PERGAMON

GEOTHERMICS

Geothermics 30 (2001) 591–616

www.elsevier.com/locate/geothermics

Inferring injection returns from chloride monitoring data

Ma. Michelle Sullera, Roland N. Horne*

Stanford Geothermal Program, School of Earth Sciences, Stanford University, Stanford, CA 94305-2220, USA

Received 7 June 1999; accepted 1 November 1999

Abstract

Reservoir chloride concentration and injection rate data are used to examine injector-producer connectivity in two reservoirs, Dixie Valley, Nevada and Palinpinon-I, Philippines. General trends of chloride and injection rate with time are isolated from their respective short-term variations using the wavelet transformation approach. Multiple regression techniques are 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. Communication between specific injector-producer pairs, as implied by analysis of data from Palinpinon-I, was also verified by comparison with tracer test data and qualitative field observations. Results of analysis of data from Dixie Valley demonstrate that multilinear modeling is not suitable for analyzing data sets that lack sufficient time variability. In contrast, adequate time variability is observed in data from Palinpinon-I, and qualitative field observations and tracer test data agreed best with the results of regression on changes in chloride concentration over four-month periods (wavelet detail level 3). Improvements in the analysis could result from increased data collection frequency of both chloride and injection rate as well as accounting for the nonlinearity of chloride with injection rate.

Keywords: Tracer; Injection; Regression analysis; Wavelet

1. 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.

* Corresponding author. Tel.: +1-650-723-9595; fax: +1-650-725-2099.
E-mail addresses: horne@stanford.edu (R.N. Horne), sullera@stanfordalumni.org (M.M. Sullera).

0375-6505/01/$20.00
PII: S0375-6505(01)00016-5

Nomenclature

a_0	a constant associated with local initial chloride concentration
a_n	linear coefficient of well I_n
b	linear time term coefficient
t	time elapsed since first recorded chloride concentration data
Cl_{In}	chloride concentration in injector well, In
Cl_P	chloride concentration or chloride concentration detail in production well, P
m	number of predictors
N	number of data points
Q_{In}	mass flow rate or 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
$\overline{Y}$	average value of Y

On the other hand, there are substances occurring naturally in the reservoir that can behave as tracers. One such substance is chloride. In the Palinpinon-I 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-I are 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. We may therefore be able to extract the same, if not more, information from chloride data as we could from traditional tracer tests and at lower cost.

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. The permeability of these paths was then ranked by quantifying the degree of connectivity between injectors and producers.

2. 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 from Palinpinon-I. Of the models tested by Macario (1991), the linear combination model came closest to reproducing field observations. In the first phase of this project, therefore, we chose to expand on that model and test it further.

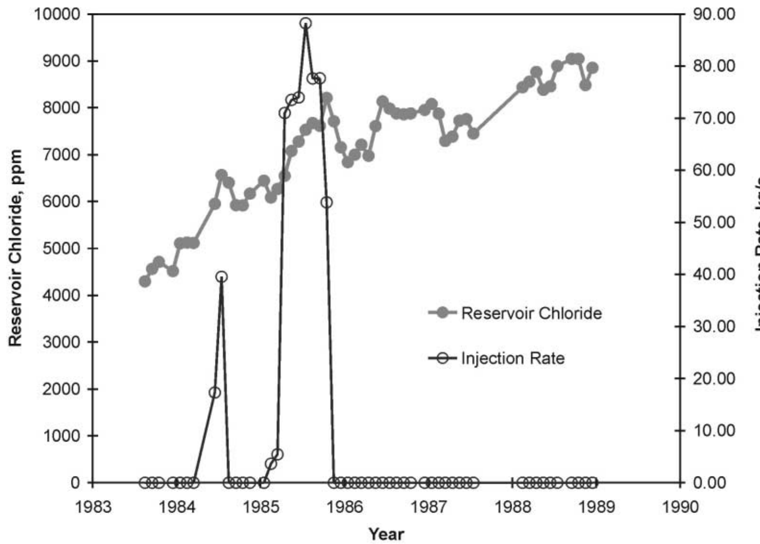

Fig. 1. Example of chloride and injection data from Palinpinon-I.

The following is the original linear combination model proposed by Macario (1991):

$$Cl_{\mathrm{p}} = a_0 + a_1 Q_{\mathrm{I1}} + a_2 Q_{\mathrm{I2}} + a_3 Q_{\mathrm{I3}} + \dots\dots + a_{\mathrm{n}} Q_{\mathrm{In}} \tag{1}$$

Based on this model, the strength of the connection between the modeled producer P and an injector $\mathrm{I_i}$ 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 from 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:

$$Cl_{\mathrm{p}} = a_0 + a_1 Q_{\mathrm{I1}} + a_2 Q_{\mathrm{I2}} + a_3 Q_{\mathrm{I3}} + \dots\dots + a_{\mathrm{n}} Q_{\mathrm{In}} + bt \tag{2}$$

More than anything, it was simplicity that guided our choice of the form (linear) of the time term. As the reservoir pressure drops with production, more steam is produced and the chloride salts that are left in the brine get more concentrated. Solution saturation limits could then 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 time will be addressed in more detail in a later section.

We also hypothesized that the effect of the reinjection return 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):

$$Cl_p = a_0 + a_1 Q_{I1}\mathrm{Cl}_{I1} + a_2 Q_{I2}\mathrm{Cl}_{I2} + a_3 Q_{I3} Cl_{I3} + \dots\dots + a_n Q_{In} Cl_{In} + bt \tag{3}$$

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 chloride 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 that best fits 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 from 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.

2.1.1. Dixie Valley case

At Dixie Valley, injection rates were recorded daily while chloride concentrations were measured much less frequently. Only simultaneously measured chloride concentrations and injection rates were 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 a 2–35% increase in R^2 while inclusion of injectate chloride concentration in model (3) did not result in any significant change in R^2 values. The small effect of the injectate chloride term is due to its nearly constant value (the injectate chloride concentration is strongly a function of the separator pressure, which is controlled so as to remain more or less constant). Fig. 2 shows the effect of a 35% difference in R^2 on the quality of data fit for well 84-7. It also illustrates the very minor effect that the injectate chloride term

Table 1
R^2 values for Dixie Valley wells

Well name	R^2, Model (1)	R^2, Model (2)	R^2, Model (3)
27-33	0.917	0.963	0.965
28-33	0.852	0.936	0.940
45-33	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

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 from the regression. Model (2) was then used to predict these values using the coefficients calculated on the basis of the truncated data set. Fig. 3 plots the results of the truncated series analysis for well 27-33, which had a 9% maximum deviation of predicted chloride from 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. In fact, the coefficient of the time term is several orders of magnitude (3 to 5, even 8 orders!) greater than the injection rate coefficients. This discrepancy was enough to render the injection rate terms trivial and excluding an injection rate term from model (2) resulted in only tiny changes in the quality of the data fit. Fig. 5 shows the chloride match for well 74-7 when the chloride is predicted using model (2) but with the injection rate term corresponding to injector INJ5218 excluded. That the 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 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 at Dixie Valley.

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 that lack temporal variability.

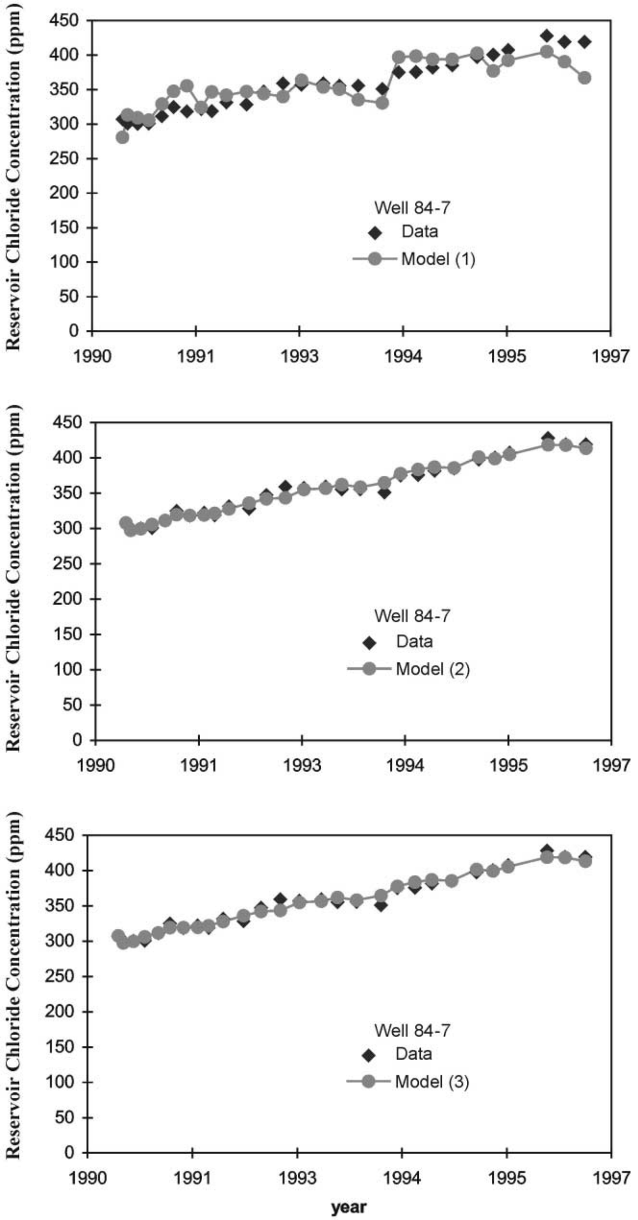

Fig. 2. Predicted vs. measured chloride concentration for well 84-7, Dixie Valley, using models (1), (2), and (3).

2.1.2. Palinpinon-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

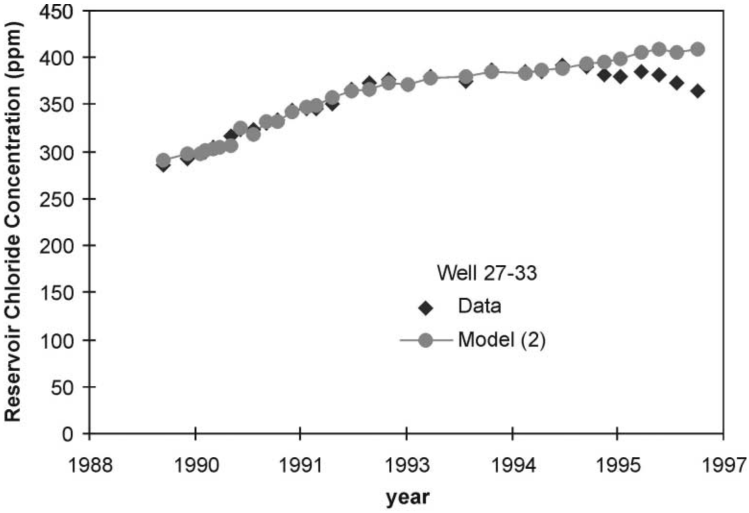

Fig. 3. Predicted vs. measured chloride concentration for well 27-33, Dixie Valley. Model (2) coefficients were calculated with the last six data points excluded.

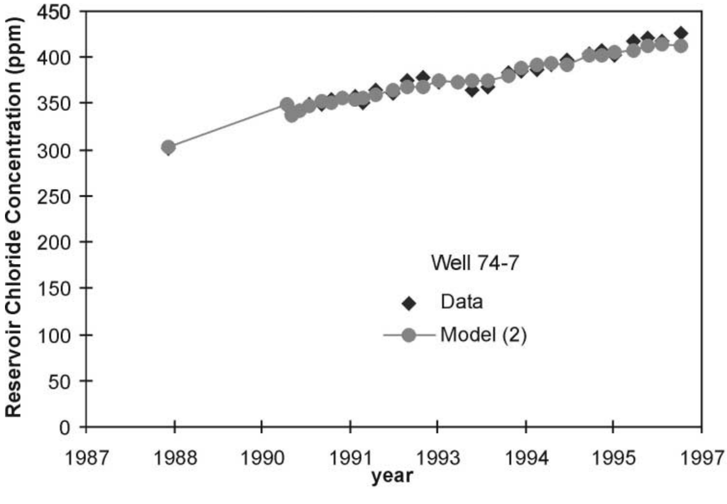

Fig. 4. Predicted vs. measured chloride concentration for well 74-7, Dixie Valley. Model (2) coefficients were calculated with the last six data points excluded.

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 from 1983 to 1989 was initially available for use in the preliminary 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).

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

	Production wells								
Model parameter	27-33	28-33	45-33	63-7	73-7	74-7	76-7	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	-6.61E-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
INJSWL1	-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
INJSWL3	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
b	8.48	10.37	7.61	2.30	17.44	12.95	4.82	26.25	19.04

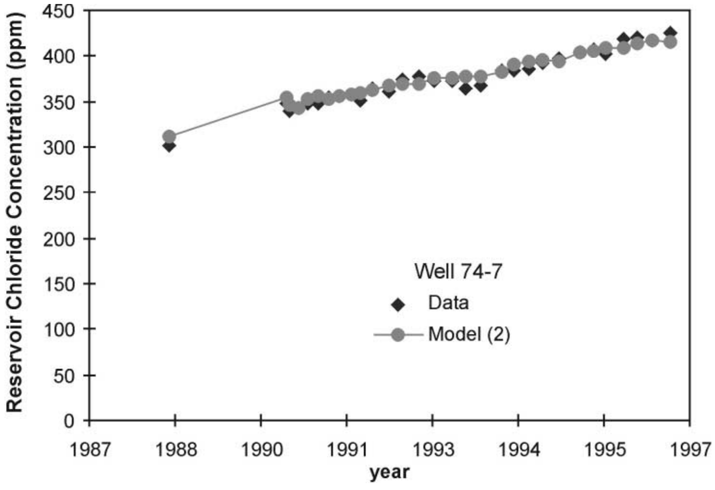

Fig. 5. Predicted vs. measured chloride concentration for well 74-7, Dixie Valley. Predicted values were calculated using model (2) with one injection rate term excluded.

Table 3
R^2 values for Palinpinon-I wells

Well name	R^2, Model (1)	R^2, Model (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

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, the last six points in the chloride series were not considered in the calculation of the linear coefficients in the subsequent regression using model (2). The excluded chloride values were then predicted using the coefficients calculated based on the truncated

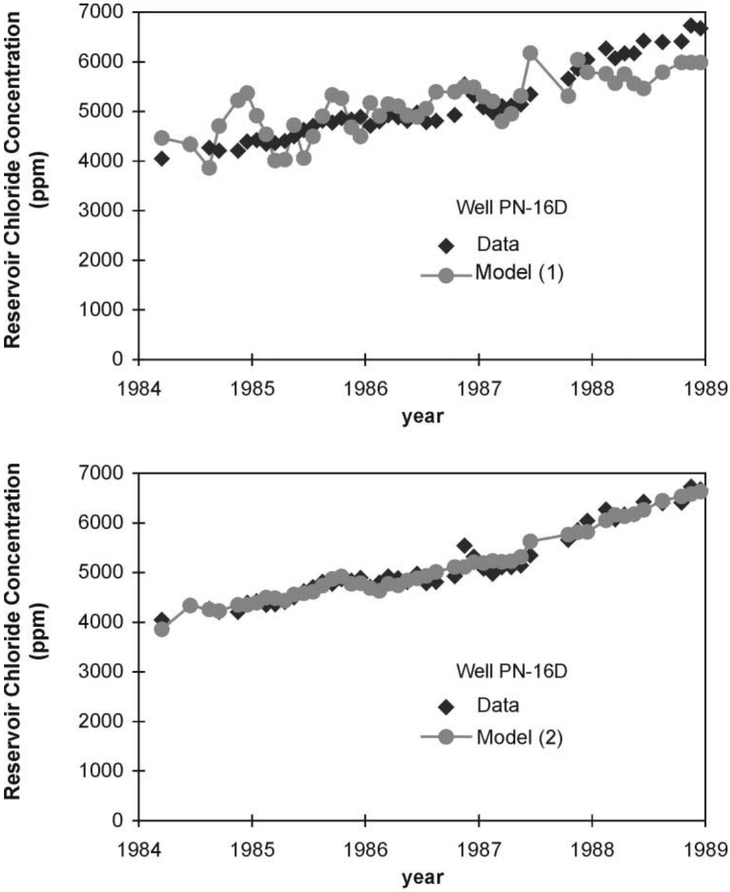

Fig. 6. Predicted vs. measured chloride concentration for well PN-16D, Palinpinon-I, using models (1) and (2).

data set. Deviations of predicted chloride values from actual data for the Palinpinon-I data set were relatively high (with a maximum of about 20%) compared to those of Dixie Valley. Model (2) overpredicted the chloride for well OK-9D (Fig. 7) and underpredicted it 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, which are characterized by marked dips

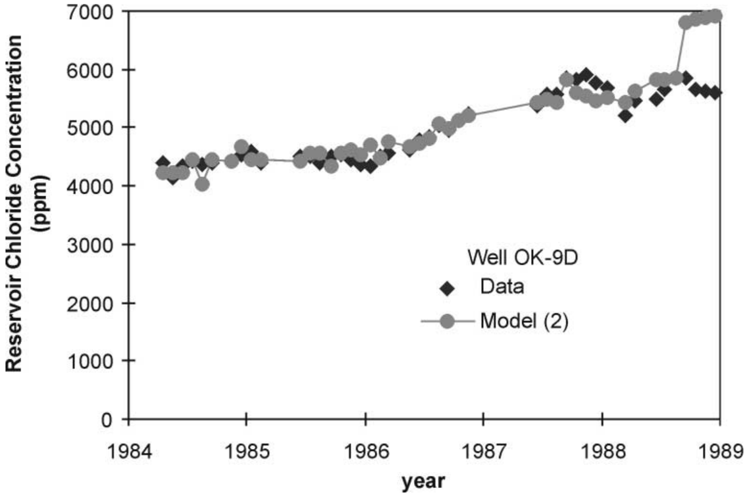

Fig. 7. Predicted vs. measured chloride concentration for well OK-9D, Palinpinon-I. Model (2) coefficients were calculated with the last six data points excluded.

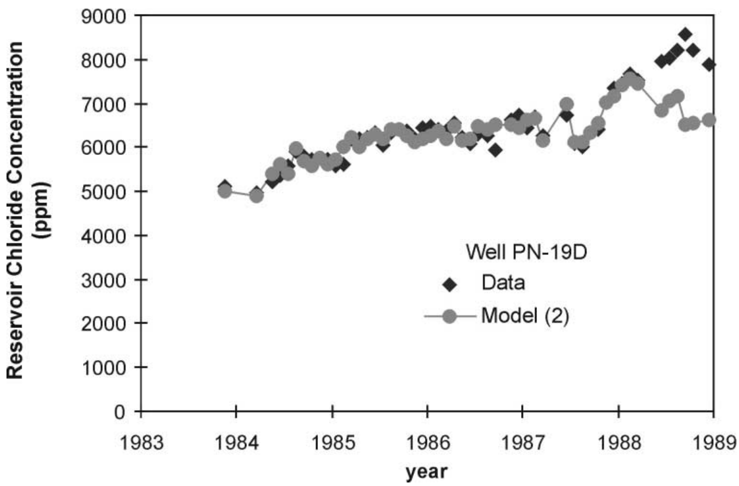

Fig. 8. Predicted vs. measured chloride concentration for well PN-19D, Palinpinon-I. Model (2) coefficients were calculated with the last six data points excluded.

and bumps that are superimposed on the general increasing trend in chloride. Since the dips and bumps that 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 being produced. One explanation is that the injectors with negative coefficients

Table 4
Model (2) coefficients for Palinpinon-I production wells

	Production wells							
Model parameter	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	5043.93
Injection wells								
PN-1RD	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
b	683.31	322.12	123.82	707.56	590.02	881.05	670.13	621.24
Model parameter	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	1.31E+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
b	475.82	661.45	685.95	569.91	708.20	836.91	189.52	677.22

could be diverting the flow from 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 an indication of how much the time variable accounted for variations in chloride. For both Dixie Valley and Palinpinon-I, the high values of time term coefficients showed that the time term contributed very significantly to the chloride model. These observations provided the motivation to identify the correct form of the time term or, equivalently, the general trend in chloride with time. Wavelet analysis was chosen for this task.

Based on the results of the analysis of the Dixie Valley data set, it was 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 reflected by the magnitude of its coefficient in the regression equation? Would the comparison of those small coefficients from 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, model (2) yielded regression coefficients that could not be generalized from the early to the later portion of the data set; that is, the model had poor predictive capability. Possible reasons for these discrepancies are discussed in the following section.

3. Multiple regression

The high values of R^2 coupled with the poor predictive capacity of model (2) suggested that this was due to having a sample size that is too small relative to the number of variables in the linear model (e.g. overfitting, Tabachnick and Fidell, 1996). To illustrate the point, consider the case of bivariate regression where a straight line ($y = mx + b$) is fitted through the data points. When calculating the parameters m and b, the square of the prediction error (graphically, the deviation of the data points from 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 and Fidell (1996) suggest the following generalizations:

$$\begin{aligned} &N \geqslant 50 + 8m \quad \text{for testing } R^2 \\ &N \geqslant 104 + m \quad \text{for testing individual coefficients} \end{aligned} \tag{4}$$

Tables 5 and 6 show that for both Dixie Valley and Palinpinon-I, too few data points were used in the regression analysis. This means 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 prevented analysis of the Dixie Valley data set. Fortunately, a larger data set from Palinpinon-I was made available by PNOC-EDC. In instances where even

Table 5
Number of data points in Dixie Valley data set

$m=9$, model (2)[a]			
Well name	Actual N[a]	$(50+8\ m)$	$(104+m)$
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

[a] m is the number of IVs and N is the number of data points included in the analysis.

Table 6
Number of data points in Palinpinon-I data set[a]

$m=10$, model (2)[a]			
Well name	Actual N[a]	$(50+8\ m)$	$(104+m)$
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

[a] m is the number of IVs and N is the number of data points included in the analysis.

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.

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. This accounts for the observed increase in R^2 values described above.

A second issue was the suitability of inferring the contribution of injection rate terms to the regression solution from the sizes of the coefficients alone. According to Tabachnick and Fidell (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 (IVs; injection rates and time in the case of model 2) are independent of each other. Disregarding the issue of 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 macro 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 in the linear model to meet the data requirement as discussed previously and taking care not to exclude IVs whose importance is masked by their interdependence with other IVs, we have proposed the following procedure for subsequent applications 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 that 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. Wavelet analysis

An alternative approach is to use wavelet analysis, which has recently been applied to the analysis of production data from oil fields (Jansen and Kelkar, 1997). Wavelet analysis allows examination of the 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).

Since the effect of changing injection rates is expected to manifest itself as short-term variations in reservoir chloride concentrations, analysis of the *detail* functions instead of the *approximation* functions is more appropriate. 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:

$$\mathrm{Cl_p} = a_1 Q_{\mathrm{I1}} + a_2 Q_{\mathrm{I2}} + a_3 Q_{\mathrm{I3}} + \ldots\ldots + a_{\mathrm{n}} Q_{\mathrm{In}} \tag{5}$$

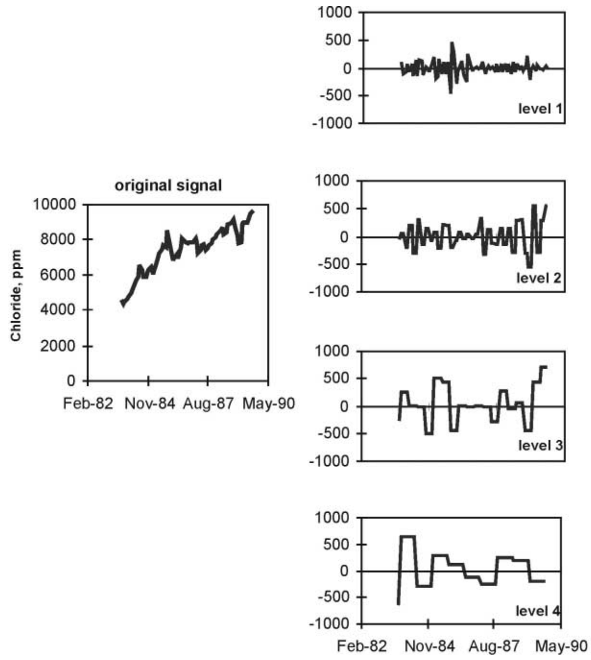

Fig. 9. Wavelet decomposition: breaking a function down into a very coarse approximation, with an ordered sequence of detail functions making up the difference. Here, the chloride concentration of well OK-7D is decomposed into an approximation and four detail functions.

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 remains 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.

Visual 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 is high. Some degree of correspondence at levels 1 and 2 is also apparent from Fig. 11 though not quite as obviously as in level 3. Thus, all three levels of detail were analyzed.

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, only the longest continuous portion of the data series was analyzed.

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

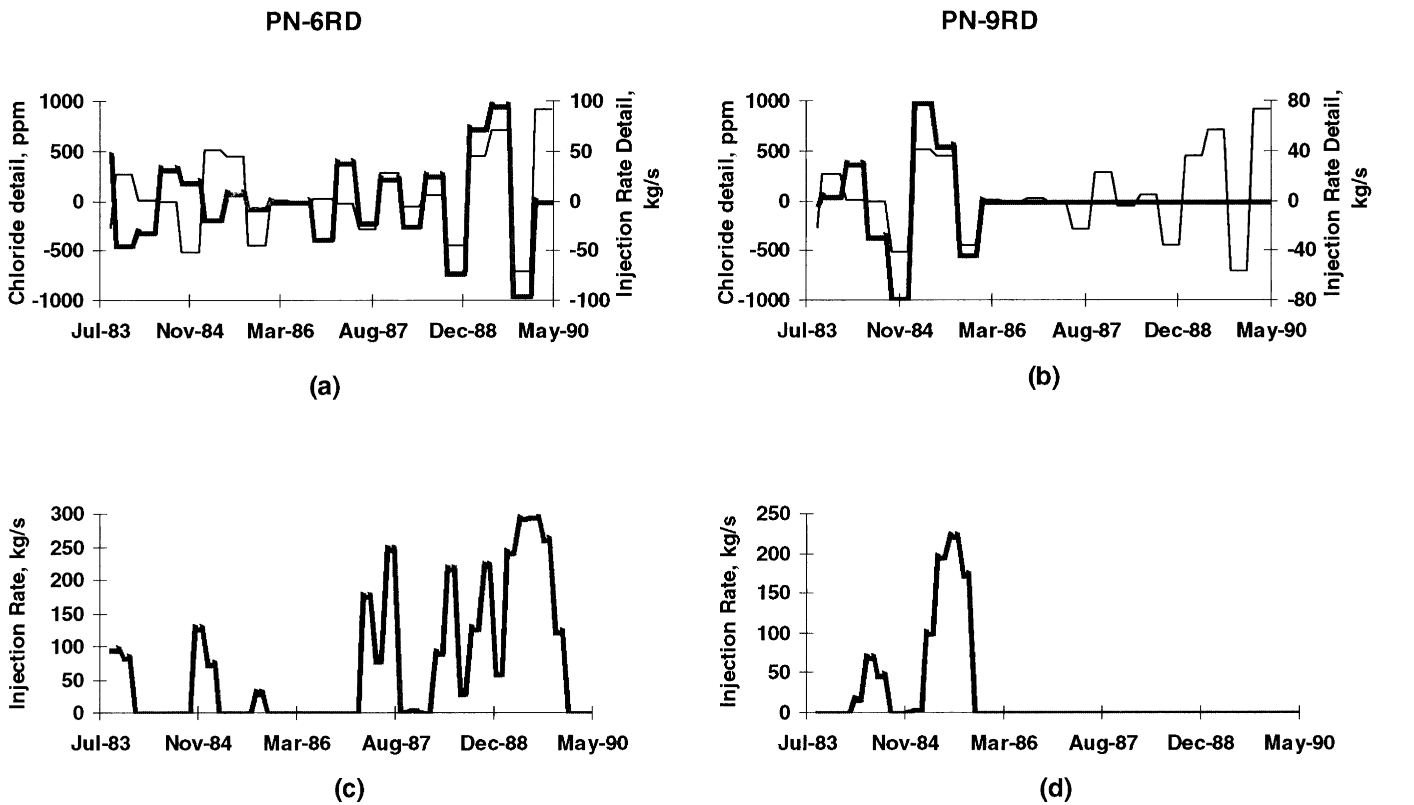

Fig. 10. (a) Level 3 detail of OK-7D chloride—thin line; level 3 detail of PN-6RD injection rate—thick line. (b) Level 3 detail of OK-7D chloride—thin line; level 3 detail of PN-9RD injection rate—thick line. (c) PN-6RD injection rate. (d) PN-9RD injection rate. Level 3 details of wells PN-6RD and PN-9RD closely follow the detail of OK-7D chloride during intervals when injection to these wells is high.

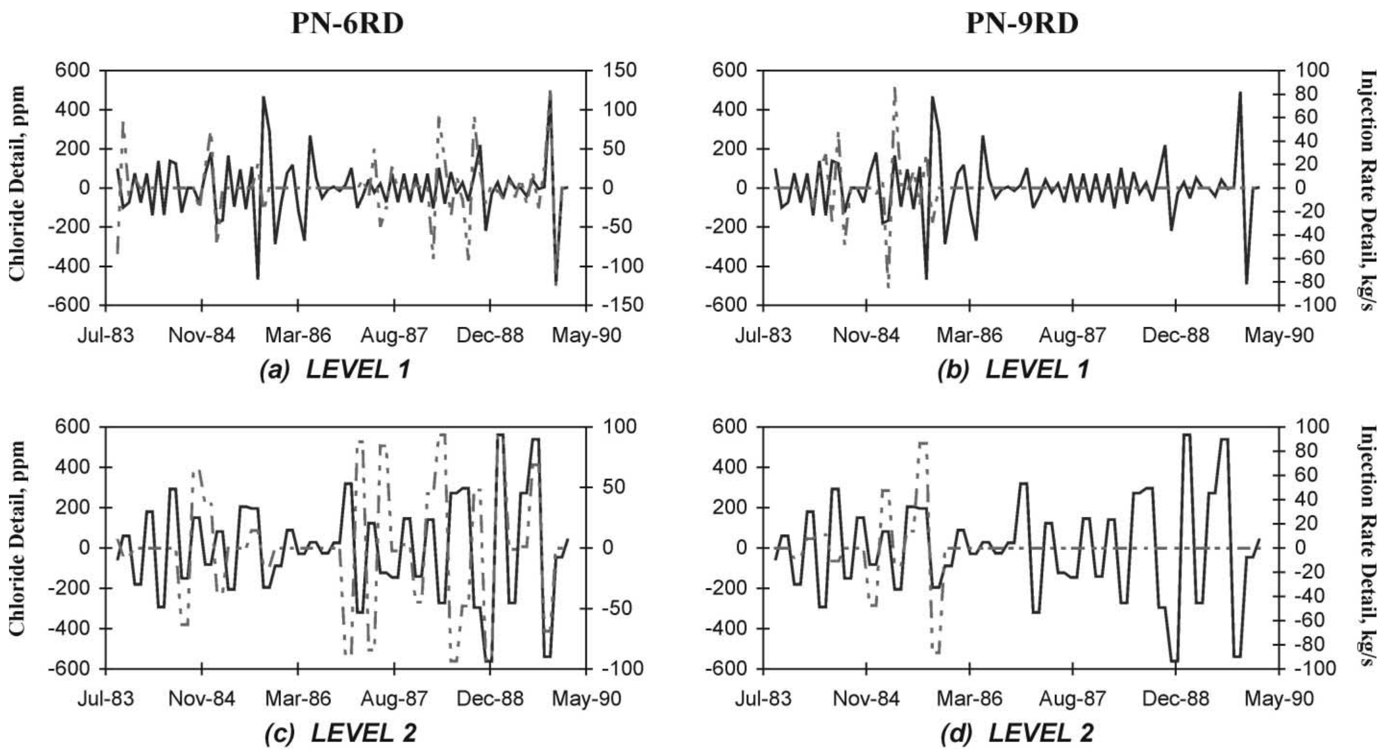

Fig. 11. (a) Level 1 detail of OK-7D chloride—solid line; level 1 detail of PN-6RD injection rate—dashed line. (b) Level 1 detail of OK-7D chloride—solid line; level 1 detail of PN-9RD injection rate- dashed line. (c) Level 2 detail of OK-7D chloride—solid line; level 2 detail of PN-6RD injection rate-dashed line. (d) Level 2 detail of OK-7D chloride—solid line; 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). Matlab, the program used to take the wavelet transform of our data, apparently utilizes 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 three to six data points were eliminated from the details series.

In about 1990, the bulk of the injection in Palinpinon-I was moved farther away from 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 those that operated between 1983 and 1990, those that started to inject in 1990 and are still injecting, and those that 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. The regression analysis included only wells that 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 that 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 the results of this analysis. One test was 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 list 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 regression analysis to be affected

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

Monitored well[a]	Mean transit time, days
PN-9RD tracer test	
OK-7D	5.4
PN-26D	13
PN-28D	14
PN-29D	15.4
PN-30D	15.7
PN-23D	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

[a] Only wells which have chloride data are reported here.

Table 9
Regression results for OK-12RD

Affected well	Coefficient
Detail level 1	
OK-10D	12.38
Detail level 2	
PN-23D	2.46
PN-29D	−4.05
PN-31D	−10.82
Detail level 3	
PN-15D	125.27
PN-16D	−7.40
PN-29D	−3.34
PN-30D	6.15

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

Table 10
Regression results for PN-9RD

Affected well	Coefficient
Detail level 1	
PN-30D	5.74
PN-29D	3.99
PN-16D	1.47
Detail level 2	
PN-19D	4.87
PN-18D	4.06
OK-7D	2.96
PN-16D	2.02
PN-29D	−11.65
Detail level 3	
OK-7D	9.4
PN-29D	1.83
PN-16D	0.92
PN-23D	0.43

well OK-12RD. Four of the seven wells were not monitored during the tracer test. The well which is most connected to PN-9RD based on the tracer test had the highest coefficient at level 3 regression; the same is true for OK-12RD.

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 from OK-7D but no injection rate data from OK-7D were available to allow verification with regression results.

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 considered 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 (Tables 12–14), however, show that OK-10D was not affected by TC-2RD. Rather, it was 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 1996. 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

Table 11
Level 3 regression results for PN-8RD

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

Table 12
Level 3 regression statistics for OK-10D (1986–1990)

Regression statistics	
Multiple R	0.839369312
R^2	0.704540842
Standard error	79.20406691
Observations	48

	Coefficients	r (simple)	Standard error	P-value[a]
PN-1RD	1.0860989	0.742898166	0.432894148	0.015779816
PN-2RD	−3.710875222	−0.725961532	1.775342121	0.042277591
PN-3RD	−4.411471297	−0.789481133	1.783793986	0.017237615

[a] P-value is the probability that an independent variable is not important to the regression solution.

Table 13
Level 3 regression statistics for OK-10D (1990–1996)

Regression statistics	
Multiple R	0.835160095
R^2	0.697492385
Standard error	150.1670809
Observations	80

	Coefficients	r (simple)	Standard error	P-value[a]
PN-3RD	−11.49790761	−0.771033019	1.485833395	3.48317E-11
TC-3R	1.704839524	0.522611064	0.404330736	6.77595E-05
N3	30.87773863	−0.653934071	6.498963063	9.35588E-06
OK-3R	−7.361847973	−0.360888493	2.134566707	0.00092135

[a] P-value is the probability that an independent variable is not important to the regression solution.

(−0.79 and −0.77). The large positive coefficient of well N3 is suspect as it conflicts with its negative r value. PN-20D was also determined to be affected by PN-3RD. The effect of TC-2RD and TC-4RD on PN-20D could not be ascertained from regression analysis due to insufficient chloride data from PN-20D after 1990.

Table 14
Level 3 regression statistics for PN-20D (1983–1989)

Regression statistics				
Multiple *R*	0.635764265			
R^2	0.404196201			
Standard error	495.2038572			
Observations	80			
	Coefficients	*r* (simple)	Standard error	*P*-value[a]
PN-1RD	10.21445053	0.346132276	1.741845857	1.0654E-07
PN-3RD	12.28448348	0.357197703	2.153358367	2.06197E-07
PN-6RD	4.493768629	0.105136293	1.637086662	0.007527849

[a] *P*-value is the probability that an independent variable is not important to the regression solution.

Table 15
Level 3 regression statistics for PN-13D (1990–1996)

Regression statistics				
Multiple *R*	0.962568329			
R^2	0.926537788			
Standard error	74.57273176			
Observations	70			
	Coefficients	*r* (simple)	Standard error	*P*-value[a]
TC-2RD	4.826890146	0.403207897	0.459363951	1.44516E-15
TC-3R	1.989170863	0.37912479	0.16882133	1.11144E-17
TC-4R	52.80156771	−0.473050095	4.657792303	5.98442E-17
ML-1RD	−191.3866835	−0.731292263	16.95558741	7.1995E-17
N3	11.67741779	−0.769449454	1.849731659	2.93281E-08
OK-3R	27.49754181	−0.472419422	2.227450499	1.38165E-18

[a] *P*-value is the probability that an independent variable is not important to the regression solution.

Pamatian (1997) reported that reinjection fluid from TC-2RD neutralized the fluid acidity in 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 15). Terms with conflicting *r* and coefficient signs still posed interpretation problems.

5. Conclusions and recommendations

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

A closer look at multiple regression techniques showed that what seemed to be highly encouraging results (high R^2 values) from prior multilinear modeling were 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 whereas 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 detail in the data.

Emphasis is also placed on the need for continuous data measurements when doing wavelet analysis. Highly intermittent measurements result in data loss. Because 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.

Acknowledgements

This work was funded by the US 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 from Dixie Valley.

References

Amistoso, A.E., Orizonte Jr., R.G., 1997. Reservoir response to full load operation, Palinpinon production field, Valencia, Negros Oriental, Philippines. Proceedings of 18th PNOC-EDC Geothermal Conference, pp. 55–66.

Bowerman, B.L., O'Connell, R.T., 1990. Linear Statistical Models: An Applied Approach, 2nd Edition. PWS-KENT Publishing Company, Boston, MA.

Harper, R.T., Jordan, O.T., 1985. Geochemical changes in response to production and reinjection for Palinpinon-I geothermal field, Negros Oriental, Philippines. Proceedings of 7th New Zealand Geothermal Workshop, pp. 39–44.

Jansen, F.E., Kelkar, M.G., 1997. Application of wavelets to production data in describing inter-well relationships. Proceedings of 1997 SPE Annual Technical Conference and Exhibition. SPE 38876.

Macario, M.E., 1991. Optimizing Reinjection Strategy in Palinpinon, Philippines, Based on Chloride Data. MS thesis, Stanford University, Stanford, CA.

Ogden, R.T., 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, pp. 121–133.

Tabachnick, B.G., Fidell, L.S., 1996. Using Multivariate Statistics. Harper Collins Publishers, New York.