ELSEVIER

Journal of Hydrology 239 (2000) 162–178

Journal of Hydrology

www.elsevier.com/locate/jhydrol

A quantitative precipitation forecast experiment for Puerto Rico

M.M. Carter[a,*], J.B. Elsner[b], S.P. Bennett[c,1]

[a]*Department of Meteorology, Florida State University, Tallahassee, FL 32306, USA*
[b]*Department of Geography, Florida State University, Tallahssee, FL 32306, USA*
[c]*NOAA National Weather Service, Brownsville, TX 78521, USA*

Received 23 November 1999; revised 12 July 2000; accepted 20 September 2000

Abstract

A quantitative precipitation forecast (QPF) experiment is conducted for the island of Puerto Rico. The experiment ranks the utility of six objective rainfall models and an operational forecast issued by the National Weather Service Forecast Office, San Juan. This is believed to be the first experiment to rank the utility of rainfall forecast schemes in the tropics.

Using an analysis of variance tool called common factor analysis (CFA), the island of Puerto Rico is divided into convective rainfall regions. These regions are statistically independent and represent the forecast domains for the experiment. All forecasts and realizations are area-averaged over each convective region.

The QPF experiment is conducted in real-time over three 6-week periods in 1998. The periods fall in three separate rainfall seasons. All seven forecast schemes are configured to produce an area-averaged 24 h rainfall forecast. Forecasts are realized through a network of 114 data rain gauges, whose 24 h values are also area-averaged within convective region. We conduct this experiment in a Bayesian framework. Users may determine the ex ante value of forecast products through the Bayesian correlation score (BCS). Over each of the three seasons, the climatology forecast held the highest ex ante utility for users. Although objective forecast utility scores for heavy rain events are low, they yield higher BCS values than operational forecasts.

Keywords: Atmospheric precipitation; Prediction; Tropical; Model studies

1. Introduction and motivation

Making a quantitative precipitation forecast (QPF) is among the most difficult operational tasks at the San Juan National Weather Service Forecast Office (WFO). Not only is making a QPF more challenging than it is on the mainland, the consequences are greater. Puerto Rico has the greatest recurring threat to life and property due to flash flooding of any state or territory under the jurisdiction of the National Weather Service (Carter, 1997). By the time an unexpected heavy rain event becomes imminent, it is often too late to take protective actions against resultant flash flooding (Krzysztofowicz et al., 1993).

The San Juan WFO currently makes QPFs at 0000 and 1200 UTC every day. At 0000 and 1200 UTC, gridded output from the National Centers for Environmental Prediction (NCEP) 48 km Eta and 48 km nested grid models (NGM) provide objective QPFs. Forecasters qualitatively assimilate the model output and large-scale weather features to issue a subjective QPF. The forecast takes the form of a contoured

* Corresponding author. Worldwide Weather Trading Co., 20 Exchange Place, 33rd Floor, New York, NY 10005, USA. Fax: +1-212-785-2270.

E-mail address: mcarter@weathertrading.com (M.M. Carter).

[1] Previously at: NOAA National Weather Service, Carolina 00979, Puerto Rico.

0022-1694/00/$ - see front matter
PII: S0022-1694(00)00349-8

isohyet map, and is valid for the 12 h immediately following 0000 and 1200 UTC. The subjective QPF is sent to the NWS River Forecast Center in Atlanta, GA, where river responses are forecast using hydrological models tuned to Puerto Rico's river basins. The river response forecast is sent back to the San Juan WFO where, in conjunction with the subjective isohyet forecast, the QPF product is issued.

Technically, quantitative precipitation forecasting is the projection of precipitation amount at a point or region. QPFs are single-valued forecasts over a fixed period of time, usually 12 or 24 h. Since 1960, The NCEP Meteorological Operations Division (MOD) has produced QPF guidance for its WFOs and river forecast centers (RFCs) (Olson et al., 1995).

The relative accuracy of the Eta QPF to other NCEP products and subjective QPFs is difficult to ascertain, despite studies chronicled in meteorological literature. Several researchers have concluded that subjective QPFs, when forecasters are properly conditioned, outperform objective model guidance (Olson et al., 1995). Murphy and Winkler (1977) demonstrated the reliability of subjective precipitation forecasting in the presence of rugged terrain. Rugged topography dominates the weather physics of many locales in Puerto Rico, so we anticipate a similar level of skill by operational forecasters in San Juan.

Subjective QPF experiments do not always reveal operational forecasting accuracy, however. Murphy and Daan conducted a pair of experiments in which forecast probabilities consistently exceeded observed frequencies (Murphy and Daan, 1984; Murphy et al., 1985). The resulting error is called value-induced bias, where forecasters opt on the side of life and property protection.

The discovery of value-induced bias in previous experiments provides the impetus for developing a QPF experiment in Puerto Rico. Heavy rainfall events cause destructive flash floods, but economic costs also exist for over-forecasting as well. The hypothesis of this experiment is for heavy rain events in Puerto Rico above a predetermined threshold, an objective QPF will provide a higher Bayesian correlation score (BCS) than a subjective QPF. The BCS is a utilitarian measure of forecast skill. It is a score of association that enables us to rank the value of each forecast system in this experiment.

The hypothesis is based on the posit that value-induced bias will be omnipresent in subjective QPF forecasts, and may be further amplified in heavy rainfall situations. To test the hypothesis, a QPF experiment is conducted in real time. This experiment contains seven forecast schemes, each producing the same product: a 24 h area-averaged QPF. It was run concurrently with QPF procedures in San Juan, enabling us to assimilate current operational QPFs into the experiment. In this way, we test the hypothesis directly.

2. Tropical rainfall over Puerto Rico

2.1. Geography

The commonwealth of Puerto Rico encompasses the easternmost islands of the Greater Antilles. Although most know Puerto Rico by its main island, the US protectorate also includes Isla Mona east of Dominican Republic, and Vieques and Culebra west of the US Virgin Islands (USVI). All three outlying islands are sparsely populated and have a gently rolling topography. These islands comprise a small fraction of the geography of Puerto Rico. Unless denoted otherwise, Puerto Rico will henceforth refer to the main island of the commonwealth.

The Caribbean Sea and the Atlantic Ocean border Puerto Rico to the south and north, respectively. The mountain range Cordillera Central is the island's spine, reaching a height of 1338 m at its highest point (Pico, 1974). The windward-facing eastern slopes of the Luquillo range in the northeast are very steep, and the resulting orographic lift leads to local annual rainfall amounts in the hundreds of centimeters. Coastal plains describe Puerto Rico's land–sea border on the north and west coasts. The San Juan metropolitan area encompasses a significant part of the northeast, and is home to the island's National WFO. The island's southern coast is carved out as a valley by fingers of the Cordillera Central that extend to the Caribbean Sea. Ponce is the largest city on Puerto Rico's southern coast, and contains an automated surface observing system (ASOS). Other ASOS units are found at Mayaguez, Aquadilla and Roosevelt Roads. A map of elevation and ASOS stations on Puerto Rico is shown in Fig. 1.

Vegetation on Puerto Rico reflects the diverse

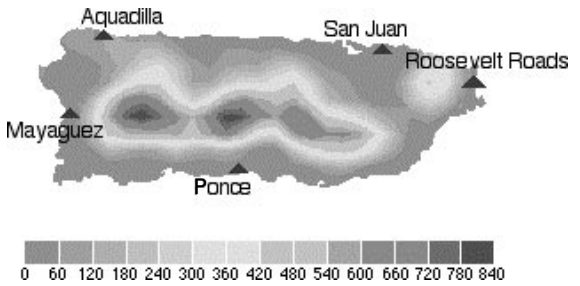

Fig. 1. ASOS plotted on an elevation map of Puerto Rico. ASOS stations measure atmospheric variables at the surface for a number of variables. Elevation contours are provided in the legend and given in meters.

rainfall regimes over very small areas. Luquillo's heavy rainfall has already been mentioned. Large annual rainfall totals are also found in the forested western Cordillera Central. Rainfall runoff into the southwest plain nurtures the island's abundant sugar cane fields. In contrast, the southern coastal plain to the east is sparsely vegetated and shadowed by the brown mountain faces to the north.

2.2. Rainfall mechanisms

Puerto Rico's rugged terrain and land–sea interfaces give rise to locally driven weather events (Carter, 1995). Although small diurnal and annual variations make temperature forecasting tractable, rainfall prediction is much more challenging. Diurnal effects strongly influence Puerto Rico's rainfall regime (Hamilton, 1981). Early studies picked up on what is true throughout the island: rainfall intensity is greatest in the afternoon while frequency varies in time and space (Ray, 1928). Updated histograms of diurnal rainfall amounts and frequencies can be found in Carter (1995). In one of the earliest studies of rainfall in Puerto Rico, Fassig (1916) noted that the average duration of rain events in San Juan, over all months of the year, is less than 1 h.

Regions of Puerto Rico subject to sea and land breezes exhibit a low diurnal frequency; few hours of the day are more disposed to rain events than are others. San Juan typifies such a regime. Thermal circulations converge with prevailing trade winds leading to late evening showers. The reverse land breeze circulation leads to early morning rain events. Showers may occur throughout the day, however, which flattens the diurnal frequency histogram.

Mountainous regions of Puerto Rico give rise to valley breezes that precipitate rainfall on the island's interior (Pico, 1974). Mountain peaks heat rapidly during the day and drive upward-sloping thermal circulations that converge at the summit. These peaks act as foci for solar insolation, releasing energy from convectively unstable conditions. In fact, the tropical atmosphere over Puerto Rico is nearly always convectively unstable (Riehl, 1947). The perpetual convective instability borne of the moist troposphere cannot itself explain the occurrence of rainfall on the island. Yet it is a necessary ingredient for the occasion of rain.

The mountains of Puerto Rico also act to block and redirect prevailing easterly winds. The wind flow is forced up and over the eastern slopes of the Cordillera Central and Luquillo ranges. It undergoes buoyancy oscillation intrinsic to internal gravity waves. As these waves propagate past the western slopes of the Cordillera Central, it forces rising motion into a progressively unstable atmosphere. The resulting rainfall focus east of Mayaguez is nearly stationary and leads to very high rainfall totals. The Mayaguez standing gravity wave was discovered in a mesoscale-modeling project (Bennett et al., 1997), and the attending rain focus can be clearly witnessed from a nearby dry vantage.

It is common for the downslope coastal areas of the island to experience subsidence at the time of interior thunderstorm activity. Through an experiment utilizing four helical aircraft soundings, Malkus (1955) showed that streamlines calculated from momentum equations closely matched the observation that interior convection induces compensating subsidence in the coastal areas. Furthermore, the fact that this subsidence is a result of efficient vertical moisture transport in the interior may lead to both local drought conditions and a disruption of the larger scale flow over the island.

The above descriptions of rainfall mechanisms represent just the mesoscale forcings. Synoptic scale phenomena of longer duration contribute significantly to Puerto Rico's rainfall patterns. Summertime easterly waves pass off the West Coast of Africa and propagate over the island of Puerto Rico with a frequency of 3–5 days. Localized phenomena,

however, make it difficult to attribute rainfall to them. Large-scale cloudiness may accompany the passage of easterly waves over the island, suppressing insolation and weakening the sea breeze. On the other hand, upper level divergence may provide a rainfall-releasing mechanism (Carter, 1995).

Of more rainfall-inducing certainty is the development of easterly waves into tropical cyclones. Every summer, Puerto Rico is at risk from tropical storms and hurricanes. Associated heavy rainfall with such storms owes both to convective rain bands as well as intense orographic lifting. Thus, tropical cyclones need not make direct landfall on Puerto Rico to cause heavy rain events and flash flooding (Carter, 1997).

The tropical upper-tropospheric trough is a climatological feature of the western Atlantic that influences rainfall in Puerto Rico. The TUTT provides a preferred zone for the spawning of cold-core upper-level cyclones (Frank, 1970; Kousky and Gan, 1981). These cyclones behave similarly to passing easterly waves, though they may linger for longer periods of time. Upper level mid-latitude troughs and surface cold fronts occasionally impinge on the eastern Caribbean during the winter. They may linger to the north and west for days at a time, providing a focus of upper-level divergence and low-level moisture convergence. An upper-level trough was the culprit for the Three Kings Day flood in January 1992 that claimed 23 lives. The island of Puerto Rico has never experienced a freezing temperature and has never seen even a trace of snow (Pico, 1974). Therefore, all precipitation modeled in this study is rainfall.

2.3. Complications

The challenge of diurnal rainfall prediction is framed in its sensitivity to the release of moist static energy in the surface layer. The frequency of this releasing mechanism is not periodic or uniform. This fact complicates the statistical modeling of rainfall. Dynamic models, such as NGM and Meso Eta used in this experiment, utilize the parameterization of first-principle atmospheric conservation laws, thereby drawing a contemporaneous relationship between physics and rainfall. Statistical models are subject to a different set of assumptions, primarily the normality of frequency distributions, which do not exist for rain on a small time scale.

Although Puerto Rico receives rainfall virtually every day of the year, at any given station the most frequent daily rainfall amount is zero. Over 95% of all hours are rain free. Convective rains spawned from mesoscale island features develop and decay in the span of a few hours. Since these rain events are spatially small, few hours in total receive any rain. This behavior decays on the diurnal time scale. It is almost equally likely that a rain gauge will receive no rain over the course of a day as will receive a small amount (less than 0.51 cm). These two events together comprise over 55% of all event frequencies. On the monthly and annual time scales, however, rainfall is near-normally distributed. Even for a point location, it is very rare for at least one day out of each month to receive no rain. Since zero is no longer a significant frequency of rainfall, monthly and annual rainfall distributes symmetrically about a mean.

3. Convective rainfall regions

Because rainfall in Puerto Rico is convective in nature, it is much more tractable to forecast for a small region than an individual gauge. Convective regionalization is a means to represent shared rainfall characteristics among nearby stations. These common forcings are manifest in the variance structure of daily rainfall records. Stations within a convective region exhibit similar variance structures and are assumed to share responses to antecedent atmospheric conditions. As such, convective regions in Puerto Rico serve as forecast domains in this QPF experiment.

Because river flow rates are dependent on rainfall, QPFs are often made for river basins instead of point gauges (Krzysztofowicz and Sigrest, 1997). The demarcation of common rainfall signatures, however, may not coincide with river basin boundaries. Topography is a consideration in defining a rainfall region, but may induce a local effect on one part of the island while also affecting a region downstream. Convective regionalization takes into account geographically diverse effects.

As in a previous regionalization of rainfall in Puerto Rico, common factor analysis (CFA) is used to define region boundaries (Carter and Elsner, 1996). Factors

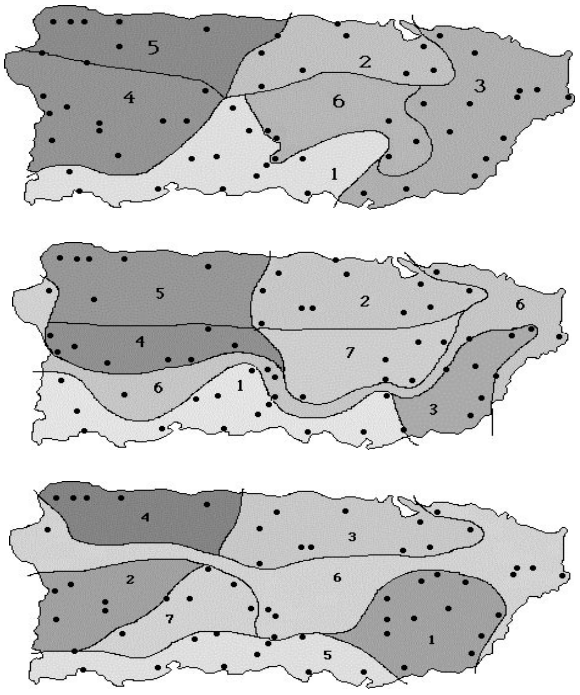

Fig. 2. Regionalization of Puerto Rico's daily rainfall based on a common factor analysis model for winter (top), transition (middle) and summer (bottom) seasons. Each region is a forecast domain in the real-time QPF experiment.

are intangible entities, much like principal components, that cannot be observed but act as surrogates to that which can be observed. Variables represented by a common factor are highly correlated with each other while remaining disparate from other variables. Common factors are extracted from a correlation matrix of observed variables, and number between 1 and p, where p is the number of variables (Johnson and Wichern, 1982). In this study, the variables of interest are station daily rainfall amounts and the resulting regions are given in Fig. 2.

We construct three sets of regions, one for each season in the experiment. During the winter, intrusive mid-latitude systems interact with mesoscale land–sea interactions. Regions from this season depict the large-scale blocking of wind flow from these systems by the Cordillera Central. Regions on the lee side of the mountain spine remain dry while regions on the windward side are wet. Transition season regions reveal considerable interaction between winter systems from the northwest and incepting summertime easterly waves. Summertime regions are defined by the near-absence of mid-latitude troughs as well as the northern coastal plain sea breeze circulation and its climatological coastal subsidence (Malkus, 1955). The standing gravity wave denotes the split of regional boundaries on the island's western interior.

4. Experiment design

We present a framework in which the real-time QPF experiment is carried out. The methodology for this experiment begins with the assembly of data from diverse resources. The procedures of the experiment design follow, and we provide the details of the Bayesian framework in which this study is conducted.

4.1. Data description

This QPF experiment integrates an extensive collection of surface, upper air, rainfall and model data. These data represent both real-time observations as well as 20 years of historical records. They range in spatial resolution from a pair of rain gauges a mere 200 m apart to upper air sounding stations 1900 km from San Juan. They include a predictor pool of 162 upper air and 84 surface variables, both derived and observed. Despite the breadth of these variables in the experiment, lack of data is the largest obstacle to successful QPF prediction in the tropics. Many island nations and territories comprise the Caribbean region and each has its own approach to atmospheric data collection. Some stations launch radiosondes only once a day. Others do not report sea-level pressure. Furthermore, entire months of data may be missing from historical records and until recently, hours in which everybody reported valid data were rare. The assembly of these data was a formidable task of this experiment borne from a number of different sources.

4.1.1. Rainfall data

The National Climatic Data Center (NCDC) archives daily rainfall data for Puerto Rico collected by the United States Geological Survey (USGS). These data are measured by the data collection platform (DCP), a network of 114 rainfall gauges distributed throughout the island. Gauges are added and taken away, so the number of gauges whose rainfall data is available differs in time. For the

years 1974–1993, 61 stations in the DCP have data records in the same location of at least 10 years in continuation. Of these stations, 55 have continuous records for the entire 20 year period with missing daily values of not more than 10%. The gauges report their information electronically to the USGS in the San Juan suburb of Guayama four times daily at or near 0300, 0900, 1500 and 2100 UTC.

Because the DCP data were also available to this QPF experiment in real time, they have not undergone a rigorous quality-control procedure. Close inspection of the data and frequent interaction with San Juan WFO served as the operational means of validation. DCP data and their attendant errors were further smoothed through a Thiessen area averaging. The experimental assumption is that instrument-based errors are random and form a gaussian frequency distribution over time. This assumption is infeasible to verify because an archive of current DCP data does not extend back to the most recent issuance by NCDC of their quality-controlled data. We recognize the fragility of the assumption of random instrument errors and seek to investigate this matter in the future when such data comparison is possible.

4.1.2. Surface and upper air data

The role of surface data pervades the QPF experiment, from statistical model building to climatology stratification. Hourly surface data include wind, temperature and dew point information that contributed to the pool of possible rainfall predictors. These data include stations from across the Caribbean, as well as five surface observation sites on the island of Puerto Rico. The National Weather Service maintains an ASOS platform at each of the five sites on the island as well as two in the USVI. Separate governments are responsible for hourly surface observation sites found on island nations throughout the Caribbean.

We retrieved surface data from 14 stations in the Caribbean (including Puerto Rico and USVI). Retrieval was made possible through the National Center for Atmospheric Research. The stations include Antigua, Aquadilla, Charlotte Amalie, Christiansted, Guadeloupe, Martinique, Mayaguez, Montego, Ponce, Roosevelt Roads, St. Kitts, St. Lucia, St. Martin and San Juan. Because of the discrepancy in retrieval schedules, the number of missing data varies from station to station. Under the auspices of the National Weather Service, the stations in Puerto Rico and the USVI possess the most complete records. Records from island nations are less comprehensive, report fewer variables, and have larger data gaps than their NWS counterparts. Surface temperature, dew point temperature and winds are available at all of the stations, with surface pressure available only at Aquadilla, Christiansted, Roosevelt Roads and San Juan.

In Puerto Rico, only the San Juan WFO possesses a comprehensive upper air data history. In addition to San Juan, four stations throughout the Caribbean provide sufficient historical upper air records for use in model building. These stations include Key West, Florida; Montego, Jamaica; Princes Juliana, St. Martin; and Grantley Adams, Barbados. Their records encompass the years 1977–1993 and were selected for their completeness of data as well as their geographical range.

4.1.3. Objective model data

Three of the prediction schemes in this experiment came from operational forecast models at the NCEP. The 48 km Eta, 48 km NGM and 29 km Meso Eta run twice daily. Each of these models forecasts precipitation amounts over a fixed time period. The Eta and NGM models forecast 12 h precipitation amounts for each inception at 0000 and 1200 UTC. The 0000 UTC run is not available to San Juan operational forecasters in time to issue the 0000 UTC QPF product. Therefore, to emulate product availability to these forecasters, the 1200 UTC previous day runs of the NGM and Eta models were used in this experiment. The models forecast precipitation amounts for the periods of 12–24 h and 24–36 h after inception. By summing these totals, we get a 24 h QPF from 0000–2359 for the forecast day.

The 29 km Meso Eta has replaced the 48 km Eta model as the NCEP “early look” run (Black, 1994). Delivery of the Meso Eta to San Juan WFO, however, has been difficult to implement. The 1500 UTC run coincides with heavy afternoon Internet traffic, and electronic infrastructure on the island lags the continental United States. Preferential ingesting of the Meso Eta model “bumps” more vital NCEP products from the local data manager platform (LDM). Dropping forecast products superfluous to

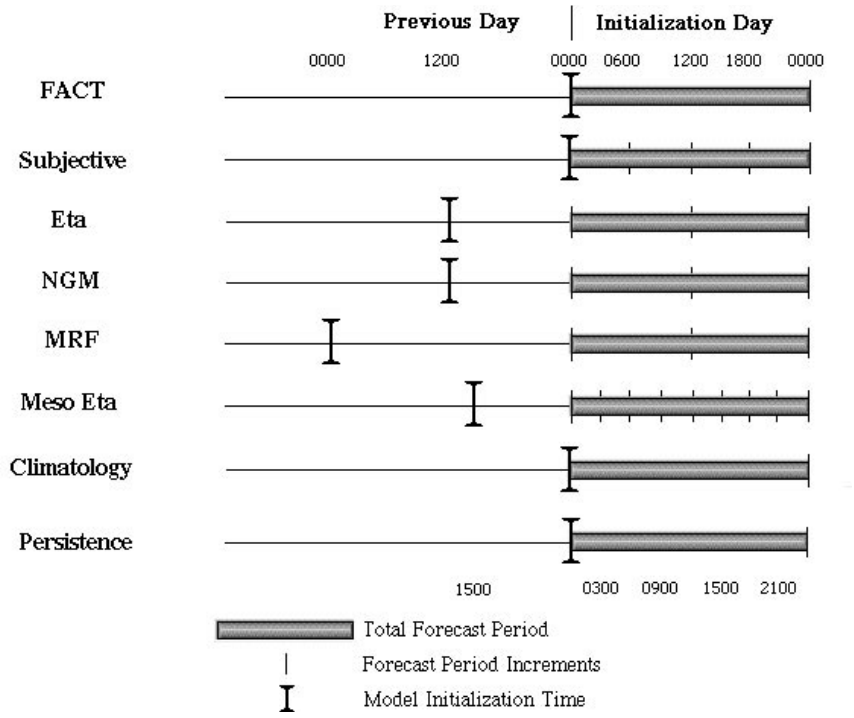

Fig. 3. Model inception and forecast periods for seven forecast schemes in the QPF experiment. Capital "I's" indicate model initialization times and shaded bars indicate total forecast periods. Small tics represent the period segments for which each model forecasts. Predictand is 24 h area-averaged rainfall beginning at 0000 UTC.

operational duties has recently mollified this dilemma. The Meso Eta is available to San Juan WFO beginning April 1, 1998. It was therefore included in the transitional season and summer forecast periods of this QPF experiment. NCEP runs the Meso Eta twice a day at 0300 and 1500 UTC. The 1500 UTC previous-day run was used in this experiment and forecasts 3 h rainfall amounts up to 36 h after inception. Forecast periods were 12, 15, 18, 21, 24, 27, 30 and 33 h following the 1500 UTC inception. The eight 3 h precipitation totals were summed to get a 24 h QPF. Fig. 3 shows a timeline of the forecast schemes in this experiment, with their inceptions and encompassing forecast periods.

4.1.4. A statistical model

In a previous study, a statistical approach performed skillfully in predicting two-category daily rainfall for Puerto Rico (Carter and Elsner, 1997). The motivation for constructing a statistical model for this experiment lies in the assumption that over time, tropical rainfall mechanisms exhibit preferred behavior under like conditions. Furthermore, operational forecast guidance available to the San Juan WFO does not include a stand-alone statistical model. Armed with direct observations and a derivation of meteorological fields, a statistical approach allows for an objective selection of those atmospheric variables that tend toward rainfall.

In constructing a model for this experiment, we seek a methodology that is both useful in its interpretability and subject to statistical significance testing by way of cross validation. Moreover, a model that allows for nonfunctional relationships between predictor and predictand reveals subtle bifurcation in the forecast system that may be lost in a single, best-fit equation. This bifurcation may separate classes of rainfall. Building on a previous study in statistical rainfall prediction (Carter and Elsner, 1997), we choose a classification algorithm known as partially adaptive classification trees. An algorithm, fast action classification

trees (FACT), was developed at the University of Wisconsin and is used in this study (Loh and Vanichsetakul, 1996).

A classification tree is named as such because the structure of bifurcated classification rules geometrically resemble the branches of a tree. A classification tree is fit to predictors/categorical rainfall for each convective region during each season. There are 20 trees in all, and therefore, one set for each 24 h area averaged categorical forecast (six for each winter day, seven for each transitional day and seven for each summer day). From a set of predictors and categorical rainfall, FACT fits classification trees to historical data for each region in each season. The resulting tree produces a set of rules that may be followed with real-time observations. The obedience of these rules produces a 24 h area-averaged categorical QPF.

4.2. Experiment parameters

Three NCEP model products (Eta, Meso Eta and NGM models), FACT, climatology, persistence and subjective forecasts comprise the components of the experiment. All products are formatted in this study to produce a 24 h categorical forecast beginning at 0000 UTC for each convective region on the island. Because forecasts have different ex ante values for disparate users, this experiment is conducted in a Bayesian framework.

An initial decision of import is whether to carry out the experiment as hindcasts, or to conduct them in real time. The advantage of the hindcast approach is control over experiment elements. A day in which a product is missing or unavailable may be discarded in favor of a day with complete information. Unfortunately, the discarded day may have been exceedingly wet or very dry in a spell. In fact, an extreme event is often the cause of a missing data point. In summarily disposing a hindcast day, the forecaster assumes to assign utility to the forecast. This is not consistent with Bayesian methodology. Moreover, an operational forecaster does not have the luxury of omitting a forecast for a day deprived of model guidance. The forecaster must carry out her/his duties in recognition that all users place utility in forecasts. Conducting an experiment in real time emulates the challenge of operational forecasts, and describes utility better than hindcasts. No days are dismissed from the experiment. The previously mentioned experiments by Krzysztofowicz et al. (1993); Murphy and Winkler (1977) and Murphy and Daan (1984) measured operational forecast success in real-time, providing impetus for us to follow suit. Accordingly, the QPF experiment for Puerto Rico described in this study is conducted in real-time.

4.2.1. Bayesian framework

Metrics of distance that measure accuracy, such as Brier score, are insufficient for improving a forecast system (Sanders, 1963). Distance metrics and normalized distance metrics inform forecasters of skill, but ensuing system improvement may not imply improvement to the user (Krzysztofowicz, 1992). Moreover, these metrics say nothing about forecast certainty. In fact, a forecast system that does not take into account uncertainty as well as historical information and experience will never exceed the economic value of a system that does. This is the primary thesis of Bayes Theorem (Krzysztofowicz, 1983).

By not accounting for uncertainty, an opportunity loss, or the expected risk of making an uncertain forecast, will always be incurred regardless of the forecast system. Therefore, the value of the forecast could actually be negative. From Krzysztofowicz (1983), Fig. 4 shows the values of several forecast procedures. Suboptimal forecasts systems ($\bar{V}_C$) do not consider forecast error. Optimal forecasts (V_C) take uncertainty into account. Naive (V_N), or strictly climatological, forecasts in an optimal setting always have a value of zero. It provides prior information to a forecaster, which neither adds to the forecast value nor subtracts from it. The three systems measure the same predictand: temperature extrema at the Denver WFO. The ex ante value of the suboptimal forecast is less than the optimal naive forecast for error $\sigma^2 > (91°\mathrm{F})^2$. More generally, the forecast error variance is greater than the difference between the sample variance of the variable and the mean forecast error squared. Consequently, the opportunity loss may be eliminated if the forecaster disseminates posterior means of the variable instead of just the categorical forecast output (Krzysztofowicz, 1983). Because a Bayesian system relates system uncertainty to forecast uncertainty, it is optimal.

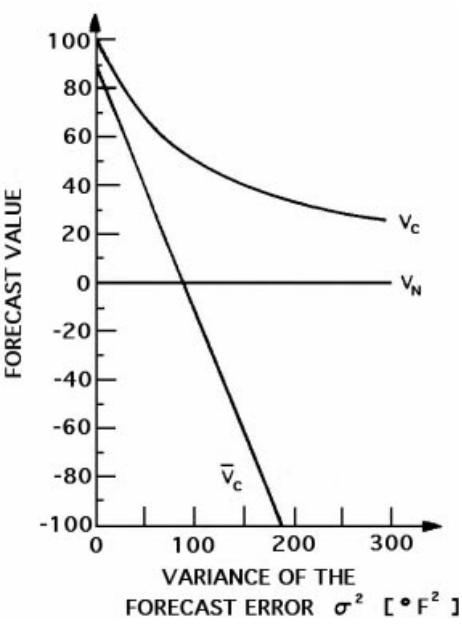

Fig. 4. Ex ante forecast value of three systems. Shown are optical categorical (V_C), suboptimal categorical ($\bar{V}_C$), and optimal naïve (V_N), For systems in which forecast error is not considered, ex ante value may be negative for large-variance systems. The figure is adapted from Krzysztofowicz (1983) and refers to forecasts of temperature extrema at the Denver WFO. The units on the abscissa are degrees Farenheit squared.

4.2.2. Thiessen averaging

Except for FACT, none of the forecast systems produce a convective region area-averaged QPF. NCEP QPF values are assigned to grid points. Operational QPFs are computed as basin averages. Climatology and persistence are calculated directly from individual DCP rain gauges, as are realization values. Each of these algorithm products must be area-averaged within each convective region.

We employ the Thiessen method for all area-averaging tasks in this experiment. In recognition that rain gauges are not distributed uniformly, Thiessen method assigns a weight to each station based on a representative area within a boundary (Thiessen, 1911). Within each convective region, stations are plotted on a map, and proximal stations are perpendicularly bisected. These bisectors are connected to form a polygon whose area may be expressed as a percentage of total area enclosed by the boundary. These fractions serve as the station weights.

Actual rainfall values for each station are multiplied by their attendant weight, and summed over the entire region. In this way, we calculated an area-average realization and persistence forecast for each region.

NCEP model QPF values are uniformly spaced, unlike the DCP stations. Because the convective region boundaries are complex polygons, however, the grid points are not uniform within region. Again, a Thiessen analysis is employed. For the 48 km Eta and 48 km NGM, adjoining grid points were bisected and given an interpolated value based on a center-differencing application. The Meso Eta requires no such bisection, as gridspace resolution is sufficiently high to carry out a Thiessen averaging. The grid points were treated as stations in the previous Thiessen averaging, inhabiting unique polygons. Weights were assigned to each grid point and each interpolated grid point. Gridpoint values were multiplied by their corresponding weights and summed to produce a convective region area-average.

Finally, the dense network of DCP stations was used to convert operational QPF river basin area-averages to convective region area averages. The Thiessen method was applied in reverse: polygons were constructed for DCP stations within each river basin, and unique station values were calculated by multiplying their Thiessen weight by the basin average precipitation. Next, these stations were placed in convective regions, and new Thiessen polygons were formed. Values calculated from the inverse-Thiessen for each DCP station were multiplied by their new weights, and summed over each convective region. The result is a convective area-average operational QPF. We now have convective area-averaged QPFs, in inches, for each of the forecast systems.

4.2.3. Additional considerations

The experiment predictand is categorical rainfall. Following the Pittsburgh experiment (Krzysztofowicz et al., 1993), rainfall is categorized into six intervals as shown in Table 1. Except for FACT, the forecast algorithms produce a 24 h QPF amount. Rainfall intervals are assigned to the training data set and programmed into the model parameters.

The experiment is divided into three 6-week seasons. Each is chosen as a representation of the evolution of annual rainfall forcing mechanisms. Henceforth, the term season will refer to a 6-week

Table 1
Rainfall predictand categories and their corresponding amount intervals

Category	Rainfall interval (cm)
1	< 0.025
2	0.025–0.63
3	0.64–1.26
4	1.27–2.53
5	2.54–5.07
6	> 5.07

period in the forecast experiment unless expressed alternatively. The experiment took place in 1998. The winter season began on February 15. QPFs were made daily for 42 consecutive days ending on March 28. The transition season took place in April and October, encompassing the 42 days of April 1–28 and October 1–14. The summer season ran from July 12 through August 22, and like the two previous seasons, encompassed 42 days.

All forecasts were attempted for each day in the experiment. Missing forecasts are an inevitable byproduct of data collection. Occasionally, an NCEP model run was not made for a particular day. In addition, a surface or upper air observation critical to the FACT decision trees may have been missing. Data transmission problems also led to the unavailability of the subjective QPF product to this experiment. These are hazards with working in real-time. Table 2 shows the number of missing products for each forecast system. If a realization was missing, all forecasts were considered missing, as full verification is impossible. Forecast seasons were designed of sufficient length to ensure statistical sampling in lieu of sporadic missing products.

Table 2
Percent of present forecast/realization pairs for each scheme by season

Forecast scheme	Winter	Transition	Summer
Eta	7.1	2.4	7.1
Meso Eta	N/A	33.3	16.7
NGM	10.9	4.8	4.8
FACT	9.5	3.1	24.1
Climatology	9.5	2.4	2.4
Persistence	9.5	2.4	2.4
Subjective	42.9	31.0	23.8

The experiment was designed to be non-intrusive on operational duties by San Juan WFO staff. Visits between San Juan WFO staff and FSU researchers provided insight on how to balance the utility of the forecast system with realistic expectations from operational forecasters.

5. Results

We present the results of this experiment stratified by region and forecast scheme. The utility of each forecast system in this QPF experiment is measured with the BCS. By dividing all events into light and heavy categories, we may test the research hypothesis. A discussion of individual forecast utility reveals which QPF schemes hold ex ante value for users.

5.1. Verification calculation

Forecast improvements via standard measures of skill, such as metrics of distance and statistical association, do not generally reflect improvement by users. Decisions in planning and management may depend on system facets that a traditional skill metric does not measure. A utilitarian measure of skill, however, allows the user to choose a forecast system that assures the highest economic value. Krzysztofowicz (1992) orders forecasts according to ex ante economic values to the user, and has accordingly derived a skill measure. The score that associates this ex ante economic value to a Bayesian estimator of correlation is called the BCS. Expressed as,

$$\mathrm{BCS} = \left(\frac{\sigma^2}{a^2 S^2} + 1 \right)^{-1/2} \tag{1}$$

it incorporates σ, the ratio of forecast error standard deviation (noise), to a, the standard deviation of the forecast error conditional on the realization of the predictand. The ratio is standardized by sample standard deviation S.

BCS is bound by 0 and 1. The latter represents a perfect forecast and the former a random guess. A forecast system that exceeds another in BCS is sufficient to describe that system. Insofar as correlation is an absolute scale, BCS is a utilitarian forecast score. Moreover, BCS rewards low forecast error variability for high system variability. For instance, a rainfall

Table 3
BCS by forecast scheme for the six winter regions

Forecast scheme	1	2	3	4	5	6
Eta	0.03	0.35	0.29	0.24	0.39	0.22
NGM	0.41	0.47	0.29	0.44	0.48	0.45
FACT	0.08	0.17	0.01	0.08	0.08	–
Climatology	0.66	0.66	0.65	0.67	0.65	0.68
Persistence	0.32	0.42	0.42	0.40	0.45	0.33
Subjective	0.57	0.52	0.41	0.48	0.55	0.35

forecast system with a low forecast error standard deviation will yield a higher BCS than a temperature forecast system with an identical σ value. This "reward" is not an ex ante assumption for the end-user. An accurate forecast of a difficult system will always have high utility.

BCS is a robust measure of utility provided the frequency distribution of the predictand is near-normal (Krzysztofowicz, 1992). Using a goodness-of-fit *G*-test (Sokal and Rohlf, 1981), the gamma distribution is adequate to represent the frequency of individual storm events. We applied a transformation to represent rainfall with a gaussian distribution, thereby fulfilling the requirement to employ BCS utility analysis. For the gamma distribution, the square root of twice the log-likelihood ratio statistic is approximately normally distributed (McCullagh and Nedler, 1989), and the gamma-to-log-likelihood transformation is given as,

$$\hat{y} = 3\left(\sqrt[3]{\frac{y}{\bar{y}}} - 1\right), \tag{2}$$

where $\hat{y}$ is a transformed value of daily rainfall, y an untransformed value and $\bar{y}$ the mean of the untransformed daily time series. It follows that the persistence and climatology forecast systems exhibit the same relative frequency distribution. For the purposes of comparing the utilities of all forecast schemes, we transformed the FACT forecasts and realizations in the same manner as the others.

Table 4
BCS by forecast scheme for the seven transition regions

Forecast scheme	1	2	3	4	5	6	7
Eta	0.10	0.06	0.34	0.31	0.38	0.28	0.19
NGM	0.26	0.22	0.28	0.29	0.33	0.31	0.21
FACT	0.01	0.04	0.21	0.17	0.13	0.13	0.10
Climatology	0.64	0.64	0.61	0.70	0.66	0.64	0.63
Persistence	0.32	0.34	0.34	0.35	0.34	0.32	0.35
Subjective	0.44	0.55	0.66	0.47	0.45	0.52	0.47
Meso Eta	0.22	0.13	0.21	0.29	0.24	0.26	0.25

Table 5
BCS by forecast scheme for the seven summer regions

Forecast scheme	1	2	3	4	5	6	7
Eta	0.61	0.36	0.23	0.26	0.38	0.48	0.40
NGM	0.61	0.56	0.49	0.61	0.60	0.68	0.42
FACT	0.10	0.09	0.01	0.03	0.20	0.28	–
Climatology	0.68	0.41	0.49	0.60	0.36	0.62	0.60
Persistence	0.37	0.37	0.35	0.32	0.38	0.40	0.37
Subjective	0.68	0.41	0.49	0.60	0.36	0.62	0.60
Meso Eta	0.52	0.32	0.39	0.62	0.60	0.68	0.42

Having transformed the dependent variable such that the frequency distribution is approximately normal, we calculated the BCS for each forecast system. Each forecast system produced a 24 h area-averaged rainfall forecast. There were *n* forecasts per day, one for each of the convective regions in Puerto Rico. The assemblies of daily forecasts within the dates of each of the three seasons comprised a seasonal time series. We transformed each daily value in this time series. The dependent time series of actual rainfall realizations was constructed in the same manner.

The BCS value provides a measure of utility of forecast system *i* for region *j* during each of the three rainfall seasons. BCS values for each forecast system during the winter season are shown in Table 3. The BCS values for the transition season and summer season are shown in Tables 4 and 5, respectively.

5.2. Discussion

The climatology forecast system has the highest BCS for each region during the winter season. BCS values lie between 0.65 and 0.70. For each calendar day, the rain that fell in all previous years builds a distribution about a mean value. The climatology forecast is fixed at this value. Since we have transformed the variable, the mean and median for each calendar day are very close to each other. Over many years, the forecast errors will be much smaller than if we had fixed a forecast value in the distribution tails instead of at the mean (climatology). Underpinning

BCS is the ratio of noise to signal. Since the variance of the forecast errors is small, the BCS values are relatively high.

The subjective, or NWS operational QPF, ranks second in BCS utility. Their BCS values generally exceed all of the other forecast systems. The exception is region six, which proves to measure half as much utility as climatology but is less than the NGM. Overall, persistence and NGM forecast systems demonstrate proximal utility, while Eta model lags behind.

The FACT model is clearly insufficient for all other forecast systems. Noise dominates this forecast system, casting aspersion on its ability to select and bifurcate predictor variables. Region six demonstrates this the best. Every forecast for the winter period was a category two. FACT selects 12 h temperature tendency at Ponce (TT_{Ponce}) as the first node variable, performs a one-way ANOVA and split it as follows (values in °C):

$TT_{ponce} < -4.54 \rightarrow$ Category 6 rainfall

$-4.54 \leq TT_{Ponce} < -1.61 \rightarrow$ Split on another variable

$-1.61 \leq TT_{Ponce} < -1.36 \rightarrow$ Category 3 rainfall

$-1.36 \leq TT_{Ponce} \rightarrow$ Category 2 rainfall

Because the 12 h temperature tendency at Ponce exceeded −1.36°C for every day in the winter period, every decision is shunt to the terminal node category two. Transformed, every value becomes zero, thereby forcing the linear regression slope between forecast and actual to be zero as well. Therefore, the BCS value goes to infinity. This feature is an artifact of the transformation, not the BCS (although the transformation was necessary to invoke the BCS). The qualitative utility of preventing a forecast from ever deviating from a single value (in this case, category two) is analogous to making a null forecast. FACT is not very robust in this instance, and not very utilitarian either.

Climatology is the forecast system with the highest BCS for each region during the transition season. The one exception is region 3, which indicates that the subjective system has the greatest utility for southeast Puerto Rico. Persistence shows similar utility to the NGM and Meso Eta models. The Eta model is the least sufficient of the dynamic models, and again FACT demonstrates little Bayesian value for reasons indicated.

During the summer season, climatology proves to be the most utilitarian forecast system. Climatology BCS scores are the highest of all systems for each of the seven regions in Puerto Rico. Unlike the previous two seasons, however, the subjective forecasts do not perform significantly better than the dynamic models. On the whole, BCS scores for the subjective forecast systems are similar to the NGM and Meso Eta models. During the summer, the dynamic models are able to spatially resolve the easterly waves that pass over the island every few days. Although Puerto Rico lies on the southeast corner of the NGM, Eta and Meso Eta domains, initialization data from oceanic buoys to the east are assimilated into model runs. These data include sea level pressure changes associated with summertime easterly waves.

5.2.1. Traditional measures of skill

The forecast verification for this experiment measures utility, not skill. It is more traditional in both the research literature and operational setting to verify QPFs through metrics of skill. We provide a brief contrast on how skill differs from utility by verifying our results with a traditional measure of success, the Heidke skill score (HSS). This brief analysis of skill does not speak to our research hypothesis. Instead, it provides a comparison on our results for users more familiar with verification studies of forecast skill.

Heidke skill score is given as

$$\mathrm{HSS} = \frac{H - E}{N - E} \tag{3}$$

It is the ratio of correct forecasts using the current algorithm (H) minus the correct forecasts using climatology (E) to the number of realizations (N) minus the number of correct forecasts using climatology (E) (Barnston, 1992). HSS is a climatology-relative forecast. If climatology is used for the current algorithm (whose number of correct forecasts are measured by H), $\mathrm{HSS} = 0$. Forecast algorithms that are more skillful than climatology yield positive HSS. Forecasts less skilled than climatology yield negative

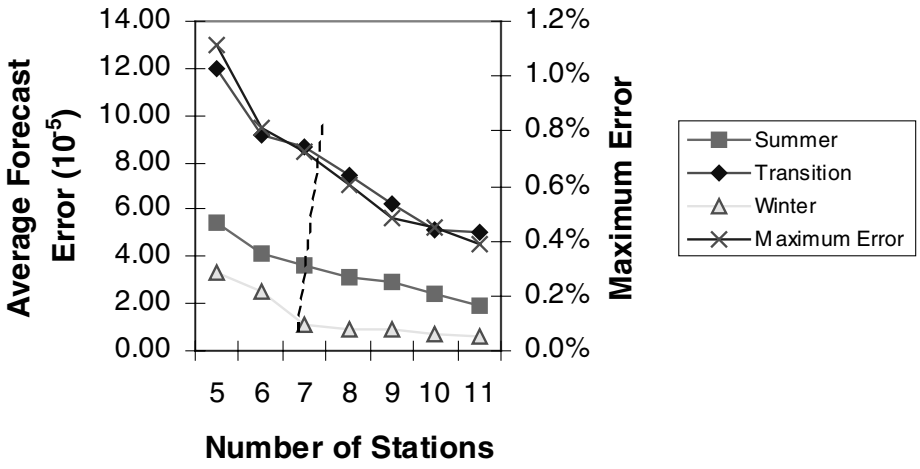

Fig. 5. Average and maximum forecast errors for a sequential number of stations in the San Juan region. One station at a time is removed without replacement and forecast errors are recalculated. The slope of errors gets smaller as more stations are added to the region.

HSS values. In this analysis, a "hit" occurs when the categorical forecast matches a categorical realization for the same day.

Unlike BCS, some of the forecast schemes show HSS success with respect to climatology. In fact, for the winter season, all seven schemes exceed climatology in skill for at least one region. The persistence forecast is consistently more skillful than climatology over all three seasons. The skill of the dynamic models varies from region to region. The subjective forecasts are more skillful than climatology for the winter and transition seasons but less skillful during the summer. FACT continues to perform with the least success of the seven forecast schemes. This analysis demonstrates how skill differs from utility in forecast verification. HSS does not take into account the forecast or system variability. All things being equal, it will reward temperature forecasts (which have low variability) and rainfall forecasts (which have high variability) the same. HSS intrinsically values rain, temperature, snow, and all other underlying variables of equal importance to the user. As such, HSS is not consistent with the Bayesian framework of this QPF experiment.

5.2.2. Limitations and forecast error

It is important not to underestimate the role of Thiessen averaging errors and user–computer interfaces on the verification results. Limitations in the software used for drawing isohyets coupled with the difficulty inherent in using a PC mouse to draw lines make it extremely difficult to be precise over small river basins or geographic domains. The final product resembles the forecaster's design, but the awkwardness of computer input distorts the final product. This is perhaps amplified in the summer season when small-scale features like the sea breeze and mountain convection are more active.

Errors in spatially averaging rainfall may also cloud statistics of skill and utility. The procedure of Thiessen averaging assigns relative importance to stations within a region. If a station with a large weight reports a missing value, however, the remaining stations may degrade in their ability to represent the region. To test the stability of our area-averaging procedure, we measure the sensitivity of forecast errors to number of stations within a region.

Following Nkemdirim and Meller (1982), we calculate the error of estimate of mean rainfall for the collection of stations within the San Juan convective region (Fig. 2, region 3). This is done for each day in the period. The error of estimate is given by,

$$\lambda = \sigma n^{-1/2}, \tag{4}$$

where σ is the standard deviation of daily station rainfall amounts and n is the number of stations in the region. Next, one station is randomly removed from the region, and the standard error is recalculated. Another station is randomly removed, without repla-

Table 6
BCS stratified by light (categories 1, 2 and 3) and heavy (categories 4, 5 and 6) 24 h rain. Scores are computed for each forecast scheme over all regions and seasons

Forecast scheme	Categories 1, 2 and 3	Categories 4, 5 and 6
Eta	0.23	0.43
NGM	0.22	0.00
FACT	0.17	0.02
Climatology	0.69	0.13
Persistence	0.31	0.06
Subjective	0.44	0.04
Meso Eta	0.20	0.00

cement, and the procedure is repeated. This exercise is repeated for each day in the summer period. Fig. 5 shows the average and maximum error of estimate for each day in the period by number of stations sequentially left in the region. The slope of each curve drops sharply between 5 and 6 stations and then becomes constant until all 11 stations are present. This implies that the addition of another station will near-linearly reduce area-averaging errors. More likely, however, the errors will continue to diminish with the addition of 12 or more stations (Nkemdirim and Meller, 1982). For the complete region of 11 stations, the maximum daily error of estimate was less than 0.4%. If we look at an error interval ($x + \lambda x$) instead of a point rainfall value, the error of estimate was sufficient to change the category of rainfall realization once in the 41 days. A "hit" becomes a "miss", and the BCS and HSS degrade. This degradation is limited by the fact that we regionalized the group of stations by their variance signature.

5.3. Value-induced bias

It is a tenet of this experiment that subjective forecasts will contain value-induced bias. Because these rain events are destructive and life threatening, QPFs may be inadvertently inflated in the interest of public safety. The Bayesian approach cares not for the public safety, but for individual users to make their own decisions. This is in no way meant to denigrate the role of the operational forecaster and her/his charge on public safety. Instead, it is meant to say that even for the heaviest events, not all users will assign a high utility on these forecasts.

Each forecast system for each day presents a forecast made and an attendant realization. Since BCS is utilitarian, we may compare forecasts across season and across region. We may group them as we wish and perform a BCS analysis on them. Here we split each forecast system time series in two: all days that realized category 1, 2 or 3 rain (less than 1.27 cm area-averaged) and all days that realized category 4, 5 or 6 rain (1.27 cm area-averaged or more). Over all regions and seasons, we construct these time series, and calculate the BCS. There are 14 BCS values in all, two for each forecast system (Table 6). The dominant feature of the stratified BCS analysis is the deterioration in utility for heavy rain events. BCS values are much higher for light rain events than for heavy rain events for six of the seven forecast systems, and these values are all less than 0.15. Although the sample size for heavy rain events is small ($n_2 = 69$) compared with light rain events ($n_1 = 751$), it is sufficiently large to claim this deterioration in utility.

The exception to this claim is the Eta model, which actually doubled its utility for heavy rain events. The NGM model did not replicate this. The NGM uses a modified Kuo parameterization scheme, which triggers precipitation via low-level moisture convergence, an important mechanism for releasing convective instability in Puerto Rico. This criterion of moisture convergence is easier to satisfy in the NGM than the two-tiered convective instability criteria are to satisfy in the Eta. As such, it is more apt to develop rainfall on the island. NGM's Kuo scheme does not stratify deep and shallow convection (unlike the Eta), and its QPF utility drops for heavy rain events (COMET, 1999). The Eta model, on the other hand, increases its utility for heavy rain events. For rainfall amounts that define the heavy threshold, the ambient physics pervade to induce the Eta's Betts–Miller–Janjic (BMJ) scheme into producing rainfall. Once rainfall is established, the Eta model is superior to the NGM in its ability to place isohyets due to its more refined gridspace resolution. It is not surprising to see an increase in the QPF utility of the Eta model for heavy rain events.

Despite possessing the same underlying physics, the Meso Eta has lower QPF utility than the Eta model for heavy rain events. Furthermore, the Meso Eta does not replicate the Eta's improvement in utility or heavy rain. This is primarily an artifact of model availability. The Meso Eta was not available to this

experiment in the winter season. The BMJ scheme in both models is more successful at modeling rainfall for synoptic-scale weather systems than for inducing mesoscale circulations. Middle latitude synoptic-scale systems are most likely to occur during the winter, the season the Meso Eta was not available. The Eta model's improved utility for heavy rain events is principally due to its QPF effectiveness in the winter season. Though not tested in this experiment, had the Meso Eta been available during the winter, it would have ostensibly shown parallel behavior to the Eta model.

Overall, for heavy rain events, subjective forecasts are sufficient for the FACT, NGM and Meso Eta models. It is insufficient for the persistence and climatology systems and considerably insufficient for the Eta model. Although most of the forecast systems were similar in BCS to the subjective forecasts, based upon the strength of the Eta model, the relative utility of the suite of objective guidance is higher than subjective forecasts for heavy rain events. (The average BCS for objective guidance is 0.11. The BCS for the subjective forecast is 0.04.) Therefore, the research hypothesis is accepted. We assert this with the caveat that all guidance, save for the Eta, possess little utility for heavy rain events. From a Bayesian standpoint, we are careful not to claim that this guidance should be improved for heavy rain events.

6. Conclusions

A QPF experiment for independent rainfall regions in Puerto Rico reveals relative utility for seven prediction schemes. Three NCEP dynamic models, climatology, persistence, classification tree model and the subjective NWS operational forecast predict area-averaged 24 h rainfall amounts over three rainfall seasons: winter, transitional and summer. The results are analyzed using the BCS, a metric of association measuring utility and the experiment is conducted in a Bayesian framework of methodology.

Climatology, or the historical arithmetic mean for a particular calendar day, demonstrated the highest utility of the seven forecast schemes. For only one region during the transition season did another forecast system (subjective) prove to be sufficient for climatology. Climatology is often considered a benchmark for developmental forecast schemes. Because its calculation and initialization are very straightforward, climatology is a simple preference to resource-intensive dynamic models if no competitive advantage is gained from the latter. It is rare to find a study, especially in the middle latitudes, where a dynamic model does not demonstrate skill over climatology.

Rainfall in the tropics gives rise to such a possibility. Each convective region of Puerto Rico contains several rain gauges, and the historical rainfall measured in these gauges determines climatology. Climatology (along with persistence and FACT), therefore, has the highest spatial resolution of all forecast systems. The dynamic models, including Meso Eta, possess a coarse gridspace resolution and do not resolve convective scale rainfall. Instead, they resolve large-scale atmospheric fields, such as sea level pressure, temperature and wind. The island of Puerto Rico is topographically complex, however, and rainfall is forced by these atmospheric variables on a local scale. A wind component value at a grid point, for instance, may measure the low-level mean flow while a local sea breeze occurs between grid points. This sea breeze may force rain near the coast, and the only manifestation of this event is the rain that falls in the gauge.

Because Puerto Rico has a dense network of rain gauges, it may seem that climatology forecasts have an inherent advantage over gridded-model forecasts. Each climatology forecast benefits from spatially comprehensive historical data, which acts as the forecast initialization. On the other hand, dynamic models do not have good spatial resolution over Puerto Rico. Even the highest-resolution dynamic models have only a few grid points on the island. These grid points do exist, however, and reside within one or more convective regions. Each grid point contains a quantitative precipitation forecast amount determined by the model's forecast algorithm. The difficulty of the algorithm to resolve antecedent atmospheric conditions into rainfall renders a lower ex ante value to the user than climatology forecast. This is shown for every region and every season.

Persistence and FACT (which includes persistence in its variable pool) also make use of the comprehensive network of rain gauges on the island. It is the

previous day's rainfall that initializes these models, however, and tropical rainfall was shown to have a statistically significant autocorrelation to only one day. In a supporting study to this experiment conducted by the authors (not shown) this autocorrelation was determined to be less than 0.10. In Puerto Rico, yesterday's rain can only explain about 1% of the variance of today's rain. Persistence does not fare as well as climatology.

The utility of forecast guidance deteriorates with increasing rainfall. The Eta model is the only forecast scheme in this experiment to show an improvement in utility with increasing rain amount. This is due to a stringent set of criteria in which the BMJ convective parameterization scheme develops rainfall (COMET, 1999). Once triggered, the Eta model is accurate at placing rain fields, thereby enhancing its utility for heavy rain forecasts. On the strength of the Eta model, and to a much less extent climatology, the overall suite of objective guidance showed greater utility than the subjective operational forecasts for rain events of category 4, 5 and 6 amount. As a result of the BCS analysis of experiment data on the stratified (light and heavy) 24 h rainfall, we accept the research hypothesis that for heavy rainfall events in Puerto Rico, objective guidance will provide higher ex ante value to the user. We accept the hypothesis on the caveat that three of the objective forecasts (NGM, Meso Eta and FACT) showed marginally less utility than the subjective forecasts. The overall utility for six of the seven forecast schemes is low for heavy rain situations.

This QPF experiment shows that objective guidance has a long way to go to match the utility of climatology in predicting short-term tropical rainfall. Both dynamical and statistical models suffer from the absence of a comprehensive network of data collection platforms. With five hourly surface instruments and 114 rain gauges, Puerto Rico has the densest network of continuous data collection in the Caribbean. Still, even small-scale dynamic models such as the Meso Eta and advanced-methodology statistic models such as FACT have difficulty resolving convective rainfall. A 20 km Eta model prototype is now available to forecasters in Puerto Rico, and an even smaller scale regional atmospheric modeling system is under study. Each convective region may now have its own grid points, or set of grid points, with dozens of atmospheric variables in three dimensions. The current drawback to such an approach is the lack of data available to fine-resolution model assimilation. These already require interpolation and mapping of actual observations to equally spaced grid points. Data observing instruments must equal or outpace the shrinking of distance between dynamic model grid points or interpolation error will grow. The consequence of not investing in data collection is that mesoscale dynamic models will continue to miss the convective scale processes despite a finer grid resolution.

In the short term, climatology will continue to be a benchmark, but shouldn't be aspersed as it is in the middle latitudes. Climatology has demonstrated BCS up to 0.70. This is a high measure of utility that should not be denigrated because multi-million dollar forecast schemes come up short. Instead, it represents a lofty goal for the advancement of the science of rainfall prediction in the tropics, which is one of the most difficult challenges in weather prediction. This experiment has shown the current utility of a variety of forecast products and hopefully points to promising avenues for the future of rainfall prediction.

Acknowledgements

The study was made possible through a National Weather Service Cooperative Project under the auspices of the Cooperative Program for Operational Meteorology, Education, and Training (COMET). The Scientific Services Division (SSD) of the Southern Region Headquarters of the NWS has supported this project through the Cooperative Institute for Tropical Meteorology (CITM). We especially acknowledge Dan Smith and Bernard Meisner of SSD for their support of our research in both principle and deed. The National Center for Atmospheric Research (NCAR) Scientific Computing Division (SCD) provided to this project access to atmospheric data and supercomputer facilities. We are grateful to Mark Laufersweiler and the FSU Department of Meteorology computing staff for providing their expertise and assistance and to X.N. for his help with statistics. We thank Ethan Gibney for his help with the regionalization maps.

References

Barnston, A.G., 1992. Correspondence among the correlation, RMSE, and Heidke forecast verification measures: refinement of the Heidke score. Weather Forecasting 7, 699–709.

Bennett, S., Grusbisic, V., Rasmussen, R., 1998. Gravity waves, rainbands, and deep convection induced by trade wind flow past Puerto Rico. Preprints 12th Conference on Numerical Weather Prediction, 11–16 January 1998, Phoenix, AZ.

Black, T.L., 1994. The new NMC mesoscale eta model: description and forecast examples. Weather Forecasting 9, 265–278.

Carter, M.M., 1995. Convective Rainfall Regions in Puerto Rico. Masters thesis, Florida State University, Tallahassee, 75pp.

Carter, M.M., 1997. A climatology of tropical cyclone rainfall over Puerto Rico. Preprints of the 22d Confernece on Hurricane and Tropical Meterology, Fort Collins, American Meteorological Society, 1997, pp. 596–597.

Carter, M.M., Elsner, J.B., 1996. Convective rainfall regions of Puerto Rico. Int. J. Climatol. 16, 1033–1043.

Carter, M.M., Elsner, J.B., 1997. A statistical method for forecasting rainfall over Puerto Rico. Weather Forecasting 12, 511–521.

Cooperative Program for Operational Meteorology Education and Training, 1999. Eta Model Physics. Online Training Materials for the Symposium on Numerical Weather Prediction (available online at http://www.comet.ucar.edu/nwplessons/etalesson2/physics.htm).

Fassig, O., 1916. Tropical rains: their duration, frequency, and instensity. Mon. Weather Rev. 44, 329–336.

Frank, N., 1970. On the Nature of Upper Tropospheric Cold Core Cyclones over the Tropical Atlantic. PhD thesis, Florida State University.

Hamilton, K., 1981. A note on the observed diurnal and semidiurnal rainfall variations. J. Geophys. Res. 86, 12122–12126.

Johnson, R., Wichern, D., 1982. Applied Multivariate Statistical Analysis. Prentice Hall, Englewood Cliffs, NJ (594pp.).

Kousky, V.E., Gan, M.A., 1981. Upper tropospheric cyclone vortices in the tropical South Atlantic. Tellus 33, 538–551.

Krzysztofowicz, R.A., 1983. Why should a forecaster and decision maker Use Bayes theorem. Water. Resour. Res. 19, 327–336.

Krzysztofowicz, R.A., 1992. Bayesian correlation score: a utilitarian measure of forecast skill. Mon. Weather Rev. 120, 208–219.

Krzysztofowicz, R.A., Drzal, W.J., Drake, T.R., Weyman, J.C., Giordano, L.A., 1993. Probabilistic quantitative precipitation forecasts for river basins. Weather Forecasting 8, 424–439.

Krzysztofowicz, R.A., Sigrest, A.A., 1997. Local climatic guidance for probabilistic quantitative precipitation forecasting. Mon. Weather Rev. 125, 305–316.

Loh, W.-Y., Vanichsetakul, N., 1996. User Guide for FACT Version 4.0, University of Wisconsin Board of Regents, 35 pp.

Malkus, J., 1955. The effects of a large island upon the trade-wind air stream. Q. J. R. Meteorol. Soc. 81, 538–550.

McCullagh, P., Nedler, J., 1989. Generalized Linear Models. Chapman and Hall, London (511 pp.).

Murphy, A.H., Winkler, R.L., 1977. Experimental point and area precipitation probability forecasts for a forecast area with significant local effects. Atmosphere 15, 61–78.

Murphy, A.H., Daan, H., 1984. Impacts of feedback and experience on the quality of subjective probability forecasts: comparison of results from the first and second years of the Zierikzee experiment. Mon. Weather Rev. 112, 413–423.

Murphy, A.H., Hsu, W., Winkler, R.L., Wilks, D.S., 1985. The use of probabilities in subjective quantitative precipitation forecasts: some experimental results. Mon. Weather Rev. 113, 2075–2089.

Nkemdirim, L.W., Meller, B.C., 1982. Assessment of true variability of small rainfall events in a small mountainous watershed in the summer. Nordic Hydrol. 13 (4), 205–212.

Olson, D.A., Junker, N.W., Korty, B., 1995. Evaluation of 33 years of quantitative precipitation forecasting at the NMC. Weather Forecasting 10, 498–511.

Pico, R., 1974. The Geography of Puerto Rico. Aldine Publishing, Chicago, IL (439 pp.).

Ray, C.L., 1928. Diurnal variation of rainfall at San Juan, P.R. Mon. Weather Rev. 58, 140–141.

Riehl, H., 1947. Diurnal variation of moisture and stability aloft in the vicinity of San Juan, Puerto Rico. Bull. Am. Meteorol. Soc. 28, 137–143.

Sanders, F., 1963. On subjective probability forecasting. J. Appl. Meteorol. 2, 191–201.

Sokal, R., Rohlf, F., 1981. Biometry. Freeman, New York (859 pp.).

Thiessen, A., 1911. Climatological data for July 1911, District No. 10, Great Basin. Mon. Weather Rev. 39, 1082–1084.