reanalysis climatology as well as multimodel means for the 20<sup>th</sup> and 21<sup>st</sup> centuries. The GCM simulations match the reanalysis fairly closely in shape, although the latter has a slightly greater amplitude. Fig. 2b also shows that the amplitude of the seasonal cycle decreases in going from the 20<sup>th</sup> to 21<sup>st</sup> centuries. This is also a robust feature of the multimodel dataset, as illustrated in Fig. 3. This figure shows the distribution across GCMs of the difference between 21<sup>st</sup> and 20<sup>th</sup> century WSII values, by month.

The largest decreases occur in July and August. This is the core of the wet season, when rainfall is greatest. Shifts for June and September would have a reduced influence on overall rainfall change, but are still negative for most models. Like the spatial pattern of the shear field, the seasonal cycle shape is rather stable across centuries, weakening slightly but showing no phase shift. The NHMM assumes a degree of stationarity in the large-scale flow structure; both the spatial and temporal stability of the WSII cycle support this assumption.

WSII is not unique among climate variables in having a well-defined seasonal cycle. What is important



**Figure 3.** Intermodel spread in  $\Delta\text{WSII}$ , going from the 20<sup>th</sup> to the 21<sup>st</sup> century. Distribution shown separately for each month in JJAS.

for the exogenous predictor, rather, is that it link monsoon circulation strength and rainfall over the Indian subcontinent in a physically sensible manner. Its change, in going from the 20<sup>th</sup> to 21<sup>st</sup> centuries, then becomes an indicator of dynamically-mediated shifts in monsoon rainfall. It is this linkage that imbues its use in the NHMM with meaning.

## 5. Methodology

### 5.1. Theoretical basis

The NHMM is similar to the homogeneous model described in Greene *et al.* (2008), in that it decomposes the daily precipitation field over a network of locations into a small set of discrete “states,” that proceed in time according to a first-order Markov process. In both of these models rainfall at each of the locations is conditionally independent: given the state, it does not depend on rainfall at other locations (or at other times). Spatial and temporal dependence are then represented implicitly, the former through the existence of multiple states, the latter via the Markov property of the state transition process. For each location, conditional on state, the rainfall “emitted” by the model on a given day is drawn from a three-component mixture, a delta function at zero (representing no rain) and two exponentials. Two-component exponential mixtures have been found effective in the modeling of daily rainfall distributions (Woolhiser and Roldán 1982). The model comprises these exponential

parameters and mixing weights as well as the matrix of probabilities governing day-to-day transitions among the states.

The present model differs from that of Greene *et al.* (2008) by the inclusion of an exogenous predictor. Lacking such a predictor the homogeneous model is not suitable for prognosis, as there is no covariate that might assume differing values during different climate epochs. It will be seen that the inclusion of this predictor has little influence on the NHMM’s inferred hidden states. Here the WSII index is taken as the covariate; its value is related to the state-to-state transition probabilities according to the polytomous logistic regression (Robertson *et al.* 2004):

$$P(S_t = j | S_{t-1} = i, X_t = x) = \frac{\exp(\sigma_{ij} + \rho_j x)}{\sum_{k=1}^K \exp(\sigma_{ik} + \rho_k x)}. \quad (1)$$

Here,  $t$  is time, discretized in terms of days, and  $P$  the probability of an  $i \rightarrow j$  transition in going from time  $t - 1$  to  $t$ , given value  $x$  of predictor  $X$  at time  $t$ . The  $\sigma_{ij}$  and  $\rho_j$  are parameters to be fit and  $K$  is the number of modeled states, here taken as four in correspondence with the model of Greene *et al.* (2008).

Expression (1) refers to transition probabilities  $P$  among the states but makes no reference to the state characteristics themselves, which include rainfall probabilities and intensity distributions for each of the stations. Through the agency of the transition matrix, however,  $X$  may influence state occupation statistics, thus, *inter alia*, total seasonal rainfall, shape of the seasonal cycle or the fraction of rainfall deriving from each of the states. The underlying assumption is that future rainfall will arise from the same assemblage of large-scale weather patterns as presently exists, but that their relative occurrence frequencies, timing or persistence characteristics may shift as climate changes.

This stationarity of weather states, with the forced response taking the form of shifts in occupation statistics,

is consistent with a nonlinear-dynamical view of climate, in which the states represent basins of attraction toward which trajectories in the climate phase space are drawn (Lorenz 1963; Palmer 1999). There is also a connection with the idea of fluctuation dissipation, i.e., that the response of a nonlinear system to small perturbations can be understood by studying fluctuations of the unperturbed system (Leith 1975).

A final element in the modeling scheme involves the scaling of NHMM intensity distributions so as to bring simulated regional-mean precipitation change into agreement with the GCM ensemble mean. This is performed outside the NHMM fitting process, and reflects a reliance on the GCMs for large-scale climate information. From a more process-oriented perspective this procedure, which conditions rainfall amounts on wet days, may be thought of as representing variation due to changes in atmospheric moisture, the ultimate source of wet-day amounts, given a dynamical predisposition to moisture flux convergence.

Use of multimodel mean fields follows from work showing that such fields as a rule provide more consistently faithful representations of the observed climate than do individual models (Gleckler *et al.* 2008), and provides a natural way of estimating uncertainties arising from differences in GCM formulation.

## 5.2. Model fitting

WSII values (Fig. 2b, “20C3m” and “A1B”) were linearly interpolated to daily resolution for ingestion by the NHMM. Since daily, or even interannual variations of model WSII cannot be expected to correspond with observed rainfall changes, these cycles were utilized in the form of climatologies, repeated for each year for both training and simulation. One may think of the historical sequences as equally likely draws, given the WSII climatology, in the conditioning of a statistical forecast model. This is similar to the model output statistics (MOS) idea, in which localized

weather forecast statistics are conditioned on more coarsely resolved model output.

The model is fit using the observed rainfall and interpolated 20<sup>th</sup>-century WSII, producing both the decomposition of the rainfall by states and the parameters of the logistic regression (1). As in Greene *et al.* (2008), maximum-likelihood estimates of the model parameters are obtained using the expectation-maximization algorithm (Dempster *et al.* 1977).

In Sec. 5.3 it will be seen that the spatial patterns of rainfall probability and mean daily amount associated with each of the hidden states are quite similar to those inferred using the homogeneous model of Greene *et al.* (2008), in which precipitation from a smaller and somewhat more localized network of stations was fit. This suggests that (a) the smaller network is in fact extensive enough to capture most of the spatiotemporal variability present over the 24 locations of the present study, and (b) inclusion of the exogenous variable has little influence on the state decomposition. As we will see, it is rather in the simulation step that the effect of WSII, realized through its conditioning of the state transition matrix, becomes manifest.

## 5.3. State descriptions

The four states from the fitted model are summarized in the four columns of Fig. 4, with rainfall probabilities in the top row, mean intensities below (cf. Greene *et al.* 2008, Fig. 4). Because rainfall sequences include many dry days, HMM decomposition typically identifies a state having both low occurrence probabilities and intensities; here this is state 4 (Figs. 4d, 4h). State 3 exhibits the uniformly highest probability of rain (Fig. 4c), along with relatively high intensities in both the west and adjacent to the Bay of Bengal (Fig. 4g). This reflects enhanced convection over most of the subcontinent. In state 1 convection occurs preferentially toward the Himalayan foothills, while state 2, with enhanced probabilities toward the southeast but moderate intensities, may represent an early stage in

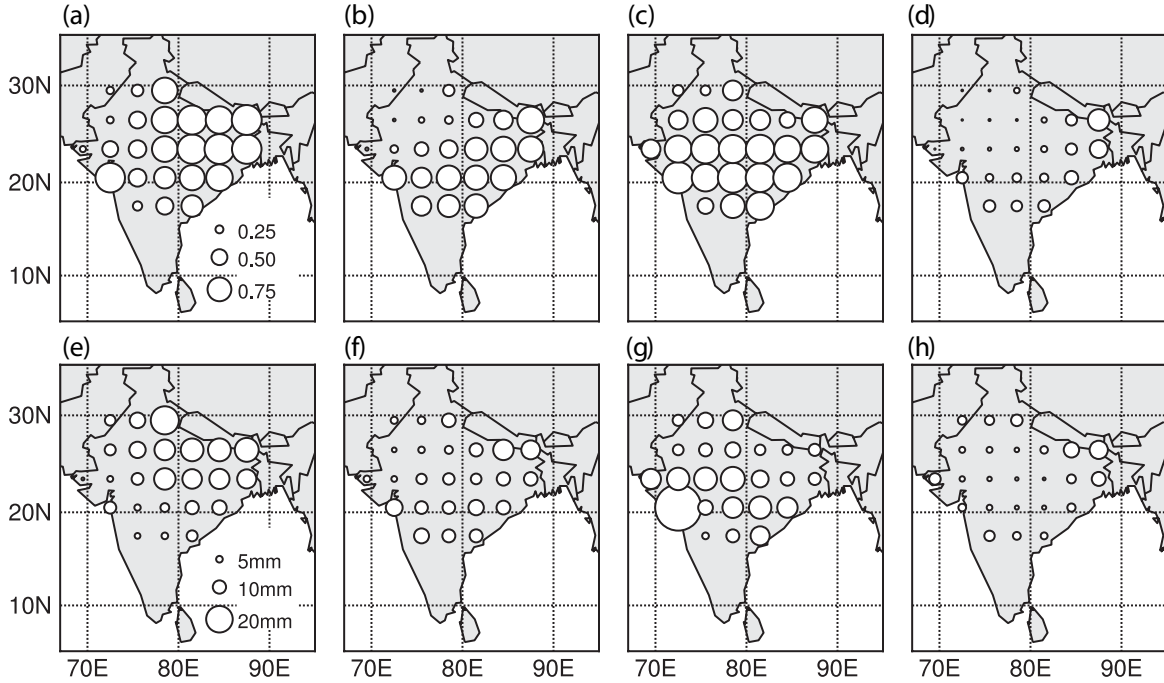


Figure 4. Probability of rain (a-d) and mean daily intensity (e-h), for states 1-4, shown in the corresponding columns.

the propagation across the subcontinent of intraseasonal oscillations (Goswami and Mohan 2001). State 2 differs more than the others in the earlier analysis, but in that study the southeastern sector, where the probability of rain is seen to be concentrated, was not sampled. Since the states described here are quite similar to those inferred in Greene *et al.* (2008), the reader is referred to that work for more detailed discussion.

#### 5.4. Effect of exogenous predictor

The learned relationship between WSII and transition probabilities is summarized in figure 5a, which shows the four probabilities associated with the equilibrium state distribution as a function of WSII. The vector of state probabilities  $\mathbf{p}$  is advanced in time by the linear transformation

$$\mathbf{T}'\mathbf{p}_{t-1} = \mathbf{p}_t, \quad (2)$$

(Norris 1997), where  $\mathbf{T}'$  is the transpose of the transition matrix. Identification of a state probability vector that is invariant under this transformation corresponds to the

eigenvalue problem

$$\mathbf{T}'\mathbf{p} = \lambda\mathbf{p}, \quad (3)$$

where we seek the eigenvector having  $\lambda = 1$ . The existence and uniqueness of such a vector is implied by the stochastic character of  $\mathbf{T}$  (its rows each sum to unity) and the ergodic property of the Markov chain modeled by the NHMM (it is aperiodic, with finite return times).

To generate Fig. 5a the matrix  $\mathbf{T}$  was computed on a grid of  $x$ -values spanning the seasonal range of WSII, using Eq. 1. For each fixed  $\mathbf{T}$  thus obtained, the eigendecomposition suggested by (3) was performed and the eigenvector having unit eigenvalue identified. The figure shows, for each  $x$ , the four components of this vector. The distributions so identified represent points of stable equilibrium (Cox and Miller 1965).



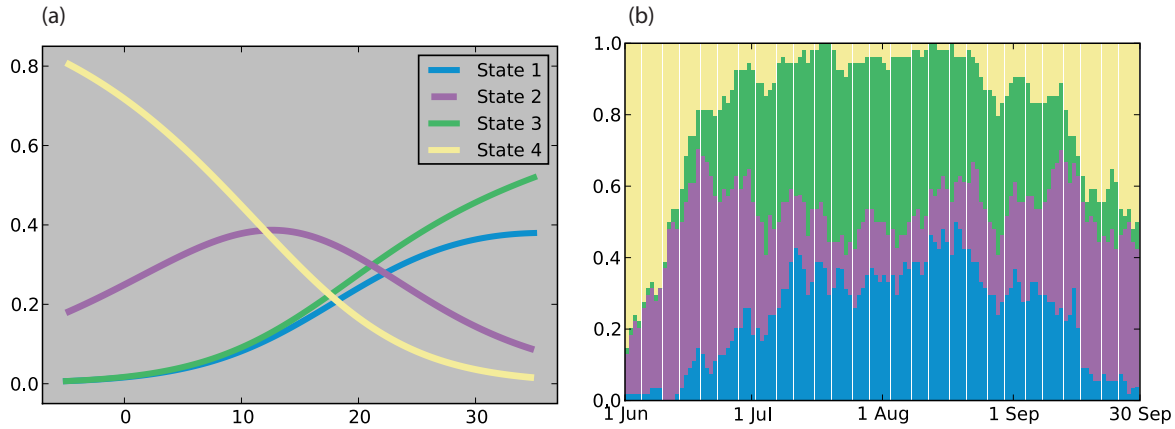


Figure 5. (a) Dependence of equilibrium state probabilities on WSII, (b) Seasonal climatology of the Viterbi sequence, using same colors.

5.5. Viterbi sequence

Complementary to the equilibrium state probability distributions shown in Fig. 5a is the sequence of most-likely states. Computed using the Viterbi algorithm (Forney 1973), this is the most likely day-by-day trajectory through state space, given the data (IMD observations and 20<sup>th</sup>-century WSII) and the fitted model. A daily climatology of the Viterbi sequence, in terms of relative state occupation frequencies, is shown in Fig. 5b. (Note that in this figure relative state frequencies correspond to the height of the bands representing the respective states, not to their absolute positions along the ordinate.)

For low WSII values, Fig. 5a indicates that state 4, the dry state, is most likely. This is consistent with Fig. 5b, where early in the season state 4 occupies most of the vertical probability space. Fig. 5b also indicates that this state occurs infrequently during the Jul-Aug core of the rainy season, corresponding to its relatively low equilibrium probabilities at high values of WSII. The core of the rainy season, on the other hand, is dominated by states 1 and 3, the equilibrium probabilities for which rise correspondingly with increasing WSII. Conditioning of the transition probabilities by WSII thus acts as an effective control on the subseasonal evolution of precipitation.

6. Simulations

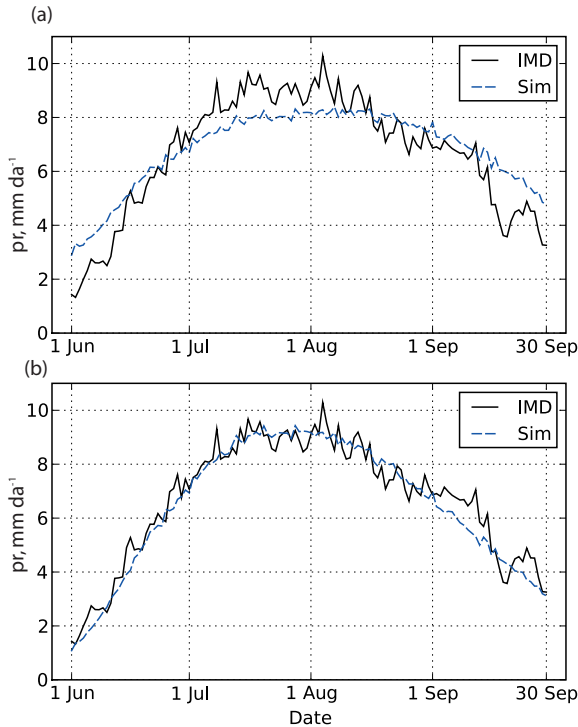
6.1. Twentieth century

6.1.1. Seasonal cycle and bias correction

Using the learned NHMM parameters and 20<sup>th</sup>-century shear index, an ensemble of ten 54-yr simulations was generated. Biases were noted in these simulations in regional mean rainfall (overestimated by ~ 2%) and in the shape of the seasonal cycle. Small additive biases are not atypical (Robertson *et al.* 2009); to minimize possible effects, changes in rainfall characteristics were assessed by comparing 20<sup>th</sup>- and 21<sup>st</sup>-century simulations.

The climatological seasonal cycle of spatially-averaged observed rainfall is shown in Fig. 6a, along with the initial simulation. It can be seen that the latter overestimates rainfall in the early and late parts of the season, underestimating during July and the first part of August. Since the state frequency distribution differs considerably between the central and peripheral parts of the season (Fig. 5b), such a mismatch could have significant effects on the spatial pattern of simulated rainfall change; daily rainfall statistics would also likely be biased. A correction was therefore undertaken.

From the modeling it is inferred that circulation-driven rainfall changes account for only a portion of the total change. The remainder is assigned to thermodynamic effects, essentially the moistening of the lower atmosphere



**Figure 6.** (a) Seasonal cycle of location-averaged precipitation in the observations (“IMD”) and initial 20<sup>th</sup>-century simulations. (b), as in (a), with simulations corrected for seasonal-cycle shape bias.

owing to climatic warming. This idea is applied in the climate change sense by scaling intensity distributions so as to bring NHMM-simulated rainfall change into agreement with the GCM-simulated change. A similar compositing of effects is invoked here for bias correction.

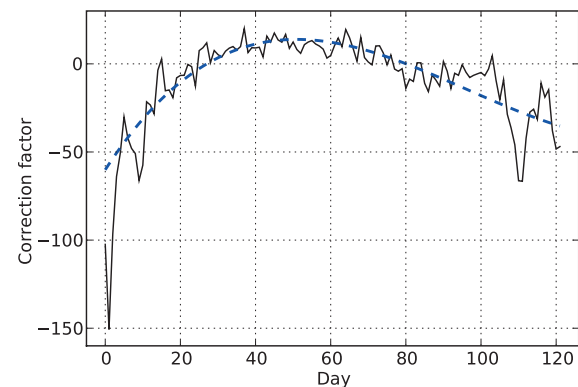
Lower tropospheric moisture exhibits a seasonal cycle quite similar in shape to the shear cycle (Fig. 2b), but this significant component of intraseasonal variation has so far been ignored. The seasonal variation shown in the “SIM” trace of Fig. 6a is due only to shifts in relative state occupation frequencies. These represent changes in the statistics of large-scale flow anomalies, but do not capture the contribution to the seasonal cycle of intensity changes due to atmospheric moistening. It is perhaps not surprising that the simulated seasonal cycle is less “peaked” than the observed, to which the seasonality of lower tropospheric moisture undoubtedly contributes as well.

The correction was implemented by first computing, for each day in the seasonal cycle climatology, the factor

required to compensate for the simulation excess or deficit, as shown in Fig. 6a. A third-order polynomial was fit to the resulting “compensation curve” (Fig. 7) and the fitted values then used to rescale the simulated rainfall within each simulated season.

The climatology of the bias-corrected 20<sup>th</sup>-century simulation is shown in Fig. 6b, where it can be seen that a considerably better fit to observations has now been obtained. Even for September, where high-frequency features render the correction somewhat less effective than during the remainder of the season, the RMS difference between series is reduced by 36%. (For the season as a whole the reduction is 44%.) Since the shear cycle does not undergo any significant change in shape as we go the 21<sup>st</sup> century, it seems reasonable to assume that seasonal-cycle bias will also persist. The correction, computed as above, is thus applied uniformly, across both 20<sup>th</sup>- and 21<sup>st</sup>-century simulations.

As a test, an 8-state model was fit to the data, the supposition being that with more degrees of freedom the NHMM might better be able to “tune” the seasonal cycle, by fitting a more precisely calibrated sequence of evolving states (thus exponential parameters). This did not result in an improved representation, suggesting that a good fit to seasonal-cycle shape would not likely be obtained by



**Figure 7.** Daily correction factor required to remove seasonal-cycle bias in the 20<sup>th</sup>-century simulations, and the 3<sup>rd</sup>-order polynomial fit used to scale the simulations. The first five entries, with values  $< -50$ , were influential, thus not used in fitting.

driving the model with WSI1 alone. This result also tends to validate the procedure utilized.

The correction process introduces a small negative shift, so that the corrected 20<sup>th</sup>-century series now underestimates regional mean rainfall by 2.7%. The sign of this shift is related to the larger negative than positive area lying between the compensation curve and the abscissa ( $y = 0$ ) in Fig. 7. Means of both 20<sup>th</sup>- and 21<sup>st</sup>-century rainfall simulations were shifted similarly, so the correction procedure should not compromise inferences regarding climatological rainfall changes.

6.1.2. Daily rainfall distributions

Daily rainfall distributions are an area of pervasive GCM bias, with typical parameterizations producing too much drizzle and too few heavy rainfall events. Figure 8 compares a typical GCM distribution with that simulated by the NHMM, for the central location 23.5N, 78.5E. (The closest GCM gridpoint is utilized.) GCM daily amounts are sharply truncated around 70 mm, while the NHMM captures the much longer tail of the observed distribution. There is a small degree of mismatch at levels above about 200 mm, suggesting that close attention to distribution extremes might be a fruitful path for extending this work. Vrac and Naveau (2007) have implemented an NHMM utilizing extreme-value mixtures that is possibly of relevance, but this potential refinement of distributional form lies beyond the scope of the present study.

6.1.3. Spatial pattern

Figure 9 shows spatial patterns of the GCM ensemble-mean climatological JJAS rainfall, the observations and the bias-corrected 20<sup>th</sup>-century simulations. The first of these (Fig. 9a) captures the general NW–SE precipitation gradient seen in the observations (Fig. 9b), but misses various small-scale features, including the sharp maximum on the western coast, as well as variations along SW–NE diagonals. The multimodel mean is gridded  $2^\circ \times 2.5^\circ$ , fine enough to discern these details if they were present (although the

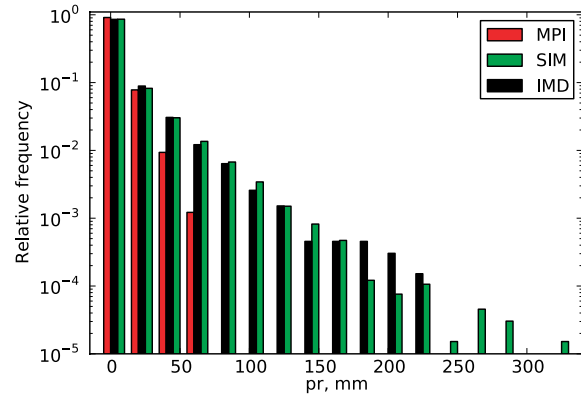


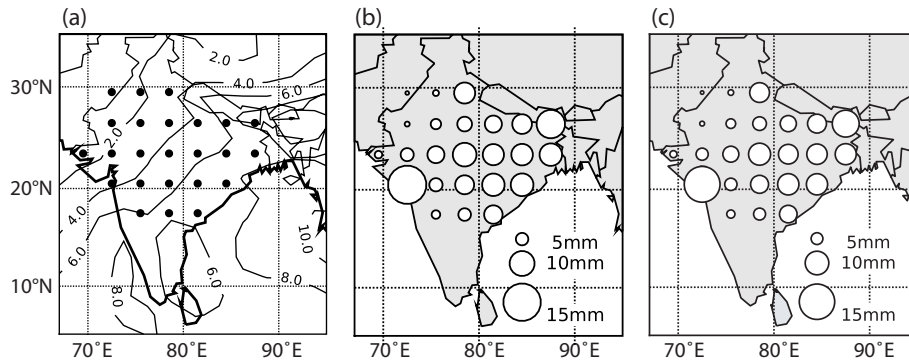
Figure 8. Daily rainfall distribution for the centrally-located point 23.5°N, 78.5°E, for observations (“IMD”) the ECHAM5 model (“MPI”) and the 20<sup>th</sup>-century NHMM simulations (“Sim”).

resolution of some of the GCMs is coarser). The NHMM simulation, constrained as it is by the observed patterns, captures these details faithfully.

6.2. Projections

A second ensemble of simulations was generated using the A1B shear index. Because the seasonal shear cycle weakens in going to the 21<sup>st</sup> century (Fig. 2b), and owing to the WSI1 dependence of state distributions (Fig. 5a), the balance in these simulations is tilted away from the wetter states and toward the dry, resulting in an overall reduction in seasonal precipitation. Figures 10a and 10b show the mean JJAS rainfall and the decrease owing to this effect, respectively. Averaged over locations, seasonal rainfall is reduced via this mechanism by 4.3%.

Figure 10d indicates that the multimodel mean rainfall changes in the *opposite* sense, with seasonal values everywhere increasing. Thus, while monsoon drying may be consistent with dynamical GCM behavior it is not consistent with the total GCM response. As discussed previously, the discrepancy is here attributed to thermodynamic effects, and resolved by scaling the NHMM daily intensity distributions so that the simulated regional mean fractional precipitation change (or equivalently, precipitation sensitivity) comes into agreement with the multimodel mean. This scaling (not to be confused with the bias correction described in Sec. 6.1.1) is applied directly to the means of the



**Figure 9.** (a) Multimodel mean climatology, (b) observations and (c) 20<sup>th</sup>-century simulations. Plots show average daily rainfall.

exponential distributions in the NHMM, from which wet-day amounts are drawn during the simulation step. All 192 distributions (two at each location and for each state) are scaled identically; given sufficient predictive information it might be possible to implement a more granular scaling scheme, but this was judged impractical for the present study.

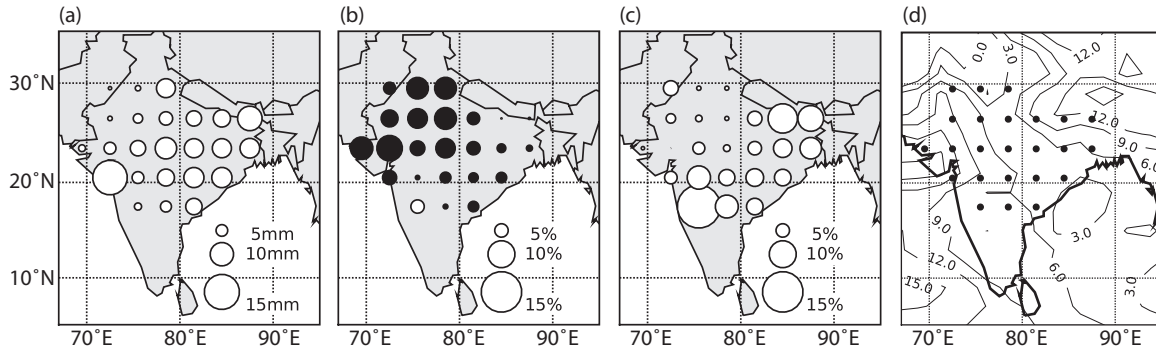
A third set of simulations was then generated, using both the A1B shear and the scaled rainfall distributions. The resulting daily sequences reflect the dynamically-induced shift in state occupation statistics but are now consistent with the GCM-inferred total precipitation change; they also incorporate the corrected seasonal-cycle pattern.

The multimodel mean fractional JJAS rainfall change (i.e., the average of the models' fractional changes) for the study area (15–30N, 65–90E) amounts to an increase of  $\sim 5.7\%$  (the averaging was performed in this way in order to reduce potential effect of mean rainfall biases within the ensemble). The required increase for the rainfall intensity distribution means was then found to be 10.4%, close to the simple difference between total and circulation-induced changes. The overall picture is thus one in which only about half of the thermodynamically-driven potential increase in regional mean precipitation is realized, owing to a compensating weakening of the large-scale circulation.

It is perhaps of interest to note that the increase in monsoon rainfall, as represented in the IPCC ensemble (and in the A1B scenario), occurs principally during 2030–2060.

Before this period there is little apparent change in level from that of the second half of the 20th century, while afterward there are some significant decadal swings but again, little change in overall level. Conditional on shifts in dynamics (as in the above analysis), adjustments to the scaling factor utilized would be expected to follow such fluctuations. We do not speculate here about the causes or likely realism of rainfall variations in the multimodel ensemble, but the presence of decadal-scale fluctuations provides, *ex post facto*, some justification for the use of 30-year averaging periods in computing the “time-slice” statistics described herein.

The change in seasonal mean rainfall in the third ensemble of simulations is shown in Fig. 10c, and can be compared with the multimodel mean change in Fig. 10d. The large-scale patterns are similar, with a broad minimum extending from northwest to southeast, and higher values toward the southwest and northeast, although, as expected, there are differences at finer scales. Given that the patterns of Figs. 10c and 10d are the products of quite different analytical processes — the simple multimodel mean vs. the disaggregation into states, logistic regression of Eq. 1 and the constraints imposed by observed patterns of variability — the agreement is noteworthy.



**Figure 10.** (a) The mean simulated JJAS 20<sup>th</sup>-century precipitation field, (b) Percent changes for 2070-2099, using only WSII, (c) Percent changes using WSII and scaled daily intensity distributions, (d) the multimodel mean percent precipitation change. In (b) and (c) filled markers show values <math>< 0</math>, unfilled markers > 0.

**7. Fine-scale rainfall changes**

*7.1. Spatial patterns*

Figure 10b indicates that the largest fractional rainfall reductions occur toward the northwest. These magnitudes are large in part because the 20<sup>th</sup>-century values are small, but also because of the shift in relative state frequencies: States 1, 2 and 4 are all drier in this region, in the sense of both rainfall probabilities and daily intensities, than state 3. A shift away from state 3 (such as would be expected from a weakening of WSII) would thus tend to dry the northwest more than it would subregions where states 1, 2, and 4 (which substitute for state 3 to varying degrees as WSII weakens) agree less well.

The largest fractional rainfall increase occurs at 17.5N, 75.5E. Figure 10b shows that circulation changes alone bring about an increase in seasonal rainfall at this location. This occurs because of the low probability and intensity values in state 3, the intensity actually being less than that of the dry state. State 2, meanwhile, exhibits considerably higher rainfall frequency and intensity at this point. A dynamically-mediated increase here, unlike at most other locations, is thus consistent with a reduction in state 3 occurrence frequency. This increase increments that mediated by the intensity scaling.

These observations confirm the basic functioning of the NHMM, modulating relative state occupation frequencies according to the dependencies shown in Fig.

5a while increasing intensity-distribution means uniformly across stations within states. Figure 11a shows changes in state frequencies, and indicates that the core monsoon wet states 1 and 3 become less frequent, the dry state (4) more frequent, in going from the 20<sup>th</sup> to 21<sup>st</sup> centuries.

Fig. 11b shows daily 20<sup>th</sup>- and 21<sup>st</sup>-century rainfall distributions pooled across stations. For almost all bins frequencies are greater in the 21<sup>st</sup> century, the notable exception being the lowest, which includes dry days. Note that the log scale of Fig. 11b causes the substantial difference in counts in the lowest bin to appear quite small, when in fact it is sufficient to balance all the differences of opposite sign in the higher bins. The uniformity of response is no doubt related to the uniform application of distribution scaling.

*7.2. Temporal aspects of precipitation change*

As a downscaled example, we consider the central location 23.5N, 75.5E. This point receives most of its rainfall from state 3, with secondary contributions from states 1 and 2 and very little from state 4. The rainfall increase at this location is 4.5%, somewhat less than the regional mean change. This increase occurs with a concurrent reduction in the frequency of rain, the number of dry days in the 122-day JJAS season increasing from an average of 55.7 to 58.7, or about 5%. The change in mean intensity must more than compensate this decrease, and  $\bar{I}$  goes from 12.2 to 13.3 mm, an increase of 9.5%.

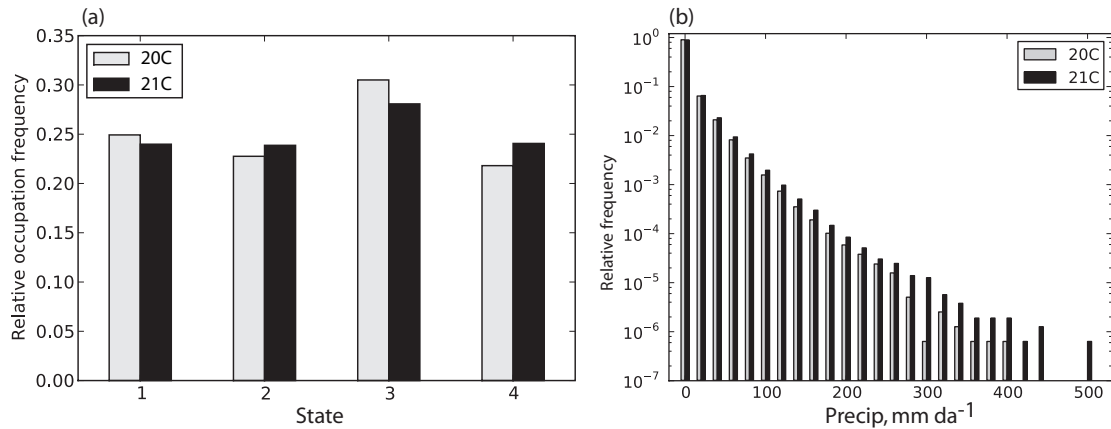


Figure 11. Change in (a) relative state occupation frequencies and (b) precipitation, pooled over stations.

The length of dry spells, averaged over stations, increases by about the same percentage that rainfall decreases,  $4.5 \pm 2.5\%$  ( $1\sigma$ ). This amounts to only a small fraction of a day, since the mean length of 20<sup>th</sup>-century dry spells is  $\sim 2.5$  da. The percentage increase is smaller if one considers only breaks longer than seven days, with a mean increase of  $1.1 \pm 1.7\%$ , the noisier statistics owing to the reduced number of samples. The mean dry-spell length, for spells longer than seven days, is  $\sim 11$  da.

Possible changes in the date of monsoon onset were investigated using the agronomic definition of Moron *et al.* (2009). As employed here, onset is defined locally, as the first wet day ( $> 1$ mm) of the first five-day wet spell in the season for which accumulated rainfall is least as great as the climatological five-day wet spell, provided that this period is not followed, during the succeeding 30 days, by a spell longer than 15 days in which accumulated rainfall is less than 5 mm. (The latter condition is imposed in order to avoid “false positives,” since a significant dry spell would negate the effects of the earlier rains.) Onset is thus sensitive not only to accumulated rainfall but to the temporal details of how this rainfall is received.

With this definition it was found that monsoon onset is delayed in going to the 21<sup>st</sup> century, but only by an average of one day, with a standard deviation over locations of 1.1 da. The interannual standard deviation in the classical onset time, over Kerala state, is about 6 days, (Pai and Rajeevan

2007), so a shift of one day would in fact be difficult to discern, and by itself would be expected to have little agronomic impact. This minimal change is in some contrast to the shift that might be expected from changes in WSII alone (Fig. 2b). Based on the June 15 value from the daily interpolation, the change in WSII suggests an onset delay of four days, as well as a three-day advance in monsoon withdrawal, leading to a net reduction in wet-season length of about seven days. (This is still not large, compared with the interannual standard deviation, however.) Evidently the increase in intensities is sufficient to compensate for a delay in “dynamical” monsoon onset. The agronomic calculation also suggests a delay in monsoon withdrawal of  $0.2 \pm 0.8$  da, reducing the estimated net reduction in wet-season length to 0.8 da. It should be borne in mind that other factors, including temperature and atmospheric carbon dioxide concentrations, may also have appreciable effects on crop yields (see, e.g., Challinor *et al.* 2009).

The above result may be contrasted with that of Naylor *et al.* (2007) for Indonesian rice agriculture. Those authors found significant increases, for year 2050, in the probability of onset delay exceeding 30 days, particularly in Eastern Java and Bali. Methodology, onset definition, region and, not incidentally, monsoon system differ from those of the present study, but the considerably greater climate sensitivity implied is notable.

A shift toward higher intensities may be expected to increase the frequency of heavy rainfall events. This tendency is apparent in the projected distributions shown in Fig. 11b, but for the A1B scenario the changes are not dramatic. The 90<sup>th</sup> percentile for daily intensity during the 20<sup>th</sup> century ranges from 19 to 51 mm across locations. The intensity increase at this percentile is  $9.5 \pm 2\%$ , fairly uniform across stations. Holding amounts constant, the shift in percentiles is only about -1.5, meaning that events corresponding to the 20<sup>th</sup>-century threshold occur only slightly more often during 2070-2099. It should be kept in mind that the present analysis is not optimized in particular for inference regarding extreme events, for which other distributional forms may be more appropriate (see, e.g. Coles 2001).

**8. Dynamics and thermodynamics**

Although the experimental design does not contemplate a direct estimation of the thermodynamic contribution to monsoon rainfall changes, the use of a circulation index as covariate implies at least an approximate partitioning of effects along dynamic/thermodynamic lines. An attempt is made here to validate this partitioning by comparing the components of rainfall change estimated in this way with results obtained via an independent method.

Bony *et al.* (2004) performed such a decomposition for cloud radiative forcing, by first inferring an empirical relationship between the forcing and the probability distribution of 500-mb vertical velocity,  $\omega_{500}$ . Shifts in cloud radiative forcing due solely to changes in this distribution were then identified as dynamically-induced, while independent changes were assigned thermodynamic origin. (A small component was found to result from covariation.) Since rainfall is strongly coupled to the vertical velocity, such an analysis is also appropriate here. Accordingly, probability distributions were computed for  $\omega_{500}$  and rainfall using pooled monthly JJAS values for the region and time periods of the preceding analysis. The computations were carried out using a sub-ensemble of 18

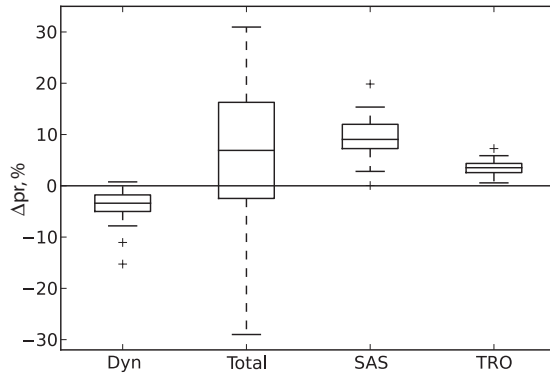
models from the 23-member group for which the requisite vertical velocity fields were available. The multimodel mean fractional rainfall change, now weighted by the probability distribution function of  $\omega_{500}$ , was 5.3%, quite similar to the multimodel mean as previously computed. This indicates that the inferred quantitative association between  $\omega_{500}$  and precipitation is a reasonable one.

The dynamically-induced regional mean rainfall change computed in this way and averaged over GCMs is -2.3%, the thermodynamic component 7.2%. There is an additional small contribution of 0.4% arising from covariation between changes in the probability distribution function of  $\omega_{500}$  and precipitation. (In the foregoing analysis this would be assigned to the non-dynamical component.) The dynamical shift is only about half of that inferred with the NHMM, but in both analyses the thermodynamically-mediated increase is compensated by dynamical effects to a similar degree: by 32% in the  $\omega_{500}$  decomposition, 41% in the NHMM analysis. Thus the overall picture does not differ greatly between the methods. The quantitative discrepancy may be attributable to use of daily data in the NHMM analysis, as opposed to monthly values for the  $\omega_{500}$  decomposition..

**9. Uncertainty**

The tendency of model representations to improve with increasing spatial scale is a recognized characteristic of GCMs (Solomon *et al.* 2007), and can at least partly be ascribed to the finite resolution of model grids. The analysis presented above utilizes GCM information on the largest spatial scales consistent with the statistical and physical constraints of the problem: If a significantly larger region were chosen the degree of applicability to the study area would come into question, since there are significant local variations in rainfall change within the larger Indian Ocean domain.

Figure 12 illustrates intermodel spread for the dynamic component and for the total JJAS precipitation change for the study region (left two boxes), as well as the spread



**Figure 12.** Distributions of area-mean rainfall change. “Dyn” refers to the shear-induced component, “Total” to the rainfall change for the study area (15–30N, 65–90E), “SAS” to the IPCC South Asia region (5–30N, 65–100E) and “TRO” to the global tropics (30S–30N). Values are percent changes for JJAS rainfall over land.

in total JJAS precipitation change for two larger regions. To obtain the first of these distributions the NHMM was fit to the IMD data using each GCM’s 20<sup>th</sup>-century shear cycle. Simulations were then generated for the 20<sup>th</sup> and 21<sup>st</sup> centuries, as was done previously using the multimodel mean values. The third and fourth boxes show distributions for total JJAS (land-only) precipitation change for the IPCC South Asia region (“SAS”) and for the global tropics (“TRO”). Means for the four distributions are -4.0%, 5.7%, 9.5% and 3.5%, respectively. For the study area uncertainty is dominated by the spread in total precipitation change, for which the standard deviation is about 14%. The corresponding value for region SAS is 4.2% and for global tropical land 1.5%.

The relatively large intermodel spread in rainfall change over India is intrinsic to the CMIP3 ensemble, and for the present analysis represents irreducible uncertainty. This is consistent with deployment of the NHMM as a downscaling tool, rather than as a method of superseding the GCMs with respect to large-scale climate projection. There is less uncertainty associated with the rainfall change in region SAS, so one might consider applying the NHMM to this larger region, but the daily observational data at hand is limited to India, precluding such an option for now. The necessity of a homogeneous shear cycle also limits the

extent the modeled domain. The area of region SAS is about double that of the study region.

## 10. Summary and conclusions

A nonhomogeneous hidden Markov model was utilized to investigate potential changes in Indian monsoon summer (JJAS) rainfall, comparing the 2070–2099 period in the IPCC A1B scenario with the second half of the 20<sup>th</sup> century. A dynamical index of monsoon circulation was utilized as the exogenous predictor; rainfall observations derive from a 1°-gridded precipitation product produced by the India Meteorological Department.

The dynamical index, in the form of a 23-model ensemble mean, conditions the NHMM transition matrix, and by extension the timing and relative occupation frequencies of the NHMM’s hidden states. A weakening of the circulation in going to the 21<sup>st</sup> century, as translated by this mechanism, induces rainfall reductions at most locations, with an area-averaged JJAS decrease of  $\sim 4.3\%$ . At the same time, multimodel mean precipitation increases by  $\sim 5.7\%$ . A “non-dynamical” increase of  $\sim 10.4\%$ , obtained by scaling the daily rainfall intensity distributions, is required to bring the area-averaged NHMM and multimodel mean changes into agreement. This factor is assigned to thermodynamic processes. An independent method of decomposing precipitation change into dynamic and thermodynamic components yields a similar proportionality. The scaling procedure is performed outside the NHMM fitting process, owing to the quite different timescales on which the relevant processes operate and the fact that conditions corresponding to the late 21<sup>st</sup>-century basic state fall outside the range of 20<sup>th</sup>-century variability.

Shifts in subseasonal rainfall statistics arise from the state decomposition effected by the NHMM, as conditioned on the dynamical predictor, and from the concomitant scaling of intensity distributions. These shifts include small increases in the number of dry days, in the average length of dry spells, in mean daily intensity, and in the



exceedance frequency of nearly all daily rainfall percentiles. No significant delay in monsoon onset was indicated, owing to a compensation of “dynamical” delay by increased daily intensities.

These inferences are drawn from stochastic daily rainfall sequences generated by the NHMM. Such sequences have realistic daily distributions and, given the appropriate observational data, may be generated for arbitrarily precise locations. As such, they are expected to be useful in a variety of applications, such as the driving of streamflow or agricultural models, and should be helpful in testing, thus potentially increasing, the resilience of proposed adaptations.

Owing to significant subregional variations in precipitation change, GCM information must be extracted at a scale no larger than that of the study area. Although there exists a reasonably strong model consensus regarding the sign of projected change for this region, agreement is not universal and there is considerable intermodel spread. In considering the results presented, this uncertainty should be borne in mind.

The modeled shifts tend to be smooth, giving no indication of possible abrupt changes in rainfall statistics. This smoothness should not be taken as a guarantee that such changes are outside the realm of possibility, but rather, arises as a consequence of model assumptions regarding structural stationarity of the large-scale flow, including the anomalous departures associated with the NHMM’s hidden states. The smooth dependence of the equilibrium state vector on WSI1 (Fig. 5a) is implicit in this assumption. The uniformity with which the thermodynamic scaling is applied may also play a role in the generally “incremental” nature of the modeled response. None of this is to say that such a response is unrealistic, however; the stationarity assumption may well be a reasonable one, at least for modest departures from the 20<sup>th</sup>-century basic state. But is well to keep in mind the role of model structure and assumptions, in contemplating the results of this study.

The NHMM fits a two-exponential mixture to the daily intensity distributions, permitting a degree of independence in fitting distribution means and tails. Scaling is applied uniformly, across locations, states and both exponentials. While this is consistent with the use of GCM information only at large spatial scales, a more complex statistical model (accompanied by suitable conditioning information) might permit some spatial disaggregation of distribution scaling, or differentiation in its application to the two exponential parameters.

The feasibility of the modeling enterprise described depends to a significant degree on the existence of specific regional dynamics, which we have shown to be robust among the GCMs of the multimodel ensemble. We believe that the methodology is in fact more broadly applicable, but extension to other settings would necessarily depend on regional details, including the identification of a suitable predictor field (or fields). Ongoing work includes exploration of this possibility.

A key assumption is that the multimodel ensemble is the optimal source of quantitative information regarding expected climatic changes. The NHMM is not utilized for this aspect of the projections, but rather, for the disaggregation of daily rainfall variability over a network of precise locations, for which the GCMs provide only “generalized,” i.e., large scale, information. These complementary tools are thus fused, in a process that exploits the most appropriate information from each. While the merging of such data sources *per se* is not a unique idea, we believe that the deployment presented herein, in the regional setting of the Indian monsoon and respecting basic theoretical expectations through the use of a dynamical exogenous variable and the implied thermodynamic scaling, represents a novel application.

#### Acknowledgements

Many colleagues from the International Research Institute for Climate and Society, Lamont-Doherty Earth Observatory of Columbia University and the NASA Goddard

Institute of Space Studies have provided the authors with helpful insights and comments. The NHMM computations were performed using the MVNHMM toolbox (<http://www.stat.purdue.edu/~skirshne/MVNHMM/>), written and maintained by Sergey Kirshner, Purdue University. We are grateful to M. Rajeevan of the India Meteorological Department for providing the rainfall data, and acknowledge the modeling groups, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the WCRP's Working Group on Coupled Modelling (WGCM) for their roles in making available the WCRP CMIP3 multi-model dataset. Support of this dataset is provided by the Office of Science, U.S. Department of Energy. This research was supported by US Department of Energy grants DE-FG02-02ER63413 and DE-FG02-07ER64429.

## References

- Ailliot P, Thompson C, Thomson P. 2009. Space-time modelling of precipitation by using a hidden Markov model and censored gaussian distributions. *J. Roy. Stat. Soc. C-App.* **58**(3): 405–426.
- Bates BC, Charles SP, Hughes JP. 1998. Stochastic downscaling of numerical climate model simulations. *Environ. Modell. Softw.* **13**(3–4): 325–331, doi:10.1016/S1364-8152(98)00037-1.
- Bony S, Dufresne JL, Treut HL, Morcrette JJ, Senior C. 2004. On dynamic and thermodynamic components of cloud changes. *Clim. Dyn.* **22**: 71–86, doi:10.1007/s00382-003-0369-6.
- Challinor AJ, Ewert F, Arnold S, Simelton E, Fraser E. 2009. Crops and climate change: progress, trends, and challenges in simulating impacts and informing adaptation. *J. Exp. Bot.* **60**(10): 2775–2789, doi:10.1093/jxb/erp062.
- Coles S. 2001. *An introduction to statistical modeling of extreme values*. Springer series in statistics, Springer: London.
- Cox DR, Miller HD. 1965. *The theory of stochastic processes*. Chapman and Hall: London.
- Dai A. 2006. Precipitation characteristics in eighteen coupled climate models. *J. Climate* **19**: 4605–4630.
- Dempster AP, Laird NM, Rubin DB. 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc. B MET.* **39**(1): 1–38.
- Forney GD. 1973. The Viterbi algorithm. *Proc. IEEE* **61**(3): 268–278.
- Gelati E, ssing Christensen OB, Rasmussen PF, Rosbjerg D. 2010. Downscaling atmospheric patterns to multi-site precipitation amounts in southern Scandinavia. *Hydrol. Res.* **41**(3): 193–210.
- Gleckler PJ, Taylor KE, Doutriaux C. 2008. Performance metrics for climate models. *J. Geophys. Res.* **113**, doi:10.1029/2007JD008972. D06104.
- Goswami GN, Mohan RSA. 2001. Intraseasonal oscillations and interannual variability of the Indian summer monsoon. *J. Climate* **14**: 1180–1198.
- Greene AM, Robertson AW, Kirshner S. 2008. Analysis of Indian monsoon daily rainfall on subseasonal to multidecadal time-scales using a hidden Markov model. *Q. J. R. Meteorol. Soc.* **134**: 875–887.
- Held IM, Soden BJ. 2006. Robust responses of the hydrological cycle to global warming. *J. Climate* **19**: 5686–5699.
- Hughes JP, Guttorp P. 1994a. A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena. *Water Resour. Res.* **30**(5): 1535–1546.
- Hughes JP, Guttorp P. 1994b. Incorporating spatial dependence and atmospheric data in a model of precipitation. *J. Appl. Meteorol.* **33**(12): 1503–1515, doi:10.1175/1520-0450(1994)033<1503:ISDAAD>2.0.CO;2.
- Hughes JP, Guttorp P, Charles SP. 1999. A non-homogeneous hidden Markov model for precipitation occurrence. *J. Royal Stat. Soc. C* **48**(1): 15–30, doi:10.1111/1467-9876.00136.
- Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y, Chellah M, Ebisuzaki W, Higgins W, Janowiak J, Mo KC, Ropelewski C, Wang J, Leetmaa A, Reynolds R, Jenne R, Joseph D. 1996. The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Met. Soc.* **77**(3): 437–471.
- Leith C. 1975. Climate reponse and fluctuation dissipation. *J. Atmos. Sci.* **32**: 2022–2026.
- Lorenz EN. 1963. Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**: 130–141.
- Meehl GA, Covey C, Delworth T, Latif M, McAvaney B, Mitchell JFB, Stouffer RJ, Taylor KE. 2007. The WCRP CMIP3 multi-model dataset: A new era in climate change research. *Bull. Am. Met. Soc.* **88**: 1383–1394.
- Mehrotra R, Sharma A. 2005. A nonparametric nonhomogeneous hidden Markov model for downscaling of multisite daily rainfall occurrences. *J. Geophys. Res.* **110**(D16108), doi:10.1029/2004JD005677.
- Moron V, Robertson AW, Boer R. 2009. Spatial coherence and seasonal predictability of monsoon onset over indonesia. *J. Climate* **22**: 840–850.
- Naylor RL, Battisti DS, Vimont DJ, Falcon WP, Burke MB. 2007. Assessing risks of climate variability and climate change for Indonesian rice agriculture. *Proc. Natl. Acad. Sci. USA* **104**: 7752–7757, doi:10.1073/pnas.0701825104.

- Norris J. 1997. *Markov chains*. Cambridge University Press: Cambridge. 237 pp.
- Pai D, Rajeevan M. 2007. Indian summer monsoon onset: Variability and prediction. National Climate Centre Research Report 6, India Meteorological Dept., Pune.
- Palmer TN. 1999. A nonlinear dynamical perspective on climate prediction. *J. Climate* **12**(2): 575–591, doi:10.1175/1520-0442(1999)012<0575:ANDPOC>2.0.CO;2.
- Rajeevan M, Bhate J, Kale J, Lal B. 2006. High resolution daily gridded rainfall data for the indian region: Analysis of break and active monsoon spells. *Curr. Sci. India* **91**(3): 296–306.
- Richardson C, Wright D. 1984. Wgen: A model for generating daily weather variables. Technical Report ARS-8, Agriculture Research Service, United States Department of Agriculture.
- Robertson AW, Kirshner S, Smyth P. 2004. Downscaling of daily rainfall occurrence over northeast Brazil using a hidden Markov model. *J. Climate* **17**: 4407–4424.
- Robertson AW, Kirshner S, Smyth P, Charles SP, Bates BC. 2006. Subseasonal-to-interdecadal variability of the Australian monsoon over North Queensland. *Q. J. R. Meteorol. Soc.* **132**: 519–542.
- Robertson AW, Moron V, Swarinoto Y. 2009. Seasonal predictability of daily rainfall statistics over Indramayu district, Indonesia. *Int. J. Climatology* **29**: 1449–1462.
- Sikka DR, Gadgil S. 1980. On the maximum cloud zone and the ITCZ over India longitude during the Southwest monsoon. *Mon. Wea. Rev.* **108**(11): 1840–1853.
- Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt K, Tignor M, Miller H (eds). 2007. *Climate change 2007: The physical science basis. contribution of working group i to the fourth assessment report of the intergovernmental panel on climate change*. Cambridge University Press: Cambridge, United Kingdom and New York, NY, USA.
- Vrac M, Naveau P. 2007. Stochastic downscaling of precipitation: From dry events to heavy rainfalls. *Water Resour. Res.* **43**, doi: 10.1029/2006WR005308. W07402.
- Wang B, Fan Z. 1999. Choice of south Asian summer monsoon indices. *Bull. Am. Met. Soc.* **80**(4): 629–638.
- Wilby R, Dawson C, Barrow E. 2002. SDSM — a decision support tool for the assessment of regional climate change impacts. *Environ. Modell. Softw.* **17**: 147–159.
- Wilks D. 1998. Multisite generalization of a daily stochastic precipitation generation model. *J. Hydrol.* **210**: 178–191.
- Wilks DS. 1992. Adapting stochastic weather generation algorithms for climate change studies. *Clim. Change* **22**: 67–84.
- Woolhiser DA, Roldán J. 1982. Stochastic daily precipitation models 2. A comparison of distributions of amounts. *Water Resour. Res.* **18**(5): 1461–1468.