

Causality analysis of groundwater dynamics based on a Vector Autoregressive model in the semi-arid basin of Gundal (South India)

S. Mangiarotti, M. Sekhar, L. Berthon, Y. Javeed, P. Mazzega

► **To cite this version:**

S. Mangiarotti, M. Sekhar, L. Berthon, Y. Javeed, P. Mazzega. Causality analysis of groundwater dynamics based on a Vector Autoregressive model in the semi-arid basin of Gundal (South India). *Journal of Applied Geophysics*, Elsevier, 2012, 83, pp.1-10. 10.1016/j.jappgeo.2012.04.003 . ird-01062682

HAL Id: ird-01062682

<https://hal.ird.fr/ird-01062682>

Submitted on 10 Sep 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Causality Analysis of Groundwater dynamics based on a Vector Autoregressive model in the semi-arid basin of Gundal (South India)

Mangiarotti^{a,1} S., M. Sekhar^{b,c}, Berthon^{a,d} L., Javeed^e Y., Mazzega^d P.

^aCentre d'Etudes Spatiales de la Biosphère, CNRS-IRD-CNES-Univ. Toulouse III, Observatoire Midi-Pyrénées, 18 av. É. Belin, 31401 Toulouse, France, sylvain.mangiarotti@cesbio.cnes.fr

^bIndian Institute of Science, 560 012 Bangalore, India, muddu@civil.iisc.ernet.in

^cIndo-French Cell on Water Sciences, 560 012 Bangalore, India

^dGéosciences Environnement Toulouse, CNRS-IRD- Université Toulouse III, Observatoire Midi-Pyrénées, 14 av. É. Belin, 31400 Toulouse, France

^eNational Institute of Engineering, 570008, Mysore, India.

Abstract:

Causal relationships existing between observed levels of groundwater in a semi-arid sub-basin of the Kabini River basin (Karnataka state, India) are investigated in this study. A Vector Auto Regressive model is used for this purpose. Its structure is built on an upstream/downstream interaction network based on observed hydro-physical properties. Exogenous climatic forcing is used as an input based on cumulated rainfall departure. Optimal models are obtained thanks to a trial approach and are used as a proxy of the dynamics to derive causal networks. It appears to be an interesting tool for analysing the causal relationships existing inside the basin. The causal network reveals 3 main regions: the Northeastern part of the Gundal basin is closely coupled to the outlet dynamics. The Northwestern part is mainly controlled by the climatic forcing and only marginally linked to the outlet dynamic. Finally, the upper part of the basin plays as a forcing rather than a coupling with the lower part of the basin allowing for a separate analysis of this local behaviour. The analysis also reveals differential time scales at work inside the basin when comparing upstream oriented with downstream oriented causalities. In the upper part of the basin, time delays are close to 2 months in the upward direction and lower than 1 month in the downward direction. These time scales are likely to be good indicators of the hydraulic response time of the basin which is a parameter usually difficult to estimate practically. This suggests that, at the sub-basin scale, intra-annual time scales would be more relevant scales for analysing or modelling tropical basin dynamics in hard rock (granitic and gneissic) aquifers ubiquitous in south India.

Keywords: groundwater, time space analysis, causality, VAR model, semi-arid region

1. Introduction

Large amounts of water are stored in the ground: fresh groundwater represents among one third of total fresh water (Gleick 1996). However, it is one of the most poorly known reservoirs on the Earth. It is firstly because groundwater access is difficult: the precise monitoring of the water table depth requires digging many wells. The borders of each reservoir (when these borders exist) are generally ignored or poorly known, their interactions either. Moreover, vertical and horizontal couplings between the various reservoirs of the water cycle make the integrated dynamics of whole system difficult to monitor. Indeed, integrated dynamics include coupling between atmosphere and surface, soil and vegetation, surface and underground, etc.

¹ Corresponding author: Sylvain Mangiarotti, 18 avenue Edouard Belin, 31401 Toulouse, France. Tel: 00-33-(0)5-61-55-85-24, fax: 00-33-(0)5-55-85-00, sylvain.mangiarotti@cesbio.cnes.fr

The dynamics of the reservoirs can be influenced by long time scales (during temporary droughts or associated with climatic trends) as well as by short time scales (due to floods or extreme events) and at both large and short spatial scales associated with the size and the shape of the water catchments or due to the erratic distribution of rainfall. At large scale, ground water reservoir content will depend on the interannual and seasonal precipitation and evaporation regimes, on the vegetation cover and possibly on the water demand. At smaller scale, geomorphological controls (surface topography, vertical heterogeneity, type of soil and in the case of rock substratum the lineaments, dykes and their characteristics such as spatial distribution, numbers, length and orientation) and interactions between underground reservoirs may also play an important role.

Many studies have been dedicated to the relation between groundwater level and climate control. Annual precipitation is generally significantly correlated to groundwater level (Chen et al. 2004). In the Illinois area, Eltahir & Yeh (1999) revealed that compared to a flood and at the regional-scale, a drought can cause a more significantly persistent signature. Groundwater level provides a kind of averaged measure of the overall impact of both natural and anthropogenic changes to groundwater resources. The groundwater storage is generally clearly connected to the agriculture water use and storage changes in groundwater observed suggest a need for groundwater management in the semi-arid basin (Gleick 2000).

India has experienced repeated droughts during the last two decades. An increasing anthropogenic pressure on groundwater has also taken place through increasing pumping for agricultural needs. Recently, Rodell et al. (2009) have shown the unsustainable development of groundwater in the Northwest region of India using the data from the GRACE satellite. Similarly, in another example in the state of Orissa (India), the recharge was not sufficient enough to balance the groundwater discharge due to the anthropogenic and natural processes (Panda et al. 2007). South India is a strongly anthropogenic region in which demand for water is highly increasing. The limited resources in water play a primordial role in land-use, industry-use and domestic-use. Javeed et al. (2009) reported the extensive use of groundwater in the Kabini river basin, a part of the upper Cauvery river basin, in south India. The Gundal basin considered in this study has also experienced (but less drastically) the droughts reported for the state of Orissa. The Gundal basin is a 1270 km² basin located in the Kabini basin (Southwest of Karnataka). The drainage pattern of the Gundal basin is made of a main stream (the Gundal River) joined by many tributaries. The geology of the area is mostly characterized by granitic gneiss and is traversed by number of East-West dykes ranging from 0.5 to 1.5 km length and 5 to 15 m width (see Sekhar et al. 2006; 2004 for maps and details). Regionally, the climate is dominated by a regime of monsoon that generates a strong gradient of precipitation along the Western / Southwestern axis of the area under study. Along this gradient, the precipitations vary from a humid zone of the rain to a semi-arid zone. Both the humid and the semi-arid areas are characterized by a mono-modal distribution of the rainfall whereas the transition zone shows a bi-modal distribution at the limit of the Gundal basin.

The Gundal basin is mainly located in a semi-arid portion of the Cauvery river basin where big changes have occurred. The agriculture has moved from an agriculture low water consuming (sorgho and millet), to an intensive agriculture highly dependant in irrigation and water consuming, with several crops a year. Both the surface water and the groundwater reservoirs are used to meet the needs of irrigation. A dramatic decline of the groundwater has occurred in some places of the Gundal basin that may be due to the impact of land use. This decline may have achieved at the whole system scale due to the importance of the lateral fluxes (Sekhar et al. 2006).

Physical approach (numerical modelling) is a very useful tool for analyzing the groundwater dynamics. In the Gundal basin, several efforts have already been done in this direction (Sekhar et al. 2004; 2006). However, the ignorance of the boundary conditions and the interactions of various processes does not allow for resolving completely the problems. Especially in hard rock aquifers such as in south India, while the spatial aquifer extent is not clearly identified, it is not clear how the various spatial aquifer units interact along a topographic relief of a sub-basin due to the non uniform changes in the piezometric levels and due to anthropogenic effects taking place. Indeed, a careful accounting of all

the system parameters and their geographical distribution, even though not exhaustively available, is often required for applying numerical models dealing with partial differential equations and for analysing the spatial fluxes (Sanford 2002). Therefore, complementary approaches may be useful in association with the numerical modelling.

Statistical approaches can contribute to complement physical approaches. In a previous study based on an Empirical orthogonal functions (EOF) decomposition (Sekhar et al. 2011), it is shown that a multivariate analysis including cumulative rainfall departure and groundwater levels can help in discriminating the control of groundwater recharge by rainfall and the control groundwater discharge by topography, land use practices and in some extent of climatic changes. In the Singular Spectrum Analysis performed in Javeed et al. (2009) based on rainfall, stream flows and groundwater levels observed in the whole Kabini basin, it is shown that the presence of strong decadal climatic variations associated with changes in the land use provide distinct groundwater signatures in the system.

Most of the traditional statistical tools aim to find correlations between measured variables with a given level of confidence. However, as such, correlation does not imply any structural link between the variables and does not suggest any causality. Moreover, even assuming an existing causal relationship between two variables, a high correlation does not give any information about the causal direction between these variables. The concept of causality implies a direction: the idea is that a cause can not come after the effect. However, a causal relation is easier to define theoretically than to reveal practically especially when reciprocal causality may occur. A practical concept of causality has been introduced by Granger in the 1960s aiming at clarifying the direction of causality between two related variables and detecting whether or not a feedback is occurring (Granger 1969). The technique consists in determining whether a time series can be useful in forecasting another through a spectral approach. The Vector Auto Regressive (VAR) models (Lütkepohl, 2008; Lütkepohl & Krätzig, 2004; Krätzig, 2007) allow for another practical method to approach the concept of causality which has been largely unexplored in the domain of hydrology and geophysics (see also Salvucci et al. 2002).

In this paper, a structural VAR model is defined and used in order to investigate the causal relationship of groundwater levels observed in the Gundal basin. The period 1979-2000 corresponding to a relatively moderate (although monotonically increasing) level of pumping for most of the stations is considered. The following questions are addressed: Is it possible to confirm, infirm, reveal or point out some features concerning the interactions taking place inside the Gundal basin? And more especially, is it possible to identify some reciprocal actions between groundwater variables and to establish some link with the geomorphological properties of the basin?

2. Rainfall and Groundwater Data

2.1 Groundwater data

Information about groundwater comes from a network of wells distributed in the Gundal basin where the level of water is measured with a monthly time step. Ten stations are considered in this study (from G1 to G10), their geographical distribution (Figure 1) is representative of the main topographic situations within the basin. Stations cover the main parts of the basin (centre, upper and lower basin including the junction of the main upper tributaries and the closest vicinity to the Kabini River. Land use changes associated with increased groundwater pumping have occurred over the period of study (Javeed 2010). Several stations are located at places where an increased anthropogenic effect of pumping is qualitatively known (see Table 1). The period considered in this study starts on January 1979 and finishes on May 2000, which corresponds to 257 individual monthly values for each series.

2.2 Rainfall data and CRD

Information about rainfall comes from 8 rain gauge stations which distributions allows for a good description of the Southwest-Northeast climatic gradient changing from a humid to a semi-arid climate

(see Figure 1 and Table 2). The initial daily series are summed up into monthly time series in order to obtain a dataset of coherent sampling with the dataset of groundwater. The more far away station from the zone of study is R1 (~50 km), located in the humid zone. One station belongs to the transition zone (R2). Other stations are located in the semi-arid zone among which four ones (R3 to R6) are in the Gundal basin itself and two other ones (R7 and R8) are in the Northeastern part of the basin. The Northwest-Southeast direction which is orthogonal to the climatic gradient is also represented. At larger scale, a complementary spatially Averaged Rainfall (AR) information (1 degree x 1 degree monthly aggregated time series) from the Indian Meteorological Department is used to represent time variability of precipitation in the humid (AR-H) zones which is more poorly represented. This series was obtained by aggregating daily time series available in the 1°x1° pixel centred on 76.5°E, 12°N into monthly time series.

The concept of Cumulative Rainfall Departure (CRD) is often used to analyse the links between rainfall and groundwater variability. In the present study, CRD is used as a forcing. It is calculated for each of the rainfall time series, on the same time period 1979-2000. For each time series $\{z_i\}_{i=1..N}$ of N samples, the CRD is computed as follows: the average \bar{z} is estimated from the full time period (1979-2000); this average is removed from the time series leading to a centred time series; each term of the CRD time series, hereafter noted R_j , is obtained by summing the centred series from the first to the j^{th} value. This can be expressed mathematically as follows:

$$R_j = \sum_{i=1}^j (z_i - \bar{z}) \quad (1)$$

As it was pointed by Weber & Stewart (2004), CRD should be used with some caution. Indeed, CRD is very sensitive to the uncertainty associated with the value subtracted to the original time series in order to get the centred time series (here the average \bar{z}). Indeed, a misestimate of this value will generate over- or under- accumulations giving rise to spurious linear trends. For this reason, applications dedicated to slow variability should be definitively avoided. Contrarily, short term predictions may be only marginally sensitive to this effect. The causal approach used in the present study is based on one step ahead predictions and thus clearly falls into the second case. Moreover, as it is designed, the modelling approach (see section 3) mathematically allows the elimination of spurious accumulations. Finally, it is pointed out that all the variables considered in the analysis are considered on the same period of time (1979-1980) in order to keep the analysis coherent. CRD can thus reasonably be considered as adequate in the present context.

3. Theoretical background and experiments

The main objective of this paper is to analyse the causal network of a real set of groundwater time series. The approach is based on the reconstruction of a Vector Auto Regressive model. The principle consists in building an optimal VAR model from a structural-VAR originally built based on available a priori knowledge and to deduce the causal network directly from the causal relationships at work in the optimal model.

3.1 VAR model and causality

The general form of the vector autoregressive (VAR) model used in this study is given by:

$$\mathbf{y}(t) = \mathbf{A}_1 \mathbf{y}(t-1) + \dots + \mathbf{A}_p \mathbf{y}(t-p) + \mathbf{B} \mathbf{x}(t) + \mathbf{C} + \mathbf{u}(t) \quad (2)$$

where $\mathbf{y}(t)$ is a vector of K modelled variables at time t , $\mathbf{x}(t)$ is a vector of M forcing variables, and $\mathbf{u}(t)$ is a K -dimensional zero mean white noise process with positive definite covariance matrix $E(\mathbf{u} \cdot \mathbf{u}') = \Sigma_u$. Because of the forcing (i.e. exogenous) component, this model can also be called a VARX model. The number of lags p used in the model defines the order of the VAR model and also corresponds to the maximum time delay. Matrix \mathbf{A}_1 to \mathbf{A}_p are $K \times K$ - dimensional, matrix \mathbf{B} is $K \times M$ - dimensional and \mathbf{C} is a K -dimensional vector. Due to the low sampling of the groundwater dataset (monthly sampling), the forcing of the model is assumed to be instantaneous. In the present application, $\mathbf{y}(t)$ will contain the $K=10$ variables of groundwater levels $G_1(t)$ to $G_{10}(t)$ corresponding to stations G1 to G10, respectively; $\mathbf{x}(t)$ will contain the $M=9$ CRD forcings $R_1(t)$ to $R_{H-AR}(t)$ corresponding to the meteorological stations R1 to R8 plus the aggregated station AR-H, respectively.

In practice, the effective VAR model can be derived in two stages. The first stage consists in defining a model structure in order to take advantage of the a priori knowledge available about the system. The description of the structural shape used in the present study will be explained in section 3.2. The full description of the structural model also requires the definition of the lag number p . Practically, the maximum lag p was chosen such as $p=2$ (in month) since, in practice, it was noted that the precision did not increase significantly for orders larger than 2.

The second stage consists in exploring optimal sub-models of the structural model by applying successive restrictions. In optimization approaches, it is invariably found that the larger the size of the model is (i.e. the number of non zero parameters), the better the model fitness is. However, good fitness cannot ensure good prediction skills. Contrarily, over-parameterized models generally lead to non robust predictions and to spurious effects, potentially. Models of intermediate order should thus clearly be preferred. A criterion accounting additionally for model fitness and model size can be used for this purpose. In the present context, the Akaike Information Criterion (AIC, Akaike 1974) is used, which can be written as:

$$AIC(p) = \log \det(\tilde{\Sigma}_u(p)) + \frac{2}{T} pK^2 \quad (3)$$

where T is the time length, K is the number of variables and p is the order of the VAR model. The first term accounts for the model fitness, the second term for the model's size.

A strategy for visiting the structural model's sub-space was thus chosen in order to apply optimal restrictions to the structural model, based on the AIC criterion. A sequential elimination of regressors based on a top-down strategy was applied (Lütkepohl, 2008). Details about this selecting process are given in Appendix A. Broadly speaking, the approach consists in estimating the parameters of \mathbf{A}_1 to \mathbf{A}_p , \mathbf{B} and \mathbf{C} through a least square approach associated with a trial method based on an AIC criterion. For one current variable, two cases are considered: one case including the current variable in the parameters estimation, one case setting the parameterization of this variable to zero. Depending on the situation of best AIC, the variable will be retained or rejected from the model. In such a selection process, the order in which variables will be considered may lead to different results. The selecting process should thus be considered precisely in order to avoid misinterpretations. The top-down strategy used in the present work consists in considering the variables of equation 2 one by one starting from the last one down to the first one. In other words, this strategy will privilege terms located upper in the equation than lower. Therefore, variables of short time delays will be preferred rather than variables of long time delays. In order to obtain the more parsimonious model satisfying the AIC criterion, the top-down strategy is repeated until no variables can be removed anymore.

The final restricted model that is obtained is considered as optimal against AIC criterion and is used as a proxy of the dynamics to analyse causal relationships. The causal network is directly derived from

the structure of the final restricted model. Considering each variable of $\mathbf{y}(t)$ one by one, it will be said that variable $G_i(t)$ has a causal effect on variable $G_j(t)$ if G_j helps predict G_i . This will be noted graphically as $G_j \rightarrow G_i$. Details about how causal links are deduced from the VAR model equations are given in Appendix C.

3.2 The model structure

By their genericity, VAR models allow for a large panel of situations. However, this genericity may also be misleading. When poorly constrained, VAR models may lead to mathematically equivalent but practically untrue results. This issue can be bypassed bringing external information to the system based on a priori knowledge or external controls. Three types of constraints are identified, here.

In order to constrain the model satisfyingly, some modelling scenarios have been investigated concerning (1) the series auto-prediction, (2) the geomorphological control and (3) the climatic forcing. A scenario allowing for series auto-prediction has first been tested. It showed relatively good forecasting skills, but led to avoid the circulation of information between the variables, and thus to prevent the visibility of the interactions involved. To avoid this issue, auto-prediction was prohibited from the structural models by forcing the diagonal of the \mathbf{A} matrices to zero (see equation 4 in Appendix B).

Another constraint concerns the hydrological topography of the underground table. As shown in Figure 2 the gradient of the underground table is closely linked to the surface topography. This indicates that lateral interactions between sub-basins of the Gundal basin must be quite low, and that the morphological control is another important constraint of the groundwater fluxes. Interactions between Gundal's sub-basins should thus preferably be forbidden in the model structure. Therefore, only the causal relationships following the topographic slope are allowed in the structural model (see Figure 3a).

A scenario without forcing has been tested in order to evaluate the importance of the climatic forcing that demonstrated the necessity to include this information into the model. Indeed, when climatic forcing is not provided to the model, the approach tends to retrieve this information by developing spurious causal links between the stations. As a consequence, information about rainfall was systematically included in the structural models under study. Corresponding structure of this forcing is given by matrix \mathbf{B} in equation 5 (see Appendix B). In order to avoid spurious effect, it was preferred to limit the number of forcing at one or two close rainfall stations (see Figure 3b). However, when climatic influences were ambiguous or not clearly identifiable, additional forcing were then allowed. It is the case for station G10 for which 4 rainfall contributions are allowed in the structural model. Station G5 being located along the Kabini River, it is mostly located in the humid zone. For this reason, closer rainfall stations were associated with aggregated station AR-H that better represents the rainfall of the humid area. Dynamically, the forcing is defined as an instantaneous forcing variable (with no retroaction). The following forcing are applied: R1 on G2; R2 on G1, G2 & G10; R3 on G4, G7 & G5; R4 on G1, G3, G8 & G10; R5 on G10; R6 on G9 & G10; R7 on G5 & G6; R8 on G6 & G9; AR-H on G5.

The structure obtruded from these three types of constraint leads to the structural model shown graphically in Figure 3 in which lateral sub-basin interactions are not allowed (its formal definition is given by equations 2, 4 and 5 (see section 3.1 and Appendix B). The direction of the causal links is not imposed by the structural model (both causal directions are *a priori* allowed) since the object of the present work is to investigate the causal directions. The effectiveness of the causal links will be deduced through the restriction process establishing the optimal models (see Appendix A).

4. Results and discussions

Starting from the structural model presented in Figure 3, two restricted models were obtained. Their formulation is provided in Appendix C. These restricted models are 'optimal' against the AIC fitting criterions and are used as a proxy of the underlying dynamics for the causality analysis. The causal

network is directly interpreted from the interaction network exhibited between the variables of the model. Two cases are considered, corresponding to 1-lag and 2-lags models. Causality diagrams corresponding to these two models are shown in Figure 4. Models fitness computed for these two optimal models are given station by station in Table 3. Two criteria are considered, the percentage of explained variance and the Nash-Sutcliffe Efficiency (NSE). See for example Legates & McCabe (1999) for a critical analysis of various validation criteria used in hydrologic model. The percentage of explained variance for the two models is moderate (45% and 47% respectively). Without forcing, this level would fall to 25%, illustrating the preponderant role of precipitation forcing. The NSE is computed for each variable from the one step ahead prediction. Values of the NSE show a large range including very low values at station G4, G5 and G7 (lower than 0.25). Skewness and kurtosis which measure the degree of asymmetry and peakedness of a distribution, respectively, are also estimated for the residual signals and reported in Table 3. Qualitatively, these results do not suggest that data are so far from Gaussian. Most of the skewness are in the range [-1; 0.4]. The larger magnitudes are negative indicating that the tail on the left side of the distribution is longer. These include the three stations for which high pumping started to increase in the early 1990s. Most of the kurtosis are in the range [2.3; 5] (for a normal distribution, kurtosis equals 3). In the following discussion, interpretation is started from the 1-lag model (Figure 4a), whereas the 2-lags model (Figure 4b) is used to confirm, to clarify, to detail or to tighten the interpretations.

Most of the forcing links of the structural model represented by the endogenous variables are conserved in the 1-lag model (if not significant, they would have been rejected during the restriction process). Two exceptions are noted: R4, R5 & R6 which were a priori acting on well station G10 (see the structural model, Figure 3) have disappeared. The disappearance of these three causal links may be interpreted by the dominant influence of the humid climatic zone in this area (in the vicinity of the transition zone) which contribution, represented by rain station R2, is conserved. However, this interpretation is not confirmed by the 2-lags model in which all the forcing are conserved. This underlines the importance of the 2 months delay process in the upper part of basin, and the necessity to include it in the model.

Groundwater level at station G5 is influenced by cumulated contributions. Being at a very close vicinity to the Kabini River, station G5 is directly and largely influenced by the dynamics of the Kabini River's height. Note that the draining area of the Kabini basin being mostly located in the humid zone, variations in Kabini's flow are thus connected to the rainfall of the humid zone that can be represented by rain station AR-H (see Figure 1). Information provided by AR-H is representative of large scale behaviour (1 degree x 1 degree resolution) and mixes a large spectrum of time-scale influences. Moreover, the geographical cover of this aggregated gauge does not strictly correspond to the Kabini basin. Despite these imperfections, influence of the humid area (as it can be represented by AR-H station) is effectively seen in G5's signal as confirmed by the causal analysis (forcing link between AR-H and G5 is conserved). Moreover, being at the outlet of the Gundal basin, G5's level is also influenced by the contribution of all the sub-basins of the Gundal. These cumulated contributions represent an intricate signal of complex time-space integrated behaviours that cannot be caught by the fully linear model and that leads to a very poor fitness with the 1-lag model (close to 6%, NSE close to 0.06). The fitness is slightly improved with the 2-lags model (close to 17%, NSE close to 0.2), see Table 3 which suggests an important influence of process occurring at a 2-months time scale.

Despite the intricate integration of these various influences, causal relations between the groundwater level observed at the outlet of the basin (by G5) and the levels observed in the other parts of the Gundal basin are not prevented. From a direct (simplistic) point of view, a causal effect may be expected as simply following the stream direction: broadly speaking, the variations observed upper should propagate downstream and lead to similar downstream variations. This is not what is observed in practice. Indeed, the analysis of the causal diagram shows that causal links do not simply follow the flow orientation. The following interpretation is suggested: when the groundwater level is increasing at one of these sub-basins, exceeding water will flow downstream toward the outlet of the basin before being

drained out through the Kabini River. If the flow is too high to be efficiently drained through the river, water will accumulate and a saturated zone will develop, possibly propagate or at least generate a feedback in the reverse direction (upstream). It is the case for G4↔G5 for which reciprocal causality is explained by the proximity of G4 to the Kabini River. This coupled causality expresses the role of the border outlet areas as a damper: when the groundwater level is changing at the outlet (controlled by the river height dynamics or by other sub-basin contributions), the causality goes from G5 to G4; when the region corresponding to G4 receives recharge as well as lateral fluxes from the upper relief regions, G4's groundwater storage rises up and water accumulation will transfer toward the outlet leading to a G4 to G5 causality. Although explained variance at G4 is close to 20% (NSE=0.21), which is also low, the 2-lags model confirms the reciprocal causality and exhibits an optimal time scales of 2 months (NSE=0.23). Although much farther away in space, a very similar behaviour is observed for G7↔G5 with a similar fitness (Explained variance of 20% and NSE = 0.2). Only the upstream causality can be considered as robust since confirmed by the 2-lags model (see Figure 4b).

Contrarily, causality relations obtained between G8 and G5 is clearly monodirectional and follows the stream direction (i.e. G8→G5). Station G8 is located at the outlet of the upper part of the Gundal. The contribution of the upper basin to the groundwater flow is important in volume due to its location closer to the humid influence. By its location, station G8 is likely to receive a cumulated signature from the whole upper regions. This explains the clear downstream direction of causality whereas the farther distance from G5 can explain the nonreciprocal link, possibly associated with topographic thresholds that may prevent from reciprocal effect. This single downstream causality is emphasized by the 2-lags model for which both 1 month and 2 months delayed causalities are obtained.

No interactive links are found in groundwater dynamics at stations G6 and G9. At these two stations, for which mono modal-West climate is dominant, most of the signal is dynamically explained by the rainfall control (explained variance and NSE are high at both stations, see Table 3). Again, this strengthens the necessity to include the exogenous climatic information into the structural model. Moreover, their contribution to the outlet flow is small comparatively to other regions. The hypothesis that may explain this behaviour is the very active groundwater pumping (see Table 1), leading to the declining of the local groundwater storage. Due to this declining, the piezometric gradient along the drainage directions from these station areas towards the main stream valley has declined. As a result, the flow, which is relative to the piezometric gradient also could have declined reducing the effectiveness of the causal interactions. Water pumping starting in these regions in the 1990s may also emphasize this effect, by reducing the contribution and thus the detectability of the causal relations. However, in spite of these limitations of temporal changes, when considering the 2-lags model, a downstream causality between G9→G5 with a 2 months delay becomes clearly detectable.

The upper part of the basin is characterized by the systematic presence of upstream oriented causalities centred on G8. This feature can be explained by the small size of the watersheds (short distances between the stations) that leads to facilitate the retroactions effects and thus their detection. The intense pumping taking place in G8's zone since the beginning of the 1990s may also contribute to strengthen this reverse effect. Upstream orientated causalities are confirmed by the 2-lags model, time delays of larger magnitudes are sometimes preferably detected. Downstream causalities are more difficult to be detected, especially when using the 1-lag model; these links appear more clearly with the 2-lags model, though corresponding causality delays are likely to be expected with a time delay closer to 1 month. This observation is interpreted as follows: on the one hand, time delays corresponding to the downstream causality are associated with over-accumulation of water in upper regions that leads to a quick response downstream and thus to a short time delay between the cause and the effect (to be noted that strong gradients of the groundwater table are observed in this part of the Gundal basin that may lead to quick downstream oriented reaction, see Figure 2). Typically, this delay should be smaller than 1 month in the present case, making its detection tricky at a 1 month sampling. On the other hand, upstream causality is associated with depletions or over accumulations occurring downstream that will propagate upstream with a longer time delay. In the present case, this time delay is likely to be closer to

1 to 2 months and easily detectable from a monthly data set although (a priori) shorter in magnitude. The contrast observed between 1-lag and 2-lags models can thus be explained as follows: when the 1-lag model is used, upstream causalities (easy to detect with an appropriate 1 month sampling) are obtained with the 1 month delay whereas downstream causalities are mostly hidden; when the 2-lags model is used, upstream causality are adequately detected at a 2 months delay (because closer to the real time delay), and 1 month delay causalities of lower magnitude (upstream 'residual' causalities and downstream poorly informed causalities) become detectable.

Several time scales are generally involved in groundwater dynamics. One corresponds to the mass transport of water in the aquifer, its magnitude relates to the velocity of groundwater flow that relies itself on the medium porosity. Another one corresponds to the storage and transmissivity parameters of the aquifer that can be termed as the hydraulic response related to the diffusion of pressure wave characterised by the diffusion equations describing the groundwater system. The characteristic time scales of both these two processes can cover a huge range of values (several orders of magnitudes) (Alley et al. 2002) and it is generally difficult to estimate these values practically. The present study focuses on short time scales (1 to 2 months) and suggests that an intra-annual time scales are operating in the interconnections in the basin even though the system in general is forced by the yearly monsoon and hence it appears that for numerical analyses of such tropical hard rock (granitic and gneissic) aquifers the time scale should be properly selected for capturing the upstream-downstream controls. Further the national data bases currently in operation are collecting groundwater at the scale of a month and perhaps this may also need refinement based on the results of this study. The approach used in the present paper is empirical and therefore does not explicitly separate these two processes. However, the relatively short time scales obtained here compared to the basin and sub-basin lengths (5 to 40km), together with the upward causalities found between several stations of the Gundal, strengthens the idea of a time scale characterizing the latter process corresponding to diffusion rather than mass transport. The differential time scales revealed by the study and relative to the upward and downward directions are not likely to distinguish these two processes (diffusion from mass transport) but rather to represent the anisotropy of the diffusion process. Since the causal approach is based on one step ahead prediction, contributions of longer time scales may not be expected from the present results.

5. Summary and conclusions

Due to the low accessibility of the underground structures and due to their multi-scale heterogeneity, the dynamic of groundwater is poorly observable. The more direct way to monitor spatial variations of the groundwater table consists in measuring the water levels in a network of wells distributed inside an area of study. In the present work, the network of dynamical links existing between the various variables is investigated through a causality approach based on a Vector Auto Regressive model that does not require a strong a priori knowledge of the underground structure. Contrarily to most of the physical approach that will fix the causal processes at a given scale and analyse the link between input and output, the present study only formulates soft hypotheses about the variables and investigates the links existing between the variables. The objective is not to replace physically based approach, but rather to bring a complementary point of view on the groundwater dynamics.

In practice, a network of ten stations of groundwater level is considered in the present study. The causal relations existing between these ten variables are investigated (reconstructed), analysed and discussed. Three constraints are applied to the model in order to strengthen the robustness of the results. These are included into the structural model: (1) variables are not auto-predictable, (2) topography is considered as an important control of the causal paths (but not of the causal direction), and (3) rainfall is used as an exogenous control of the dynamics.

The analysis is applied to the Gundal basin which is characterized by a strong climatic gradient, with a complex mix of mono-modal and bi-modal regimes. Time series of groundwater level are considered

at ten stations of the Gundal basin during the period 1979-2000 during which pumping effects are still limited compared to more recent years. Two optimal models are obtained for the basin by applying model restrictions. The resulting restricted models are used as a proxy of the underlying dynamics of the basin to analyse the causal relations existing between the variables.

The analysis of the resulting models confirms or reveals the following features:

- Rainfall plays a predominant role for the overall model structure with, however, unequal contributions in space. Its contribution is especially important in the Northwestern part of the Gundal characterized by a mono modal-West climatic conditions.
- The regions located in the Northeastern part of the Gundal exhibit a strong coupling with the outlet dynamics.
- The upper part of the Gundal does not act as a coupling element of the basin but rather as a forcing element of the outlet dynamics. The upper part of the Gundal basin is thus mostly independent from the lower part of the basin (it can therefore be analysed separately).
- A 1-2 months delay appears to be an efficient time scale for detecting the causal links among a network of groundwater measurements in a small basin such as the Gundal. However, it is insufficiently precise to detect causal links of shorter time scales, especially when shorter distances are considered inside the basin. Indeed, for the upper part of the basin, downstream causality delays are likely to be lower than 1 month, making their detection tricky at a 1 month sampling. Contrarily, although supposedly weaker in magnitude, upstream causality delays are estimated to be in the range 1 to 2 months and easier to detect due to the adapted monthly sampling of the dataset. These characteristic times can be considered as reasonable indicators of the hydraulic response time of groundwater inside the water basin of Gundal and also support the more general idea that intra-annual time scales would be more relevant for a dynamical analysis, while modelling the tropical basins such as this in hard rock (granitic/gneissic) aquifers ubiquitous in south India.

Appendix A

For a given structural model associated with a lag order p , the subset model selection aims at reducing the size of the parameter space for a proper representation of the dynamics. The top-down strategy was used for this purpose (Lütkepohl, 2008). Since variables do not have feedback effect at a one step ahead prediction, the K equations of the structural VAR model can be considered separately. For each equation, variables are considered one by one starting from the last variable of the equation up to the first variable. Following this order, each current variable is considered as follows: (1) the equation is estimated by a least square which fitness is evaluated based on a chosen criterion (the AIC criterion is used here, see eq. 3); (2) the coefficient corresponding to the current variable is then set to zero leading to a restricted equation. This restricted equation is also estimated and its criterion evaluated; (3) The criterion values obtained with the unrestricted and restricted models are then compared, and the current variable is removed when leading to improve the criterion. In order to get the more parsimonious models, the top-down strategy associated with the AIC criterion is repeated until none variables can be removed anymore.

The causal network is then directly deduced from the matrices **A** and **B** of the final model, and variable $G_j(t)$ is said to cause $G_i(t)$ if $G_i(t) = \alpha \cdot G_j(t) + \dots + u_i(t)$ with $\alpha \neq 0$. This causal link is noted graphically $G_j \rightarrow G_i$, the direction of the arrow corresponding to the causal direction.

Appendix B

Like for the general model, the structural model can be defined by matrices **A**, **B** and vector **C** (see eq. 1). Three types of constraint are considered to reconstruct the structural model used in the present study. In practice, matrices **A** is set following the following rule: if the causality of variable j on variable i is forbidden, then the i^{th} line and j^{th} column is set to zero. By contrast, other values are kept unset which will be estimated in the optimization process (see Appendix A). The auto-prediction of the variables can be excluded by setting the diagonal of **A** matrices to zero. In the present context, this rule leads to the following definition of matrix **A**:

$$\mathbf{A}_1 = \begin{pmatrix} \mathbf{0} & 0 & 0 & 0 & 0 & 0 & 0 & ? & 0 & 0 \\ 0 & \mathbf{0} & ? & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & ? & \mathbf{0} & 0 & 0 & 0 & 0 & ? & 0 & 0 \\ 0 & 0 & 0 & \mathbf{0} & ? & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & ? & \mathbf{0} & ? & ? & ? & ? & 0 \\ 0 & 0 & 0 & 0 & ? & \mathbf{0} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & ? & 0 & \mathbf{0} & 0 & 0 & 0 \\ ? & 0 & ? & 0 & ? & 0 & 0 & \mathbf{0} & 0 & ? \\ 0 & 0 & 0 & 0 & ? & 0 & 0 & 0 & \mathbf{0} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & ? & 0 & \mathbf{0} \end{pmatrix}, \quad \begin{cases} \mathbf{A}_2 = 0 \text{ if } p = 1 \\ \mathbf{A}_2 = \mathbf{A}_1 \text{ if } p = 2 \end{cases} \quad (4)$$

where questionmarks denote the parameters kept unset.

Similarly, forcing corresponding to rain gauge R_i onto gauge G_i is forbidden by setting the i^{th} line and j^{th} column of matrix **B** to zero. Linear trends being intentionally allowed in order to remove potential drift, all the values of vector **C** are kept unset:

$$\mathbf{B} = \begin{pmatrix} 0 & ? & 0 & ? & 0 & 0 & 0 & 0 & 0 & 0 \\ ? & ? & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & ? & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & ? & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & ? & 0 & 0 & 0 & ? & 0 & ? & ? \\ 0 & 0 & 0 & 0 & 0 & 0 & ? & ? & 0 & 0 \\ 0 & 0 & ? & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & ? & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & ? & 0 & ? & 0 & 0 \\ 0 & ? & 0 & ? & ? & ? & 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{C} = \begin{pmatrix} ? \\ ? \\ ? \\ ? \\ ? \\ ? \\ ? \\ ? \\ ? \\ ? \end{pmatrix} \quad (5)$$

Appendix C

The two VAR models obtained by applying the top-down strategy (see Appendix A) on the structural model presented in Appendix B are defined hereafter. Their definition relies on equation 1 and requires the parameterization of matrices and vector **A**, **B** and **C**. For the 1-lag model, these matrices and vector read:

$$\mathbf{A}_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.54 & 0 & 0 & 0 & 0 & 0 & -0.07 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.55 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.14 & 0 & 0 & 0 & -0.05 & -0.03 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.47 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.99 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.13 & 0 & 0 \end{pmatrix}$$

$$\mathbf{B} = \begin{pmatrix} 0 & 3.60 & 0 & -3.15 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.29 & 2.08 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1.87 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3.11 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2.56 & 0 & 0 & 0 & -0.99 & 0 & 0.17 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 6.77 & 1.63 & 0 & 0 \\ 0 & 0 & 10.21 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 18.23 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 7.34 & 0 & 6.56 & 0 & 0 \\ 0 & 1.13 & 0 & 0 & -0.95 & 1.01 & 1.67 & 0 & 0 & 0 \end{pmatrix}$$

$$\text{and } \mathbf{C} = \begin{pmatrix} -3.93 \\ -4.06 \\ -0.09 \\ -4.42 \\ -3.89 \\ -12.48 \\ -11.88 \\ -11.94 \\ -8.75 \\ -3.94 \end{pmatrix} \quad (7)$$

Similarly, the analysis can also be applied to the 2-lags model. Taking the third variable $G_3(t)$ as example, variables having a causal effect on $G_3(t)$ are found by considering the third lines of matrices \mathbf{A}_1 , \mathbf{A}_2 and \mathbf{B} . In practice, contributions associated to 1 month and 2 months delay can be distinguished. At a 1 month delay, one endogenous causal contribution is found (second line of \mathbf{A}_1) corresponding to G2→G3. At a 2 months delay (matrix \mathbf{A}_2), two contributions are found G2→G3 and G8→G3. Finally, one instantaneous forcing is found, reading R4→G3.

References

- Akaike, H., 1974. A new look at the statistical model identification, *IEEE Trans. on Automatic Control*, AC-19, 716–723.
- Alley, W.M., Healy, R.W., LaBaugh, J.W., Reilly, T.E., 2002. Flow and storage in groundwater systems. *Science*, **296**, 1985–1990.
- Chen, Z., Grasby, S. E., Osadetz, K. G., 2004. Relation between climate variability and groundwater levels in the upper carbonate aquifer, southern Manitoba, Canada. *Journal of Hydrology* **290**, 43–62.
- Eltahir, E. A. B., Yeh, P. J. F., 1999. On the asymmetric response of aquifer water level to floods and droughts in Illinois. *Water Resour. Res.* **35** (4), 1199–1217.
- Gleick, P. H., 1996. Water resources. *In: Encyclopedia of Climate and Weather*, ed. by S. H. Schneider, Oxford University Press, New York, vol. 2, 817–823.
- Gleick, P., 2000. The changing water paradigm: looking at twenty-first century water resources development. *Water International* **26**(1), 127–138.
- Granger, C.W.J., 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37** (3), 424–438.
- Javeed, Y., Sekhar, M., Banyopadhyay, S., Mangiarotti, S., 2009. EOF and SSA analyses of hydrological time series to assess the climatic variability and land use effects: A case study in the Kabini river basin of South India. *IAHS Publication*. 329, 167–177.
- Javeed, Y., 2010. Analysis of groundwater dynamics in semi-arid regions: Effect of rainfall variability and pumping. *Doctoral Thesis*, Indian Institute of Science, Bangalore, India.
- Krätzig, M., 2007. A software framework for data analysis, *Computational Statistics & Data Analysis*, **52**(2), 618–634.
- Legates, D.R., and G.J. McCabe Jr, 1999. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation, *Water Resour. Res.*, **35**, 233–241.

- Lütkepohl, H., 2008. *New Introduction to multiple time series analysis*, Springer Verlag, Berlin.
- Lütkepohl, H., Krätzig M., 2004. *Applied Time Series Econometrics*, Cambridge University Press, Cambridge, DOI: 10.2277/0521547873.
- Panda, D.K., Mishra, A., Jena, S.K., James, B.K., Kumar, A., 2007. The influence of drought and anthropogenic effects on groundwater levels in Orissa, India. *J. of Hydrology*, **343**, 140–153.
- Rodell, M., Velicogna, I., Famiglietti, J.S., 2009. Satellite-based estimates of groundwater depletion in India, *Nature*, **460**, 999–1002.
- Salvucci G.D., Saleem J.A. & Kaufmann R., 2002. Investigating soil moisture feedbacks on precipitation with tests of Granger causality. *Advances in water Resources*, 25, 1305–1312.
- Sanford W.E., 2002. Recharge and groundwater models: An overview. *Hydrogeology Journal*, 10(1), 110–120.
- Sekhar, M., Javeed, Y., Bandyopadhyay, S., Mangiarotti, S., Mazzega, P., 2011. Groundwater management practices and emerging challenges: Lessons from a case study in the Karnataka State of South India. In: *Groundwater Management Practices* (Eds. Findkakis, A. N. and Sato, K.), Chapter 5, IAHR Monograph, ISBN: 9780415619875.
- Sekhar, M., Rasmi, S.N., Sivapullaiah, P.V., Ruiz, L., 2004. Groundwater Flow Modeling of the Gundal Sub-basin in Kabini River Basin, India. *Asian Journal of Water, Environment and Pollution*, 1(1 and 2), 65–77.
- Sekhar, M., Rasmi, S.N., Javeed, Y., Gowrisankar, D., Ruiz, L., 2006. Modelling the groundwater dynamics in a semi-arid hard rock aquifer influenced by boundary fluxes, spatial and temporal variability in pumping/recharge. *Advances in Geosciences*, **4**, 173–181.
- Weber, K., Stewart, M. 2004. A critical analysis of the Cumulative Rainfall Departure concept, *Ground Water*, 42, 935–938.

TABLES:

TABLE 1: **Groundwater stations.** Basic information about the groundwater stations: reference name, station name and qualitative indicator of the pumping increase Low, High (H) or Very High (VH), minimum and maximum values, mean value and standard deviation. Their geographic location of the stations is shown in Figure 1.

| # | Groundwater Stations | Qualitative indicator of increase in pumping | Minimum (m) | Maximum (m) | Mean (m) | Standard Deviation (m) |
|-----|----------------------|--|-------------|-------------|----------|------------------------|
| G1 | Siddainapur | Low | -11.0 | -3.0 | -7.4 | 2.1 |
| G2 | Kaggalahundi | Low | -12.2 | -3.2 | -6.9 | 1.8 |
| G3 | Bheemanbeedu | Low | -12.0 | -0.5 | -6.4 | 2.8 |
| G4 | Kalale | Low | -9.4 | -2.6 | -5.9 | 1.5 |
| G5 | Nanjangud | Low | -5.1 | -2.1 | -3.3 | 0.6 |
| G6 | Harve | H (since 1990) | -27.6 | -4.8 | -10.6 | 3.4 |
| G7 | Hasaguli | VH (since 1990) | -27.7 | -8.2 | -13.7 | 3.9 |
| G8 | Gundulpet | VH (since 1990) | -45.5 | -8.4 | -20.8 | 8.2 |
| G9 | Terakanambi | H (since 1998) | -29.0 | -1.0 | -13.1 | 6.5 |
| G10 | Bommalpura | H (since 2000) | -6.9 | -1.0 | -4.0 | 1.2 |

TABLE 2: **Rainfall stations.** Basic information about the rainfall stations: reference name, station name, qualitative indicator of the climate Mono Modal-West (MM-W), Bimodal (BM) and Mono Modal-East (MM-E) and the maximum values of the daily water height over period 1979-2000. Their geographic location of the stations is given in Figure 1.

| # | Rain (R) gauge Stations or Average Rain (AR) location | Qualitative indicator of climate zone | Maximum minus minimum (mm) |
|------|---|---------------------------------------|----------------------------|
| AR-H | Humid zone | MM-W | 4888. |
| R1 | Ambalvayal | MM-W | 2440. |
| R2 | Bandipur | BM | 1736. |
| R3 | Begur | BM | 628. |
| R4 | Gundulpet | BM | 1149. |
| R5 | Kundakere | BM | 1223. |
| R6 | Terakana | BM | 1236. |
| R7 | Ummatur | MM-E | 1790. |
| R8 | Haradhanalli | MM-E | 1920. |

TABLE 3: **Models fitness.** Models fitness obtained with the 1-lag and the 2-lags models. Percentage of explained variance at groundwater stations G1 to G10. The Nash-Sutcliffe efficiency is also reported together with the skewness and kurtosis of the residues. Note that the normalize model bias is estimated lower than 10^{-12} for both models and for any variables.

| Structural Model | | Stations | | | | | | | | | |
|------------------|---------------------------|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Lags | | G1 | G2 | G3 | G4 | G5 | G6 | G7 | G8 | G9 | G10 |
| 1 | Explained variance (in %) | 60.40 | 57.71 | 62.49 | 19.73 | 6.18 | 41.73 | 19.80 | 66.74 | 85.08 | 29.30 |
| | Nash-Sutcliffe Efficiency | 0.61 | 0.61 | 0.66 | 0.21 | 0.07 | 0.43 | 0.20 | 0.68 | 0.86 | 0.31 |
| | Skewness | 0.32 | -0.69 | 0.40 | -0.03 | -0.11 | -0.68 | -0.89 | -0.81 | -0.00 | -0.23 |
| | Kurtosis | 2.36 | 5.03 | 3.19 | 2.72 | 2.68 | 4.15 | 3.68 | 3.24 | 4.59 | 2.47 |
| 2 | Explained variance (in %) | 62.61 | 55.34 | 64.36 | 21.70 | 16.81 | 41.29 | 20.73 | 69.60 | 85.36 | 33.75 |
| | Nash-Sutcliffe Efficiency | 0.64 | 0.58 | 0.66 | 0.23 | 0.18 | 0.42 | 0.22 | 0.71 | 0.86 | 0.35 |
| | Skewness | 0.31 | -0.70 | 0.18 | -0.02 | -0.27 | -0.63 | -0.90 | -1.02 | -0.02 | -0.22 |
| | Kurtosis | 2.36 | 4.82 | 2.61 | 2.86 | 2.90 | 4.06 | 3.69 | 3.59 | 4.71 | 2.68 |

FIGURES CAPTIONS

All the figures are required to be published in colour on the web and in black-and-white in print.

FIGURE 1: **Region of study.** Geographical distribution of the groundwater stations (trapezes G1 to G10) and rainfall stations (empty triangles R1 to R9) and rainfall aggregated station (AR-H). The area of study is characterized by a strong climatic gradient. Semi-arid region is separated from the humid region by a transition zone. The Kabini River is indicated by a curved dashed arrow.

FIGURE 2: Groundwater piezometric map. Grey (color) gradient in the background indicates the piezometric map of groundwater levels (relative to the mean sea level) in the upper part of the Gundal basin. Bore wells are indicated with (red) dots. Darker (blue) regions in the North correspond to lower groundwater levels (between 690 m. and 710 m.) whereas shallower dark (orange) colors in the South correspond to higher levels (between 910 m. and 930 m.). Local drainage is indicated by thin (blue) lines whereas drainage at larger scale is highlighted by large black arrows. Groundwater stations available in this part of the basin are also reported (black trapezes, see Fig. 1). The piezometric levels are closely linked to the topography indicating the important connectivity of groundwaters between upper and lower regions of the basin (i.e. along the topographic gradient) and the low connectivity between sub-basins. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

FIGURE 3: **The structural model.** Structural model that indicates the network of links allowed in the causality analysis. This structure is given in two parts: endogenous interactions between the groundwater levels (left panel) in which sub-basins are connected through an upstream / downstream relation; and exogenous actions corresponding to rainfall forcing (right panel).

FIGURE 4: **Causal networks.** Causality network deduced from the 'optimal' models obtained after restrictions have been identified. Results are presented for the 1-lag model (left pattern), and the 2-lags model (right pattern) in which both exogenous (thin arrows) and endogenous (thick arrows: plain and dashed arrows for causality of 1 month and 2 months delays, respectively).

FIGURE 1:

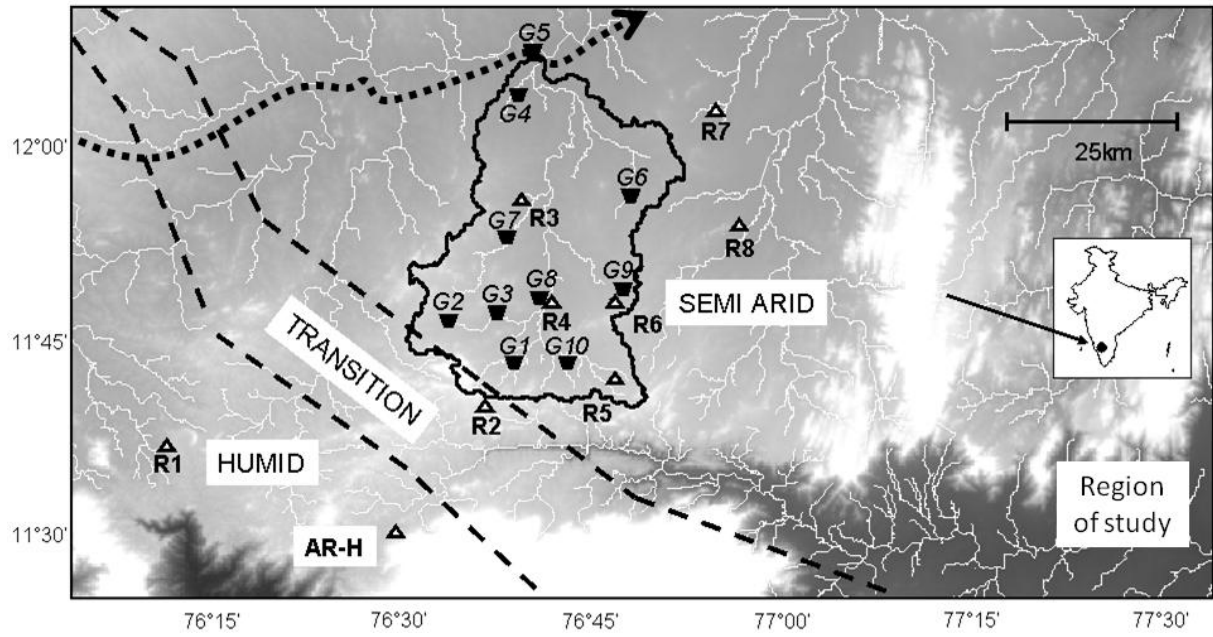


FIGURE 2:

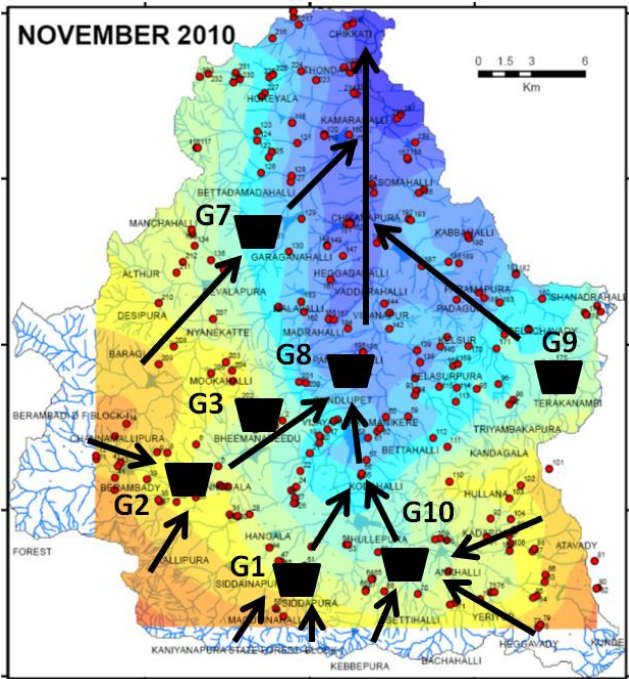


FIGURE 3:

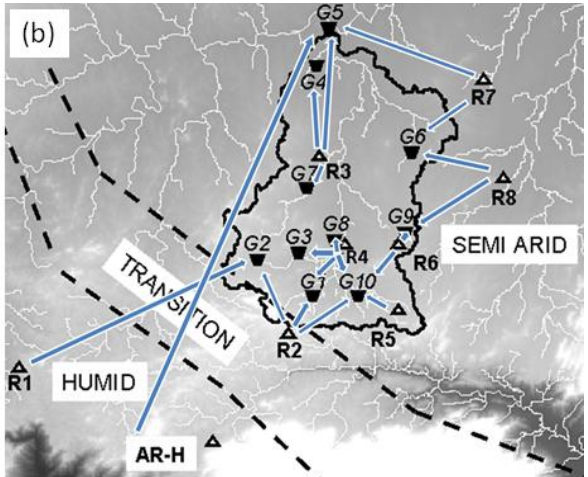
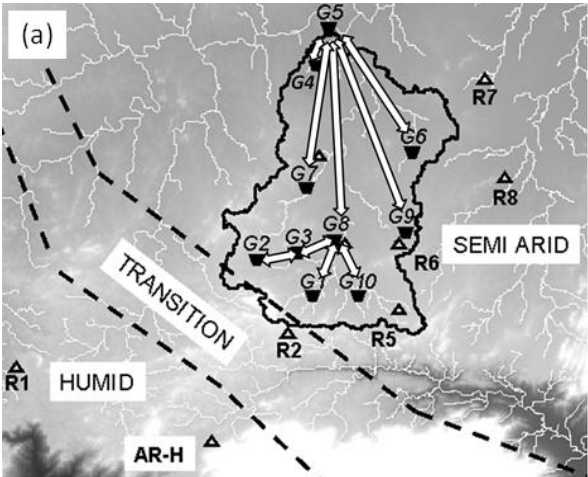


FIGURE 4:

