P 2.10

# USE OF SMALL BASIN NETWORKS FOR MONITORING CONTINENTAL SCALE CHANGES IN THE WATER CYCLE

J. Sheffield* and E. F. Wood.
Princeton University, Princeton, NJ

## 1. INTRODUCTION

Detection of continental scale changes in the water cycle involves monitoring a range of variables over large space and time scales. Because monitoring water cycle variables for entire continents is not feasible at this time, the use of large-scale basins, such as those under the GEWEX umbrella (Morel 2001), for detection of change is attractive because they cover a significant part of their respective continents. To do so, however, we must assume that changes occurring in the GEWEX basin are representative of those occurring on the continent as a whole. Numerous GCM studies of global warming suggest, however, that the ensuing hydrological cycle changes will vary from region to region (IPCC 2001, chapter 9). Thus, the changes occurring in any one region (up to some threshold spatial scale) could differ substantially from that occurring for the continent as a whole. In this study, the use of existing small scale monitoring basins within the USA as indicators of change in the continental hydrologic cycle is investigated. Using a genetic algorithm we select a small number of "indicator stations" that are statistically representative of the USA as a whole using observed precipitation and streamflow records.

## 2. DATA

The dataset used as the basis for this work is the hydroclimatology dataset of Wallis, Lettenmaier and Wood (1991). The dataset consists of long term records (1948-1987) over the continental United States of precipitation, taken from 1036 National Oceanic and Atmospheric Administration (NOAA) meteorological stations, and stream flow records from 1009 U.S. Geological Survey (USGS) stations. The observations have been compiled into a consistent daily database with missing data estimated using simple nearest neighbour algorithms.

Figure 1 shows the distribution of precipitation stations over the United States. The locations of the streamflow stations were chosen to be as free of human influence as possible and the distribution of these stations is shown in Figure 2. Streamflow stations are located in all regions but are most dense in the more humid areas and less dense in arid and semi-arid areas where regulation is more prevalent.

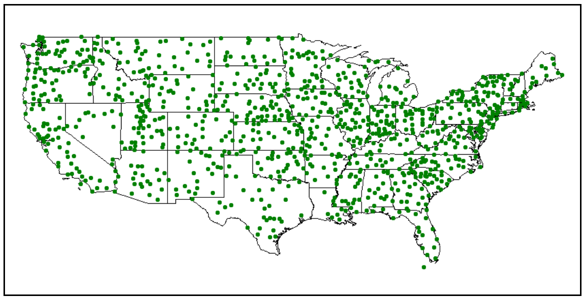

Figure 1. Distribution of precipitation stations.

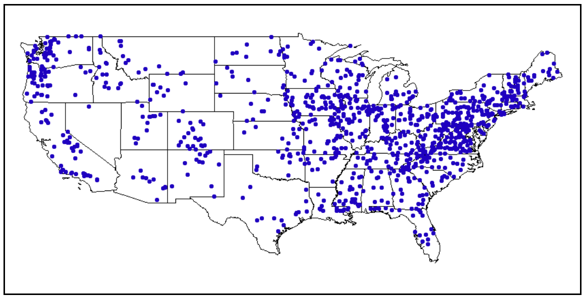

Figure 2. Distribution of streamflow stations.

## 3. METHOD

To investigate the existence of hydrological indicator stations, we select *m* stations (equal to 5% of the entire population), for which the mean of the annual time series is "closest" to the population mean. This is carried out separately for the precipitation and streamflow data. Closeness is quantified by calculating the mean squared error for all the years in the time series:

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(\overline{p}_{all,t} - \overline{p}_{m,t})^2 \qquad (1)$$

where $\overline{p}$ is the mean precipitation or streamflow at the stations comprising either the *m* indicator stations or the entire population (*all*). The *m* indicator stations are determined using a genetic-type algorithm that generates a set of solutions to the objective function represented by Equation (1). Each solution is analogous to a binary strand of genetic material, having an associated MSE value. New generations of solutions are created by "mating" two of the better solutions, and by allowing mutations (solutions chosen at random). The potential indicator stations are those producing the lowest MSE after 10,000 generations.

---

**Corresponding author address*:
Justin Sheffield, Dept. of Civil Engineering, Princeton University, Princeton, NJ, 08544; e-mail: justin@princeton.edu.

## 4. RESULTS

Two sets of stations, which mimic the mean behaviour of the whole population of stations, were selected using the genetic algorithm, separately for the precipitation and the streamflow stations. Maps of the selected stations are shown in Figures 3 and 4 for precipitation and streamflow, respectively. It can be seen that the indicator stations are distributed across the country and cover the major climatic zones of the USA.

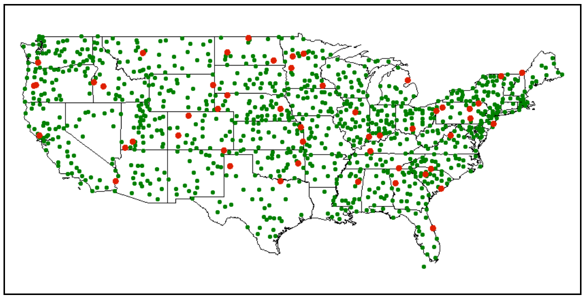

Figure 3. Distribution of indicator stations (red circles) and all stations (green circles) for precipitation.

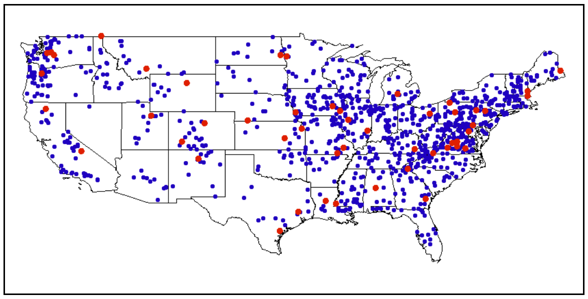

Figure 4. Distribution of indicator stations (red circles) and all stations (blue circles) for streamflow.

The indicator stations are representative of the larger spatial area in terms of the their comparative MSE values. However, to be useful in detecting changes in the hydrological cycle, the stations must also mimic the variability of the larger scale as well as the mean behaviour. Table 1 shows the mean and variance calculated for the total population and the indicator stations. The statistics show a close match in the mean and variance for both precipitation and streamflow.

| | | Mean | Std Dev | Variance |
|---|---|---|---|---|
| P | Population | 823.2 | 63.9 | 4085.1 |
| | Indicator | 822.9 | 64.0 | 4096.0 |
| Q | Population | 500.9 | 66.1 | 4363.9 |
| | Indicator | 492.4 | 66.6 | 4434.7 |

Table 1. Comparison of the statistics of the population and indicator stations for precipitation and streamflow.

## 5. CONCLUSIONS

A genetic algorithm was used to identify two sets of indicator stations that collectively mimic, respectively, the precipitation and streamflow regimes of the United States as characterized by observations from 1036 precipitation and 1009 streamflow stations. Results indicate that a small number of representative stations scattered over a range a climate zones can be used to collectively mimic the behaviour of the USA as a whole by encompassing the mean and variability. As many of these stations are still operational and may be so for the foreseeable future, they may provide an efficient monitoring network at no extra cost that is capable of providing a surrogate view of the continental hydrologic cycle.

Such networks of stations may provide the key to rapid detection of possible changes in the continental hydrological cycle. The length of records required to identify significant changes in the hydrological cycle is determined by the strength of the signal and the level of noise or variability within the data. Ziegler et al. (2001) determined that decades to centuries of records were required to detect statistically significant trends in the continental hydrological cycle, based on GCM predictions of future climate. They also found that by using networks of small indicator “basins”, which were representative of continental behaviour, a similar amount of time to detection was required. One could envisage that networks of small numbers of stream gauging and meteorological stations, distributed over the continents, and augmented with remote sensing and modeling, could provide a framework for rapid detection of changes in the in continental water cycle.

## ACKNOWLEDGEMENTS

This work was funded by NASA grants NAG5-9486, NAG5-9414

## REFERENCES

IPCC, 2001: *Climate Change 2001: The Scientific Basis*, Houghton, J. T., Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, and D. Xiaosu, Eds., Cambridge University Press, 944 pp.

Morel, P., 2001: Why GEWEX? The agenda for a global energy and water cycle research program. *GEWEX News*, Twitchell, P., Ed., **11**, 1, 7–11.

Wallis, J. R., D. P. Lettenmaier and E. F. Wood, 1991: A daily hydroclimatological data set for the continental United States. Water Resources Research, **27**, 1657-1663.

Ziegler, A.D., J. Sheffield, E.F. Wood, B. Nijssen, E.P. Maurer and D.P. Lettenmaier, 2002, Detection of intensification of the global water cycle: the potential role of FRIEND, International Association of Hydrological Sciences (IAHS) Redbook publication No. 274, 51-57.