INFERRING INJECTION RETURNS FROM CHLORIDE MONITORING DATA

Ma. Michelle Sullera and Roland N. Home

Stanford Geothermal Program, Stanford University; Stanford, California 94305-2220, U.S.A.

ABSTRACT

Reservoir chloride concentration and injection rate data have been used to identify injection return flow paths in the reservoir. General trends of chloride and injection rate with time were isolated from their respective short-term variations using the wavelet transformation approach. Multiple regression techniques were then used to correlate the isolated short-term variations in chloride with corresponding short-term fluctuations in injection rates and subsequently to quantify the degree of connectivity between injectors and producers. Data from Palinpinon-I field in the Philippines were analyzed using the outlined method and results were verified successfully against tracer test data and qualitative field observations.

Results of analysis of chloride and injection rate data fiom Dixie Valley field in Nevada illustrated that multilinear modeling is not suitable for analyzing data *sets* that lack sufficient time variability or "texture".

1.0 INTRODUCTION

Traditionally, tracer tests are used to establish the degree of connectivity **between** wells. However, for wells that are only weakly connected these tests may need to be conducted over long periods of time using huge amounts of tracer of sufficient stability to obtain meaningful data. In such cases tracer tests *can* be too costly and impractical.

On the **other** hand, there are substances occurring naturally **in** the reservoir that ***can*** behave **as** tracers. One such substance is chloride. In Palinpinon geothermal field in the Philippines, some injectors and producers are strongly connected **so** that changes in injection rates result in corresponding increase **or** decrease in chloride concentrations measured in production wells. Data from one such injector-producer pair in Palinpinon is shown in Fig. **1**. The magnitude of the changes in chloride concentration thus reflects the degree of communication between wells. Moreover, Chloride is stable, reasonably conservative and it is free. Therefore, we may be able to extract the same, if not more, information from chloride **data as** we could **from** traditional tracer tests and at lower cost.

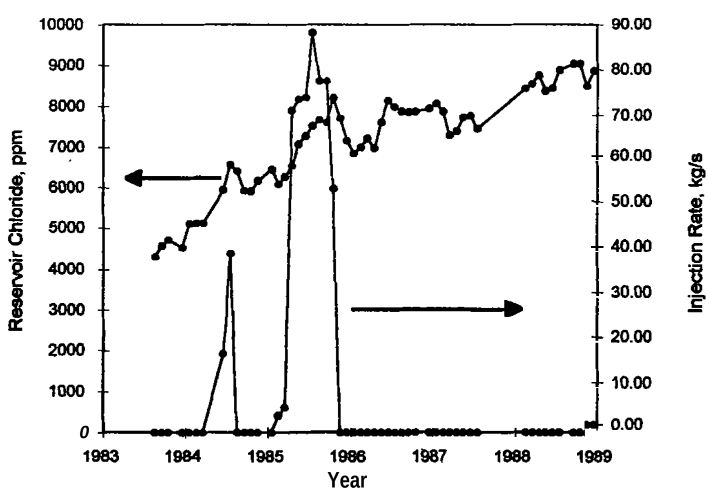


Figure **1**. Example chloride **and** injection **data** from Palinpinon-I.

The following sections summarize how the method of wavelets and multiple regression techniques were used to analyze chloride and injection data and consequently, identify injection return flow paths; and, how the permeability of these paths were ranked by quantifying the degree of connectivity between injectors and producers.

2.0 PRELIMINARY LINEAR MODELS

As part of an optimization problem, an earlier work by Macario (1991) proposed several correlations for modeling the reservoir chloride and applied these models to data fiom Palinpmon-I geothermal field in the Philippines. Of the models tested by Macario (1991), the linear combination model came closest to reproducing field observations. In the first phase of this project therefiore, we chose to expand on that model and test it further.

Following is the original linear Combination model proposed by Macario (1991):

$$Cl_p = a_0 + a_1Q_{I1} + a_2Q_{I2} + a_3Q_{I3} + \dots\dots + a_nQ_{In} \tag{1}$$

Based on this model, the strength of the connection between the modeled producer P and an injector I_i is assessed by the magnitude of the coefficient, a_i, of that injector in the model; high values of a correspond to strong connections.

Aside fiom the extent of reinjection fluid returns, other factors could also affect the chloride level in the reservoir. Extensive boiling and steam separation within the reservoir and natural recharge of higher mineralized fluid are processes that could increase chloride concentrations (Harper and Jordan, 1985). The first process, boiling and steam separation, is a natural reservoir response to exploitation. The chloride concentration may therefore be expected to increase with time as the reservoir is produced. To model this variation with time, a linear time term was added to model (1), thus:

$$Cl_p = a_0 + a_1Q_{I1} + a_2Q_{I2} + a_3Q_{I3} + \dots\dots + a_nQ_{In} + bt \tag{2}$$

More than anything, it was simplicity that guided our choice of the form (linear) of the time term. Solution saturation limits could be expected to put a cap on the maximum chloride concentration and cause it to level off late in the life of the reservoir. For practical purposes, however, we assumed that the chloride concentrations being modeled were far from the maximum limit and increased linearly with time. The question of how chloride concentration in the reservoir actually varies with respect to rime will be addressed in more detail in a later section.

We also hypothesized that the reinjection returns' effect on reservoir chloride is governed not just by the rates of injection but also by the chloride concentration of the reinjected fluid. Hence, we have proposed the following modification to model (2):

$$Cl_p = a_0 + a_1Q_{I1}Cl_{I1} + a_2Q_{I2}Cl_{I2} + a_3Q_{I3}Cl_{I3} + \dots\dots + a_nQ_{In}Cl_{In} + bt \tag{3}$$

The additional parameter Cl_{In} refers to the chloride concentration of the fluid being injected to injector I_n.

2.1 Results and Discussion

The original and extended models were applied to analysis of both the data set from Palinpinon-I previously used by Macario (1991) and another data set from the Dixie Valley field in Nevada. Qualities of the fit to the data were assessed by inspecting both the calculated values of the multiple regression coefficient, R^2, and plots of model predicted chlorides against actual data. The multiple regression coefficient, R^2, represents the proportion of variation in the modeled variable (in this case, chloride concentration) that is predictable from the model. It is therefore desirable to have high values of R^2. Only the model which best fitted the data or equivalently, had the highest value of R^2 was subjected to further tests.

For any model to be considered relevant it was deemed necessary that that model be able to account for variations in chloride at any time interval in the data set regardless of which portion of the data set was used to calculate the linear coefficients. Thus we assessed model relevance by examining how well the model predicts later chloride measurements using coefficients that were calculated fiom earlier portions of the data set.

The following section discusses the results of application of models (1), (2), and (3) to the Dixie Valley and Palinpinon-I data sets. Model (3) was not used to analyze the Palinpinon-I data set due to the lack of injectate chloride data from that field

Dixie Valley **Case**

At Dixie Valley, injection rates were recorded daily while chloride concentrations were measured much less fiquently, thus, it was the amount of chloride data that set the limit to the number of data points (cases in which chloride concentrations and injection rates were available simultaneously) used for regression.

Table **1** lists R^2 values for models (1), (2), and (3) obtained for each production well. Except for wells **27-33** and **28-33,** model (2) gave the highest value of R^2 for all production wells. Addition of the time term to model **(1)** resulted in 2 to 35% increase in R^2 while inclusion of injectate chloride concentration in model (3) did not result in any significant change in R^2 values. Figure 2 shows the effect of a 35% difference in R^2 on the quality of data fit for well **84-7;** also, it illustrates the very minor effect that the injectate chloride term had on the quality of the match. Based on these results we chose to subject model (2) to further testing.

Subsequently, the last six data points were excluded **fiom** the regression. Model **(2)** was then used to predict these values using the coefficients calculated based on the truncated data set. Figure 3 plots the results of the truncated series analysis for well **27-33** which had a 9% maximum deviation of predicted chloride fiom actual data - the highest deviation observed among all the production wells. *Other* wells had as little as 1% deviation (Fig. **4).**

Inspection of the calculated coefficients revealed one possible reason for the relatively good predictive capacity displayed by model (2) (see Table 2). For this data set, the time term dominates the correlation; the coefficient of the time term is several orders of magnitude (3 to **5,** even **8** times!) greater than the injection rate coefficients. This discrepancy was enough to render the injection rate terms trivial: excluding an injection rate term from model (2) resulted in only tiny changes in the quality of the data fit. Figure **5** shows the chloride match for well **74-7** when the chloride is predicted using model (2) but with injection rate term corresponding to injector INJ5218 excluded. That the

Table 1. R^2 values for Dixie **Valley wells.**

	R^2		
27-33	0.917	0.963	0.965
28-33	0.852	0.936	0.940
4533	0.935	0.970	0.966
63-7	0.826	0.828	0.815
73-7	0.774	0.952	0.952
74-7	0.755	0.968	0.967
76-7	0.930	0.947	0.943
82-7	0.764	0.969	0.967
84-7	0.716	0.978	0.978

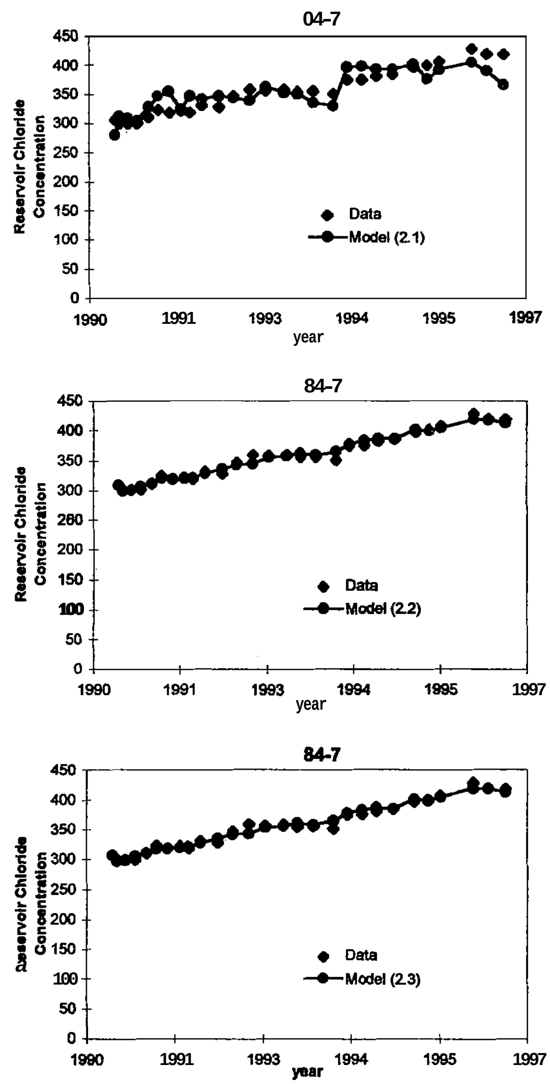


Figure 2. Predicted vs. measured chloride concentration for well 84-7, Dixie Valley.

Table **2.** Model **(2)** coefficients **for** Dixie Valley production wells.

Model Parameter	Production Well								
	27-33	28-33	45-33	63-7	73-7	74-7	767	82-7	84-7
a_0	314.58	354.39	271.46	283.77	272.43	318.32	384.12	254.39	271.60
Injection wells:									
INJ255	8.61E-05	-4.84E-04	5.06E-04	1.80E-03	2.94E-04	1.51E-05	3.60E-04	-3.05E-06	3.99E-04
INJ455	-2.20E-04	-5.56E-04	3.33E-04	1.33E-03	2.70E-04	-1.62E-04	-1.37E-04	7.57E-05	3.89E-05
INJ3218	-1.21E-03	-4.30E-04	-1.54E-03	2.69E-03	2.12E-04	-5.89E-04	-3.59E-04	6.67E-04	-1.24E-03
INJ4118	4.58E-04	6.02E-04	3.54E-04	-4.40E-04	4.35E-05	7.31E-05	2.87E-04	-2.31E-04	9.76E-07
INJ5218	-1.20E-03	-7.39E-04	-1.18E-03	-2.95E-04	4.37E-04	-2.50E-04	-5.42E-04	7.43E-04	-1.42E-04
INJ6518	3.56E-03	3.40E-03	2.65E-03	-2.77E-03	-6.31E-04	9.02E-04	-5.63E-04	-3.33E-04	2.07E-03
INJ_SWL1	-3.99E-03	-3.74E-03	-3.33E-03	8.45E-04	5.39E-04	1.68E-04	4.18E-04	6.27E-04	-7.24E-04
INJ_SWL3	1.49E-03	9.69E-04	1.33E-03	-4.17E-03	-1.35E-03	-2.18E-04	-3.49E-05	-1.43E-03	-3.27E-04
t	8.48	10.37	7.61	2.30	17.44	12.95	4.82	26.25	19.04

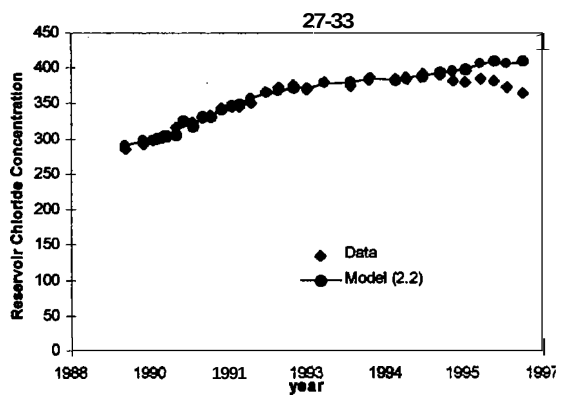


Figure **3.** Predicted vs. measured chloride concentration for well **27-33,** Dixie Valley; model **(2)** coefficients calculated with last six data points excluded.

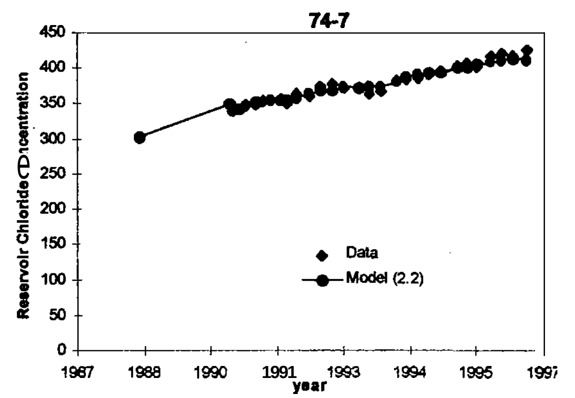
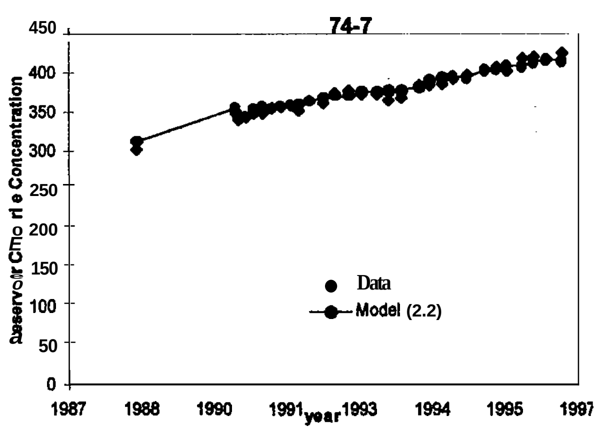


Figure **4.** Predicted **vs.** measured chloride concentration for well **74-7,** Dixie Valley; model **(2)** coefficients calculated with last **six data** points excluded.



Figures. Predicted vs. measured chloride concentration for well **74-7,** Dixie Valley; predicted values calculated using model **(2)** with one **injection** rate ***term*** excluded injection rate terms are inconsequential to chloride prediction was also evident from inspection of the chloride data: for the most part, the chloride increased linearly with time and response to changes in injection rates was not readily evident. Hence, once the variation of chloride with time was captured in the analysis of the early portion of the data set and there was little deviation observed in the succeeding predictions. It was noted however, that although the deviations were small some of them showed a tendency to increase (Fig. 3). This was true for wells whose chloride ceased at some point to vary linearly with time.

At this point, it is worthwhile to recall that the goal of this project was not prediction but rather, correlation. Although for this specific data set model (2) matched and predicted chloride data relatively well, the dominance of the time term rendered the injection rate coefficients meaningless and ultimately made this model unsuitable for comparing the effects of injection wells on production wells in the field being considered here.

The preceding results lead us to conclude that for the purpose we have set for this project, multiple regression is not a suitable analysis tool for chloride data sets which lack texture.

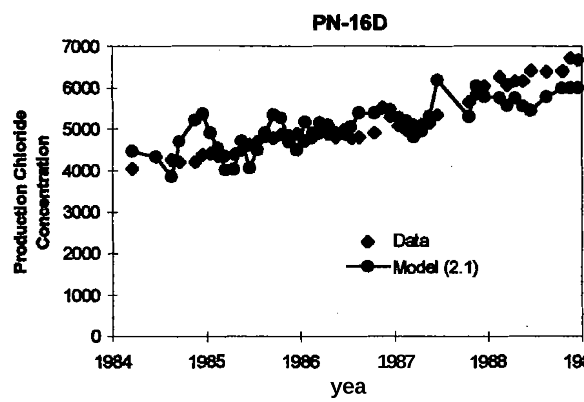
Polinpinon-I Case

In this case, injection rate data were available as monthly average values; thus, chloride data were converted to monthly average values prior to analysis. As with the Dixie Valley data set, the amount of chloride data set the limit on the number of data points used for regression. Only the portion of the data set fiom 1983 to 1989 was initially available for use in the initial inspection of the linear models; thus, the following results pertain to the analysis of that early portion of the data set.

The effect on R^2 of adding the time term to model (1) was even more drastic for the Palinpinon-I data set: a maximum increase of 80% in R^2 was observed (Table 3) The effect of a 60% increase in R^2 on the quality of the match for well PN-16D is shown in Fig. 6.

As was done previously in the analysis of the Dixie Valley data set, in the subsequent regression using model (2) the last six points in the chloride series were not considered in the calculation of the linear coefficients. The excluded chloride values were then predicted using the coefficients calculated based on the truncated data set. Deviations of

Well Name	R^2	
	Model (2.1)	Model (2.2)
OK-7D	0.783	0.956
OK-9D	0.717	0.902
OK-10D	0.490	0.535
PN-15D	0.824	0.993
PN-16D	0.606	0.964
PN-17D	0.519	0.939
PN-18D	0.718	0.930
PN-19D	0.559	0.903
PN-23D	0.736	0.958
PN-24D	0.706	0.895
PN-26D	0.728	0.922
PN-27D	0.696	0.944
PN-28D	0.643	0.895
PN-29D	0.817	0.948
PN-30D	0.710	0.832
PN-31D	0.625	0.946



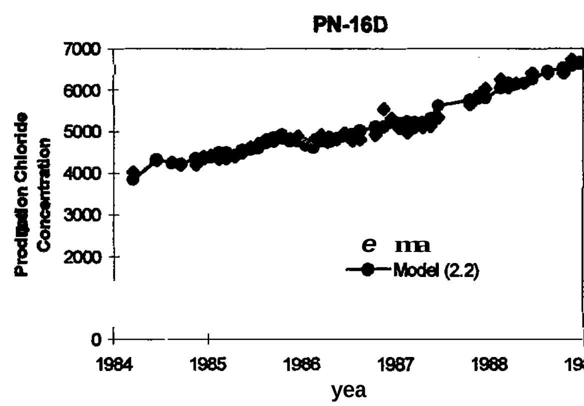


Figure 6. Predicted vs. measured chloride concentration for well PN-16D, Palinpinon-I.

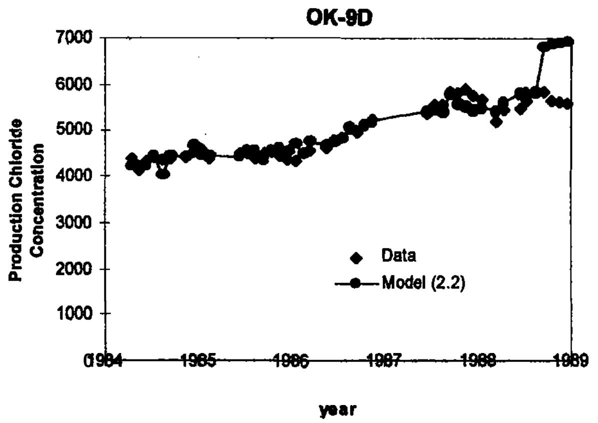


Figure 7. **Predicted vs. measured chloride concentration for well OK-9D, Palinpinon-I; model (2) coefficients calculated with last six data points excluded.**

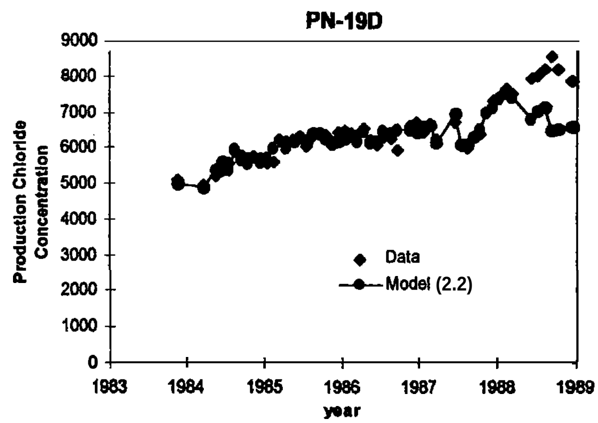


Figure 8. **Predicted vs. measured chloride Concentration for well PN-19D, Palinpinon-I; model (2) coefficients calculated with last six data points excluded.**

predicted chloride values from actual data for the Palinpinon-I data set were relatively high compared to those of Dixie Valley with a maximum of about 20%. Model (2) overpredicted the data for well OK-9D (Fig. 7) and underpredicted the chloride for well PN-19D (Fig. 8).

As with the Dixie Valley data set, the increasing deviations may be explained by the fact that the linear form of the time term does not account properly for the general trend in chloride with time. Moreover, the relatively high values of the deviations suggest that the injection rate terms contribute significantly to the model but that their contribution has not been assessed adequately.

Table 4 shows that the time term coefficients for this data set are only one to two orders of magnitude higher than the injection rate coefficients, as compared to five to eight orders of magnitude in the Dixie Valley data set. This is due to the more textured nature of the Palinpinon-I data; superimposed on the general increasing trend in chloride are marked dips and bumps. Since the dips and bumps which are accounted for by the injection rate terms are of substantial magnitude, the injection rate coefficients had high absolute values compared to those calculated for the relatively untextured Dixie Valley data set.

It is also important to note that contrary to expectation, some of the injection rate coefficients had negative values. This implies that the operation of injection wells corresponding to those negative coefficients would actually lessen the percentage of injectate beiig produced. One explanation is that the injectors with negative coefficients could be diverting the flow fiom the other injectors away from the production well. It is also possible that increased injection to the well with negative coefficient prevents inflow of natural recharge fluids with higher chloride concentration.

2.2 Improvements

The linear form of the time term in model (2) was a very convenient assumption we made despite the nonlinear trend in chloride that was readily apparent from the data. Use of the linear time term in the previous section gave us an idea of how much the time variable accounted for variations in chloride. For both Dixie Valley and Palinpinon-I the previous results showed that the time term contributed a very significant part to the chloride model as implied by the high values of time term coefficients. From these observations came the motivation to set aside convenience and identify the correct form of the time term or equivalently the general trend in chloride with time. We have chosen to use wavelet analysis for this task.

Based on the results of our analysis of the Dixie Valley data set, we concluded that the effects of individual injection rate terms on chloride were trivial because their corresponding coefficients were very small compared to the coefficient of the time term. But is the significance of a variable's contribution to the regression solution really

Table 4. Moc :l (2) Coefficients for Palinpinon-I production wells.

Model Parameter	Production Wells							
	OK-7D	OK-9D	OK-10D	PN-15D	PN-16D	PN-17D	PN-18D	PN-19D
a_0	5292.15	4290.85	3875.44	4761.51	3824.35	4917.71	4758.59	5043.92
Injection wells								
PN-IRD	4.49E+00	1.15E+00	3.31E+00	2.56E+00	1.09E+00	1.04E+01	1.81E+00	-8.37E-01
PN-2RD	-2.49E+01	-7.06E+00	-2.27E+00	-7.73E+00	-5.82E-01	-1.13E+01	-2.27E+01	1.01E+00
PN-3RD	1.37E+00	5.97E+00	-2.64E+00	1.88E+00	-3.81E+00	-3.48E+01	-1.40E+00	-9.77E+00
PN-4RD	-1.75E+00	6.28E+00	9.97E-01	-1.63E+01	-5.25E-01	1.04E+00	-6.84E+00	-3.38E+00
PN-5RD	1.38E+01	-1.46E+01	-3.48E+00	1.01E+01	-4.68E-01	2.47E+01	7.15E+00	7.57E+00
PN-6RD	-4.51E+00	8.91E-01	3.62E+00	1.97E+00	-4.20E-01	2.72E+00	-1.75E+00	-4.68E+00
PN-7RD	5.69E+00	-5.69E+00	1.15E+01	1.67E+01	7.65E+00	9.72E+00	-4.55E+00	4.97E-01
PN-8RD	2.56E+00	-1.64E-01	6.77E-01	9.05E-01	2.36E+00	-1.49E+01	4.69E+00	4.92E+00
PN-9RD	1.07E+01	5.81E+00	-1.99E+00	-9.65E+00	9.92E-01	-1.04E+01	2.15E+00	-1.61E+00
t	683.31	322.12	123.82	707.56	590.02	881.05	670.13	621.24

Cont.

Model Parameter	Production Wells							
	PN-23D	PN-24D	PN-26D	PN-27D	PN-28D	PN-29D	PN-30D	PN-31D
a_0	4434.14	3770.09	5552.84	3949.51	5843.37	5233.71	4360.31	4365.67
Injection wells								
PN-1RD	9.54E-02	-2.48E-01	7.56E+00	9.03E-01	6.16E+00	2.01E+00	-4.14E-01	1.33E+00
PN-2RD	-6.39E+00	-4.68E+00	-1.22E+01	-6.89E+00	-1.90E+01	-1.99E+01	-2.95E+00	2.01E-01
PN-3RD	4.71E+00	-5.60E+00	1.04E+00	1.54E+00	-1.06E+00	3.81E+00	3.96E+00	-2.84E+00
PN-4RD	3.02E+00	-7.31E+00	-1.15E+00	1.06E+01	-2.98E+00	-1.92E+00	8.92E-01	-5.57E+00
PN-5RD	-1.46E+00	2.85E+00	1.36E+01	-1.42E+01	5.99E+00	3.28E+00	3.24E+00	5.36E+00
PN-6RD	-1.92E+00	4.09E+00	4.53E-01	8.73E-01	-4.56E+00	-4.24E+00	-2.65E-01	-1.63E+00
PN-7RD	-3.10E+00	2.32E+01	6.82E+00	-5.24E+00	6.31E+00	-3.36E+00	-3.00E+00	-2.40E+00
PN-8RD	-1.86E+00	7.24E+00	3.87E+00	-6.15E-01	2.28E+00	I.31E+00	-3.07E+00	4.79E+00
PN-9RD	1.71E+00	9.74E+00	3.93E+00	1.11E+01	6.52E-01	9.01E+00	-1.15E-01	6.85E+00
t	475.82	661.44	685.94	569.911	708.20	836.911	189.521	677.22

reflected by the magnitude of its coefficient in the regression equation? Would the comparison of those small coefficients fi-om Dixie Valley result in as meaningful **and** valid conclusions as **those** derived from comparison of the bigger coefficients in Palinpinon-I?

Regression using model (2) gave us R^2 values that are very close to unity, signifying that the variation in chloride is almost entirely predictable **from** the model. Ironically, we also obtamed regression coefficients that could not be generalized from the early to the later portion of the **data** *set;* that **is,** we observed **poor** predictive capacity of the model. **How** could we reconcile these results?

To answer these questions we turned to **statistics** and the next section **outlines** our findings.

3.0 MULTIPLE REGRESSION

The high values of R^2 coupled with **poor** predictive capacity observed in the previous chapter lead us to doubt the validity of the results of the previous regression work. A ***survey*** of materials/texts ***on*** multivariate statistics suggested that what we had was a case of overfitting: we had a solution that provided a wonderful fit to the sample (the early portion of the **data** ***set)*** but did not generalize to the population (the entire chloride ***series).*** According to Tabachnik **(19%)** overfitting is a result of having **a** sample size that is too small relative to the number of variables in the linear model. To illustrate the point, let us take the case of bivariate regression where a straight line ($y = mx + b$) is fitted through the data points available. When calculating the parameters m and b, the square of the prediction error (graphically, the deviation of the data points 60m the 'best fit' line) is minimized. In the extreme case where only two data points are available, the minimization problem reduces to a deterministic problem; m and b are calculated exactly based on the two data points and the solution becomes perfect (and meaningless).

Tabachnick (1996) suggests the following rules of thumb for the required sample size for multiple regression:

$N \geq 50 + 8m$ *for testing* R^2

$N \geq 104 + m$ *for testing individual coefficients*

Tables 5 and 6 show that for both Dixie Valley and Palinpinon-I the number of data points used in the previous regression analysis was significantly below the required amount of data suggested by Tabachnick (1996). This asserts that the R^2 values and coefficients calculated previously were only artifacts of the data analyzed and do not generalize to extensions of the chloride series. The restriction on the amount of data required for analysis prevented us fiom using the Dixie Valley data set. Fortunately, however a larger data set fiom Palinpinon-I was made available to us by PNOC-EDC and we used this extended data set for succeeding analyses. In instances where even the extended data set was short of the required amount of data, the only possible solution was to reduce the number of terms in the model to include only those which contribute significantly to the regression solution. The procedure for choosing the important terms is discussed later in this section.

Table 5. Nuniber of data points in Dixie Valley data set.

	m = 9, model (2.2)		
Well Name	actual N	$(50 + 8m)$	$(104 + m)$
27-33	32	122	113
28-33	31	122	113
45-33	36	122	113
63-7	44	122	113
73-7	39	122	113
74-7	31	122	113
76-7	56	122	113
82-7	42	122	113
84-7	28	122	113

able 6. N nber of data points in Palinpinon-I data set

	m = 10. model (2.2)		
Well Name	actual N	$(50 + 8m)$	$(104 + m)$
OK-7D	53	130	114
OK-9D	44	130	114
OK-10D	55	130	114
PN-15D	25	130	114
PN-16D	47	130	114
PN-17D	24	130	114
PN-18D	46	130	114
PN-19D	52	130	114
PN-23D	54	130	114
PN-24D	30	130	114
PN-26D	37	130	114
PN-27D	37	130	114
PN-28D	36	130	114
PN-29D	54	130	114
PN-30D	52	130	114
PN-31D	50	130	114

It is worth noting here that the addition of the time term to model (1) pushed the regression problem towards the deterministic region as it lowered the data-to-parameter ratio; hence, the increase in R^2 values observed previously.

Another issue raised earlier was that of the suitability of inferring the contribution of injection rate terms to the regression solution fiom the sizes of the coefficients alone. According to Tabachnick (1996), interpretation of the multivariate solution based on the sizes of the coefficients alone is strictly possible only in the case where all the independent variables or IVs (injection rates and time in the case of model 2) are indeed independent fiom each other. Disregarding the interdependence between IVs, there are statistical tests that allow us to tell whether the unique contribution of an IV as represented by its coefficient is significantly different from zero or not; that is, it tells one whether to accept or reject the hypothesis that the coefficient of an IV is zero. One such test is the probability or P-test. According to this test, there is a (100-x)% probability that an IV is important to the regression solution or equivalently, its coefficient is not equal to zero if its P-value is less than or equal to x%. It is common practice to set x to 5%; hence, there is a 95% certainty that the coefficient of an IV is not equal to zero if its P-value is less than or equal to 0.05. Calculation of P-values is discussed by Bowerman and O'Connell (1990) and is done automatically by the Microsoft Excel regression ma00 that we used. Note again that the P-test does not take into account the interdependence between IVs.

Considering the need to eliminate unimportant terms m the linear model to meet the data requirement as discussed previously and taking care not to exclude IVs whose importance are masked by their interdependence with other IVs, we have proposed the following procedure fix succeeding application of multiple regression analysis:

1. To economize on IVs, temporarily *set* aside variables with P-values higher than 0.05;
2. *Also,* eliminate IVs with P-values lower than 0.05 and low values of simple correlation, *r*;
3. Inspect IVs which were eliminated in step 1 and put those with high r back to the model;
4. Perform another regression using the reduced model and interpret the results of this *regression.*

There are several possible variations to the preceding procedure and the one outlined above may not be the best but the important point to consider is the need to be aware of the possible complications that prevent straightforward interpretation of regression results based *on* the magnitude of coefficients alone.

4.0 WAVELET ANALYSIS

Recent work on the use of wavelet analysis (Jansen and Kelkar, *1997)* in analyzing production data from oil fields prompted us to look into its applicability to our problem. In the course of our investigation we found *out* that wavelet analysis had a capability that *serves* our purpose of isolating the general trend of a signal (in *our* case, chloride concentration and injection rate) from its short-term variations. Wavelet analysis allows us to examine features of a signal of any size by decomposing the signal to different *detail* levels and a coarse *approximation* The *approximation* retains the general trend with time while the details bear information on the signal's fluctuations at different time scales. Fig. 9 illustrates the concept using the chloride concentration signal from well OK-7D from Palinpinon-I. It is worth emphasizing that the approximation to OK-7D chloride shown in Fig. 9 demonstrates that the general trend in chloride is nonlinear, contrary to the assumption in model (2).

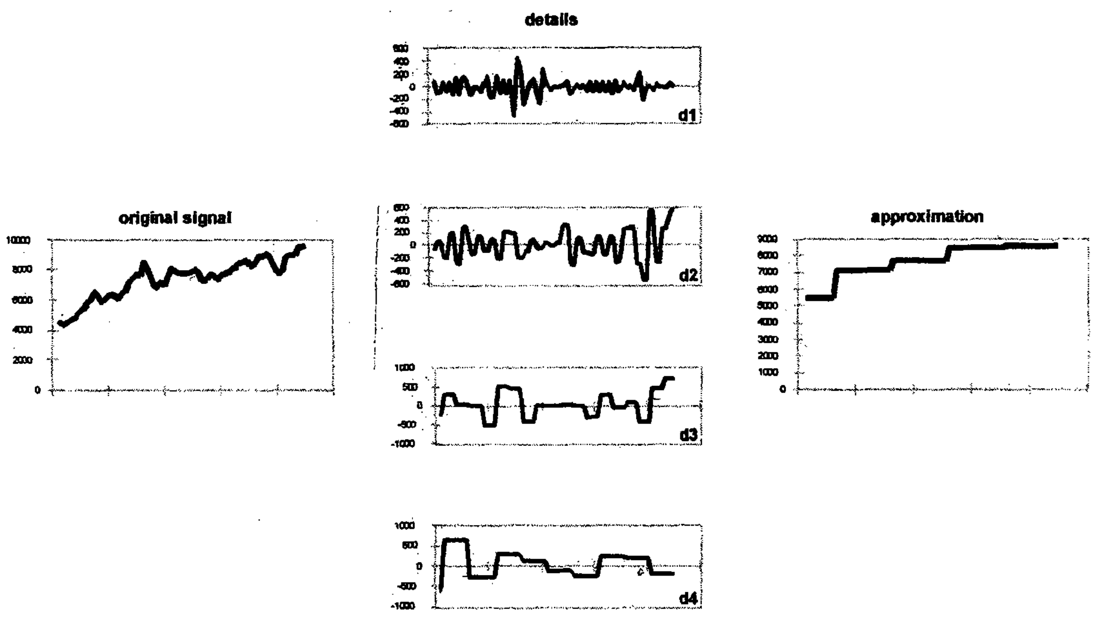


Figure 9. Wavelet decomposition: breaking a function down into a very coarse approximation, with an ordered sequence of detail functions making up the difference. (Chloride concentration, well OK-7D).

Since the effect of changes in injection rates is expected to manifest as short-term variations in reservoir chloride concentrations, it is more appropriate to analyze the *detail* functions described in the previous chapter instead of the *approximation* functions. Also, because the approximation functions isolate and retain the general trend in chloride with time, multiple regression of the details does not require a time term in the linear model. Thus, we used the following model:

$$Cl_p = a_1Q_{I1} + a_2Q_{I2} + a_3Q_{I3} + \dots \quad \dots + a_nQ_{In} \tag{5}$$

where Cl_p = chloride concentration detail in production well, P, Q_{In} = injection rate detail in injection well I_n, a_n = linear coefficient of well I_n

Comparison of coefficients obtained by using model (5) allows us to differentiate the degree of connectivity of different injectors to a given producer. Since details are deviations from local averages multiple regression using details ignores the differences in base chloride levels between producers. Regression results for different producers may therefore be intercorrelated; more specifically, the coefficients obtained may be used to compare the contributions of an injector to different production wells and consequently, to verify any conclusions drawn from the analysis against tracer test results.

The choice of modeling details over approximations was an obvious and straightforward decision. The appropriate detail level to model was less obvious, however. It seemed reasonable at first to assume that the best choice is the one that will give the highest R^2 value. Investigation of the R^2 values obtained from modeling the chloride details of OK-7D invalidated that assumption. Table 7 shows that at level 4 the regression coefficient becomes unity signifying a perfect correlation; and, correlation at succeeding levels remain perfect. As the decomposition level goes up, the detail will have longer time intervals with constant values. This effectively reduces the amount of data to be modeled and results to a perfect, meaningless correlation. The choice is thus narrowed down to levels 1, 2, and 3.

Table 7. R^2 values for multiple regression of OK-7D chloride detail.

Detail Level	R^2
1	0.202
2	0.530
3	0.853
4	1.000
5	1.000

Inspection of injection and chloride details showed that the correspondence between changes in chloride concentration and changes in injection rate is more readily visible at level 3. In Fig. 10 the level 3 details of injection wells PN-6RD and PN-9RD closely follow the detail of OK-7D chloride during intervals when injection to these wells are high. Some degree of correspondence at levels 1 and 2 is also apparent from Fig. 11 albeit not quite as obviously as in level 3. Thus, we have decided to analyze all three levels of detail.

As was done in previous analyses, the chloride data that were recorded at irregular time intervals were converted to monthly average values to put them in the same time basis as the injection rates. Since wavelet analysis requires that data be available in the entire time interval being analyzed, missing chloride data were linearly interpolated. Interpolation was done over maximum intervals of six months and only when no drastic fluctuations were apparent within six months of the interval where interpolation was to be done. Where interpolation was not possible, we analyzed only the longest continuous portion of the data series.

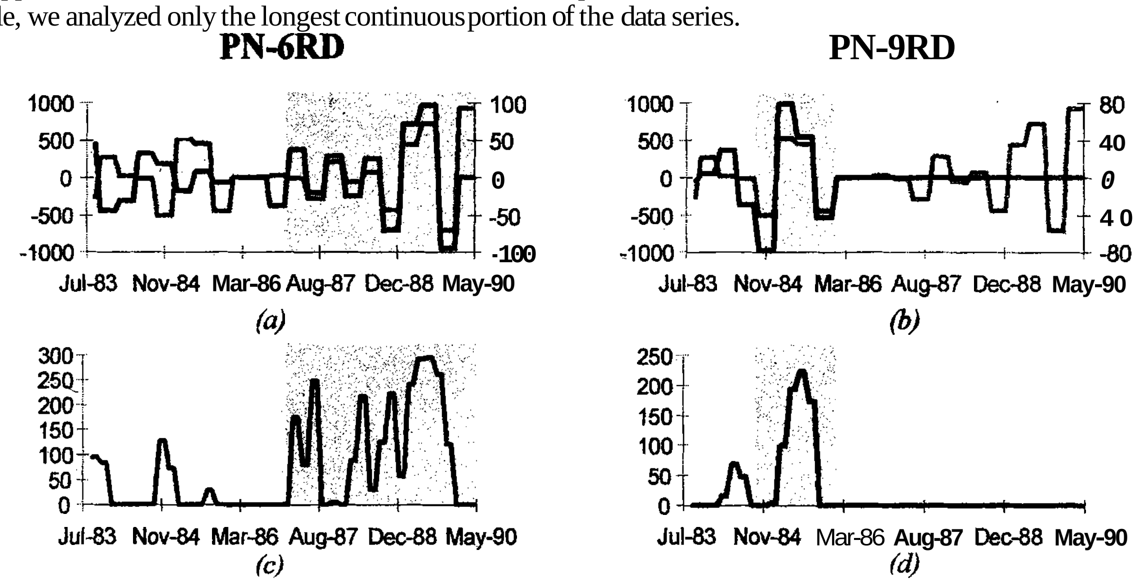


Figure 10. (a) Level 3 detail of OK-7D chloride - light line; level 3 detail of PN-6RD injection rate - dark line. (b) Level 3 detail of OK-7D chloride - light line; level 3 detail of PN-9RD injection rate - dark line. (c) PN-6RD injection rate. (d) PN-9RD injection rate.

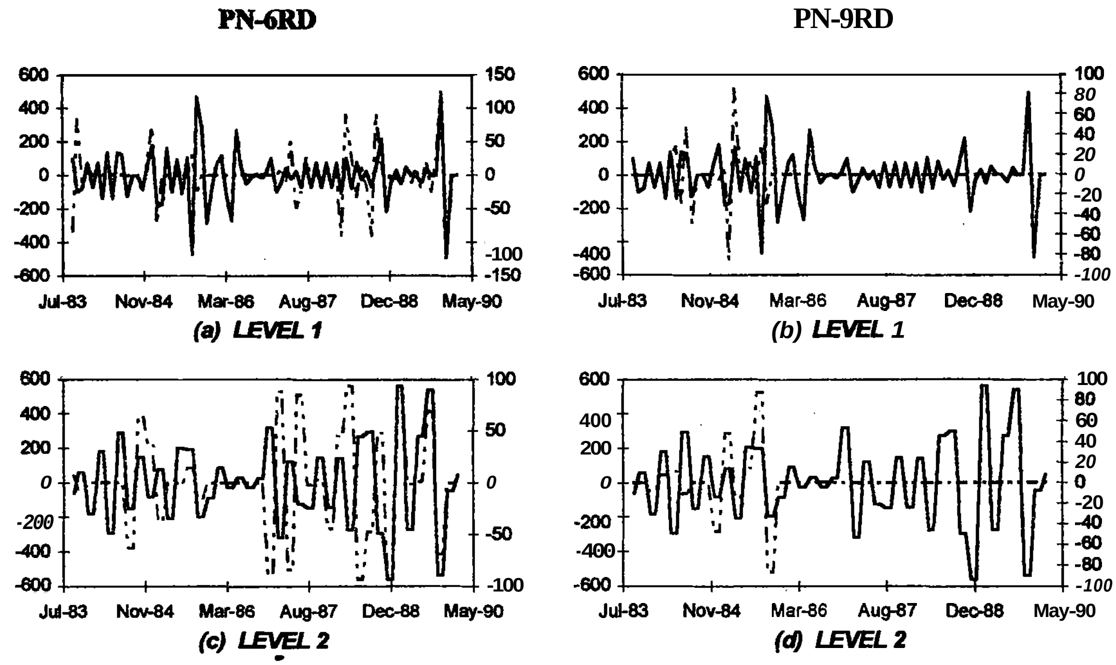


Figure 11. **(a) Level 1 detail of OK-7D chloride - solid l i e ; level 1 detail** of PN-6RD **injection** rate **-dashed line.**
(b) Level 1 detail of **OK-7D chloride - solid h e ; level 1 detail of** PN-9RD **injection** rate - dashed **line.**
(c) Level 2 detail of OK-7D chloride - solid line; level 2 detail of PN-6RD **injection** rate -dashed **line.**
(d) Level 2 detail of OK-7D chloride - solid h e ; level 2 detail of PN-9RD **injection** rate - dashed **line.**

When taking wavelet transforms of discrete **data,** the algorithms used require that the data set **size** be a power of two. A **common** way to precondition the **data** when this is not true is to "pad with zeroes," that is, to increase the size of the **data** ***set*** to the next larger power of ***two.*** Although **this** is a reasonable approach, it is problematic in that it "dilutes" the signal near the end of the original **data set** since wavelet coefficients will have zeroes averaged into their computation (Ogden, **1997).** We believe that Matlab, the program we used to take the wavelet transform of our **data,** practices **this data** preprocessing procedure **as** evidenced by the inaccurate reconstruction near the end of the **data series.** It was therefore necessary to truncate the detail component functions **used** in regression analysis to eliminate the end effects of padding with zeroes. Plots of wavelet component functions retain the end effects but in the analysis, **3** to **6 data** points were eliminated fiom the details **series.**

In about **1990,** the bulk of the injection in Palinpinon-I was moved farther away **fiom** the production sector resulting in some injection wells being shut off and new ones **being** operated. **Thus,** based on operating time, injection wells in Palinpinon-I *can* be grouped into three: those which operated between **1983** and **1990,** those which started to inject in **1990** and are **still** injecting, and those which have **been** injecting since **1983** and are **still** injecting. It is logical to assume that regression analysis will be best able to **assess** the degree of contribution of injection wells if all the wells are operating during the time interval over which the regression is done. So, where possible, the chloride series was divided into two time intervals, **1983** to **1990 and 1990** to present; and, regression analysis included only wells which were operating during those periods.

Regression was done for levels **1, 2,** and 3 of the detail functions using the procedure outlined earlier. In some ***cases,*** that procedure had to be applied repeatedly until the number of terms left in the linear model is such that the **data** size requirement is met **(or** almost met). **When** eliminating injection well terms that had small r values, care was taken not to remove wells which operated only for very short **periods** of time (the small r values in these ***cases*** are unnatural effects of the **scarcity** of correlatable **data).**

4.1 Checking Results Against Tracer Test Data

Two **sets** of radioactive *tracer* test results were available for comparison with results of our analysis: that of the test conducted ***on*** well **PN-9RD** and one ***on*** OK-12RD. Both ***sets*** were reported by **Macario (1991)** and are reproduced in Table **8. Macario (1991) defined mean** transit **time as** the time it **takes** for half of the **tracer return** to reach the production well. Assuming that the **mean** transit time **measures** the degree of connectivity **between** the **injector** tested **and** a **producer** (lower transit times **corresponding** to stronger **connections)** Table 8 lists the production wells in **order** of *decreasing* connectivity with the injector. Correspondingly, Tables 9 and **10** lists the wells affected by OK-12RD **and** PN-9RD, respectively, in the order of **decreasing coefficients** based ***on*** regression ***on*** all **three** wavelet **detail** levels.

Table 10 **shows, with the** exception of one well, that all wells affected by PN-9RD had positive coefficients. Comparison of Table 10 with Table 8 **shows** that tracer return was indeed monitored in all wells **shown** by

Table **8. Radioactive tracer test results for PN-9RD and OK-12RD.**

Monitored Well*	**Mean Transit Time, days**
PN-9RD Tracer Test	
OK-7	5.4
PN-26	13
PN-28	14
PN-29D	15.4
PN-30D	15.7
PN-23	15.8
PN-16D	16
PN-19D	16
PN-31D	16
PN-18D	17.2
OK-9D	monitored. no return
OK-12RD Tracer Test	
PN-15D	7.3
OK-10D	13.8
OK-7D	14.6
PN-29D	monitored. no return

**** Only wells which have chloride data are reported here.***

Table 9. Regression results for OK-12RD

OK-12RD

Affected Well	**Coefficient**
PN-23D	2.46
PN-29D	-4.05
PN-31D	-10.82
PN-15D	125.27
PN-16D	-7.40
PN-29D	-3.34
PN-30D	6.15

Table 10. **Regression results for PN-9RD.**

PN-9RD

Affected Well	**Coefficient**
d1	
PN-30D	5.74
PN-29D	3.99
PN-16D	1.47
d2	
PN-19D	4.87
PN-18D	4.06
OK-7D	2.96
PN-16D	2.02
PN-29D	-11.65
OK-7D	9.40
PN-29D	1.83
PN-16D	0.92
PN-23D	0.43

regression analysis to be affected by well PN-9RD, including PN-29D which had a negative coefficient. More importantly it shows that the order of the strength of connection **between** PN-9RD and the wells monitored in the tracer test was most closely mimicked by the results of regression ***on*** detail level 3 **with** OK-7D showing the strongest connection to PN-9RD and PN-29D, **PN-16D,** and PN-23D displaying **connections** of about the same strength.

On the other hand, comparison of Table 9 with Table 8 **shows** that tracer **return** was observed in **two** of the seven wells shown by regression analysis to be affected by well OK-12RD. **Four** of the seven wells were **not** monitored during the tracer test. As with PN-9RD, the well which is most connected to OK-12RD based **on** the tracer test had the highest coefficient at level 3 regression.

Based **on these** observations we have concluded that regression analysis of details at level 3 best **assesses** the degree of connectivity between wells: high positive coefficients correspond to **strong** connections **and,** negative and low positive coefficients correspond to weak connections.

Harper and Jordan (1985) reported the **following** observation: from May 1984 to ***October*** 1984 a large increase in reservoir chloride occurred in production wells **PN-19D, 23D,** 29D, 31D, OK-7D and OK-9D when reinjection was shifted to the wells PN-7RD and PN-8RD. This observation matches the results of level 3 detail analysis for well PN-8RD **as** outlined in Table 11: **OK-7D,** PN-19D, and PN-31D were all found to be strongly connected **with** PN-8RD. PN-23D, -29D, and OK-9D may have **been** receiving reinjection returns **fiom** OK-7D but no injection rate **data fiom** OK-7D was available to allow verification with regression results.

Table 11. Level 3 regression for PN-8RD

PN-8RD

Affected Well	Coefficient
d3	
OK-7D	3.14
PN-16D	0.64
PN-18D	2.76
PN-19D	4.93
PN-30D	-1.36
PN-31D	10.49

On the other hand, Amistoso and Orizonte (1997) reported that OK-10D and PN-20D experienced enhanced ***steam*** flows which they attributed to reinjection fluids intruding into the production sector at deeper levels. **They** cited the wells **TC-2RD,** TC-4RD, PN-3RD and PN-5RD to be wells that are providing pressure **support** to the reservoir due to deep injection but attributed the **enhanced *steam*** flow in OK-10D and PN-20D **to** TC-2RD **and** TC-4RD, specifically. Regression analysis results for *these* wells (Table 12), however, show that OK-10D is not affected by **TC-2RD;** rather it is affected by PN-1RD, **PN-2RD,** and PN-3RD between 1986 **and** 1990 and by **PN-3RD,** TC-3R, N3 **and** OK-3R between 1990 and 19%. It is worth noting that the effect of PN-3RD ***on*** OK-10D was found to be consistent between the intervals 1986-1990 and 1990–1996 **as** reflected by close r values for the **two** periods (-0.79 and -0.77). The large positive coefficient of well N3 is ***suspect*** however **as** it conflicts with its negative r value. PN-20D was also analyzed to be affected by PN-3RD. The effect of TC-2RD and TC-4RD on PN-20D could **not be** ascertained **fiom** regression analysis due to insufficient chloride data from PN-20D after 1990.

Pamatian (1997) reported that reinjection fluid **fiom** TC-2RD neutralized the fluid **acidity m** wells OK-10D and PN-13D. **Again,** the effect of TC-2RD ***on*** OK-10D was **not** substantiated **by** regression results but **its** effect ***on*** PN-13D was (Table 13). Again, terms **with** conflicting r and coefficient **signs** posed interpretation problems.

5.0 CONCLUSIONS AND RECOMMENDATIONS

Based *on* the results of regression analysis of chloride and injection rate **data** fiom Dixie **Valley,** we have concluded **that** multilinear modeling is **not** suitable **for** analyzing data ***sets*** which lack sufficient time variability **or** "texture".

A closer look at multiple regression techniques showed that what seemed to be highly encouraging results ***(high*** R^2 values) **fiom prior multilinear** modeling were **but** effects of the ***scarcity*** of **data** used in the correlation; hence, no meaningful physical interpretation may be drawn from them. Moreover, it showed that care should be **taken** not to base the interpretation of multiple regression results **on** straight-forward comparison of coefficients alone.

Wavelet analysis provided more useful results. Qualitative field observations and tracer test data agreed best with the results of regression on level 3 detail of chloride concentration and injection rates in Palinpinon-I: wells identified by tracer tests to be strongly connected had high positive coefficients and weak connections were indicated by negative and low positive coefficients at level 3 regression. This suggests that producer-injector interactions are best detected by correlating changes in chloride concentration over periods of four months (corresponding to level 3 resolution) with corresponding four-month fluctuations in injection rates. While the good correlation at such a relatively low level of time resolution may be explained as the result of the natural dispersion of chloride and injection rate signals as they propagate through the reservoir, it is also possible that this is due to the loss of information brought about by the use of monthly averaged data values in the analysis. It is therefore recommended that both chloride and injection rate data be recorded more frequently and the analysis be done on this larger data set. It is also possible that the Haar wavelet that was used in signal decomposition was too coarse in that it contributed to the loss of texture in the data. Investigation of the effects of using smoother wavelets is also recommended.

Table 12. Level 3 regression statistics for OK-10D and PN-20D.

OK-10D (1986-1990)

d3

Regression Statistics	
Multiple R	0.839369312
R Square	0.704540842
Standard Error	79.20406691
Observations	48

	Coefficients	*r (simple)*	*Standard Error*	*t Stat*	*P-value*
pn1rd	1.0860989	0.742898166	0.432894148	2.508924886	0.01577982
pn2rd	-3.71087522	-0.72596153	1.775342121	-2.09023105	0.04227759
pn3rd	-4.4114713	-0.78948113	1.783793986	-2.4730834	0.01723762

OK-10D (1990-1996)

d3

Regression Statistics	
Multiple R	0.835160095
R Square	0.697492385
Standard Error	150.1670809
Observations	80

	Coefficients	*r (simple)*	*Standard Error*	*t Stat*	*P-value*
pn3rd	11.4979076	-0.77103302	1.485833395	-7.7383559	3.4832E-11
tc3r	1.704839524	0.522611064	0.404330736	4.216448003	6.7759E-05
n3	30.87773863	-0.65393407	6.498963063	4.751179278	9.3559E-06
ok3r	-798184387	-0.36088849	2.134566707	-3.44887229	0.00092135

PN-20D (1983-1989)*

d3

Regression Statistics	
Multiple R	0.635764265
R Square	0.404196201
Standard Error	495.2038572
Observations	80

	Coefficients	*r (simple)*	*Standard Error*	*t Stat*	*P-value*
pn1rd	10.21445053	0.346132276	1.741845857	5.864152953	1.0654E-07
pn3rd	12.28448348	0.357197703	2.153358367	5.704802167	2.062E-07
pn6rd	4.493768629	0.105136293	1.637086862	2.744979074	0.00752785

****No chloride data was available from 1990 to 1993 and remaining data points were not sufficient for analysis.***

Table 13. Level 3 regression statistics for PN-13D

PN-13D (1990-1996)

d3

Regression Statistics	
Multiple R	0.962568329
R Square	0.926537788
Standard Error	74.57273176
Observations	70

	Coefficients	*r (simple)*	*Standard Error*	*t Stat*	*P-value*
tc2rd	4.826890146	0.403207897	0.459363951	10.5077687	1.4452E-15
tc3r	1.989170863	0.37912479	0.16882133	11.78269865	1.1114E-17
tc4r	52.80156771	-0.47305009	4.657792303	11.33617909	5.9844E-17
ml1rd	-191.386684	-0.73129226	16.95558741	-112875288	7.1995E-17
n3	11.67741779	-0.76944945	1.849731659	6.313033424	2.9328E-08
ok3r	27.49754181	-0.47241942	2.227450499	12.34484978	1.3816E-18

Emphasis is also placed on the need for continuous data measurements when doing wavelet analysis. Highly intermittent measurements result in data loss: since it is considered safe to interpolate only over short periods of time, the lack of data over long time intervals forces one to disregard the data collected prior to such periods when doing the analysis.

Another possible improvement to consider in future regression analyses is to take into account possible nonlinearity in the variation of chloride with injection rates. While nonlinearity does not invalidate the analysis, it certainly weakens it as the relationship between chloride concentration and injection rates is not completely captured by the coefficients of the linear model. Although regression analysis uses a linear model, effects of nonlinearity in the variation of chloride with injection rates may be incorporated into the model by using nonlinear terms: the model is kept linear even though the individual terms are not. Results of this modified analysis will be more difficult to interpret however, because the strength of interaction between producers and injectors will be measured not only by the magnitude of the coefficients but also by the exponent of each term.

6.0 NOMENCLATURE

a_0	=	a constant associated with local initial chloride concentration
a_n	=	linear coefficient of well I_n
b	=	linear time term coefficient
Cl_{In}	=	chloride concentration in injector well, In
Cl_P	=	chloride concentration/chloride concentration detail in production well, P
m	=	number of predictors
N	=	number of data points
Q_{In}	=	mass flow rate/mass flow rate detail to injection well, In
r	=	simple regression coefficient
R^2	=	multiple regression coefficient
S	=	standard deviation
SS_{reg}	=	sum of squared deviations of predicted Y from the mean
SS_Y	=	sum of squared deviations of Y from the mean
Y	=	dependent variable being modeled
Y'	=	predicted values of Y
$\bar{Y}$	=	average value of Y

ACKNOWLEDGMENT

This work was funded by the U.S. Department of Energy under grant number DE-FG07-95ID13370. We gratefully appreciate the assistance of PNOC-EDC in making the Palinpinon data available to us, and also the assistance of Oxbow Geothermal who gave access to data fiom Dixie Valley.

REFERENCES

Amistoso, A.E. and RG. Orizonte, Jr., 1997. "Reservoir Response to Full Load Operation Palinpinon Production Field Valencia, Negros Oriental, Philippines", Proceedings of 18th PNOC-EDC Geothermal Conference.

Bowman, B.L. and RT. O'Connell, 1990. Linear Statistical Models: *An* Applied Approach, 2nd Edition. PWS-KENT Publishing Company, Boston, MA.

Harper, RT. and O.T. Jordan, 1985. "Geochemical Changes in Response to Production and Reinjection for Palinpinon-I Geothermal Field, Negros Oriental, Philippines", Proceedings of 7th New Zealand Geothermal Workshop.

Jansen, F.E. and M.G. Kelkar, 1997. "Application of Wavelets to Production Data in Describing Inter-Well Relationships", paper SPE 38876, presented at 1997 SPE Annual Technical Conference and Exhibition.

Macario, M. E., 1991. "Optimizing Reinjection Strategy in Palinpinon, Philippines Based *on* Chloride Data", M.S. Thesis, Stanford University, Stanford, CA.

Ogden, RT., 1997. Essential Wavelets for Statistical Applications and Data Analysis. Birkhauser Boston, Cambridge, MA.

Pamatian, P.I., 1997. "changes in the Apparent Piezometric Levels in the Palinpinon Field as a *Response to* Twelve Years of Exploitation", Proceedings of 18th PNOC-EDC Geothermal Conference.

Tabachnick, B.G. and L.S. Fidell, 1996. Using Multivariate Statistics. HarperCollins College Publishers, New York.