



Predictability of soil moisture and river flows over France for the spring season

S. Singla^{1,2}, J.-P. Céron¹, E. Martin², F. Regimbeau¹, M. Déqué², F. Habets³, and J.-P. Vidal⁴

¹Météo-France, Direction de la Climatologie, 42 avenue G. Coriolis, 31057 Toulouse Cedex 01, France

²CNRM/GAME – URA1357 (Météo-France, CNRS), 42 avenue G. Coriolis, 31057 Toulouse Cedex 01, France

³UMR-SISYPHE, (UPMC, CNRS), Mines-Paristech, Centre de Géosciences, équipe SHR, 35 rue St. Honoré, 77305 Fontainebleau, France

⁴Irstea, UR HHLY, Hydrology-Hydraulics Research Unit, 3 bis quai Chauveau, CP 220, 69336 Lyon Cedex 09, France

Correspondence to: S. Singla (stephanie.singla@hotmail.fr)

Received: 28 July 2011 – Published in Hydrol. Earth Syst. Sci. Discuss.: 23 August 2011

Revised: 5 January 2012 – Accepted: 16 January 2012 – Published: 24 January 2012

Abstract. Sources of spring predictability of the hydrological system over France were studied on a seasonal time scale over the 1960–2005 period. Two random sampling experiments were set up in order to test the relative importance of the land surface initial state and the atmospheric forcing. The experiments were based on the SAFRAN-ISBA-MODCOU hydrometeorological suite which computed soil moisture and river flow forecasts over a 8-km grid and more than 880 river-gauging stations. Results showed that the predictability of hydrological variables primarily depended on the seasonal atmospheric forcing (mostly temperature and total precipitation) over most plains, whereas it mainly depended on snow cover over high mountains. However, the Seine catchment area was an exception as the skill mainly came from the initial state of its large and complex aquifers. Seasonal meteorological hindcasts with the Météo-France ARPEGE climate model were then used to force the ISBA-MODCOU hydrological model and obtain seasonal hydrological forecasts from 1960 to 2005 for the entire March–April–May period. Scores from this seasonal hydrological forecasting suite could thus be compared with the random atmospheric experiment. Soil moisture and river flow skill scores clearly showed the added value in seasonal meteorological forecasts in the north of France, contrary to the Mediterranean area where values worsened.

1 Introduction

Water resources are known to be unevenly distributed in space and time on Earth. Moreover, in addition to the existing climatic pressure, anthropogenic pressure is increasing as the water demands of the human population grow. Therefore, water resource managers need decision support tools in order to anticipate future water availability for human and industrial consumption, hydropower or irrigation purposes. Predicting low flows and droughts several months in advance would be a useful tool for these managers. For example, predictions in the spring period (March–April–May) can be used to detect signals of a drought onset in spring in order to help water resource managers taking decisions for the summer low-flow period.

Seasonal hydrological forecasting systems have been developed in several regions of the world in the last decade. They are based on predictions of both the hydrological system and meteorological forcing. The former is associated with the slow components of the hydrological system: soil moisture, the presence of aquifers, and snow cover (Bierkens and Van Beek, 2009; Douville, 2009; Bohn et al., 2010). The prediction skill associated with soil moisture memory may last up to two months (Koster et al., 2001). The success of seasonal hydrological forecasts also depends on the season, because of dry or wet land surface initial conditions (Wood

and Lettenmaier, 2008; Li et al., 2009). Snow cover is especially influential during the spring period as it mostly contributes to mountain river flows and is thus the main source of hydrological predictability for snowmelt dominated headwater basins, such as the South Saskatchewan River Basin in Canada (Gobena et al., 2010). The size of the river basin also has an important impact: for instance, in the Ohio River Basin with a wide range of basin sizes from a few hundred to over a ten thousand square miles, Li et al. (2009) found that the larger the basin, the stronger the influence of initial conditions. Meteorological forcings also contribute to the predictive skill of seasonal hydrological forecasts as total precipitation has a predominant effect on river flow (Materia et al., 2010; Li et al., 2009). Nevertheless, the predictive skill of a seasonal meteorological forecast depends strongly on the region and season considered and is usually weak at mid-latitudes (Kirtman and Pirani, 2008, 2009).

There are several sources of predictability of a hydrological system at the seasonal time scale according to the region of interest. Countries corresponding to different climatic regions therefore provide seasonal hydrological forecasts based on different predictors. In Senegal, it is helpful to consider the storage available at the end of the monsoon when programming water releases from a dam (Bader et al., 2006) whereas in Australia, Chiew et al. (2003) have demonstrated that a simple information based on ENSO-streamflow teleconnection and serial correlation in streamflow leads irrigators to take more informed risk-based management decisions. The North Atlantic Oscillation (NAO) or the Southern Oscillation Index (SOI) is also used to provide the magnitude of seasonal streamflow in Iran (Araghinejad et al., 2006).

Once sources of predictability have been identified, an appropriate methodology has to be chosen to provide the seasonal hydrological forecasts. For instance, studies conducted in the United States are mainly based on the macroscale semi-distributed grid based hydrological model VIC (Variable Infiltration Capacity, Liang et al., 1994). There are several ways of forcing the hydrological model. The first approach is to use statistical methods with simple or multiple linear regression between climatic phenomena (El Niño-Southern Oscillation and the Arctic Oscillation) or persistence related to soil moisture and snow cover with the mean seasonal river flow (Maurer and Lettenmaier, 2004). Wood and Lettenmaier (2006) used an ensemble streamflow prediction system with several daily hydrological model outputs provided by climate sequences resampled from previous years, taking the uncertainty of the initial atmospheric and/or oceanic conditions into account. In order to improve seasonal hydrological forecasts, more complex approaches have also been applied. Dynamic methods have been used with temperature and precipitation, with a Bayesian method merging observations with multiple seasonal forecasts (Luo and Wood, 2008). This method allowed the hydrological forecasting system to be evaluated for historical phenomena such as the 2007 US drought (Li et al., 2008).

France presents highly variable hydrometeorological conditions. A first evaluation of a seasonal hydrometeorological forecasting suite has recently been performed for the spring season (the entire March-April-May period) with an initialisation at the beginning of February (Céron et al., 2010). This study showed a higher predictive skill for hydrological variables than for near-surface atmospheric variables.

The objective of this paper was to continue the work of Céron et al. (2010) by undertaking a comprehensive assessment of the predictive skill of seasonal hydrological forecasts. This work was performed for the whole of France and included a determination of the main sources of prediction skill at the seasonal scale. The focus remained on the spring season as it is a season marked by snowmelt and is also critical for the onset of agricultural and hydrological droughts and low flows. Furthermore, thanks to the availability of a new hindcast dataset for the ARPEGE numerical climate prediction model (Weisheimer et al., 2009), the time period of the study was extended to the 1960–2005 period. A set of experiments was designed to identify the main sources of predictability of the hydrometeorological system. Then, the added value of seasonal atmospheric forecasts was assessed through the comparison with forecasts using random atmospheric forcings.

Section 2 introduces the different models and data sources used, with the description of the SIM hydrometeorological suite and the ARPEGE meteorological hindcast dataset. Section 3 describes the predictability experiments, the seasonal hydrological forecasting model and the forecast evaluation tools. Next, results in terms of soil moisture and river flows are shown in Sect. 4. Results are discussed in Sect. 5 before perspectives are provided in the last section.

2 Models and data sources

2.1 The hydrometeorological SAFRAN-ISBA-MODCOU (SIM) suite and reanalysis

The seasonal hydrological forecasting suite was the same as that used by Céron et al. (2010). It is based on the SAFRAN-ISBA-MODCOU (SIM) operational model developed by Météo-France and Mines Paris-Tech at the scale of France (Habets et al., 2008) and composed of three independent models.

First, SAFRAN (“Système d’Analyse Fournissant des Renseignements A la Neige” for “Analysis system contributing to information for snow”) is a near-surface meteorological analysis system (Durand et al., 1993; Quintana-Seguí et al., 2008; Vidal et al., 2010a). It combines meteorological model outputs with surface observations to produce hourly values of meteorological variables. SAFRAN computes seven variables (10-m wind speed, 2-m relative humidity, 2-m air temperature, incoming solar and atmospheric/terrestrial

radiation, snowfall and rainfall) over 615 climatologically homogeneous zones at several elevations, which are interpolated onto a 8-km grid covering France (total area: 544 000 km²). The long-term SAFRAN reanalysis derived by Vidal et al. (2010a) over the 1958–2008 period was used as a meteorological reference for all experiments in this study.

Next, ISBA (“Interface Sol-Biosphère-Atmosphère” for “Interaction between Soil-Biosphere and Atmosphere”, Boone et al., 1999; Noilhan and Planton, 1989) is a soil-vegetation-atmosphere transfer (SVAT) scheme, used to simulate the exchanges in heat, mass and momentum between the continental surface (including vegetation and snow) and the atmosphere. ISBA was applied here in its 3-layer force-restore version (Boone et al., 1999) with the 3-layer snow scheme of Boone and Etchevers (2001). A subgrid runoff scheme (Habets et al., 1999a) and a subgrid drainage scheme (Habets et al., 1999b) have been implemented to tackle the issue of physical processes occurring at scales smaller than the 8-km grid. ISBA thus simulates runoff through the Dunne mechanism over saturation. For soil moisture below the saturation point, the subgrid runoff is activated, its amount being smaller below the field capacity, and zero below the wilting point. Next, drainage is produced for soil moisture above the field capacity, and residual drainage is effective below this value where no aquifer layer is explicitly modelled by the MODCOU hydrogeological model. With respect to ISBA, the variables of interest for the present study are the total snow cover and the soil moisture (related to agricultural drought) described by the Soil Wetness Index (SWI) averaged over the soil depth:

$$SWI = \frac{W - W_{wilt}}{W_{fc} - W_{wilt}} \quad (1)$$

with W the soil water content, W_{fc} the water content at field capacity and W_{wilt} the water content at the wilting point.

Last, the MODCOU (MODèle COUplé for coupled model) hydrogeological model computes the temporal and spatial evolution of aquifers with several layers using the diffusivity equation (Ledoux et al., 1989). In addition to calculating the interaction between the aquifer and the river, the model routes the runoff on the surface and within rivers using an isochronistic algorithm to estimate river discharge with a time step of 3 h. The time step used to compute the evolution within the aquifer is about 1 day. In the version of SIM used here, aquifers are explicitly modelled in only two river basins: the Seine basin (three layers) and the Rhône basin (one layer).

The SIM hydrometeorology suite has previously been validated on four large French river basins: Adour (Habets, 1998), Rhône (Etchevers et al., 2001), Garonne (Voirin-Morel, 2003) and Seine (Rousset et al., 2004). It was then applied to the whole of France and validated over a 10-year period for 881 French stations to produce realistic water and energy budgets, streamflow, aquifer levels and snowpack

simulations (Habets et al., 2008). The French environment ministry uses outputs from the SIM model (snow cover, soil moisture and effective rainfall) for the Hydrological Monitoring Bulletin (<http://www.eaufrance.fr>).

The SAFRAN reanalysis has also been used to run the ISBA-MODCOU hydrological model in order to build a SIM reanalysis from 1958 to 2008 (Vidal et al., 2010b), taken here as the hydrological reference run for all experiments for the March–April–May (MAM) period. In addition, the SIM reanalysis allowed us to provide hydrological variables on 31 January for building the hydrological initial state used in all experiments.

2.2 The ARPEGE meteorological seasonal forcings

Hindcasts of the ARPEGE (“Action de Recherche Petite Echelle Grande Echelle” for “Research Project on Small Scale and Large Scale”) global coupled atmosphere-ocean climate model were used at a resolution of 2.5°. These data were produced within the ENSEMBLES project (Weisheimer et al., 2009) and covered the 1960–2005 period. Spring seasonal forecasts started on 1 February and ended on 31 May. These forcings, called ARPEGE-ENSEMBLES in the following, consisted of an ensemble of 9 runs corresponding to 9 initial conditions constructed by different realistic estimates of observed states of both the atmosphere and the ocean.

The ARPEGE-ENSEMBLES atmospheric forcing dataset was downscaled to the SIM horizontal resolution of 8 km with the simple method proposed by Rousset-Regimbeau et al. (2007) for ensemble medium-range river flow forecasts and adapted to seasonal forecasting by Céron et al. (2010). This downscaling method is explained hereafter. The original ARPEGE-ENSEMBLES temperature and total precipitation fields were first converted into anomalies, by removing their mean value, and then standardized by dividing them by their interannual standard deviation. They were then interpolated with an inverse-square weighting onto the 615 climatologically homogeneous zones considered in the SAFRAN analysis (Quintana-Seguí et al., 2008). Finally, they were combined with SAFRAN long-term means and interannual standard deviations to provide realistic 8-km atmospheric forcings that included local-scale spatial variability. The partition between snowfall and rainfall was based on a critical threshold temperature of 0.5 °C. As in Céron et al. (2010), the other atmospheric variables required by ISBA (wind speed, relative humidity, incoming solar and atmospheric radiations) were taken from the SAFRAN climatology over the same 1960–2005 period. As the ARPEGE dataset was available every 6 h for temperature and at a daily time step for total precipitation, a temporal disaggregation was also required: the total precipitation was evenly distributed throughout the day whereas temperatures were linearly interpolated between two time steps.

Céron et al. (2010) use seasonal hindcasts produced within the DEMETER project (Palmer et al., 2004) from an older version of ARPEGE, called ARPEGE-DEMETER in the following. Seasonal hindcasts were taken here from ARPEGE-ENSEMBLES runs rather than ARPEGE-DEMETER runs for two main reasons. Firstly, the ARPEGE-ENSEMBLES seasonal forecasting model is currently closer to the operational seasonal forecast model than the ARPEGE-DEMETER one. Secondly, the time period was extended from 1971–2001 to 1960–2005. Finally, the ENSEMBLES predictions were significantly better than those from DEMETER, with improved discrimination, resolution and reliability in the northern midlatitudes for the spring season (Alessandri et al., 2011). Moreover, before being used here, the ARPEGE-ENSEMBLES meteorological seasonal forecasts had been evaluated and compared with the ARPEGE-DEMETER ones. The results (not shown here) found no bias in ARPEGE runs from either project in terms of temperature and total precipitation. Moreover, we observed that the prediction skill was higher for temperature than that for total precipitation in both experiment sets. Then, we observed that there were an overestimation of rainfall and an underestimation of snowfall in both ARPEGE-ENSEMBLES and ARPEGE-DEMETER forcings. Finally, for both seasonal atmospheric forcing, lower skill scores can be found over the Mediterranean area.

2.3 Description of catchments

France presents highly variable hydrometeorological conditions with total precipitation about 500 mm yr^{-1} for dry regions and more than 2000 mm yr^{-1} for mountains. Indeed, there are two high mountain regions (Pyrenees and Alps) and several medium-elevation mountain ranges (Vosges, Jura, Massif Central and Corsica) distributed over the territory (Fig. 1). These regions are usually associated with higher amounts of precipitation and the presence of seasonal snow cover with a nival and nivo-pluvial flow regime, for example for the Durance catchment at Embrun in the Alps and the Ariège catchment at Foix in the Pyrenees (see Fig. 1 for gauging locations).

Among the four main rivers representing more than 62 % of the territory, the Rhône has the most mountainous catchment area and is strongly influenced by snowmelt in spring and summer and is subject to anthropogenic pressure with numerous dams. The Seine river basin is marked by a large and complex aquifer system with very specific hydrological behaviour, but the flow regime is essentially pluvial with floods in autumn and winter (from December to April) and a low flow period in spring and summer.

From a meteorological point of view, France is characterized by westerly flows corresponding to an Atlantic influence, with the exception of the south-east region, which has a Mediterranean climate with dry and highly variable meteorological conditions (high flows in autumn and winter

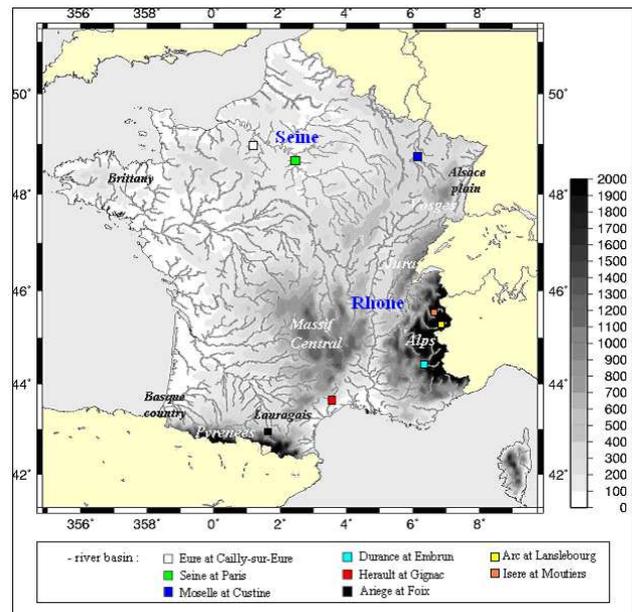


Fig. 1. Orography (m), hydrographic network over France, and location of gauging stations for catchment case studies.

contrasting with very low flows in summer). Table 1 summarizes the characteristics of the six catchments studied in Sect. 4.3.2 and identified on Fig. 1.

3 Methods and experiments

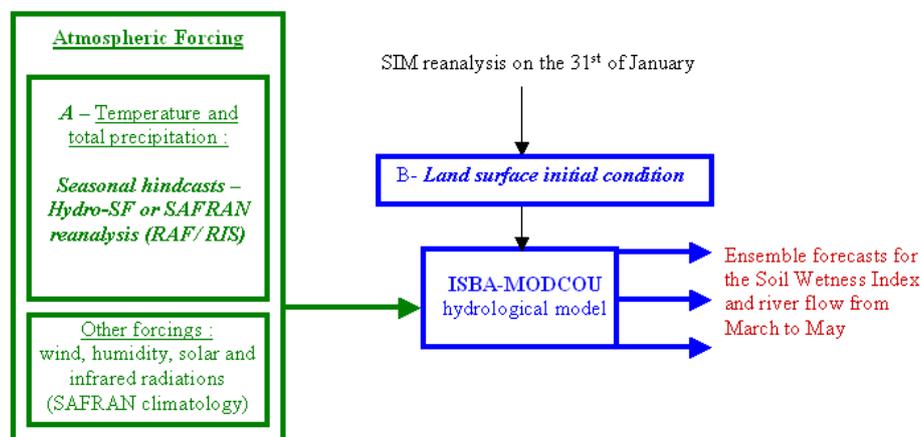
3.1 Predictability experiments

Two academic experiments were conducted, with the aim of better understanding the respective roles of the land surface initial state and the atmospheric forcings in the predictive skill of the complete hydrometeorological system. They consisted of runs initialised on 1 February for a period ending on 31 May, without considering the first month. Data constituting meteorological forcings came from the SAFRAN reanalysis over the 1960–2005 period. In order to avoid potential biases due to different ensemble sizes on probabilistic scores when comparing the results, both experiments were based on the 9-member ensembles, following the size of the ARPEGE seasonal atmospheric ensemble hindcasts used later for comparisons. All experiments in this paper followed the general scheme described in Fig. 2.

A process was designed to select 9 random years for each year simulated from the 1960–2005 period with atmospheric forcings or land surface initial conditions depending on the experiment. In order to preserve consistency between the different meteorological or land surface variables for each experiment, the process selected all variables from the same year. Moreover, the random years selected are the same for the two experiments described below.

Table 1. Characteristics of the six catchment studied and located on Fig. 1.

	Durance at Embrun	Herault at Gignac	Ariege at Foix	Eure at Cailly-sur-Eure	Seine at Paris	Moselle at Custine
Basin area (km ²)	2170	1312	1340	4598	43 800	6830
Outlet altitude (m)	787	32	375	21	26	184
Mean flow (m ³ s ⁻¹)	52	29.2	39.6	17.9	305	113

**Fig. 2.** General scheme of the ensemble seasonal hydrological forecasting suites used in this study (see Table 2).

The first experiment, called Random Atmospheric Forcing (RAF) tested the impact of a realistic land surface initial state. The land initial conditions for soil moisture, snow cover and aquifers were taken from the SIM reanalysis on 31 January. The RAF forecasts were performed using 9 members, each member corresponding to the atmospheric forcing (temperature and total precipitation) for a random year selected from the 46-year SAFRAN reanalysis.

The second experiment, called Random land surface Initial State (RIS) was complementary to the RAF experiment and evaluated the atmospheric forcings predictive skill. The atmospheric forcings used here came from the SAFRAN reanalysis for each target year and the RIS ensemble forecasts used 9 land surface initial conditions randomly chosen within the 46-year SIM reanalysis.

Table 2 summarizes the atmospheric forcings and land surface initial states used in the two experiments.

3.2 The Hydrological Seasonal Forecasting suite (Hydro-SF)

In order to perform seasonal hydrological forecasts over the 1960–2005 period and for the entire spring period, following the general scheme described in Fig. 2, the land initial conditions for soil, snow cover and aquifers were taken from the SIM reanalysis on 31 January for each year from 1960 to 2005 as in the RAF experiment (see Table 2). Then, atmospheric forcings were provided by the 9 members of the ARPEGE-ENSEMBLES meteorological seasonal forecasts initialised on 1st February of each year (see Sect. 2.2). The seasonal hydrological forecasting suite, called Hydro-SF in the following, thus provided 9 runs of soil moisture and river flow forecasts over the entire March–April–May period.

3.3 Evaluation methods

Seasonal forecasts are basically ensemble forecasts and thus provide both probabilistic and deterministic – using the ensemble mean – forecasts. They can thus be evaluated on both aspects and, consequently, the evaluations have to refer to

Table 2. Description of the RAF and RIS experiments testing the predictability of the hydrological system and the Hydro-SF suite. RAF: Random Atmospheric Forcing; RIS: Random land surface Initial State; Hydro-SF: the hydrological seasonal forecasts (see Fig. 2).

Experiment	Atmospheric forcings	Land surface initial states (soil moisture, snow cover and piezometric level)
RAF	9 random years from the SAFRAN reanalysis	Actual states from the SIM reanalysis
RIS	Actual years from the SAFRAN reanalysis	9 random states from the SIM reanalysis
Hydro-SF	9 ARPEGE seasonal hindcast runs	Actual states from the SIM reanalysis

both probabilistic and deterministic scores. In this paper, time correlation and Brier score and its decomposition are shown as examples of deterministic and probabilistic scores respectively, but many other scores (dispersion, root mean square error, standard deviation, spatial correlations, Relative Operating Characteristic curves and areas...) were computed and gave results similar to those presented below.

The prediction skill of experiments was calculated over the 1960–2005 period and the entire MAM period, with the SIM reanalysis as the reference (see Sect. 2.1). It was computed over each 8-km grid cell for the SWI (see Eq. 1) and using all 881 river gauges for river flows.

Time correlations were used to characterize the ability of the hydrometeorological suite to match the reference inter-annual variability. They were calculated from the ensemble mean as this is considered to be the best representation of a deterministic forecast from an ensemble seasonal atmospheric forecast. In the following, we consider that only time correlations higher than approximately 0.3 are significant (based on the Student test over a sample of 46 years with a significance level of 95 %).

Next, the skill of the system for a threshold exceedance was assessed through the probabilistic Brier Score (BS, Brier, 1950) using the whole ensemble distribution (Eq. A1 in Appendix A). The BS and its associated skill score (BSS) (Brier, 1950) are well known and often used as probabilistic scores for hydrological ensemble forecasts (Cloke and Pappenberger, 2009; Randrianasolo et al., 2010; Thirel et al., 2010). The lower the score the better the forecast, with a perfect forecast corresponding to a BS of 0. BS can also be decomposed as the sum of 3 terms: reliability, uncertainty and resolution (Murphy, 1973), see Eq. (A2) in Appendix A. The reliability term describes the capacity of the system to predict correct probabilities and is negatively oriented. In principle, it can be reduced by good calibration (Murphy, 1986). A small value of the reliability indicate a reliable forecast. The resolution term gives the ability of the system to correctly separate the different categories (whatever the forecast probability), i.e. it measures how much the conditional probabilities differ from the climatic average. It is positively oriented: the higher the resolution, the better the forecast. Finally, the uncertainty is exactly the BS (Eq. A1) for the sample climatology as the uncertainty is the variance of observations for

the considered event. For all hydrological variables, the SIM climatology over the 46 years was used to determine terciles and the corresponding thresholds of tercile categories. In this paper, we tested the skill of the system to predict above average (upper tercile) or below average (lower tercile) values.

In order to make comparisons between the seasonal hydrological forecasting suite and the random atmospheric forcing experiment, a bootstrapping method (Hesterberg et al., 2005) was used with a Student test on the Brier Skill Score (BSS) (Eq. B1) and the difference of time correlations (see Appendix B).

4 Results

4.1 RAF

Figure 3a shows the SWI predictive skill for spring using correlation between the RAF experiment and the reference value obtained from SIM reanalysis. About one third of France exhibited significant correlations. Correlations were maximum in the highest mountains (South and Central Alps, Pyrenees), but were also higher than 0.4 in most parts of the other mountain ranges (Vosges, Jura, Massif Central and Corsica). These high scores could be attributed to the influence of the snow cover initial state. In addition, significant correlations were found in some plain areas scattered over the country: the Alsace plain, the south-west of Paris, the Lauragais region close to the Mediterranean sea, and the lower Rhône valley. These last two regions are amongst the driest regions of France, whereas the south-west of Paris, for instance, is covered by forests and has deep root layers with an evapotranspiration/precipitation ratio exceeding 0.75 (Habets et al., 2008). Because of these diverse factors associated with the soil moisture memory, the interannual variability of initial SWI values was large enough to lead to some predictive skill during the spring season. In contrast, in more rainy areas such as western Brittany and the French part of the Basque country, the soil water content is often close to the field capacity, hence the soil moisture interannual variability in winter is low, cancelling the soil moisture predictive skill. This meaning that the interannual variability is low compared with summer periods when the interannual variability is high.

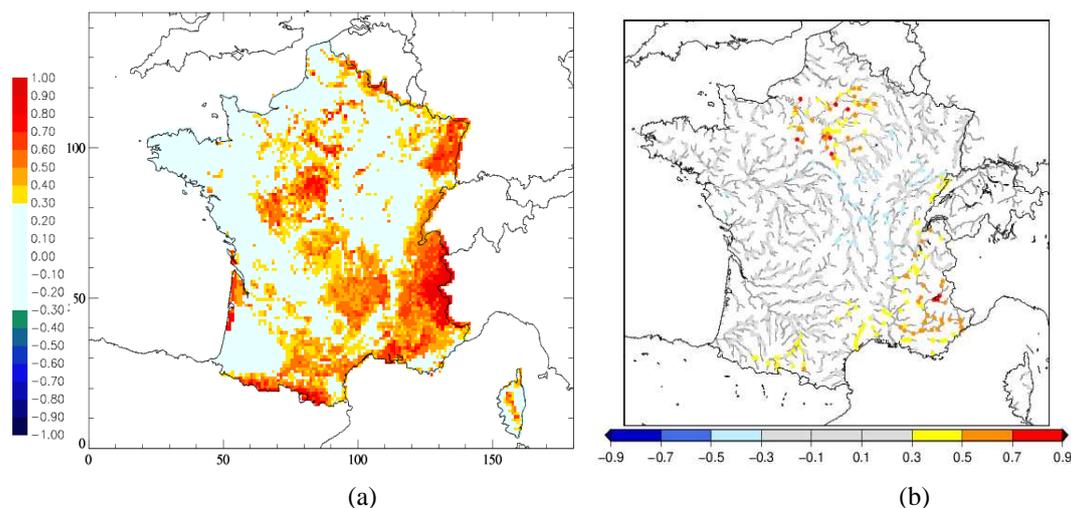


Fig. 3. Correlation maps of SWI (a) and river flows (b) between the RAF experiment and the SIM reanalysis reference run for the spring (MAM). Scores are calculated over the 1960–2005 period.

When looking at river flow forecast skill (Fig. 3b) some spatial differences can be spotted. On average (excluding the case of the Seine river basin, which will be discussed later) locations associated with a significant skill were fewer than those for soil moisture. Most areas where the soil moisture predictive skill came from the initial soil moisture did not exhibit any skill for river flow (e.g. Alsace, south-west of Paris). Indeed, below the field capacity, the bottom runoff production stopped (except for the residual drainage), cancelling the transmission of the soil moisture signal to river flows. In the case of mountains, the river flow skill was maximum in the Southern Alps. For example, the maximum value was associated with the Durance river at Embrun (cf. Fig. 1), a high mountain catchment (up to 4000 m a.s.l.). For this river, the annual snowmelt maximum occurs in May and the simulated cumulated discharge during the spring period corresponds to 47 % of the annual discharge. Hence, this experiment captured a large part of the predictive skill contained in the snow cover initialized at the end of January. In contrast, in the Northern French Alps, the annual maximum of discharge occurs mostly in June, and the spring discharge represents only around 25 % of the annual value (22 % for the Arc at Lanslebourg, 32 % for the Isère at Moutiers, see Fig. 1 for catchment locations). As the forecast ended at the end of May, the predictability associated with the snowmelt in June was not captured in this experiment. In other mountain ranges, the river flow skill was lower because of more limited snow cover due to either a warmer climate (Pyrenees and Corsica) or lower elevations (Vosges, Jura and Massif Central).

In addition, some significant river flow skill appeared in the Seine catchment, where a large multilayer aquifer system simulated by the MODCOU model influences river flows and the configuration of the river-aquifer exchanges at the scale of each sub-catchment. The time correlation varied from 0.3

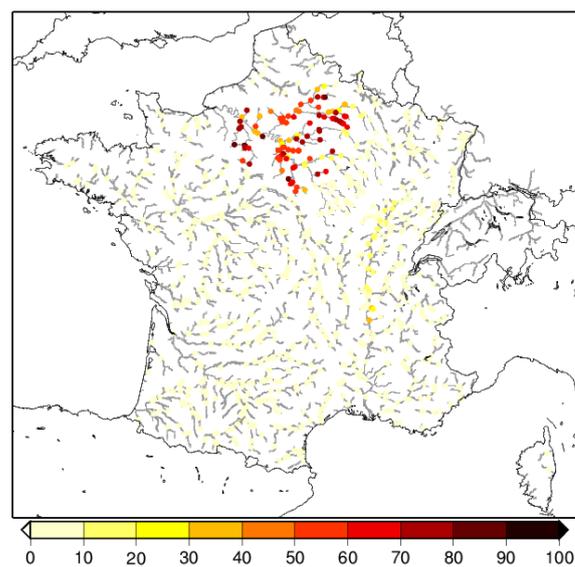


Fig. 4. Map of percentage of groundwater contribution to spring river discharge over the 1960–2005 period, calculated with the SIM reanalysis.

to 0.9 depending on the hydrogeology (Fig. 3b). The Seine hydrological features are very complex as there are several aquifer layers stacked on each other with a specific geological layout. Figure 4 presents the percentage of groundwater contribution to spring river discharge which is indeed the percentage of the amount of water transferred from the aquifer to the river compared to the amount of water flowing at a given station. This calculation is directly computed by the MODCOU model for each time step and “river” grid meshes. Indeed, if the groundwater table level is upper than the river

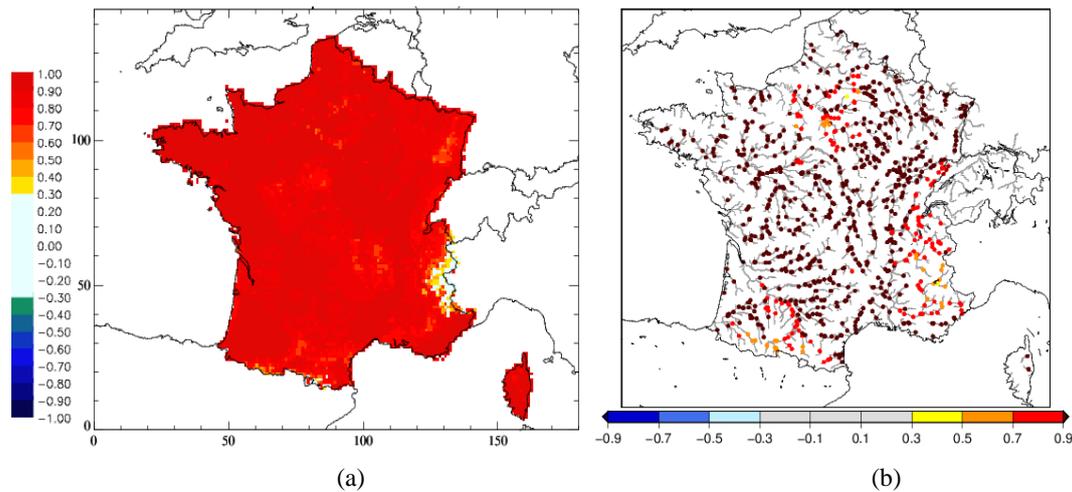


Fig. 5. Correlation maps of SWI (a) and river flows (b) between the RIS experiment and the SIM reanalysis reference run for the spring season. Scores are calculated over the 1960–2005 period.

level, the water is transferred to the river using a transfer coefficient:

$$Q = TP (H - H_0)$$

with H the river level; H_0 , the groundwater table level; TP , a transfer coefficient. The river flow Q exchanged is thus positive (negative) when the groundwater (river) gives water to the river (groundwater). The latter case is not implemented in the present version of SAFRAN-ISBA-MODCOU. Generally, we see on Fig. 4 that the skill increases with the relative importance of water coming from the aquifer in the cumulated spring discharge. However, the alluvial aquifer in the Saône/Rhône valley did not generate any significant predictability, showing that only aquifers with a sufficient delayed time response and water holding capacity can lead to predictability at seasonal scales for the spring season.

4.2 RIS

Conversely to the RAF experiment, we focused here on the reduction of hydrological prediction skill as actual atmospheric forcings from the SAFRAN reanalysis were used. The fact that the SWI prediction skill was significant and high almost everywhere was therefore not surprising (cf. Fig. 5a). The only exceptions were some parts of the Alps, and a very small region in the eastern Pyrenees, confirming the importance of the snow cover initial state in these high-elevation areas.

On Fig. 5b, the river flow prediction skill was significant everywhere. It was greater than 0.9 in most regions where the surface initial state influence was negligible. It reached a minimum in the regions mentioned above for RAF: mountainous regions (Alps and Pyrenees) and associated downstream areas (snow influence), as well as most of the Seine catchment (aquifer).

Table 3. Contingency table of biases on river flow ($\text{m}^3 \text{s}^{-1}$) of RAF and RIS experiments for Durance river basin at Embrun (Alps) over the 1960–2005 period. RAF: Random Atmospheric Forcing; RIS: Random land surface Initial State.

	RAF		
	0–20	20–40	>40
RIS 0–20	15	8	3
RIS 20–40	14	1	1
RIS >40	3	1	0

Table 3 shows the RAF error on spring discharge as a function of the RIS error in a contingency table for the Durance at Embrun, a mountain river basin (cf. Fig. 1 for location) over the 1960–2005 period. This highlights that, when river flows are well simulated for a year in the RAF experiment, the river discharge is badly simulated in RIS and vice versa. So, the contributions of the land surface initial state and atmospheric forcings vary and depend on years, introducing a predicting skill for specific years.

4.3 Added value of seasonal atmospheric forecasts

4.3.1 SWI forecasts

Figure 6a shows the time correlation between SWI forecasted using Hydro-SF experiment and its reference value obtained from the SIM reanalysis. A comparison with corresponding results for RAF (Fig. 3a) is presented in Fig. 7a, showing the impact of using the ARPEGE-ENSEMBLES seasonal forecasts instead of random forcings from the SAFRAN reanalysis. The Student variable of the difference in correlations (see

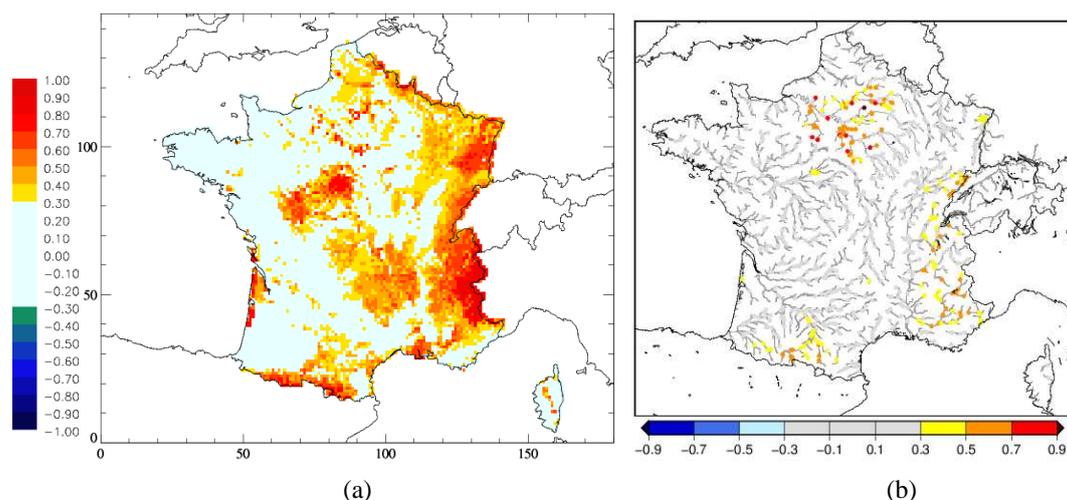


Fig. 6. Correlation maps of SWI (a) and river flows (b) between Hydro-SF and the SIM reanalysis reference run for the spring season. Scores are calculated over the 1960–2005 period.

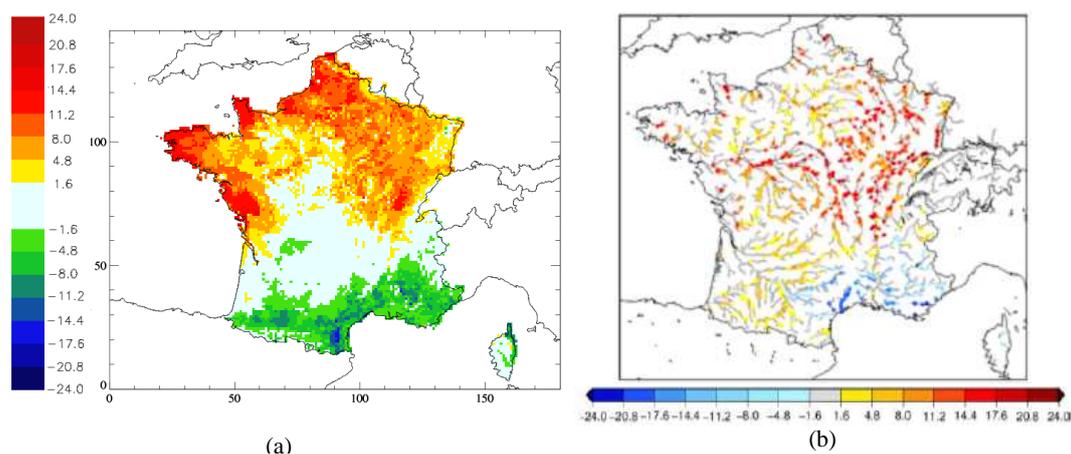


Fig. 7. Maps of Student variable of the difference of correlation (cf. Appendix B) between Hydro-SF and the RAF experiment for SWI (a) and river flows (b) for spring.

Appendix B) for spring clearly showed a north/south partition of France. The Student variable was significantly positive in the north, showing a higher skill of Hydro-SF compared to the RAF experiment. Conversely, negative Student variables in the south showed a higher skill of the RAF experiment. Between negative and positive values, a large area exhibited non-significant skill.

Differences between Hydro-SF and RAF inferred from probabilistic scores were more complex than the time correlation (Fig. 8) as no clear delineation appeared. Hydro-SF still worsened the results in the Mediterranean part of France for SWI for the upper tercile and the south of France for the lower tercile. However, it must be noted that results were improved in the south west of Paris, which still showed the highest scores for both RAF and Hydro-SF experiments (cf. Figs. 3a and 6a).

Table 4 displays values of the SWI Brier Score (Eq. A1) averaged over the whole of France. It shows that the predictive skill of Hydro-SF was similar to that of the RAF experiment (it was equivalent or lower), thus hiding the highly variable spatial patterns. This clearly highlighted the need to resort to a spatial representation in order to properly assess seasonal hydrological forecasts.

4.3.2 River flow forecasts

Figure 6b shows the time correlation between river flow forecasted using Hydro-SF experiment and its reference value obtained from the SIM reanalysis. Here again, the Alps and Pyrenees displayed higher scores (from 0.3 to 0.7), the Seine river basin had values up to 0.9, whereas the other regions showed no significant predictability of river flows.

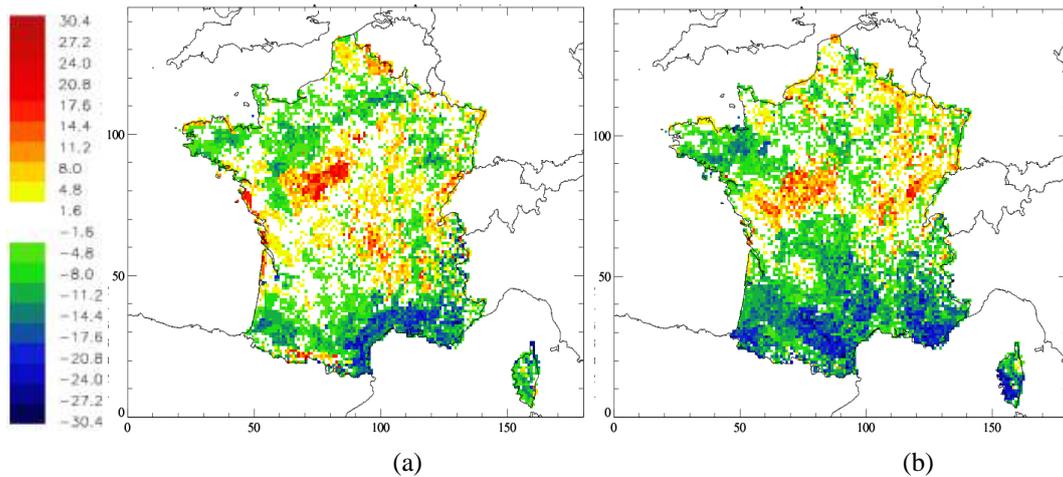


Fig. 8. Maps of Student variable of Brier Skill Score (B1) for SWI between Hydro-SF and the RAF experiment for the spring season for the upper tercile (a) and the lower tercile (b).

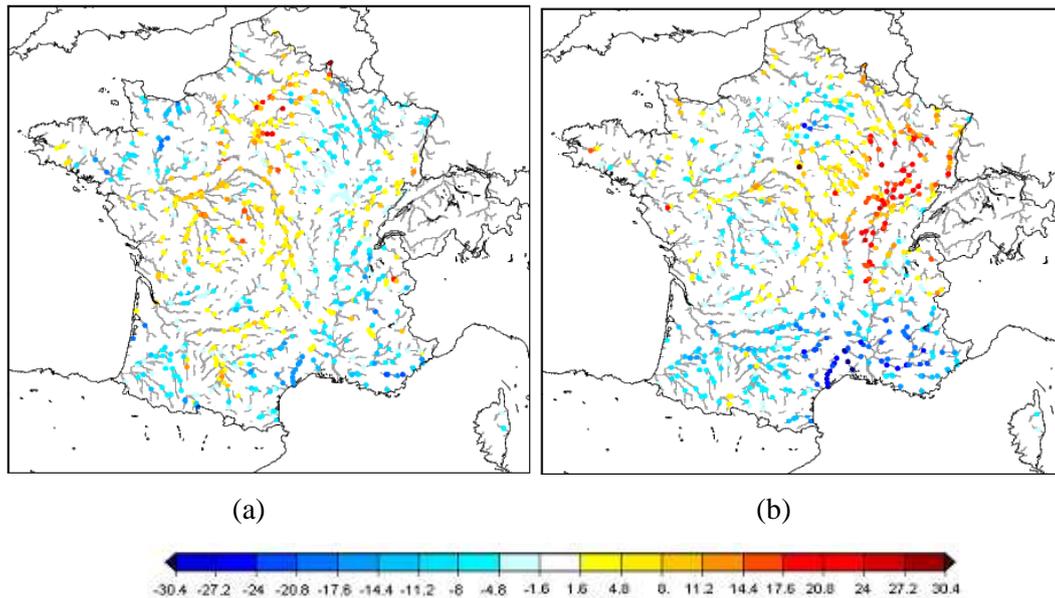


Fig. 9. Maps of Student variable of Brier Skill Score (B1) for river flows between Hydro-SF and the RAF experiment for the spring season for the upper tercile (a) and the lower tercile (b).

Consequently, at first sight, the spatial distribution of scores was quite similar to that of the RAF experiment (Fig. 3b).

Secondly, by looking at the Student variable of difference of time correlation (see Appendix B) on Fig. 7b, it can be noted that scores on river flows were significantly positive over most of France, meaning that Hydro-SF improved river flow forecasts compared to the RAF experiment, except for the Mediterranean area. Moreover, the Student variable of BSS for river flow between Hydro-SF and RAF (Fig. 9) did not show any clear skill for the upper tercile while the skill was significantly positive in the north-east of France for the lower tercile. This showed that, for probabilistic

scores, Hydro-SF was better than RAF for river flows over this region.

Finally, the BS and its three terms of decomposition (reliability, resolution and uncertainty) (Eq. A2) on river flow forecasts for Hydro-SF and RAF were shown in Fig. 10 for some catchment case studies: two catchments located in plains, the Moselle at Custine (north-east of France) and the Herault at Gignac (Mediterranean area); two catchments in the Seine river basin, the Eure at Cailly-sur-Eure (with high groundwater influence) and the Seine at Paris (less influenced by groundwater); and, finally, two catchments located in mountainous regions, the Durance at Embrun (Southern Alps) and

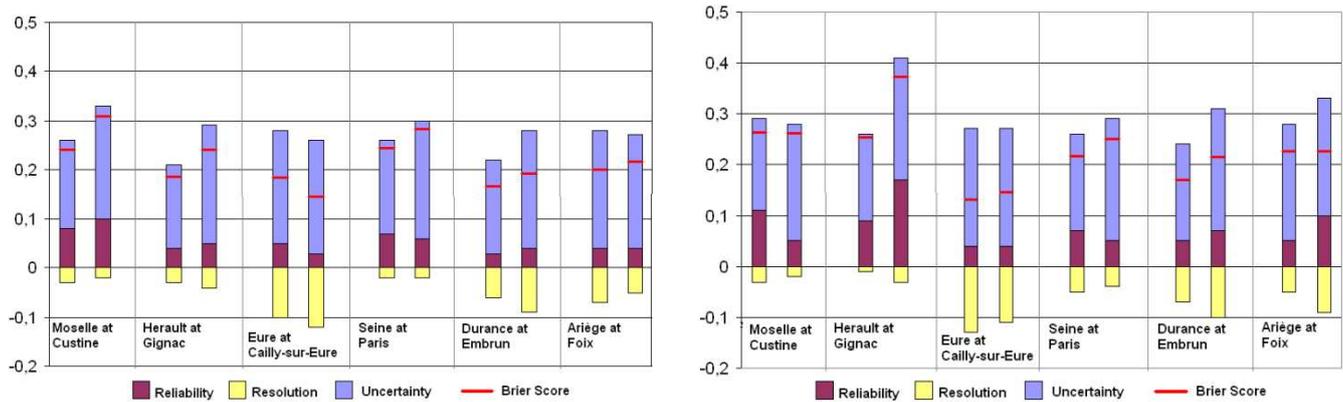


Fig. 10. Histograms of the decomposition of Brier Score (reliability, resolution, uncertainty) (A2) and Brier Score (A1) for river flow forecasts from RAF (left panel) and Hydro-SF (right panel) for Spring over the 1960–2005 period. Graphs show the results from 6 different river catchments for the upper (left bar) and lower (right bar) tercile categories.

Table 4. Brier score (Eq. A1) averaged over France from 1960 to 2005 for the spring period and the evolution for each month with the RAF experiment and Hydro-SF for SWI (Eq. 1) forecasts. RAF: Random Atmospheric Forcing; Hydro-SF: the hydrological seasonal forecasts.

	Tercile	March	April	May	spring
RAF	Upper	0.20	0.24	0.25	0.24
	Lower	0.20	0.23	0.24	0.23
Hydro-SF	Upper	0.22	0.25	0.26	0.24
	Lower	0.21	0.26	0.26	0.24

the Ariège at Foix (Pyrenees) (see Fig. 1 for gauging location). Firstly, Brier Scores showed a lower skill (higher values) for catchments located in plains than for mountainous catchments in both experiments. This observation was partly due to the resolution term as the worst resolution (smallest value) was for the river basins located over plains. Secondly, the uncertainty term was not very different from one experiment to another because it was based on the observed reference data. However, the reliability term (which should be small) was the term that changed most between the two experiments, at least for the lower tercile. For instance, for the Herault, Ariège and Durance catchments, all located in the south of France, the reliability worsened from 0.04 for RAF to 0.17, 0.1 and 0.07 respectively for Hydro-SF for the lower tercile. In contrast, the Moselle catchment in the north-west of France showed a decrease of BS from 0.1 to 0.05 for the lower tercile. This probably explained BSS features on Fig. 9. The skill worsening in the south of France for Hydro-SF thus appeared as a reliability problem, which was encouraging because we could expect to improve it using calibration of probabilities and more ensemble members in the future.

5 Discussions and conclusions

In this study, several numerical experiments covering a 46-year period were performed using the SIM hydrometeorological suite in order to investigate the sources of spring predictability of soil moisture and river flows over France. Obviously it should be relevant to use a large ensemble size for the experiments (e.g. Li et al., 2009 used 19 members and Wood and Lettenmaier, 2008 used 21 members).

However we used 9 randomly selected initial states and atmospheric forcings for RIS and RAF experiments, respectively. The objective of choosing 9 random members only is to keep those experiments fully consistent with Hydro-SF experiment that uses 9 members of the ENSEMBLES dataset. Consequently, we verified that our random selection did not bias the results toward drier or wetter year. We especially checked that dry or wet years were not over- or under-represented in the samples. Let’s assume that a year is dry if it pertains to the driest 20 % of the sample (below lower quintile). Theses years are present in the random selection 18 % of the time. For the wetter years (above upper quintile), the percentage is 19.3 %. These values are not statistically different (95 % confident interval) from the 20 %, which suggests that the random selection did not generate biased samples.

Firstly, the main conclusions of this study allowed us to confirm that the snow cover initial state was by far the most important source of spring predictability in mountain areas. But the soil moisture and river flow predictive skill varied also among regions, according to climate and elevation. For instance, for medium-elevation mountains (Massif Central, Vosges) and high mountains in dry areas (south-east of the Pyrenees, Corsica) the influence of snow cover was significant on soil moisture but not on river flows. For the southern Alps and the rest of the Pyrenees, scores for both variables were significant at the seasonal scale. But scores were not significant over the Northern Alps, the most snowy area in

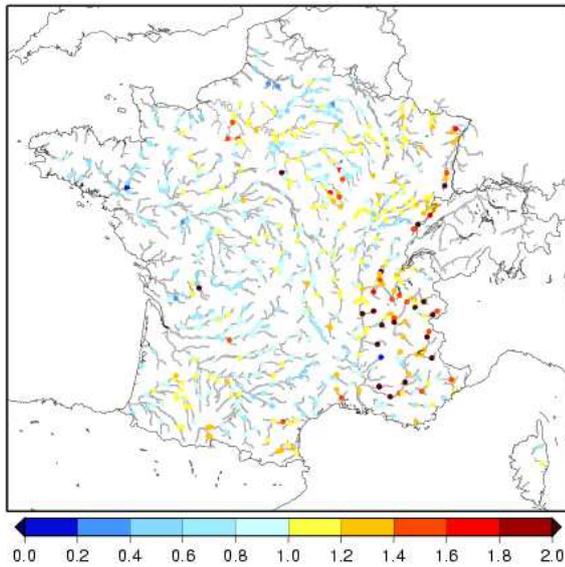


Fig. 11. Map representing ratios of river flow forecasted by the SIM model over river flow observed, calculated over the 1960–2005 period on Spring (March–April–May).

France, because of the delayed snowmelt in this region, with its maximum in June. The soil moisture and river flow predictive skill could therefore still prove significant with an extended forecast range in this region. Such a forecast, made possible by the 7-month forecast range of operational systems, is very promising, especially for the management of low-flow periods.

Secondly, the study showed that the presence of a deep aquifer could also be an important source of river flow predictability. The Seine aquifer system is the largest and deepest in France, with a great water holding capacity. While there was no impact on soil moisture forecasts over the catchment, the skill of river flow forecasts increased with the aquifer–river exchanges. In the eastern part of the basin (with higher amounts of precipitation), river flows are mostly influenced by runoff, whereas other tributaries are strongly influenced by the aquifer. It should be noted that results on the Seine cannot be generalized to other major aquifers. It is likely that the introduction of the Somme aquifer (north of Paris) into the model (Habets et al., 2010) will improve the results for this river and its tributaries because it belongs to the same hydrogeological unit. For alluvial aquifers, as shown for the Saône/Rhône aquifer already explicitly modelled, the signal will probably remain non-significant as the response of the aquifer is not delayed on a time scale relevant for spring seasonal forecasts.

Next, the present study showed that for most plains, the part of the skill associated with the soil moisture initial state was usually very low. Nevertheless, some specific regions were associated with a significant soil moisture skill. These were usually dry regions and/or regions with high vegetation and large soil reservoirs (e.g. the region south-west of Paris).

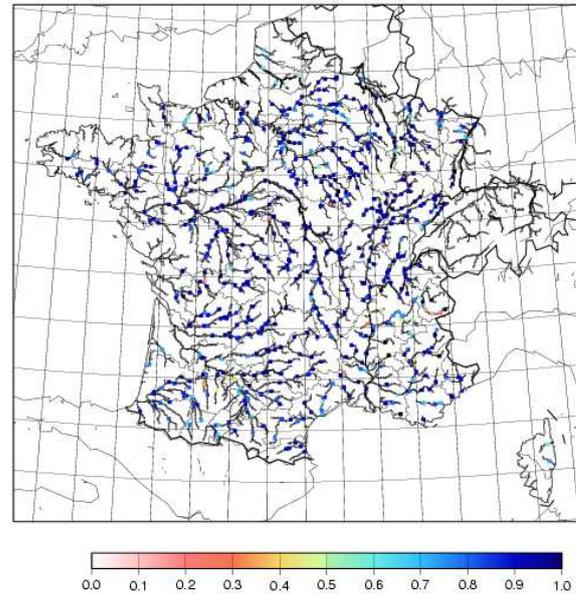


Fig. 12. Correlation maps of river flow between the SIM model and observations, calculated over the 1960–2005 period on Spring (March–April–May).

Finally, the use of meteorological forcings from ARPEGE-ENSEMBLES seasonal forecasts was then compared with a random forcing experiment. While a significant improvement of river flow skill could be observed in the north-east of France, scores reduced in the Mediterranean area. This can be explained by the worsening of seasonal atmospheric forecasts skill in the Mediterranean area of France.

In this study, we compared hydrological seasonal forecasts with its reference value obtained from SIM reanalysis, not from observations. The next step will be to compare hydrological seasonal forecasts with observations.

However, as a first step, we can study the behaviour of the SIM model compared with observed river flow in order to better characterize the reliability of the results. The discharge ratio in Spring (ratio of simulated vs. observed river flows) and the interannual correlation between simulated and observed spring mean river flows are shown on Figs. 11 and 12 respectively.

The first criterion qualifies the ability of SIM to reproduce the observed volume. Results are similar to already published comparisons over the whole year (Habets et al., 2008). The discharge ratio is generally close to 1, with some important exceptions on the Alps. It is partly the consequence of an accumulation of numerous dams used for hydropower production, thus influencing river flow observed. However as Lafaysse et al. (2011) showed, the overestimation of river flow over the Alpine region can also be explained by the grid discretization (the elevation range by each 8 km square grid is often wider than 1000 m). The consequence is

a poor estimation of meteorological variables (like snowfall), vegetation and snow cover. Moreover, in the Alpine region, the SIM model does not include water storage and release from aquifers nor ice melt from glaciers, inducing a time lag of snowmelt which occurs earlier in model results than in observations

The second criterion qualifies the ability of SIM to simulate the interannual variability. This criterion is very important in the framework of seasonal forecast. In most cases the correlation is very high (above 0.85), indicating that SIM is able to correctly predict this variability, even if the score on the discharge ratio is poorly simulated. For the Durance at Embrun, a typical Alpine river not influenced by dams, the discharge ratio is very poor (overestimation of 40 % of the discharge in Spring because of grid discretization and lack of local aquifers and glaciers in the model), while the interannual correlation on spring discharge with observation is 0.88. Hence, it is relevant to use SIM for seasonal prediction on this particular catchment.

6 Perspectives

All the above conclusions confirmed and extended the results of Céron et al. (2010) on selected basins. A number of perspectives can be envisaged based on the above conclusions.

First, this study confirmed the importance of the land surface initial state. Although we considered that the accuracy of the SIM suite was high for the simulation of the main components of the continental hydrological cycle at the scale of France (Habets et al., 2008), there is still room for improvement in its quality. This system will be completed in the future by new aquifers, which, hopefully, will lead to an improvement of scores in the corresponding regions, e.g. the Somme area (Habets et al., 2010) and the Rhine basin (Thierion et al., 2011). Improvements in the snow simulation can be achieved by taking better account of the orography over mountain catchments. Obviously, another source of improvement may lie in the assimilation of observed variables. The assimilation of remotely sensed soil moisture may be a good way to improve the soil moisture initial state (e.g. Draper et al., 2011). Concerning snow, a correct estimation of the snow cover amount is probably decisive (Wood and Lettenmaier, 2006), but space-based observations of the amount of snow in mountains are difficult to achieve and the representativeness of in situ observations is poor. The present approach based on the snow model of ISBA forced by a mesoscale meteorological analysis like SAFRAN that explicitly accounts for altitude effects may still be one of the best choices at these medium-range spatial scales.

Second, an extension to other seasons is needed. This first study was limited to the spring season in order to evaluate the skill associated with snow cover in comparison with the other sources of predictability. Spring is also a critical season for the onset of agricultural drought (Vidal et al., 2010b)

and accurate seasonal forecasts are therefore important in this time of year. However, for other seasons, the skill associated with the initial state might differ. In summer, the influence of the snow will probably remain significant – at least for June – in the Northern Alps, and large aquifers might improve the scores for river flows. In autumn and winter, the main sources of skill are hard to anticipate but we can infer that the atmospheric forcing might play a more important role than the initial state.

Third, the quality of atmospheric forcing may be improved by a refined downscaling of seasonal forecasts. In an additional experiment (not shown) based on the RAF experiment, we used all variables in the meteorological forcings of the randomly chosen year instead of only temperature and total precipitation. The only significant difference with RAF was an improvement in the shape of Talagrand diagrams (not shown), but other scores remained unchanged. This confirmed the crucial role of temperature and precipitation forecasts (including the snow/rain partition) in the forcing terms of the SVAT model. A downscaling approach based on weather types (Pagé et al., 2009) is planned in order to better account for large-scale atmospheric patterns. This method was developed by Boé et al. (2006) and validated using SIM over the Seine basin. It has also been applied for a climate impact assessment on hydrology over France (Boé et al., 2009) and is promising for applications in seasonal forecasting.

Another source of improvement of meteorological forcings would be the use of a multi-model approach, rather than the single ARPEGE model. In a second step, the multi-model approach should be expanded to the hydrological modelling step as it represents a major source of uncertainty in the forecasting suite. The ensembles technique could also be applied to the surface initial state in order to take account of the uncertainty of this component, which appears to be important, especially for mountainous areas.

Appendix A

Brier Score and its decomposition

The Brier Score (BS) quantifies the ability of an ensemble forecast to predict an exceedance (or non-exceedance) of a threshold. Indeed, BS is a quadratic measure of error in probabilistic forecasts (Mason, 2004) :

$$BS = \frac{1}{N} \sum_{k=1}^N (y_k - o_k)^2 \text{ with } 0 \leq BS \leq 1 \quad (A1)$$

with y_k the probability of the forecasted event, and o_k the actual outcome of the event at instant k (equal to 1 if the event is observed, equal to 0 if is not observed). N is the number of forecasting instances (Brier, 1950). Here, we consider three probability categories compared with the climatology as a reference (below normal, normal and above normal).

But BS can also be decomposed into three terms (Murphy, 1973): reliability, resolution and uncertainty as: $BS = \text{Reliability} - \text{Resolution} + \text{Uncertainty}$

$$BS = \frac{1}{N} \sum_{k=1}^N n_k (y_k - \bar{o}_k)^2 - \frac{1}{N} \sum_{k=1}^N n_k (\bar{o} - \bar{o}_k)^2 + \bar{o} (1 - \bar{o}) \quad (\text{A2})$$

with N , the number of forecasts issued; \bar{o} , the observed climatological frequency for the event to occur; n_k , the number of forecasts with the same probability category; \bar{o}_k , the corresponding observed frequency of the event.

Appendix B

Resampling method and student test

Following the BS description in Appendix A, we calculated y_k for the RAF experiment and y_k for Hydro-SF. Then, o_k was computed for SIM reanalysis. These calculations were done for all grid cells and stations each year over the 1960–2005 period and for both the upper and lower tercile. Secondly, 100 random samples of 40 BS from the 46 years were taken in order to calculate the BSS as written below :

$$BSS = 1 - \frac{BS_{\text{Hydro-SF}}}{BS_{\text{RAF}}} \text{ with } -\infty \leq BSS \leq 1. \quad (\text{B1})$$

The size of the bootstrapped resampling was 40 years in order to ensure a sufficient diversity of samples. So, the 100 random sample BSS were used to calculate the average and the standard deviation before applying a Student test and a significance threshold of ± 1.6 with a degree of freedom of 90 % (as the bootstrapped BSS distribution is quite symmetrical, very close to the Gaussian assumption).

In order to compare time correlations between Hydro-SF and RAF, a random resampling was performed with the same methodology as for BSS. The averages of the hydrological (soil moisture and river flow) ensemble forecasts for each experiment and random resampling were computed over each grid cell and for each river station. These 100 random ensemble means of 40 out of 46 years allowed us to compute the average and the standard deviation in order to finally calculate the Student variable with a significance threshold of ± 1.6 with a degree of freedom of 90 %.

Acknowledgements. We acknowledge Météo-France and CNRM/GAME for supporting this study which is included in a PhD thesis.

Edited by: J. Seibert



The publication of this article is financed by CNRS-INSU.

References

- Alessandri, A., Borrelli, A., Navarra, A., Arribas, A., Déqué, M., Rogel, P., and Weisheimer A.: Evaluation of probabilistic quality and value of the ENSEMBLES multimodel seasonal forecasts: Comparison with DEMETER, *Mon. Weather Rev.*, 139, 581–607, doi:10.1175/2010MWR3417.1, 2011.
- Araghinejad, S., Burn, D. H., and Karamouz, M.: Long-lead probabilistic forecasting of streamflow using oceanatmospheric and hydrological predictors, *Water Resour. Res.*, 42, W03431, doi:10.1029/2004WR003853, 2006.
- Bader, J. C., Piedelièvre, J. P., and Lamagat, H. P.: Prévision saisonnière du volume de crue du Fleuve Sénégal: utilisation des résultats du modèle ARPEGE Climat, *Hydrolog. Sci. J.*, 51, 406–417, doi:10.1623/hysj.51.3.406, 2006.
- Bierkens, M. F. P. and Van Beek, L. P. H.: Seasonal predictability of European discharge: NAO and hydrological response time, *J. Hydrometeorol.*, 10, 953–968, doi:10.1175/2009JHM1034.1, 2009.
- Boé, J., Terray, L., Habets, F., and Martin, E.: A simple statistical-dynamical downscaling scheme based on weather types and conditional resampling, *J. Geophys. Res.*, 111, D23106, doi:10.1029/2005JD006889, 2006.
- Boé, J., Terray, L., Martin, E., and Habets, F.: Projected changes in components of the hydrological cycle in French river basins during the 21st century, *Water Resour. Res.*, 45, W08426, doi:10.1029/2008WR007437, 2009.
- Bohn, T. J., Sonessa, M. Y., and Lettenmaier, D. P.: Seasonal Hydrologic Forecasting: Do multi-model ensemble averages always yield improvements in forecast skill?, *J. Hydrometeorol.*, 11, 1358–1372, doi:10.1175/2010JHM1267.1, 2010.
- Boone, A. and Etchevers, P.: An intercomparison of three snow schemes of varying complexity coupled to the same land-surface and macroscale hydrologic models, Part I: Local-scale evaluation at an alpine site, *J. Hydrometeorol.*, 2, 374–394, doi:10.1175/1525-7541(2001)002<0374:AIOTSS>2.0.CO;2, 2001.
- Boone, A., Calvet, J. C., and Noilhan, J.: Inclusion of a third soil layer in a land surface scheme using the force-restore method, *J. Appl. Meteorol.*, 38, 1611–1630, doi:10.1175/1520-0450(1999)038<1611:IOATSL>2.0.CO;2, 1999.
- Brier, G. W.: Verification of forecasts expressed in terms of probability, *Mon. Weather Rev.*, 78, 1–3, doi:10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2, 1950.
- Céron, J. P., Tanguy, G., Franchistéguy, L., Martin, E., Regimbeau, F., and Vidal, J. P.: Hydrological seasonal forecast over France: feasibility and prospects, *Atmos. Sci. Lett.*, 18, 78–82, doi:10.1002/asl.256, 2010.

- Chiew, F. H. S., Zhou, S. L., and McMahon, T. A.: Use of seasonal streamflow forecasts in water resources management, *J. Hydrol.*, 270, 135–144, doi:10.1016/S0022-1694(02)00292-5, 2003.
- Cloke, H. L. and Pappenberger, F.: Ensemble flood forecasting: a review, *J. Hydrol.*, 375, 613–626, 2009.
- Douville, H.: Relative contribution of soil moisture and snow mass to seasonal climate predictability: a pilot study, *Clim. Dynam.*, 34, 797–818, doi:10.1007/s00382-008-0508-1, 2009.
- Draper, C. S., Mahfouf, J.-F., and Walker, J. P.: Root zone soil moisture from the assimilation of screen-level variables and remotely sensed soil moisture, *J. Geophys. Res.*, 116, D02127, doi:10.1029/2010JD013829, 2011.
- Durand, Y., Brun, E., Mérindol, L., Guyomarc’h, G., Lesaffre, B., and Martin, E.: A meteorological estimation of relevant parameters for snow models, *Ann. Glaciol.*, 18, 65–71, 1993.
- Etchevers, P., Golaz, C., and Habets, F.: Simulation of the water budget and the river flows of the Rhône basin from 1981 to 1994, *J. Hydrol.*, 244, 60–85, doi:10.1016/S0022-1694(01)00332-8, 2001.
- Gobena, A. K. and Gan, T. Y.: Incorporation of seasonal climate forecasts in the ensemble streamflow prediction system, *J. Hydrol.*, 385, 336–352, doi:10.1016/j.jhydrol.2010.03.002, 2010.
- Habets, F.: Modélisation du cycle continental de l’eau à l’échelle régionale: application aux bassins versants de l’Adour et du Rhône, PhD thesis, Université Paul Sabatier, Toulouse, France, 1998.
- Habets, F., Noilhan, J., Golaz, C., Goutorbe, J.-P., Lacarrère, P., Leblois, E., Ledoux, E., Martin, E., Ottlé, C., and Vidal-Madjar, D.: The ISBA surface scheme in a macroscale hydrological model, applied to the HAPEXMOBILHY area: Part 1. Model and database, *J. Hydrol.*, 217, 75–96, doi:10.1016/S0022-1694(99)00019-0, 1999a.
- Habets, F., Etchevers, P., Golaz, C., Leblois, E., Ledoux, E., Martin, E., Noilhan, J., and Ottlé, C.: Simulation of the water budget and the river flows of the Rhône basin, *J. Geophys. Res.*, 104, 31145–31172, doi:10.1029/1999JD901008, 1999b.
- Habets, F., Boone, A., Champeau, J. L., Etchevers, P., Leblois, E., Ledoux, E., Lemoigne, P., Martin, E., Morel, S., Noilhan, J., Quintana Segui, P., Rousset-Regimbeau, F., and Viennot, P.: The SAFRAN-ISBA-MODCOU hydrometeorological model applied over France, *J. Geophys. Res.*, 113, D06113, doi:10.1029/2007JD008548, 2008.
- Habets, F., Gascoïn, S., Korkmaz, S., Thiéry, D., Zribi, M., Amraoui, N., Carli, M., Ducharne, A., Leblois, E., Ledoux, E., Martin, E., Noilhan, J., Ottlé, C., and Viennot, P.: Multi-model comparison of a major flood in the groundwater-fed basin of the Somme River (France), *Hydrol. Earth Syst. Sci.*, 14, 99–117, doi:10.5194/hess-14-99-2010, 2010.
- Hesterberg, T. C., Moore, D. S., Monaghan, S., Clipson, A., and Epstein, R.: *Bootstrap Method.*, W. H. Freeman & Company, 2005.
- Kirtman, B. and Pirani, A.: WCRP position paper on seasonal prediction. Report from the First WCRP Seasonal Prediction Workshop, 4–7 June 2007, Barcelona, Spain, WCRP Informal Report No. 3/2008, ICPO Publication No. 127. 25 pp., 2008.
- Kirtman, B. and Pirani, A.: The state of the art of seasonal prediction. Outcomes and recommendations from the First World Climate Research Program Workshop on Seasonal Prediction, *B. Am. Meteorol. Soc.*, 90, 455–458, doi:10.1175/2008BAMS2707.1, 2009.
- Koster, R. D. and Suarez, M. J.: Soil moisture memory in climate models, *J. Hydrometeorol.*, 2, 558–570, doi:10.1175/1525-7541(2001)002<0558:SMMICM>2.0.CO;2, 2001.
- Lafaysse, M., Hingray, B., Etchevers, P., Martin, E., and Obléd, C.: Influence of spatial discretization, underground water storage and glacier melt on a physically-based hydrological model of the Upper Durance River basin, *J. Hydrol.*, 403, 116–129, doi:10.1016/j.jhydrol.2011.03.046, 2011.
- Ledoux, E., Girard, G., de Marsilly, G., and Deschenes, J.: Spatially distributed modeling: conceptual approach, coupling surface water and ground-water, in: *Unsaturated Flow Hydrologic Modeling: Theory and Practice*, Kluwer Academic, Dordrecht, The Netherlands, 435–454, 1989.
- Li, H., Luo, L., and Wood, E. F.: Seasonal hydrologic predictions of low-flow conditions over eastern USA during the 2007 drought, *Atmos. Sci. Lett.*, 9, 61–66, doi:10.1002/asl.182, 2008.
- Li, H., Luo, L., Wood, E. F., and Schaake, J.: The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting, *J. Geophys. Res.*, 114, D04114, doi:10.1029/2008JD010969, 2009.
- Liang, X., Lettenmaier, D. P., Wood, E. F., and Burges, S. J.: A simple hydrologically-based model of land surface water and energy fluxes for GSMs, *J. Geophys. Res.*, 99, 14415–14428, 1994.
- Luo, L. and Wood, E. F.: Use of Bayesian merging techniques in a multimodel seasonal hydrologic ensemble prediction system for the Eastern United States, *J. Hydrometeorol.*, 9, 866–884, doi:10.1175/2008JHM980.1, 2008.
- Mason, S.: On Using “Climatology” as a Reference Strategy in the Brier and Ranked Probability Skill Scores, *Mon. Weather Rev.*, 132, 1891–1895, doi:10.1175/1520-0493(2004)132<1891:OUCAAR>2.0.CO;2, 2004.
- Materia, S., Dirmeyer, P. A., Guo, Z., Alessandri, A., and Navarra, A.: The sensitivity of simulated river discharge to land surface representation and meteorological forcings, *J. Hydrometeorol.*, 11, 334–351, doi:10.1175/2009JHM1162.1, 2010.
- Maurer, E. P. and Lettenmaier, D. P.: Potential effects of long-lead hydrologic predictability on Missouri River mainstem reservoirs, *J. Climate*, 17, 174–186, doi:10.1175/1520-0442(2004)017<0174:PEOLHP>2.0.CO;2, 2004.
- Murphy, A. H.: A new vector partition of the probability score, *J. Appl. Meteorol.*, 12, 595–600, doi:10.1175/1520-0450(1973)012<0595:ANVPOT>2.0.CO;2, 1973.
- Murphy, A. H.: A new decomposition of the Brier Score: Formulation and interpretation, *Mon. Weather Rev.*, 114, 2671–2673, doi:10.1175/1520-0493(1986)114<2671:ANDOTB>2.0.CO;2, 1986.
- Noilhan, J. and Planton, S.: A simple parameterization of land surface processes for meteorological models, *Mon. Weather Rev.*, 117, 536–549, doi:10.1175/1520-0493(1989)117<0536:ASPOLS>2.0.CO;2, 1989.
- Pagé, C., Terray, L., and Boé, J.: Dsclim: A software package to downscale climate scenarios at regional scale using a weather-typing based statistical methodology, Technical Report TR/CMGC/09/21, SUC au CERFACS, URA CERFACS/CNRS No1875, Toulouse, France, 2009.
- Palmer, T. N., Alessandri, A., Andersen, U., Cantelaube, P., Davey, M., Délécluse, P., Déqué, M., Díez, E., Doblus-Reyes, F. J., Feddersen, H., Graham, R., Gualdi, S., Guérémy, J. F., Hagedorn, R.,

- Hoshen, M., Keenlyside, N., Latif, M., Lazar, A., Maisonnave, E., Marletto, V., Morse, A. P., Orfila, B., Rogel, P., Terres, J. M., and Thomson, M. C.: Development of a European multi-model ensemble system for seasonal to interannual prediction (DEMETER), *B. Am. Meteorol. Soc.*, 85, 853–872, doi:10.1175/BAMS-85-6-853, 2004.
- Quintana-Seguí, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., Canellas, C., Franchistéguy, L., and Morel, S.: Analysis of near-surface atmospheric variables: Validation of the SAFRAN analysis over France, *J. Appl. Meteorol. Clim.*, 47, 92–107, doi:10.1175/2007JAMC1636.1, 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, 100–107, doi:10.1002/asl.259, 2010.
- Rousset, F., Habets, F., Gomez, E., Le Moigne, P., Morel, P., Noilhan, J., and Ledoux, E.: Hydrometeorological modeling of the Seine basin using the SAFRAN-ISBA-MODCOU system, *J. Geophys. Res.*, 109, D14105, doi:10.1029/2003JD004403, 2004.
- Rousset-Regimbeau, F., Habets, F., Martin, F., and Noilhan, J.: Ensemble streamflow forecasts over France, *ECMWF Newsletter*, No. 111, ECMWF, Reading, UK, 21–27, 2007.
- Thierion, C., Longuevergne, L., Habets, F., Ledoux, E., Ackerer, P., Majdalani, S., Leblois, E., Lecluse, S., Martin, E., Queguiner, S., and Viennot, P.: Assessing the water balance of the Upper Rhine Graben hydrosystem, *J. Hydrol.*, doi:10.1016/j.jhydrol.2011.12.028, 2011.
- Thirel, G., Regimbeau, F., Martin, E., Noilhan, J., and Habets, F.: Short- and medium-range hydrological ensemble forecasts over France, *Atmos. Sci. Lett.*, 11, 72–77, doi:10.1002/asl.254, 2010.
- Vidal, J. P., Martin, E., Franchistéguy, L., Baillon, M., and Soubeyrou, J. M.: A 50-year high-resolution atmospheric reanalysis over France with the Safran system, *Int. J. Climatol.*, 30, 1627–1644, doi:10.1002/joc.2003, 2010a.
- Vidal, J.-P., Martin, E., Franchistéguy, L., Habets, F., Soubeyrou, J.-M., Blanchard, M., and Baillon, M.: Multilevel and multiscale drought reanalysis over France with the Safran-Isba-Modcou hydrometeorological suite, *Hydrol. Earth Syst. Sci.*, 14, 459–478, doi:10.5194/hess-14-459-2010, 2010b.
- Voirin-Morel, S.: Modélisation distribuée des flux d'eau et d'énergie et des débits à l'échelle régionale du bassin Adour-Garonne, PhD thesis, Université Paul-Sabatier, Toulouse, France, 2003.
- Weisheimer, A., Doblas-Reyes, F. J., Palmer, T. N., Alessandri, A., Arribas, A., Déqué, M., Keenlyside, N., MacVean, M., Navarra, A., and Rogel, P.: ENSEMBLES: A new multi-model ensemble for seasonal-to-annual predictions – Skill and progress beyond DEMETER in forecasting tropical Pacific SSTs, *Geophys. Res. Lett.*, 36, L21711, doi:10.1029/2009GL040896, 2009.
- Wood, A. W. and Lettenmaier, D. P.: A test bed for new seasonal hydrologic forecasting approaches in the western United States, *B. Am. Meteorol. Soc.*, 87, 1699–1712, doi:10.1175/BAMS-87-12-1699, 2006.
- Wood, A. W. and Lettenmaier, D. P.: An ensemble approach for attribution of hydrologic prediction uncertainty, *Geophys. Res. Lett.*, 35, L14401, doi:10.1029/2008GL034648, 2008.