

Evaluation de la qualité des eaux des puits par la méthode de la régression logistique: cas du Maroc

[Quality of water wells in the region Skhirat: Evaluation by the method of logistic regression - case of Morocco]

Fadwa Ammari

Department of Econometrics,
University Mohammed V-Rabat,
Rabat, Morocco

Copyright © 2015 ISSR Journals. This is an open access article distributed under the ***Creative Commons Attribution License***, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT: This article provides a conceptual framework for measures of practical application at the Region Skhirate. The abusive use of nitrogen fertilizers and pesticides cause negative impacts on the quality of the wells, agricultural intensification can lead also to the degradation of groundwater quality resulting in problems such as salinization, codification, compaction and nitrate pollution, this study tends to affirm the results already achieved by the experts on the quality of water wells, to assess the degree of degradation of these resources.

The analysis by the method of logistic regression is a statistical technique to establish a relationship between a dependent variable and explanatory variables, to examine the associations and make predictions. One can, for example, be interested in quantifying the relationship between the quality of well water in the area Skhirate and risk factors that we will discuss later. In terms of future projections in terms of rainfall and water availability, this initiative could she manage to keep the vision of Morocco as a country that wants to farm? Can we consider irrigation as to sufficiently seen the amount of land suitable for cultivation, or rather is already a problem in the ability to see the irrigation water quality to use?

KEYWORDS: quality of water, irrigation, logistic regression, electrical conductivity (EC), absorption coefficient of the sodium (SAR).

INTRODUCTION

L'eau est un élément essentiel à la croissance économique, à la santé humaine, à l'environnement. La gestion rationnelle des ressources en eau représente pourtant aujourd'hui un défi majeur pour les gouvernements du monde entier.

Le réseau de surveillance de la qualité des eaux souterraines au Maroc comporte 535 stations d'échantillonnage réparties sur 45 nappes d'eau souterraine. La fréquence de prélèvement est semestrielle. Environ 30 000 analyses physico-chimiques et bactériologiques sont réalisées annuellement au niveau des laboratoires régionaux de contrôle de la qualité de l'eau et du laboratoire public d'essais et d'études.

L'agriculture irriguée rencontre de nos jours de nouveaux problèmes tels que le risque de salinité qui peut être apprécié par la conductivité électrique (CE) et celui de l'alcalinisation des sols. Cette dernière due aux échanges ioniques, concerne surtout le sodium, le calcium et le magnésium, entre l'eau et les argiles du sol; il est évalué par le coefficient d'absorption du sodium (SAR). Dans ce sens, la salinité de l'eau doit être aussi faible que possible afin de pallier le problème de salinisation des sols.

Le présent article offre un cadre conceptuel pour des mesures d'application pratique au niveau de la Région de Skhirate. L'emploi abusif des engrais azotés et des produits phytosanitaires entraînent des incidences négatives sur la qualité des puits, l'intensification agricole peut amener aussi à la dégradation de la qualité des eaux souterraines qui se traduit par des problèmes tels que la salinisation, la codification, la compaction et la pollution nitrique, cette étude tend donc à affirmer le résultat déjà réalisé par les experts en la matière sur la qualité des eaux de puits, pour évaluer le degré de dégradation de ces ressources

L'analyse par la méthode de la régression logistique est une technique statistique permettant d'établir une relation entre une variable dépendante et des variables explicatives, afin d'étudier les associations et de faire des prévisions. On peut, par exemple, s'intéresser à quantifier la relation entre la qualité des eaux de puits de la région de Skhirate et les facteurs à risques que nous aborderons par la suite.

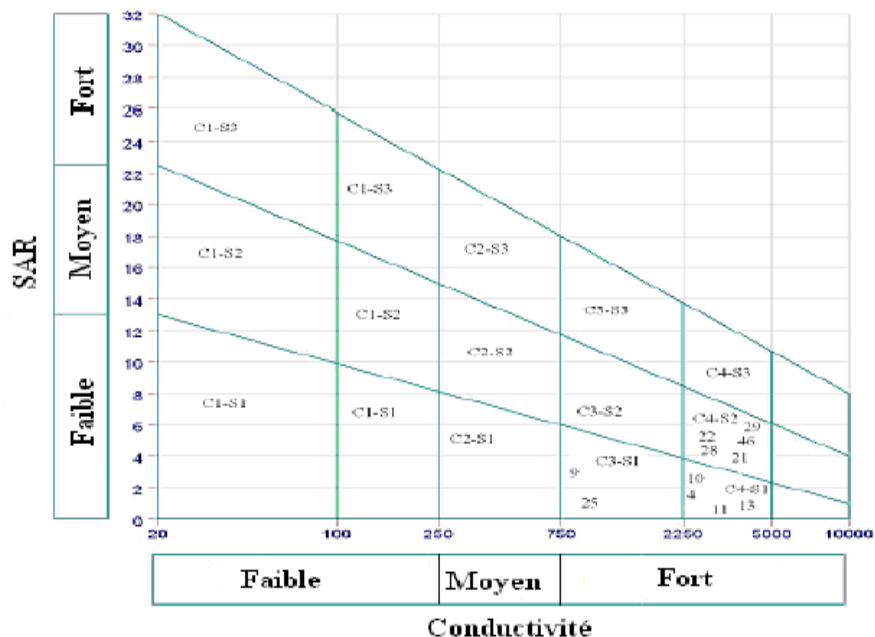
Donc la maîtrise de l'eau a tout le temps été au centre des préoccupations des communautés du Maroc. L'aménagement hydraulique moderne n'est intervenu que vers les années trente. Le Maroc a estimé après l'indépendance nécessaire de développer l'agriculture et parmi les moyens accordés à ce secteur, on trouve la politique de l'irrigation et de l'eau à partir de grandes hydrauliques.

En effet, comme il y avait une prédominance des ressources superficielles caractérisée d'un régime hydraulique des Oueds, il a été édifié la politique des barrages qui prévoyait un par an jusqu'à 2000, la capacité totale s'est trouvée le quintuple de celle de 1955. .

Vu les changements climatiques et en dehors de la problématique de la rationalisation de l'utilisation de l'eau en tant qu'élément rare et de plus en plus coûteux, ce travail portera sur l'état actuelle de l'hydraulique dans la région de Skhirate, et des projections futures des considérations hydraulique surtout que aujourd'hui malgré la politique des barrages, la mauvaise qualité des eaux est un phénomène qui mène beaucoup à réfléchir sur la reconsidération de la gestion des eaux des grandes retenues et des puits, qui influencent énormément l'orientation de l'irrigation au Maroc et mène donc à repenser à cette question de l'économie de l'eau.

LE CRITERE DE CLASSIFICATION DES EAUX DES PUIITS

Afin de déterminer le critère de qualité à établir pour ce travail, il est essentiel de suivre le critère de classification des eaux d'irrigation selon le SAR et la CE selon la méthode du « krigeage », tel que démontré par le graphique suivant :



La carte montre que les eaux se situent entre la classe bonne (S1) et la classe mauvaise (S2) où le danger d'alcalinisation du sol est appréciable dans les sols à texture fine et à forte capacité d'échange. Le C1, C2, C3, C4 sont par ordre décroissant pour la conductivité électrique.

Les eaux se distribuent entre la classe (C3) où les eaux peuvent être utilisées sans contrôle particulier pour l'irrigation des plantes moyennement tolérantes au sel ayant une bonne perméabilité et la classe (C4) où les eaux ne conviennent généralement pas à l'irrigation mais peuvent être utilisées sous certaines conditions: sols très perméables, bien lessivés et des plantes très tolérantes aux sels.

Pour rendre ce travail plus facile à la lecture, la classe moyenne est désormais intégrée dans la partie : qualité d'eau mauvaise. Ainsi, on a attribué « 1 » pour désigner la bonne qualité des eaux de puits et « 0 » pour la mauvaise qualité.

Pour ce faire, on utilisera le logiciel SPSS qui donne directement les sorties susceptibles d'être interprétées.

Pour mesure de précaution, les variables CO3 et HCO3 seront supprimées du fait que le logiciel SPSS a signalé l'existence de variables redondantes et que vu la signification de ces deux variables, il était pratique de se rendre compte qu'il s'agit de ces mêmes variables. La variable PH sera supprimée de notre analyse vu que son interprétation ne procure pas grand chose pour un chimiste si ce n'est que la répercussion de variables importantes sur l'augmentation ou la diminution du PH. D'autres variables seront éliminées ou le recours à une variable de synthèse.

LA SIGNIFICATION DES VARIABLES- PROCEDURE SPSS

CE : la conductivité électrique est l'inverse de la résistibilité, c'est la conductance d'une portion de 1mètre de longueur et de 1 m² de section, mesuré par un conductimètre, l'unité est le siemens par mètre (sm), c'est aussi le rapport de la densité du courant par l'intensité du champ électrique ;

- **PH** : il mesure l'acidité ou la basicité d'une solution. P est potenz et H est l'hydrogène. Le PH est mesuré en moles par litre. C'est le cologarithme décimal de la concentration en ions d'hydrogène $PH = -\log_{10}[H^+]$.

Le PH est un facteur logarithmique, quand une solution devient dix fois plus acide, le PH diminue d'une unité, et si la solution devient cent fois plus acide, le PH diminue de deux unités, il est aussi appelé alcalinité.

- **RS** : le résidu sec mesure la quantité de sel, l'unité est gramme par litre (g/l)
- **SAR** : c'est le rapport adaptation du sol, mesure le risque de sodium pour l'eau d'irrigation, une grande quantité d'ions de sodium dans l'eau affecte la perméabilité des sols et pose des problèmes d'infiltration. Le sodium présent dans le sol en forme échangeable remplace le calcium et le magnésium absorbés sur les argiles de sol et cause la dispersion des particules dans le sol (celui-ci est facilement cultivé et a une structure perméable et granuleuse en absence de quantité importante de sodium) le sol devient dur et compacte, ce qui réduit la vitesse d'infiltration de l'eau nécessaire pour la récolte.
- **RSC** : s'explique par la teneur en carbonate (CO₃⁻) et en bicarbonate (HCO⁻), c'est le risque de bicarbonates par l'eau d'irrigation calculé par milliéquivalents par litre.

Lorsque la concentration de calcium (Ca) et magnésium (Mg) décroît, en comparaison avec la teneur sodium (Na) et l'indice SAR deviennent plus importants, ceci provoque un effet d'alcalinisation et augmente le PH.

Par conséquent, lorsqu'une analyse d'eau indique un PH élevé, ce peut être un signe d'une teneur élevée en ions carbonates et bicarbonates.

Les autres variables appartenant au bilan ionique seront interprétables en terme d'analyse.

Le tableau figurant en annexe présente, la base de données sur laquelle ce travail a été fait.

Il a été fait sur un échantillon de soixante-dix puits.

La courbe ROC permet de représenter sous forme de diagramme les probabilités enregistrées avec la procédure de régression logistique. On peut sélectionner les variables en bloc, en les différent à chaque fois jusqu'à obtention de variables qui sont le plus adéquat avec notre cas.

La méthode choisie pour cette procédure, qui nous permet de spécifier la manière de faire entrer les variables indépendantes est : « ascendante rapport de vraisemblance » qui s'appuie sur des estimations de vraisemblance partielle maximale. c'est une méthode de sélection pas à pas.

Après entrée des variables, on obtient les résultats traités dans le point suivant.

LES SORTIES SPSS

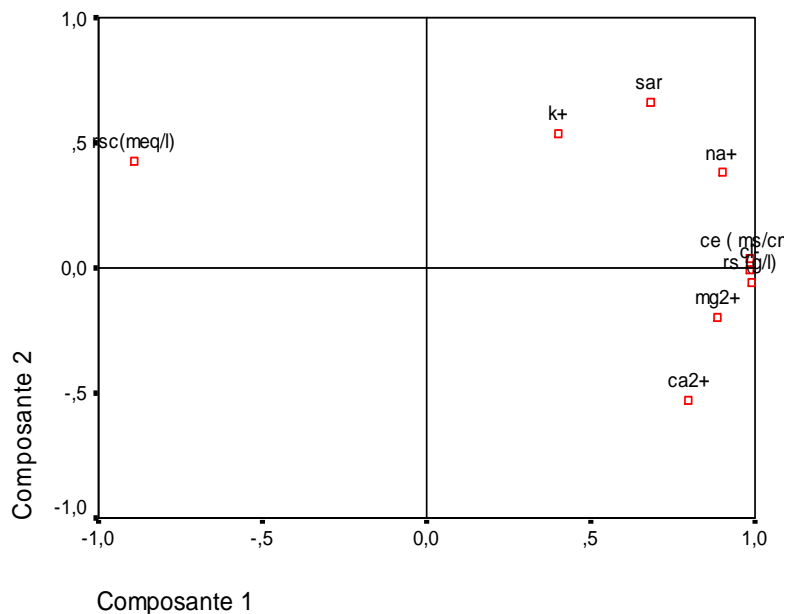
ANALYSE EN ACP

Les sorties qui découlent de la méthode d'analyse en composantes principales est une étape préliminaire à une régression dans notre cas, sur laquelle il convient de porter une analyse.

En effet, l'écart entre les moyennes et les écart- types montrent qu'il y a une hétérogénéité des comportements chimiques des différents puits, mais il y a une très forte corrélation entre d'une part Cl et CE (0,978) et d'autre part entre RS et CE (0,979). En effet le résidu sec n'est d'autre qu'une autre représentation du degré de salinité des eaux.



Diagramme de composantes



Matrice de corrélation

	CE (ms/cm)	Ca2+	Mg2+	Na+	K+	Cl-	RS (g/l)	SAR	RSC(meq/l)	
Corrélation	CE (ms/cm)	1,000	,774	,833	,919	,342	,978	,979	,722	-,845
	Ca2+	,774	1,000	,722	,491	,138	,802	,819	,175	-,937
	Mg2+	,833	,722	1,000	,722	,297	,834	,871	,437	-,900
	Na+	,919	,491	,722	1,000	,419	,883	,876	,918	-,618
	K+	,342	,138	,297	,419	1,000	,360	,336	,409	-,212
	Cl-	,978	,802	,834	,883	,360	1,000	,973	,680	-,876
	RS (g/l)	,979	,819	,871	,876	,336	,973	1,000	,646	-,895
	SAR	,722	,175	,437	,918	,409	,680	,646	1,000	-,299
	RSC(meq/l)	-,845	-,937	-,900	-,618	-,212	-,876	-,895	-,299	1,000

Les variables RS et Cl sont fortement liées avec CE, il convient donc de les éliminer pour une meilleure régression ou de régresser sur des facteurs (c'est le cas avec lequel on va procéder). C'est aussi le cas pour le RSC et partiellement de CE.

Variance expliquée totale

Composante	Valeurs propres initiales			Sommes des carrés chargées		
	Total	% de la variance ==	% cumulés	Total	% de la variance ==	% cumulés
1	6,577	73,081	73,081	6,577	73,081	73,081
2	1,367	15,191	88,272	1,367	15,191	88,272
3	,724	8,049	96,321			
4	,249	2,768	99,089			
5	3,037E-02	,337	99,426			
6	2,240E-02	,249	99,675			
7	1,621E-02	,180	99,855			
8	7,971E-03	8,857E-02	99,944			
9	5,039E-03	5,599E-02	100,000			

Méthode d'extraction : Analyse des principaux composants.

On note aussi une forte corrélation entre le Mg et RSC (-0,90) et entre Ca et RSC (-0,937), les deux agissent en effet fortement et négativement sur le risque de bicarbonates. Les deux composantes contribuent à peu près à 88,27% de l'explication apportées par les variables.

Du fait de la corrélation entre la plupart des variables, la part de la variance des variables restituée par les composantes est très élevée. Les deux composantes sont largement suffisantes pour une synthèse. Ces deux composantes contribuent respectivement à 98,1%, 97,6%, 97,3%, 97% pour RS, RSC, CE et Cl. La variable K par exemple n'est pas très prise en compte par les deux composantes.

La première composante est corrélée positivement avec RS, CE et Cl (au alentour de 0,98), mais aussi avec NA (90%) et MG (88,8%). Elle est corrélée négativement avec RSC

(-89,2%), celui-ci est négativement corrélé avec toutes les variables. On peut donc conclure que la première composante met en opposition deux différents teneurs de matières différentes. On dira des puits salins et des puits bicarbonates.

La deuxième composante est moyennant corrélée avec SAR (66,1%), K (53,5%). ceci permet de distinguer une troisième catégorie d'eau. Il s'agit bien de l'eau sodifiée, qui est en corrélation négative avec Mg et CA, comme on l'a précisé dans la partie signification des variables.

On peut dire que la mauvaise qualité des puits 21, 22,26 provient de leur haut degré de salinité. Les puits 52, 26, 45,42 et beaucoup de puits notamment sont dits bicarbonisés.

Le puit 34 est dit sodifié. Mais beaucoup de puits ne nous permettent pas de décider, ce sont ceux qui se trouvent en barycentre.

Maintenant on va traiter avec la méthode de régression logistique. On procédera d’abord par une régression sur les deux facteurs, puis on finira par une régression sur les différentes variables hormis celles qu’on devrait normalement supprimer.

ANALYSE EN RL (DEUX FACTEURS) : METHODE ASCENDANTE RAPPORTE DE VRAISEMBLANCE

Variables dans l'équation

	B	E.S.	Wald	ddl	Signif.	Exp(B)	IC pour Exp(B) 95,0%		
							Inférieur	Supérieur	
Étape 1 a	FAC1_1	-1,022	,329	9,649	1	,002	,360	,189	,686
	Constante	-,199	,270	,541	1	,462	,820		
Étape 2 b	FAC1_1	-1,481	,480	9,500	1	,002	,228	,089	,583
	FAC2_1	-,950	,453	4,397	1	,036	,387	,159	,940
	Constante	-,415	,325	1,628	1	,202	,661		

a. Variable(s) entrées à l'étape 1: FAC1_1.

b. Variable(s) entrées à l'étape 2: FAC2_1.

Pour le facteur 1, le odds d’avoir un puit de bonne qualité est 0.228 fois plus petit, la propension à avoir un puit de bonne qualité diminue de 77,2% ((0.228-1) x100).

L’effet est significatif (intervalle de confiance ne comprend pas la valeur 1).pour le facteur 2, le odds d’avoir un puit de bonne qualité est 0,387 fois plus petit, la même propension diminue de 61,3%. Le test Wald est un test statistique qui s’interprète comme le Chi square avec un seuil de signification de 0,05. Puisque le seuil est respecté, les coefficients peuvent être interprétés (p < à 5%).

Par exemple, une unité d’appréciation du facteur1 diminuerait le log des chances de 1,481. Ce qui supérieur à l’unité et donc assez grand. Ce qui approuvé par l’ACP, où le premier facteur représente au alentour de 73%.

Une façon d’interpréter les coefficients est de les insérer dans la formule des chances. Par exemple, pour savoir l’impact de la variation d’une unité de la qualité des eaux par rapport au facteur1 :

$$\text{Chances} = e^{B0} \times e^{B1 (\text{variable1})} \times e^{B2 (\text{variable2})}$$

$$1) \text{ Chances} = e^{(-0,415)} \times e^{(-1,481) \times 1,00} \times e^{(-0,950) \times 1} = 0,058$$

On neutralise l’effet du fact2 et on fait varier le fact 1 d’une unité :

$$2) \text{ Chances} = e^{(-0,415)} \times e^{(-1,481) \times 2} \times e^{(-0,950) \times 1} = 0,013$$

Donc, une unité d’augmentation de l’un des éléments du facteur1, diminue par un facteur de 4,46 (0,058 / 0,013) les chances d’avoir un puit de bonne qualité.

Faisant varier la le facteur 2 de une unité :

$$e^{(-0,415)} \times e^{(-1,481) \times 1} \times e^{(-0,950) \times 2} = 0,022$$

Donc, une unité d’augmentation de l’un des éléments du facteur2, diminue par un facteur de 2,63 (0,058 / 0,022) les chances d’avoir un puit de bonne qualité. Soit un taux moins important que celui enregistré par le premier facteur.

Statistiques		
qualité		
N	Valide	70
	Manquante	0
Moyenne		,47

La moyenne indique la probabilité. Donc, pour l'échantillon, la probabilité que la qualité d'un puit soit élevée est de 47%. La probabilité d'avoir une bonne qualité par rapport au premier facteur est : $B1 \times p (1-p)$

$$B2 \times p (1-p) = -1,481 * 0,47 (1-0,47) = -0,37$$

La probabilité d'avoir une bonne qualité par rapport au deuxième facteur : $B2 \times p (1-p)$

$$B2 \times p (1-p) = -0,95 * 0,47 (1-0,47) = -0,23$$

Donc, une unité d'appréciation du facteur 1 diminue de 37% la probabilité d'avoir une qualité élevée, et diminue de 23% par rapport au deuxième facteur.

SPSS a retenu donc les deux facteurs pour expliquer la probabilité d'avoir un puit de bonne qualité.

Tableau de classification^a

Observé			Prévu		
			qualité		Pourcentage correct
			mauvaise	bonne	
Étape 1	qualité	mauvaise	22	15	59,5
		bonne	9	24	72,7
Pourcentage global					65,7
Étape 2	qualité	mauvaise	26	11	70,3
		bonne	7	26	78,8
Pourcentage global					74,3

a. La valeur de césure est ,500

Une qualité élevée, soit une diminution de salinité ou de sodification, serait prédit avec succès 78,8% des fois, alors qu'une qualité faible sera prédite avec succès 70,3% des fois. Un taux d'erreur de classification est de 25,71% ((11+7)/70). La procédure discriminante donne :

La proportion des faux bons puits est de 29,72% (11/11+26), la proportion des faux mauvais puits est de 21,21% (7/26+7). Le taux global de bon classement est de 74,3%.

Résultats du classement^a

			Classe(s) d'affectation prévue(s)		Total
			mauvaise	bonne	
Original	Effectif	mauvaise	22	15	37
		bonne	5	28	33
	%	mauvaise	59,5	40,5	100,0
		bonne	15,2	84,8	100,0

a. 71,4% des observations originales classées correctement.

Par analyse discriminante, les résultats permettent un taux d'erreur de classification inférieur à celui obtenu par régression logistique soit : 22,8% ((15+5)/70).

Coefficients des fonctions discriminantes canoniques standardisées

		Fonction
		1
REGR factor score	1	,941
REGR factor score	2	,417

La fonction discriminante est significative à 1%, elle s'écrit : $Y = 0,941 \text{ fact1} + 0,398 \text{ fact 2}$

Matrice de corrélation

		Constante	FAC1_1	FAC2_1
Etape 1	Constante	1,000	,230	
	FAC1_1	,230	1,000	
Etape 2	Constante	1,000	,506	,479
	FAC1_1	,506	1,000	,640
	FAC2_1	,479	,640	1,000

Ce tableau nous permet de constater que les dimension fact1 et fact2 sont moyennant corrélées (64%), et donc qu'il s'agit bien de deux dimensions distinctes de la qualité des eaux de puits.

Modèle si terme supprimé

Variable	Modèle log-vraie mblance	Modification dans -2log-vrais emblance	ddl	Signification de la modification
Etape 1 FAC1_1	-48,406	12,918	1	,000
Etape 2 FAC1_1	-47,388	16,573	1	,000
FAC2_1	-41,947	5,692	1	,017

Le facteur1 n'explique que peu en l'absence du terme constant.

Recueil de tests sur les coefficients de modèle

		Khi-deux	ddl	Signif.
Etape 1	Etape	12,918	1	,000
	Bloc	12,918	1	,000
	Modèle	12,918	1	,000
Etape 2	Etape	5,692	1	,017
	Bloc	18,609	2	,000
	Modèle	18,609	2	,000

Le Chi square est plus élevé pour le Step 2, donc le modèle est amélioré par l'introduction du second facteur (12,918 et 18,609).

Le test global dans son ensemble: test significatif ($p < 5\%$) pour une meilleure qualité prédictive du modèle; est équivalent à un test que toutes les différences entre catégories sont égales à 0; autrement dit, qu'état civil n'apporte aucune information. On peut conclure que le modèle est globalement significatif, l'équation de régression logistique s'écrit :

$$Y = -0,415 - 1,481 \text{ fact1} - 0,95 \text{ fact2}$$

Si le χ^2 observé est supérieur au χ^2 théorique, à un certain seuil de signification, on peut alors rejeter l'hypothèse nulle et ainsi conclure sur le lien entre les deux variables. Le test du Chi-2 a pour limite d'être sensible à la taille de l'échantillon et au nombre de degrés.

Récapitulatif du modèle

Etape	-2log-vraisemblance	R-deux de Cox & Snell	R-deux de Nagelkerke
1	83,894	,169	,225
2	78,203	,233	,312

La valeur du -2 Log likelihood n'indique rien en soi, mais sa diminution dans le Step 2 nous indique également que le modèle est amélioré par l'introduction du second facteur. Un modèle parfait aurait un -2 Log likelihood de zéro.

Les deux pseudo R2 nous permettent d'expliquer le pourcentage de la variable dépendante binaire qui est expliqué par les deux facteurs retenus.

Le Nagelkerke est une version ajustée du Cox & Snell et est donc plus près de la réalité.

Ainsi, 31,2% de la variation dans la qualité des eaux de être expliquée par la contribution des deux facteurs.

Ce graphique est une représentation visuelle de la performance du modèle. Si le modèle était parfait, tous les 0 seraient à la gauche des 1. Dans ce cas, c'est presque le cas, mais un certain pourcentage d'erreurs existant.

Diagramme de dispersion/ facteur1 et variable dépendante

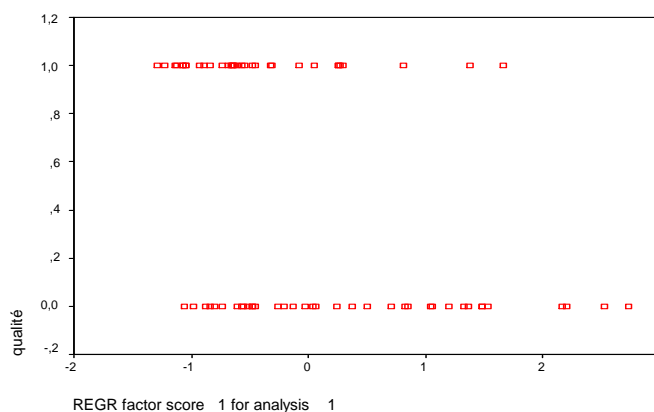
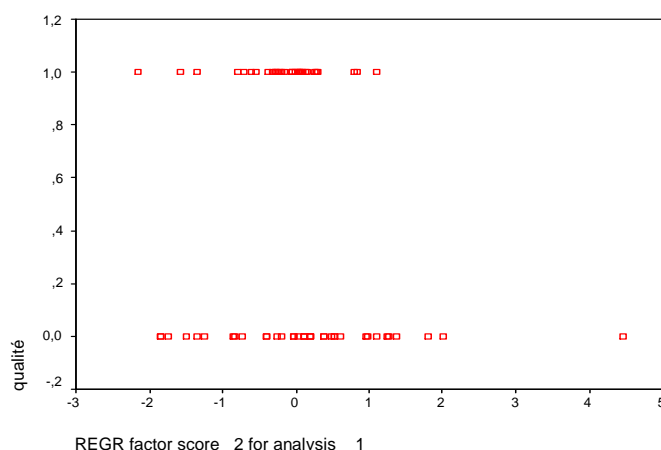


Diagramme de dispersion/ facteur2 et variable dépendante



On peut qualifier qu'un puit est effectivement de bonne qualité d'eau si la probabilité prévu dépasse 0,5. Ainsi on peut savoir le détail concernant chaque puit. Ainsi si le puit 11 est observé mauvais, il est prévu bon. La déviance joue le rôle d'un écart résiduel, mais toutefois il n'est inutile de l'interpréter vu que la statistique de Pearson, plus cette quantité est petite, plus le modèle est adéquat.

ANALYSE EN RL (VARIABLES) : MÉTHODE DESCENDANTE RAPPORT DE VRAISEMBLANCE

En faisant entrer tous les variables, hormis ceux qui nuisent à la régression (on pourrait recenser (PH, HCO3, CO3 et NA)

Variables dans l'équation

	B	E.S.	Wald	ddl	Signif.	Exp(B)	IC pour Exp(B) 95,0%		
							Inférieur	Supérieur	
Étape 1	CE	,321	1,181	,074	1	,786	1,379	,136	13,969
	CA	-,018	,242	,005	1	,942	,982	,611	1,580
	SAR	-,511	,408	1,565	1	,211	,600	,269	1,336
	RSC	-,016	,194	,007	1	,934	,984	,673	1,439
	CI	,045	,124	,133	1	,715	1,046	,821	1,334
	MG	-,320	,230	1,931	1	,165	,726	,463	1,140
	Constante	2,943	1,632	3,253	1	,071	18,980		
Étape 2	CE	,271	,964	,079	1	,779	1,311	,198	8,676
	SAR	-,497	,360	1,909	1	,167	,608	,301	1,231
	RSC	-,006	,137	,002	1	,965	,994	,759	1,301
	CI	,046	,124	,136	1	,712	1,047	,821	1,334
	MG	-,308	,164	3,527	1	,060	,735	,533	1,013
	Constante	2,855	1,087	6,899	1	,009	17,366		
Étape 3	CE	,283	,924	,094	1	,759	1,327	,217	8,116
	SAR	-,509	,227	5,048	1	,025	,601	,386	,937
	CI	,049	,098	,247	1	,619	1,050	,866	1,273
	MG	-,304	,134	5,126	1	,024	,738	,567	,960
	Constante	2,862	1,074	7,096	1	,008	17,491		
Étape 4	SAR	-,486	,213	5,195	1	,023	,615	,405	,934
	CI	,075	,052	2,092	1	,148	1,078	,974	1,192
	MG	-,296	,131	5,108	1	,024	,744	,576	,962
	Constante	3,019	,952	10,054	1	,002	20,470		
Étape 5	SAR	-,367	,200	3,375	1	,066	,693	,468	1,025
	MG	-,138	,062	4,897	1	,027	,872	,772	,984
	Constante	2,667	,912	8,554	1	,003	14,393		

a. Variable(s) entrées à l'étape 1: CE, CA, SAR, RSC, CI, MG.

La procédure s'arrête au niveau 5, on aura retenu les variables SAR, et MG. Ce résultat n'est pas le même si on procède par la méthode ascendante. Mais ceci permet d'avoir deux variables dans la fonction (MG et SAR) faite par la méthode ascendante.

Pour SAR, le odds d'avoir un puit de bonne qualité est 0,693 fois plus petit, la propension à avoir un puit de bonne diminue de 30,7%. La variable elle-même n'est pas significative.

Mais L'effet est significatif

Une façon d'interpréter les coefficients est de les insérer dans la formule des chances. Par exemple, pour savoir l'impact de la variation d'une unité de la qualité des eaux par rapport au facteur1 :

$$\text{Chances} = e^{B0} \times e^{B1(\text{variable1})} \times e^{B2(\text{variable2})}$$

$$1) \text{ Chances} = e^{(2,667)} \times e^{(-0,367) \times 1,00} \times e^{(-0,138) \times 1} = 8,69$$

On neutralise l'effet de la var1 et on fait varier la var2 d'une unité :

$$2) \text{ Chances} = e^{(2,667)} \times e^{(-0,367) \times 1,00} \times e^{(-0,138) \times 2} = 7,56$$

Donc, une unité d'augmentation de l'un des éléments de la variable2, diminue par un facteur de 1,15 (8,69 / 7,56) les chances d'avoir un puit de bonne qualité.

La matrice de corrélation indique une forte corrélation entre les variables.

Matrice de corrélation

		Constante	CE	CA	SAR	RSC	CI	MG
Etape 1	Constante	1,000	,214	-,746	-,618	-,458	,309	-,572
	CE	,214	1,000	-,579	-,580	-,245	-,416	-,412
	CA	-,746	-,579	1,000	,477	,706	,047	,703
	SAR	-,618	-,580	,477	1,000	-,147	-,356	,148
	RSC	-,458	-,245	,706	-,147	1,000	,462	,785
	CI	,309	-,416	,047	-,356	,462	1,000	,161
Etape 2	MG	-,572	-,412	,703	,148	,785	,161	1,000
	Constante	1,000	-,401		-,448	,146	,516	-,100
	CE	-,401	1,000		-,425	,283	-,477	-,008
	SAR	-,448	-,425		1,000	-,776	-,430	-,300
	RSC	,146	,283		-,776	1,000	,607	,575
	CI	,516	-,477		-,430	,607	1,000	,182
Etape 3	MG	-,100	-,008		-,300	,575	,182	1,000
	Constante	1,000	-,466		-,536		,544	-,228
	CE	-,466	1,000		-,340		-,851	-,216
	SAR	-,536	-,340		1,000		,081	,282
	CI	,544	-,851		,081		1,000	-,258
Etape 4	MG	-,228	-,216		,282		-,258	1,000
	Constante	1,000			-,836		,312	-,376
	SAR	-,836			1,000		-,417	,228
	CI	,312			-,417		1,000	-,867
Etape 5	MG	-,376			,228		-,867	1,000
	Constante	1,000			-,834			-,194
	SAR	-,834			1,000			-,296
	MG	-,194			-,296			1,000

La khi deux est amélioré d'étapes en étapes (il diminue), ici la méthode est descendante.

Le modèle est globalement significatif, L'équation s'écrit :

$$Y = 2,667 - 0,367 \text{ SAR} - 0,138 \text{ MG}$$

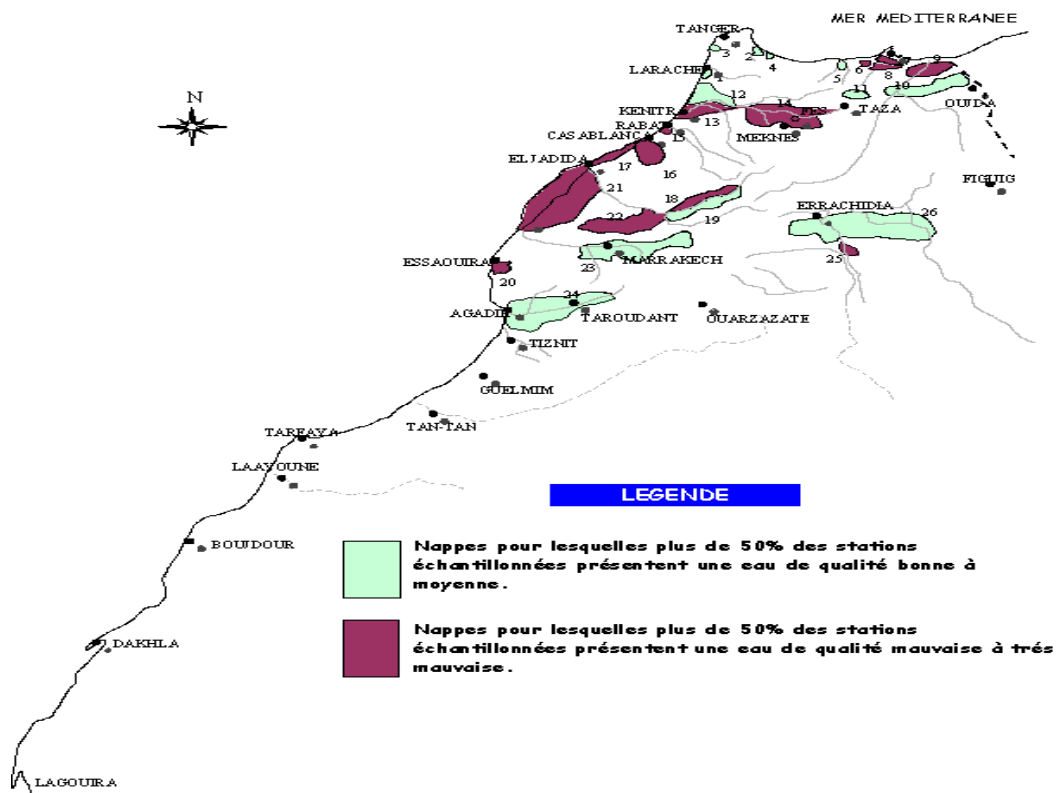
Il serait inutile de reprendre toutes les analyses précédentes. Mais le tableau suivant résume ce qui en est :

Recueil de tests sur les coefficients de modèle

		Khi-deux	ddl	Signif.
Etape 1	Etape	21,831	6	,001
	Bloc	21,831	6	,001
	Modèle	21,831	6	,001
Etape 2 ^a	Etape	-,005	1	,942
	Bloc	21,825	5	,001
	Modèle	21,825	5	,001
Etape 3 ^a	Etape	-,002	1	,965
	Bloc	21,824	4	,000
	Modèle	21,824	4	,000
Etape 4 ^a	Etape	-,094	1	,759
	Bloc	21,730	3	,000
	Modèle	21,730	3	,000
Etape 5 ^a	Etape	-2,266	1	,132
	Bloc	19,464	2	,000
	Modèle	19,464	2	,000

a. Une valeur khi-deux négative indique que la valeur du khi-deux a diminué depuis l'étape précédente.

Le taux de mauvais classement est de 28,57%. Soit un peu plus que le taux pour l'analyse par facteurs. Cette méthode laisse à omettre beaucoup de variables, ou introduit des variables supposées corrélées alors qu'une régression devrait normalement poser le postulat de non corrélation entre les variables. De plus qu'il est peu pratique d'analyser avec ces variables, car on pourrait avoir des fonctions différentes selon notre choix.



CARTE DE LA QUALITE DE L'EAU DES PRINCIPALES NAPPES D'EAUX SOUTERRAINES

CONCLUSION

L'agriculture irriguée rencontre de nos jours de nouveaux problèmes tels que le risque de salinité et qui peut être apprécié par la conductivité électrique et celui de l'alcalisation du sol. Il s'agit d'un fait scientifique prouvé. Nous remarquons que l'analyse par les facteurs laisse à prédire un modèle plus bon que celui où on prend des variables séparément.

CE, CI et NA, RS, MG et CA sont fortement corrélées avec la première composante et donc comme on l'a prouvé : une unité d'augmentation de l'un des éléments du facteur1, diminue par un facteur de 4,46 (0,058 / 0,013) les chances d'avoir un puit de bonne qualité.

Une unité d'augmentation de l'un des éléments du facteur2, diminue par un facteur de 2,63 (0,058 / 0,022) les chances d'avoir un puit de bonne qualité. Soit un taux moins important que celui enregistré par le premier facteur. On pourrait recenser SAR, RSC et K mais l'effet est moyen selon le degré de corrélation avec la composante principale2.

La régression logistique a offert un moyen de détecter les puits mal classés et de dire que les chances d'avoir un puits de bonne qualité d'eau diffèrent selon les composantes et selon, la méthode de régression logistique elle-même.

La réaction chimique des variables le sodium, le calcium, le magnésium, la quantité de sel, et le nitrate entre l'eau et les argiles du sol s'avère d'autant importante. Cette étude affirme à partir d'un modèle probabiliste l'apparition du caractère bon ou mauvais de la qualité des eaux de puits de la région de Skhirate.

Une fois encore, les probabilités calculées, le travail d'un chimiste ne s'arrête pas au niveau de détection mais au-delà de cela, il se doit d'agir vis-à-vis du problème posé, en mettant en garde contre l'usage sans précaution de l'un ou de l'autre puit.

REFERENCES

- [1] Adil Marhom, ouvrage « Analyse de données » ;
- [2] Christophe Higy, Laurent Cordey « Analyse de la qualité de l'eau de puits transformés : Un exemple d'application au Bénin », juillet 2011 ;
- [3] Rapport de l'OCDE « Qualité de l'eau et agriculture : Un défi pour les politiques publiques », 2012 ;
- [4] « Etude du risque de gastro-entérite chez les familles utilisant l'eau d'un puits domestique », direction risques biologiques, environnementaux et occupationnels institut national de santé publique du Québec septembre 2004 ;
- [5] Document de travail de l'institut national de la recherche agronomique (INRA), Maroc, portant sur les déterminants de la qualité des eaux de puits.