

# Hydrological ensemble forecasting at ungauged basins: using neighbour catchments for model setup and updating

A. Randrianasolo, M. H. Ramos, and V. Andréassian

Cemagref, UR HBAN, Parc de Tourvoie, BP 44, 92163, Antony, France

Received: 13 July 2010 – Revised: 24 September 2010 – Accepted: 13 November 2010 – Published: 25 February 2011

**Abstract.** In flow forecasting, additionally to the need of long time series of historic discharges for model setup and calibration, hydrological models also need real-time discharge data for the updating of the initial conditions at the time of the forecasts. The need of data challenges operational flow forecasting at ungauged or poorly gauged sites. This study evaluates the performance of different choices of parameter sets and discharge updates to run a flow forecasting model at ungauged sites, based on information from neighbour catchments. A cross-validation approach is applied on a set of 211 catchments in France and a 17-month forecasting period is used to calculate skill scores and evaluate the quality of the forecasts. A reference situation, where local information is available, is compared to alternative situations, which include scenarios where no local data is available at all and scenarios where local data started to be collected at the beginning of the forecasting period. To cope with uncertainties from rainfall forecasts, the model is driven by ensemble weather forecasts from the PEARP-Météo-France ensemble prediction system. The results show that neighbour catchments can contribute to provide forecasts of good quality at ungauged sites, especially with the transfer of parameter sets for model simulation. The added value of local data for the operational updating of the hydrological ensemble forecasts is highlighted.

## 1 Introduction

Predicting hydrological variables in ungauged catchments has been singled out as one of the major issues in the hydrological sciences at present. Considerable scientific effort is currently coordinated via the PUB (Prediction in Ungauged

Basins) initiative of the International Association for Hydrologic Sciences (IAHS), which has dedicated the 2003–2012 decade to focus research on this topic (Sivapalan et al., 2003). The ungauged catchment case is important from both practical and theoretical perspectives (Merz and Blöschl, 2004), and several approaches have been proposed to define hydrologically homogeneous regions around ungauged sites and to transfer information from neighbour catchments to ungauged basins. Various regionalisation methods have been proposed in the literature. One of the most frequently used techniques is regression analysis to model the relationship between the model parameters and physiographic catchment attributes (Young, 2006; Kay et al., 2007; Reichl et al., 2009). Many of these approaches hinge on spatial proximity (catchments can either be nested neighbours or adjacent neighbours) because catchments which are close to each other will also behave similarly (e.g., Merz and Blöschl, 2004; Parajka et al., 2005; McIntyre et al., 2005; Young, 2006; Oudin et al., 2008; Kjeldsen and Jones, 2007, 2009). In this paper, spatial proximity was chosen as the criteria to define homogeneous region. Spatial proximity-based approaches can be justified on explicit and implicit bases (Oudin et al., 2011):

- explicit basis: neighbours share common climate and physiographic characteristics that imprint the hydrological behaviour of a catchment;
- implicit basis: neighbours also share the unobservable or unquantifiable characteristics (underground parameter, geological attributes), which we are often unable to include in the approaches based on physical similarity.

In hydrological forecasting, local discharge data is essential for the two main operations involved in the prediction of uncertain future conditions: (1) the simulation of precipitation into discharge, for which long time series of historic discharges for model setup and calibration are needed, and (2) the updating of forecasts, which takes into account observed



Correspondence to: A. Randrianasolo  
(annie.randrianasolo@cemagref.fr)

data or observed errors between simulated and measured discharges, both available prior to the time of forecast, to adjust model inputs, internal states or outputs. It is widely acknowledged that forecast updating may improve significantly the quality of operational forecasts at short up to long lead times, and efforts to collect and explore real-time data in hydrological forecasting are crucial.

Studies on the issues of flood frequency estimation and flow simulation in ungauged basins are however more common in hydrology than those dealing with flood forecasting at ungauged sites (Ouarda et al., 2007; Oudin et al., 2008; Kling and Nachtnebel, 2009; Reichl et al., 2009; Masih et al., 2010). This paper is a contribution to improving flood forecasting in ungauged basins. Its originality lies in using a simple regionalisation procedure to both model parameterisation and updating of the forecasts.

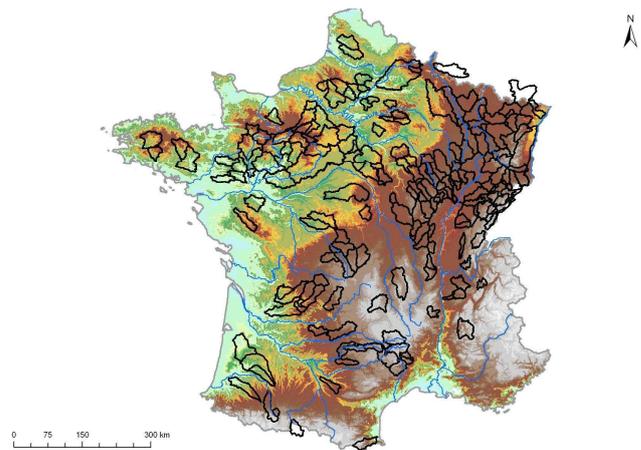
Whether the basin is gauged or not, flood forecasting remains uncertain at any site, particularly during periods of intense rainfall. Uncertainty in flood forecasting arises from many sources: precipitation observations and forecasts, initial soil moisture conditions, discharge measurements, model parameters, etc. To account for uncertainties from precipitation forecasts, hydrological ensemble forecasting approaches have been recently explored (see the review by Cloke and Pappenberger, 2009). Generally, they rely on ensemble weather prediction systems, which propose alternative scenarios for future states of the atmosphere, on the basis of perturbed initial conditions and stochastic model parameterizations during weather modelling.

The aim of this study is to evaluate hydrological ensemble forecasts at ungauged basins by using neighbour catchments to define the parameters of the hydrological model and to apply a forecast updating procedure. Neighbourhood is here defined by the criteria of simple geographical proximity. Different scenarios for the transfer of information to ungauged sites are evaluated. Hydrological ensemble forecasts are driven by an 11-member weather ensemble prediction system and flow forecasts are evaluated with the help of typical skill measurements of forecast performance. The paper is organized as follows: data and models are first presented in Sect. 2; then the methodology developed and the skill scores used in the evaluation of the forecasts are described in Sect. 3; Sect. 4 presents the results, and, finally, in Sect. 5, conclusions end this paper.

## 2 Data and models

### 2.1 Observed data and catchments

This study is based on a set of 211 catchments situated in France (Fig. 1). Meteorological and hydrological observed data are necessary for calibrating and running the hydrological model, as well as to evaluate the forecasts. For each catchment, time series of observed precipitation, daily mean



**Fig. 1.** Location of the 211 catchments studied in France.

evapotranspiration and discharge are available. Meteorological data come from the meteorological analysis system SAFRAN of Météo-France (Quintana-Seguí et al., 2008) for the period 1970 to 2006. The potential evapotranspiration was computed from temperature data using equations proposed by Oudin et al. (2005). Discharges come from the Banque Hydro (French database) and are available for a time period that varies according to each catchment, from 7 to 35 years, with 75% of the catchments with more than 27 years of data. Data were available at the daily time step.

### 2.2 Ensemble weather forecasts

The weather forecasts come from the meteorological ensemble prediction system PEARP of Météo-France, based on the global spectral ARPEGE model (Nicolau, 2002). Initial perturbations are generated by the singular vector technique and 11 future precipitation scenarios are proposed for each day of forecast. For this study, forecasts were provided with a 3-h time step, for a total forecast horizon of 60 h, at a 8-km × 8-km grid resolution. Forecast data were aggregated at daily time steps to match the observed data and spatially averaged over the studied catchments (weighted mean using the surface of each grid cell inside the catchment) to obtain the areal forecast precipitation at each lead time (i.e., for 24 and 48 h ahead). Early assessments of the PEARP system have shown good skills for short-range prediction of severe events (Thirel et al., 2008; Randrianasolo et al., 2010), even if the system still shows a certain lack of spread.

### 2.3 Hydrological forecasting model: flow simulation and forecast updating

GRP is a lumped soil-moisture-accounting type rainfall-runoff model developed at Cemagref, in France, and designed specifically for flood forecasting (Berthet et al., 2009).

**Table 1.** Synthesis of the scenarios tested and the abbreviations used for each test.

Test Reference	Acronym LRef	Updating with... local data (L)	Parameters local
<i>Case 1: Simulation with parameters from neighbours and updating with local discharge</i>			
Best neighbour	LBN	local data	Best calibrated neighbour
1 Neighbour	L1N	local data	Closest neighbour
5 Neighbours	L5N	local data	Set of 5 neighbours
10 Neighbours	L10N	local data	Set of 10 neighbours
15 Neighbours	L15N	local data	Set of 15 neighbours
20 Neighbours	L20N	local data	Set of 20 neighbours
<i>Case 2: Simulation with parameters from neighbours and no updating</i>			
Best neighbour no updating	XBN	not performed	Best calibrated neighbour
1 neighbour no updating	X1N	not performed	Closest neighbour
5 neighbours no updating	X5N	not performed	Set of 5 neighbours
10 neighbours no updating	X10N	not performed	Set of 10 neighbours
15 neighbours no updating	X15N	not performed	Set of 15 neighbours
20 neighbours no updating	X20N	not performed	Set of 20 neighbours
<i>Case 3: Simulation and updating with parameters and data from neighbours</i>			
Updating with best neighbour	NBN	discharge of the best neighbour	Best calibrated neighbour
Updating with 1 neighbour	N1N	discharge of the closest neighbour	Closest neighbour
Updating with 5 neighbours	N5N	mean discharge of 5 neighbours	Set of 5 neighbours
Updating with 10 neighbours	N10N	mean discharge of 10 neighbours	Set of 10 neighbours
Updating with 15 neighbours	N15N	mean discharge of 15 neighbours	Set of 15 neighbours
Updating with 20 neighbours	N20N	mean discharge of 20 neighbours	Set of 20 neighbours

Here, it is used at the daily time step. Input data are daily precipitation and mean evapotranspiration.

Its mechanism is classic, with a production function, which confronts the daily amounts of rainfall and evapotranspiration, and a routing function. It delays the release of effective precipitation over the next time steps. This function includes a linear routing by a unit hydrograph and a non-linear one via a routing store, which transforms the effective rainfall into flow at the catchments outlet.

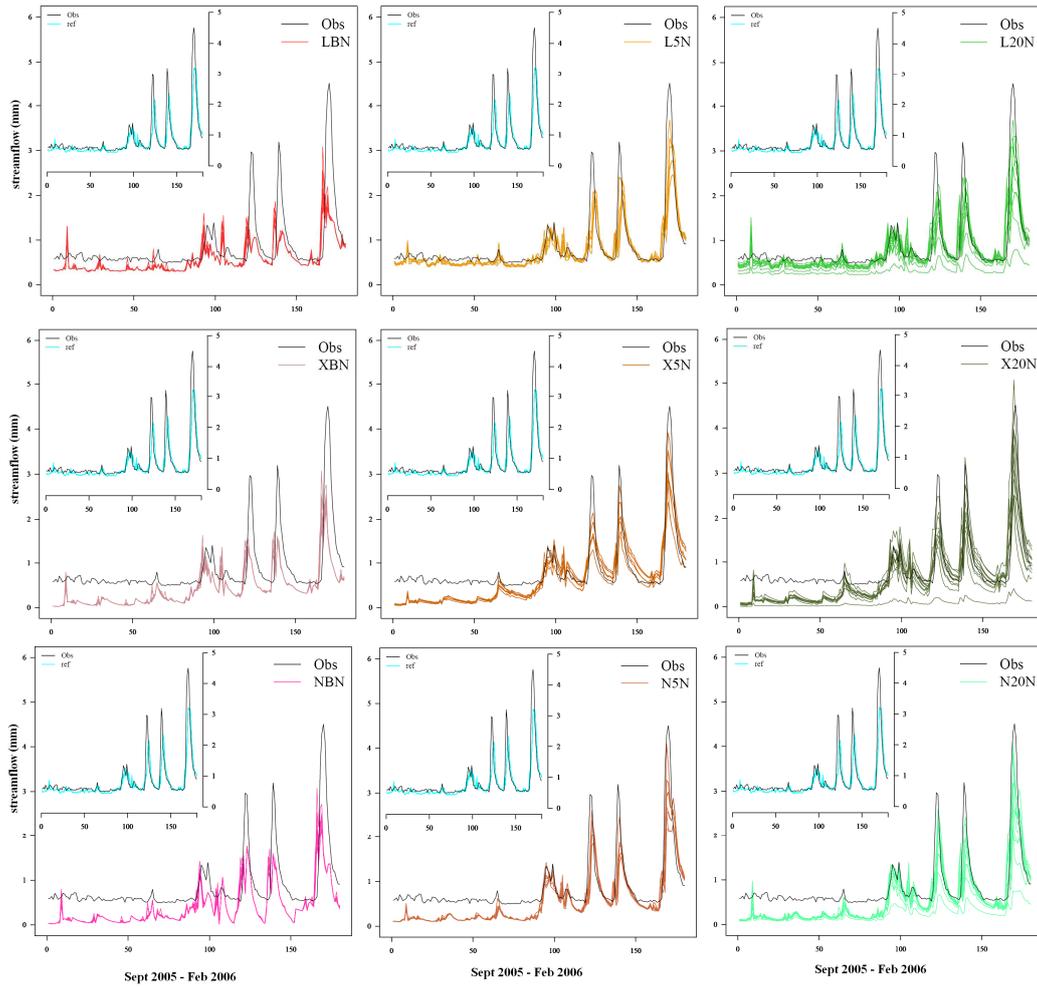
The GRP model has three parameters to be calibrated against observed data: the first one is a volume-adjustment factor, which acts over the volume of the effective rainfall, the second parameter is the maximum capacity of the routing store and the last parameter is the base time of the unit hydrograph. Note that unlike the GR4J model from which it derives (Perrin et al., 2003), in GRP, the capacity of the SMA store is a fixed parameter. Finally, the forecasting model GRP uses a combination of two assimilation (updating) functions for flow forecasting. The first integrates the last observed discharge information at the time of the forecast to update the state of the model routing store, while the second uses the last observed forecast error to correct the forecasts (Berthet et al., 2009). Only the first updating is activated in the version of the model used in this study.

### 3 Methodology

The hydrological forecasting system used in this study consists of a rainfall-runoff simulation model and a procedure for forecast updating. For gauged catchments, model parametrization is made through a calibration procedure applied to historic time series of concomitant precipitation and discharge observations. In real-time forecasting, the model also uses the last observed discharge (which usually differs from the last simulated discharge) in order to adjust the state of the routing store in the model in such a way that the output of the simulation agrees with the last observed discharge for the day preceding each day of forecast (for more details, see Berthet et al., 2009).

Thus, the use of the GRP hydrological forecasting model over ungauged catchments presents two challenges, which require transferring information from neighbour catchments:

- first, to parameterize the model before launching the forecasts;
- second, to update continuously (at each time step, and in real-time) the state of the model routing store.



**Fig. 2.** Examples of predicted ensemble hydrographs for the catchment La Saône à Auxonne (8900 km<sup>2</sup>) and 2 days of lead time. Top row: simulation with parameters from neighbour catchments and updating with local discharge. Middle row: simulation with parameters from neighbour catchments and no updating. Bottom row: simulation and updating with parameters and data from neighbour catchments. Neighbour catchments are: the best neighbour (left column; 11 members), the 5 (middle column; 55 members) and the 20 closest catchments (right column; 220 members). In each graphic, the reference scenario (11 members) is at the top left and the observed discharges are plotted.

### 3.1 Description of the different scenarios investigated

During the first step of the study, a reference situation, where local observed data were fully available, was simulated. For each catchment, the model was calibrated with historic local data and a set of optimal parameters was defined. This parameter set was then applied during a forecasting period (independent from the calibration period) and ensemble streamflow forecasts were evaluated against observations. This situation is supposed to give the best forecasts.

In the second step, a cross-validation approach was implemented: each catchment was considered as ungauged and evaluated in its turn. Different strategies for model parameter setup and forecast updating were investigated. Local data combined with data from neighbour catchments were used, as well as data from neighbour catchments alone. Proximity

of neighbours (also called “donors” using the common terminology of regionalization studies) was defined here by the Euclidean distance  $d$ , between the outlet of the ungauged site and the outlet of its neighbours (Eq. 1):

$$d = \sqrt{(X_{\text{target}} - X_{\text{neighbour}})^2 + (Y_{\text{target}} - Y_{\text{neighbour}})^2}, \quad (1)$$

where  $X_{\text{target}}$ ,  $Y_{\text{target}}$  are the geographic coordinates of the catchment considered as ungauged and  $X_{\text{neighbour}}$ ,  $Y_{\text{neighbour}}$  are the coordinates of the neighbour. On average, the closest neighbour and the 20-th farthest neighbour are about 2 km and 32 km, respectively, away from the target ungauged basin.

The different situations investigated can be grouped in three general cases described below and summarized in Table 1. An illustration of the resulting ensemble hydrographs is shown in Fig. 2.

*Case 1: Simulation with parameters from neighbours and updating with local discharge*

This case corresponds to a situation where local discharge data start to be collected at the beginning of the forecasting period. This is a hypothetical situation where no historic data is available for model calibration and a gauging station is only put in place at the outlet of the catchment at the time the forecasting system starts to operate. In this case, local discharge data can only be used in forecast updating (the case is kept simple, and does not deal with a progressive recalibration of the model while data accumulate). Parameter sets have to be transferred from gauged neighbour catchments to simulate streamflow ensembles. The alternative scenarios use model parameters from 1, 5, 10, 15 and 20 nearest neighbour catchments. A scenario using the parameter set from the neighbour catchment that gives the best performance during model calibration was also tested. On average, the best neighbour, which is among the 20 neighbours, is about 18 km away from the target ungauged basin.

*Case 2: Simulation with parameters from neighbours and no updating*

This case is similar to the previous one, except that here no local data are available at all: model parameters come either from the best neighbour catchment or from the nearest 1, 5, 15 and 20 neighbour catchments. No updating is used to correct the forecast in real-time.

*Case 3: Simulation and updating with parameters and data from neighbours*

Simulations at the target ungauged catchment are again based on model parameters transferred from the best neighbour catchment and from the nearest 1, 5, 15 and 20 neighbours. For the updating, a simple solution is tested, where the specific discharges from the neighbour catchments are used. For the best and the closest neighbour scenarios, their daily specific discharge observed prior to the day of forecast is used. For the multiple (5, 10, 15 and 20) neighbours scenarios, the mean specific discharge observed at each neighbour site is transferred to the target site.

### 3.2 Calibration of the parameters of the “donor” catchments

Even if we test here the application of a flow forecasting model to ungauged catchments, this involves the use of donor catchments for which the hydrological model must have been calibrated previously. Thus, each catchment considered in turn as ungauged can benefit from a library of calibrated parameter sets obtained from neighbouring gauged sites.

In calibration, the persistence index (Kitanidis and Bras, 1980) was used as objective function. It compares the predictions of the model with the prediction obtained by assuming that the best estimate for the future is given by the latest discharge measurement. A description of the “step-by-step” global optimization procedure used here is given in Edijatno et al. (1999).

### 3.3 Evaluation of model ensemble forecasts

For each scenario, a streamflow ensemble is generated for each day of the forecasting period and for each forecast lead time, consisting of  $11 \cdot N$  forecasts, where  $N$  is the number of parameter sets (or neighbours) used to build the scenario. Therefore, for the scenario using the 5 nearest neighbours, an ensemble of 55 ( $11 \cdot 5$ ) members was obtained for each forecast, and so on for 10 (110 members), 15 (165 members) and 20 (220 members) neighbours.

Three general types of criteria can be found as the basis for the evaluation of forecasts (Murphy, 1993):

- *consistency*, when the forecaster’s best judgment and the forecast actually issued coincide,
- *quality* or *accuracy*, which considers the correspondence between the forecast and the observation, and
- *value*, which concerns the economic worth of a forecast to its user.

In this study, the quality of the ensemble streamflows obtained from each scenario tested was assessed. Forecast values are compared to observed discharges over a 17-month period of forecast evaluation (from 10 March 2005 to 31 July 2006). Two lead times were considered: *Day 1* for the 24 h after the time the forecast is issued and *Day 2*, for the 24 h after Day 1, i.e., 48 h ahead the time of forecast. Typical scores of forecast performance were computed over the forecast evaluation period. They are summarized below and described in details in Jolliffe and Stephenson (2003):

- The *RMSE* (*root mean square error*) measures the distance between the forecasts and the observations and gives a measure of the forecast accuracy. Its advantage is being sensitive to large forecast errors and retaining the units of the forecast variable, thus being more easily interpretable as a typical error magnitude. The RMSE was calculated with the average streamflow forecast given by the ensemble mean of each scenario tested, for each lead time and over the total number of days of the evaluation period. To compare the scores over the 211 studied catchments, we computed normalized scores by dividing the RMSE of each catchment by its average observed streamflow over the evaluation period.
- The *contingency table* is a two-dimensional table that gives the discrete joint sample distribution of forecasts and observations in terms of cell counts, with two possible outcomes (yes or no) for observed and forecasted events. A perfect forecast system would only produce *hits* (events are forecasted and observed) and *correct rejections* (events are neither forecasted nor observed), and no *misses* (observed events are not forecasted) or *false alarms* (forecasted events are not observed). Thresholds were specified to separate “yes”

and “no” events. For the observed events, two streamflow thresholds were defined for each catchment: the 50-th and the 90-th percentiles of daily streamflows, computed over the evaluation period (hereafter,  $Q_{50}$  and  $Q_{90}$ , respectively). For the definition of forecasted events, two thresholds were selected: if  $p\%$  of the ensemble members forecast discharges exceeding the streamflow threshold, the event is considered a ‘forecasted event’. The values of  $p\%$  chosen in this study are 50% and 80% (hereafter,  $p_{50}$  and  $p_{80}$ , respectively). The combination of these thresholds results in four contingency tables, from which descriptive statistics can be computed. A frequently used score, particularly when the non occurrence of the event is more frequent than its occurrence, is the Threat score ( $TS$ ) or the *Critical Success Index (CSI)*. It is given by the number of hits divided by the total number of hits, misses and false alarms. The worst possible score is 0 and the best is 1.

- The *Brier score (BS)* averages the squared differences between pairs of forecast probabilities and the subsequent binary observed frequencies of a given event (e.g., exceedance of a streamflow quantile). From Eq. (2), for each realization  $j$  (here, day of forecast),  $y_j$  is the forecast probability of the occurrence of the event, given by the ratio of ensemble members forecasting the event to the size of the ensemble, and  $o_j = 1$  if the event occurs and 0 if it does not occur:

$$BS = \frac{1}{N} \sum_{j=1}^n (y_j - o_j)^2 \quad (2)$$

The same streamflow thresholds used in the contingency tables were considered to define an event: percentiles  $Q_{50}$  and  $Q_{90}$ . The Brier score is negatively oriented (smaller score better) and has a minimum value of 0 for a perfect (deterministic) system.

In order to compare the Brier score ( $BS$ ) with a reference ( $BS_{\text{reference}}$ ), it is convenient to define the *Brier skill score (BSS)*. The reference systems generally used are climatology and persistence. Here, since the aim is to evaluate the added value of alternative scenarios for ungauged catchments, comparatively to the reference situation where model calibration and updating are performed with local data, the scenario called “reference” (Table 1) is used to compute the  $BS_{\text{reference}}$ . The  $BSS_{\text{ungauged}}$  is then given by:

$$BSS_{\text{ungauged}} = 1 - \frac{BS_{\text{scenario\_ungauged}}}{BS_{\text{reference\_gauged}}} \quad (3)$$

The  $BSS$  is positively oriented (higher values indicate better performance). It ranges from  $-\infty$  to 1 (perfect deterministic system). Scores equal to 0 mean that the system does not give further information than the reference and negative scores indicate a poorer forecasting system than the reference.

- The *Discrete ranked probability score (DRPS)* is a squared measure that compares the cumulative density function of a probabilistic forecast with the cumulative density function of the corresponding observation over a given number of discrete probability categories (i.e., considering several streamflow thresholds). It is given by (Eq. 4):

$$DRPS = E \left[ \frac{1}{K} \sum_{j=1}^K (y_j - o_j)^2 \right], \quad (4)$$

where  $K$  is the number of categories of the distribution of the forecasts (here, defined by the thresholds  $Q_{10}, Q_{20}, Q_{30}, \dots, Q_{90}$  of streamflow quantiles),  $y_j$  is the forecast probability for the category  $K$  and  $o_j$  is the binary indicator for each category  $K$  ( $o_j = 1$  if the event occurs in the  $K$  category and 0, if not). The operator  $E$  denotes the average over the days of the forecasting period.

Like in the  $BS$ , skill scores were computed to compare the performance of the system given by each scenario tested with the reference scenario ( $DRPSS_{\text{ungauged}}$  derived from the application of Eq. (3) to the  $DRPS$ ).

## 4 Results

Figure 2 shows an example of hydrographs obtained for some of the different scenarios tested. The case of the Saône river at Auxonne (catchment area of 8900 km<sup>2</sup>) is illustrated for the reference scenario LRef (gauged catchment) and for the scenarios of ensemble predictions built from the best neighbour catchment, the 5 and the 20 nearest neighbours. The multiple simulations can be seen to get wider (more dispersive) as the number of ensemble members increases. Additionally, the forecast members of *Case 1* (Fig. 2, top), where simulations are performed with parameters from neighbours and updating with local discharges, are closer to the observation than the other forecasting scenarios considering the same number of neighbours. For *Case 2* (simulations with parameters from neighbours and no updating; Fig. 2, middle), a higher variability of forecast members is observed, comparatively to the scenarios of *Case 3* (simulation and updating with parameters and data from neighbours; Fig. 2, bottom), for which the ensemble spread is low and the forecast members are farther from the observation. The hydrographs presented in Fig. 2 correspond to one catchment and to a specific period in time (September 2005 to February 2006). Overall results from the scores used to evaluate the quality of the forecasts allow to better capture the average quality obtained for each scenario. They are presented in the next paragraphs for the lead time *Day 2*.

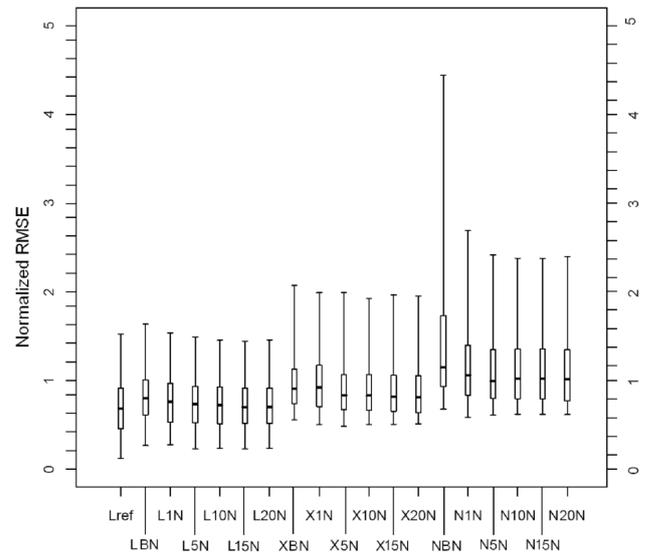
### 4.1 Normalized RMSE and CSI scores

Figure 3 shows the mean values of the normalized RMSE. The  $CSI$  for the ensemble streamflow forecasts generated by

each scenario is presented in Fig. 4, for the ensemble threshold  $p50$  and for the streamflow thresholds of  $Q50$  and  $Q90$  (also for a 2-day lead time). No significant differences were observed between the results with 50% and 80% (not shown) of ensemble members exceeding the streamflow thresholds. This is probably due to the lack of spread of the PEARP members, impacting the forecasts generated by the hydrological model (Randrianasolo et al., 2010). On the contrary, the  $CSI$  values decrease for the threshold  $Q90$ , comparatively to the values for the threshold  $Q50$ , for all scenarios. For instance, for the reference scenario, the median values of  $CSI$  vary from 0.88 for  $Q50$  to 0.69 for  $Q90$ . The performance for forecasting higher events is thus lower, although it has also to be considered that the number of events used in the computation of the scores is smaller. In general, for both RMSE and  $CSI$  scores, the ensemble scenarios built when updating is carried out with local discharges give better performance, with results closer to the reference situation. The updating with specific discharges from neighbour catchments does not improve the forecasts. On the contrary, it gives the worst average performance. The performance for the non-updated forecast scenarios is between these two cases of updating tested (with local data and with data transferred from neighbours). Small differences of performance are observed when using 1, 5, 10 and 20 neighbours, with a small improvement when increasing the number of neighbours. Distinctively, when comparing the results from the scenario based on the closest neighbour and the scenario based on the best neighbour (both scenarios have the same number of ensemble members), the results show that the former surpasses more often the latter. The scenario with the best neighbour catchment as donor, regardless its proximity to the target catchment, has in most cases the worst median score value in each group of cases considered.

#### 4.2 Skill scores

Figure 5 shows the skill scores of  $BSS$  and  $DRPSS$  when using the reference scenario as a reference forecast. The same conclusions as previously are obtained on the overall behaviour of the scores: the updating with local observations provides better performance (*Case 1*), the non-updated forecasts present intermediate results (*Case 2*), and the updating with data from neighbour catchments presents the lowest performance (*Case 3*). The scenarios that perform better are those when 5 to 20 neighbours are taken into account. It is interesting to note that, for more than 75% of the ungauged catchments, the gain is positive in the scenarios built with parameters from 5 to 20 neighbours and updating based on local observations, showing that these scenarios give forecasts with higher skill, in a probabilistic way, than the reference forecast. Particularly for the non-updating case (scenarios from *Case 2*), the average performance of the best neighbour is significantly better than the performance for the scenarios with 1, 5, 10, 15 and 20 closest neighbours. This is no longer

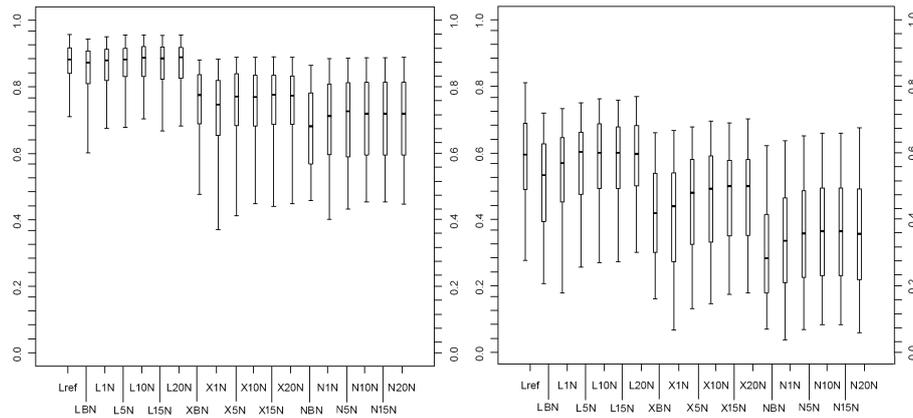


**Fig. 3.** Mean values of normalized RMSE for the reference and the scenarios tested (forecast lead time of 2 days and evaluation period from 10 March 2005 to 31 July 2006). Boxplots for the 211 studied catchments: the top and the bottom of the box represent the 75-th and the 25-th percentile, respectively, while the top and the bottom of the tail indicate the 95-th and the 5-th percentile, respectively. Median values are indicated.

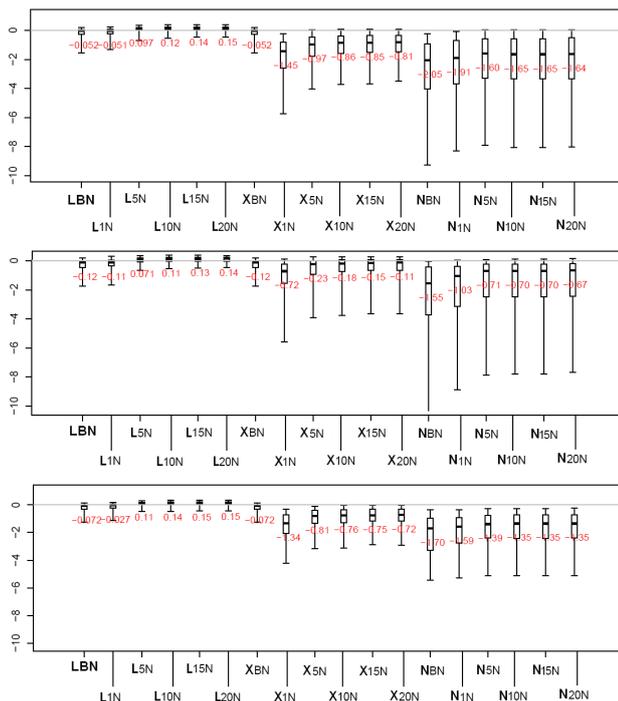
the case when updating is performed with the average of specific discharges from neighbours (scenarios from *Case 3*), for which the closest neighbour performs on average better than the best neighbour. Moreover, for the skill score  $BSS_{\text{ungauged}}$  and the  $Q50$  threshold (Fig. 5, top), around 50% of the catchments give positive skill scores (better performance than the reference scenario), while the same is observed for around 25% of the catchments for the  $BSS_{\text{ungauged}}$  and the  $Q90$  threshold (Fig. 5, middle) and the  $DRPSS_{\text{ungauged}}$  (Fig. 5, bottom).

#### 4.3 Best scenario for each catchment

For each performance measure, the best scenario obtained was plotted for each catchment (Fig. 6). Most often, the scenarios using parameters from neighbours and updating procedure with local data provide the best performance, confirming the overall results presented previously. However, in a few catchments, this is not the case for all performance measures. For instance, the maps of normalized RMSE and  $CSI$  scores (Fig. 6a–c) show some catchments where the best scenario is one of the non-updated situations (scenarios from *Case 2*): 12% of catchments for the RMSE score, 5% for the  $CSI$  exceeding  $Q50$  and 20% for the  $CSI$  exceeding  $Q90$ . Some scenarios based on updating with discharge data from neighbours (scenarios from *Case 3*) appear to be even better than the other scenarios in few catchments: as an example, for the score  $CSI$  exceeding  $Q90$  (Fig. 6c) 12 catchments (out



**Fig. 4.** Mean values of *CSI* for the reference and the scenarios tested (forecast lead time of 2 days and evaluation period from 10 March 2005 to 31 July 2006). *CSI* for  $p50$  exceeding the discharge percentile  $Q50$  (left) and  $Q90$  (right). Boxplots for the 211 studied catchments: the top and the bottom of the box represent the 75-th and the 25-th percentile, respectively, while the top and the bottom of the tail indicate the 95-th and the 5-th percentile, respectively. Median values are indicated.

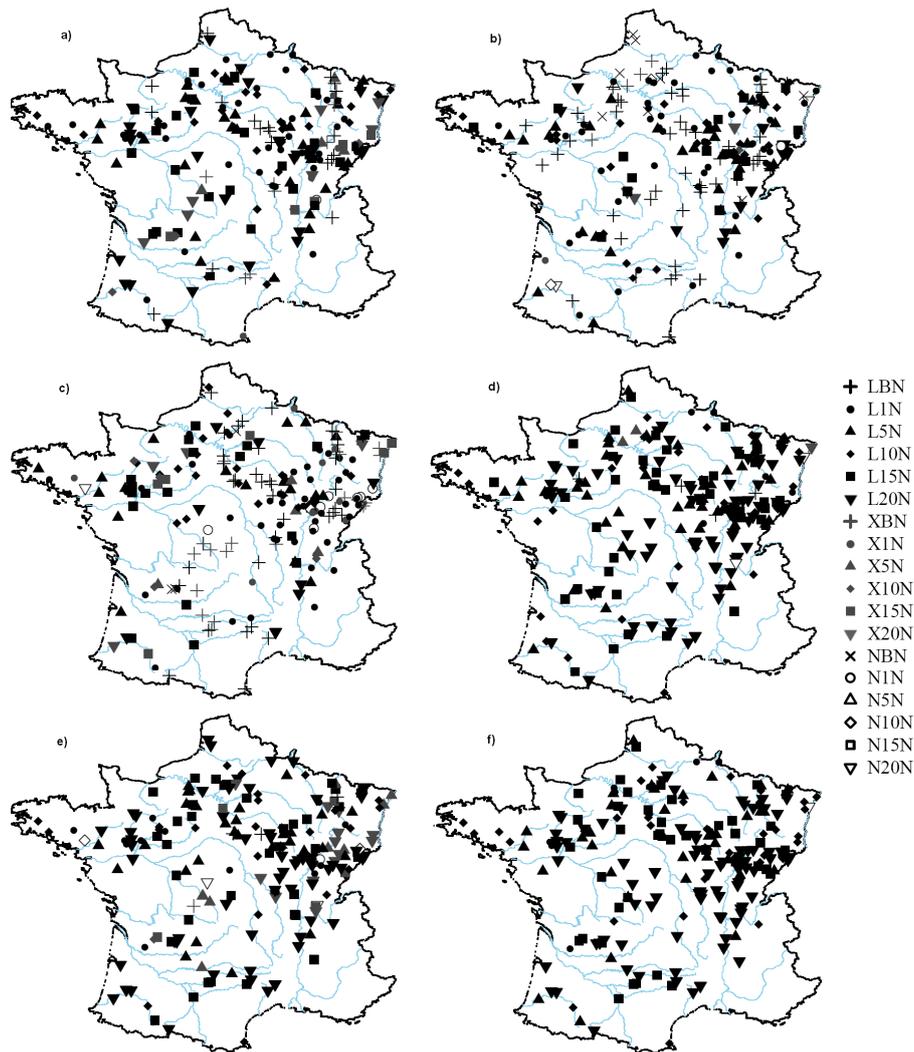


**Fig. 5.** Mean values of skill scores for the scenarios tested and relative to the reference scenario (forecast lead time of 2 days and evaluation period from 10 March 2005 to 31 July 2006). Top:  $BSS_{\text{ungauged}}$  for the discharge percentile  $Q50$ . Middle:  $BSS_{\text{ungauged}}$  for the discharge percentile  $Q90$ . Bottom:  $DRPSS_{\text{ungauged}}$ . Boxplots for the 211 studied catchments: the top and the bottom of the box represent the 75-th and the 25-th percentile, respectively, while the top and the bottom of the tail indicate the 95-th and the 5-th percentile, respectively. Median values are indicated.

of 211) are in this situation. They are particularly situated in the Eastern part of France, where the density of the studied catchments is more important. Considering the number of neighbours to be used when building the streamflow ensembles, the closest neighbour shows more often the best performance when focusing on the deterministic-based scores (*RMSE* and *CSI*). However, for the probabilistic-based scores (*BS* and *DRPS*), the scenarios that more often perform better are those when 5 to 20 neighbours are taken into account. Also, when focusing on large events, it is interesting to note that, for some catchments, the scenario that performs better becomes the scenario built with the best neighbour. This is noted when the results from the *CSI* and *BS* scores for the  $Q90$  threshold are analysed comparatively to the scores for the  $Q50$  threshold. For the higher threshold, the best neighbour scenario shows up for 2 and 19 catchments for the *BS* and the *CSI* score, respectively.

## 5 Conclusions

This study investigates the use of information from gauged neighbour catchments to run a hydrological forecasting system at ungauged catchments. Different scenarios were considered, which combine the transfer of model parameters from neighbour catchments to totally (or partially) ungauged catchments for rainfall-runoff simulation, and the transfer of observed discharges for updating at the time of forecast. A cross-validation procedure was applied to 211 catchments located in France. Hydrological models were driven by weather ensemble predictions from PEARP of Météo-France (11 members and 2 days of forecast range). Hydrological ensemble forecasts were obtained for each scenario investigated and their performance was tested against a reference situation, where the target catchment is fully gauged



**Fig. 6.** Maps of the best performing scenario for each catchment according to the score: (a) *RMSE*, (b) *CSI* for  $p80$  exceeding the discharge percentile  $Q50$ , (c) *CSI* for  $p80$  exceeding the discharge percentile  $Q90$ , (d) *BS* for the discharge percentile  $Q50$ , (e) *BS* for the discharge percentile  $Q90$ , and (f) *DRPS*. Forecast lead time of 2 days and evaluation period from 10 March 2005 to 31 July 2006.

(i.e., historic and real-time data are available for model calibration and forecast updating). Typical forecast performance measures were applied over a 17-month evaluation period. Both deterministic-focused measures (*RMSE* and *CSI*) and probabilistic-focused measures (*BSS* and *DRPSS*) were considered.

The results showed that the use of parameters from neighbour catchments can provide forecasts of good quality at the target ungauged site. This is particularly true when the transfer of parameters from gauged neighbours is accompanied by the implementation of a gauging station at the ungauged site, which provides local discharge information at the time of forecast for operational forecast updating. These scenarios, where the target site is considered partially ungauged (no historic data available for calibration, but data available at the time of forecast for updating of the system), are those that

most often show the performance that best approaches the performance of the reference gauged situation. Additionally, for these scenarios, in general, performance increases as the number of neighbours increases (from 1 to 20 neighbours as donors). This is particularly observed on the results from the probabilistic-based performance measures. The increase in spread of the streamflow ensemble predictions seems to have a positive effect on these performance measures.

In the case where local discharge data is not available at all and the target site is considered fully ungauged, the use of specific discharges data from neighbours to update the forecasting system at the ungauged site did not prove to be an overall good solution. Better performance is reached if no updating is carried out at all. Interestingly, the results from the probabilistic-based skill scores show that, while the case of scenarios including neighbour data for updating show, in

general, lower median values of performance for the scenario that uses the best neighbour catchment (the catchment that showed better performance in calibration), the opposite is observed for the case when no updating is performed: the best neighbour scenario performs better than the 1 to 20 closest neighbour scenarios. It seems that in the non-updated case, the simulation part of the forecasting model prevails and, on average, better forecasts will be achieved when the system relies on parameters from the neighbour that gives the best model performance in calibration. On the contrary, when neighbour-based updating is performed, the updating component of the forecasting system is strengthened and it is rather the closest neighbours that will give better forecast performance.

In summary, this study illustrates well the added value of having at least local data available at the time of the forecast to perform the real-time updating of a hydrological ensemble forecasting system at sites where no historic data is available for the set up (calibration) of the rainfall-runoff simulation model. For a better quality of the hydrological ensemble forecasts systems, if no local updating is possible, it is better to rely on the transfer of the parameters from the neighbour that showed best performance in calibration. If, in addition to model parameter transfer, neighbour-based updating is performed, the best solution is to rely on at least five closest neighbour donors.

*Acknowledgements.* We acknowledge the contribution of Météo-France, which gave us access to the SAFRAN data archive and to the PEARP ensemble forecasts, and the provision of the streamflow data by the MEEDM (Ministère de l'Ecologie, de l'Energie, du Développement durable et de la Mer).

Edited by: Y. He

Reviewed by: R. Turcotte and two other anonymous referees

## References

- Berthet, L., Andréassian, V., Perrin, C., and Javelle, P.: How crucial is it to account for the antecedent moisture conditions in flood forecasting? Comparison of event-based and continuous approaches on 178 catchments, *Hydrol. Earth Syst. Sci.*, 13, 819–831, doi:10.5194/hess-13-819-2009, 2009.
- Cloke, H. L. and Pappenberger, F.: Ensemble flood forecasting: A review, *J. Hydrol.*, 375, 613–626, 2009.
- Edijatno, S., Nascimento, N., Yang, X., Makhlouf, Z., and Michel, C.: GR3J: a daily watershed model with three free parameters, *Hydrolog. Sci. J.*, 44(2), 263–278, 1999.
- Haché, M., Ouarda, T.B.M.J., Bruneau, P., and Bobée, B.: Estimation régionale par la méthode de l'analyse canonique des corrélations: comparaison des types de variables hydrologiques. *Can. J. Civil. Eng.*, 29, 899–910, 2002.
- Jolliffe, I. T. and Stephenson D. B.: *Forecast Verification. A Practitioner's Guide in Atmospheric Science*, John Wiley and Sons, Chichester, West Sussex, England, 240 pp., 2003.
- Kay, A. L., Jones, D. A., Crooks, S. M., Calver, A., and Reynard, N. S.: A comparison of three approaches to spatial generalization of rainfall-runoff models, *Hydrol. Process.*, 20, 3953–3973, 2007.
- Kitanidis, P. K. and Bras, R. L.: Real-time forecasting with a conceptual hydrologic model 2. Applications and results, *Water Resour. Res.*, 16, 1034–1044, 1980.
- Kjeldsen, T. R. and Jones, D.: Estimation of an index flood using data transfer in the UK, *Hydrolog. Sci. J.*, 52(1), 86–98, 2007.
- Kjeldsen, T. R. and Jones, D.: An exploratory analysis of error components in hydrological regression modelling, *Water Resour. Res.*, 45, W02407, doi:10.1029/2007WR006283, 2009.
- Kling, H. and Nachtnebel, H. P.: A method for the regional estimation of runoff separation parameters for hydrological modelling, *J. Hydrol.*, 364, 163–174, 2009.
- Masih, I., Uhlenbrook, S., Maskey, S., and Ahmad, M. D.: Regionalization of a conceptual rainfall-runoff model based on similarity of the flow duration curve: A case study from the semi-arid Karkheh basin, Iran, *J. Hydrol.*, 391, 188–201, 2010.
- McIntyre, N., Lee, H., Wheeler, H., Young, A., Wagener, T.: Ensemble predictions of runoff in ungauged catchments, *Water Resour. Res.*, 41, W12434, doi:10.1029/2005WR004289, 2005.
- Merz, R. and Blöschl, G.: Regionalisation of catchment model parameters. *J. Hydrol.*, 287, 95–123, 2004.
- Murphy, A. H.: What is a good forecast? An essay on the nature of goodness in weather forecasting, *Weather Forecast.*, 8, 281–293, 1993.
- Nicolau, J.: Short-range ensemble forecasting. In *Proceedings WMO/CBS Technical Conferences On Data Processing and Forecasting Systems*, Cairns, Australia, 2–3 December, 4 pp., available at: <http://www.wmo.ch/pages/prog/www/DPS/TC-DPFS-2002/Papers-Posters/Topic1-Nicolau.pdf>, 2002.
- Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., and Loumagne, C.: Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 2 – Towards a simple and efficient potential evapotranspiration model for rainfall-runoff modelling, *J. Hydrol.*, 303, 290–306, 2005.
- Oudin, L., Andréassian, V., Perrin, C., Michel, C., and Lemoine, N.: Spatial proximity, physical similarity, regression and ungauged catchments: A comparison of regionalization approaches based on 913 French catchments. *Water Resour. Res.*, 44, W03413, doi:10.1029/2007WR006240, 2008.
- Oudin, L., Kay, A. L., Andréassian, V., and Perrin, C.: Are seemingly physically similar catchments truly hydrologically similar?, *Water Resour. Res.*, 46, W11558, doi:10.1029/2009WR008887, 2011.
- Ouarda, T., Ba, K. M., Diaz-Delgado, C., Carsteanu A., Chokmani, K., Gingras, H., Quentin, E., Trujillo, E., and Bobée, B.: Inter-comparison of regional flood frequency estimation methods at ungauged sites for a Mexican case study, *J. Hydrol.*, 348, 40–58, 2007.
- Parajka, J., Merz, R., and Blöschl, G.: A comparison of regionalisation methods for catchment model parameters, *Hydrol. Earth Syst. Sci.*, 9, 157–171, doi:10.5194/hess-9-157-2005, 2005.
- Perrin, C., Michel, C., and Andréassian, V.: Improvement of a parsimonious model for streamflow simulation, *J. Hydrol.*, 279, 275–289, 2003.
- Quintana-Seguí, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., Canellas, C., Franchisteguy, L., and Morel, S.: Analysis of near-surface atmospheric variables: Validation of the

- SAFRAN analysis over France, *J. Appl. Meteorol. Clim.*, 47(1), 92–107, 2008.
- Randrianasolo, A., Ramos, M.-H., Thirel, G., Andréassian V., and Martin, E.: Comparing the scores of hydrological ensemble forecasts issued by two different hydrological models, *Atmos. Sci. Lett.*, 11(2), 100–107, 2010.
- Reichl, J. P. C., Western, A. W., McIntyre, N. R., and Chiew, F. H. S.: Optimization of a similarity measure for estimating ungauged streamflow, *Water Resour. Res.*, 45, W10423, doi:10.1029/2008WR007248, 2009.
- Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J. J., Mendiola, E. M., O'Connell, P. E., Oki, T., Pomeroy, J. W., Schertzer, D., Uhlenbrook, S., and Zehe, E.: IAHS decade on prediction in ungauged basins (PUB), 2003–2012: shaping an exciting future for the hydrological sciences, *J. Hydrolog. Sci.*, 48(6), 857–880, 2003.
- Thirel, G., Rousset-Regimbeau, F., Martin, E., and Habets, F.: On the impact of short-range meteorological forecasts for ensemble streamflow prediction, *J. Hydrometeorol.*, 9, 1301–1317, 2008.
- Young, A. R.: Streamflow simulation within UK ungauged catchments using a daily rainfall-runoff model, *J. Hydrol.*, 320, 155–172, 2006.