Final Project Report:
U.S. Department of the Interior, Bureau of Reclamation (Reclamation)/ Coachella Valley Resource Conservation District (CVRCD) Cooperative Agreement #09FG340003:

A proposal to develop a new and simplified statistical modeling and analysis approach for performing EM38 / soil salinity surveys within the Lower Colorado Region.

Submitted to:
Mr. Gerald Casares
Grants & Agreements Specialist
U.S. Department of the Interior, Bureau of Reclamation – Yuma Area Office

Submitted by:
Dr. Scott Lesch
Principal Consulting Statistician
UCR – Statistical Consulting Center

Mrs. Silvia Aslan
CVRCD District Manager

Mr. Susano Duarte
CVRCD Salinity Assessment Technician

June 10, 2009

1. INTRODUCTION & EXECUTIVE SUMMARY

This report summarizes the data base compilation and statistical analysis activities for the Lower Colorado Region EM38 / Soil salinity cooperative agreement #09FG340003 between the U.S. Department of the Interior, Bureau of Reclamation's Yuma Area Office (Reclamation-YAO) and the Coachella Valley Resource Conservation District (CVRCD). For this grant, the University of California Statistical Consulting Center (UCR-SCC) has served as a sub-contractor to the CVRCD, supplying statistical modeling and analysis services. The purpose of this grant was to develop a new and simplified statistical modeling and sampling approach for performing EM38 / soil salinity surveys within the Lower Colorado Region. This final project report documents these grant research activities and summarizes both the objectives and results of our analytical study.

This report contains five additional sections beyond this Introduction (section 1), which provides an overview of the report organization. Section 2 describes and summarizes the compilation of the CVRCD database used in our study. Briefly, 82 EM38 / soil salinity survey data projects collected by the CVRCD since 2002 were retrieved, organized, and subjected to preliminary screening activities by the CVRCD and UCR-SCC. Exactly 77 of these field survey projects passed all data quality assurance / quality control (QA/QC) criteria; these 77 field survey projects were compiled into a master database to facilitate the statistical modeling and analysis. Section 2 documents these QA/QC and database development activities, in addition to a number of meta-data statistics associated with the CVRCD study data.

Sections 3 and 4 describe and summarize the proposed analysis of covariance (ANOCOVA) modeling approach. Section 3 reviews the proposed statistical methodology examined in this study, including the ANOCOVA modeling definitions, statistical assumptions, and model assessment techniques. Section 4 then documents and summarizes the modeling results. In section 4 we show that a regional ANOCOVA model successfully reduced the cross-validated, average log salinity prediction error (variance) estimate by more than 30% across these 77 fields and improved the depth-averaged prediction accuracy in 58 out of the 77 individual fields. <u>These results show that the ANOCOVA modeling approach can be used to improve the accuracy of field soil salinity predictions from EM38 signal data in most of the historical surveys conducted in the Coachella Valley, particularly in fields where only a limited number of calibration sampling locations are available.</u> The optimal ANOCOVA EM-slope parameter estimates for the Coachella Valley are also derived and presented in section 4 (see table 4.6).

Next, section 5 discusses how this ANOCOVA modeling strategy can be implemented by the CVRCD during future surveying operations. An example is presented in section 5.1 that shows how to calculate the field-specific ANOCOVA model intercept estimates and assess the reliability of these estimates. Likewise, in section 5.2 we describe how the ESAP-RSSD software program can be used to generate an optimal sampling plan for determining these intercept estimates.

Finally, in section 6 we discuss an additional analysis of the field average salinity levels associated with 70 survey projects conducted on vegetable fields (in the CVRCD database). Specifically, we examine these data to determine if different irrigation and/or leaching techniques exhibit different degrees of success at controlling the field average salinity levels. Section 6.1 presents a summary of the data examined in this additional study, while section 6.2 describes the statistical modeling methodology used to test this hypothesis. The results of this analysis (presented in section 6.3) <u>suggest that there is no specific irrigation or leaching technique that is consistently better at controlling and/or minimizing average salinity conditions in a typical Coachella Valley vegetable field</u>.

This somewhat surprising finding suggests that the currently employed sprinkler, furrow, or drip irrigation techniques and sprinkler or flood leaching techniques used across the valley are all equally effective at controlling the field average salinity level, at least on the more commonly encountered fine sand and fine sandy loam soil types.

As discussed above, this document constitutes our final report for the Reclamation / CVRCD cooperative agreement #09FG340003. In our initial grant proposal submitted to the Bureau of Reclamation, we specified that one intermediate progress report and two final reports would be produced that documented all project research activities. For purposes of compactness and clarity, these individual intermediate and final progress reports have been combined together into this single, final project report. Specifically with respect to our original grant proposal document, section 2 of this final project report constitutes our (previously proposed) intermediate progress report. Likewise, sections 3 through 5 constitute our final progress report on the ANOCOVA modeling approach, while section 6 constitutes our final progress report on the assessment of typical valley-wide irrigation and leaching techniques for salinity control.

Any questions and/or comments on the technical information contained within this report should be directed to Dr. Scott Lesch, Principal Consulting Statistician, UCR-SCC (scott.lesch@ucr.edu) and Mrs. Silvia Aslan, RCD District Manager, CVRCD (silvia@cvconservation.org), respectively.

2. SURVEY PROJECTS

2.1 Initial Project Screening and QA/QC

After initiation of this project, the CVRCD delivered electronic data files associated with more than 90 individual field survey projects to the UCR-SCC. Upon review, 82 of these projects were found to contain a complete set of EM38 survey and soil sample information. All 82 of these complete projects had been conducted between 2002 and 2008, with the majority of these performed on or after January 2005.

The ESAP EM38 signal data file and associated soil sample (profile) data file were systematically extracted from each delivered project and stored into a temporary master database. After extraction, the EM38 survey information in each signal data file was examined for internal consistency and reliability; i.e., for properly correlated and aligned EM_H and EM_V signal readings that were all positive and devoid of gross outliers, systematic instrument bias, and any obvious miss-calibration effects. Exactly 80 of the 82 signal data files associated with these complete project folders passed all internal consistency and reliability tests. (The signal files associated with fields 72 and 77 failed these tests; these field survey projects were subsequently deleted from the temporary database.)

The data associated with remaining 80 projects were then each individually assessed using a dynamic Dual Pathway Parallel Conductance (dynamic DPPC) correlation analysis (Lesch and Corwin, 2003). In these analyses, the calculated apparent soil conductivity readings (as computed from the measured soil salinity, saturation percentage, and water content measurements) were compared to the average of the EM38 signal data on a log / log basis, after adjusting for potentially low water content levels. An a priori threshold correlation level of 0.5 was specified as the minimum acceptable correlation level (for indicating reliable survey data). Exactly 77 of the 80 fields produced dynamic DPPC correlation levels > 0.5; these 77 projects were thus chosen for inclusion into the permanent (validated) EM / soil salinity database.

2.2. Survey and Soil Sample Meta-data Statistics

The meta-data statistics discussed throughout the remainder of this section summarize the general characteristics of the 77 validated field studies. Note that the EM38 signal and soil sample data from these studies has been used to develop the ANOCOVA modeling approach described in sections 3, 4 and 5 of this report.

Some general survey statistics associated with these 77 field projects are shown in table 2.1. This table quantifies general field information statistics; specifically the distribution of field sizes and the number of EM38 survey positions recorded in each field. The average field size was approximately 28.2 acres; 50% of the fields were between 17.9 to 34.8 acres and the smallest and largest fields were 2.8 and 66.2 acres, respectively. Likewise, the average number of EM38 survey positions was about 1550 and the survey data files associated with 50% of the fields contained between 1254 to 1795 positions.

Table 2.1. General field summary statistics for the 77 validated field studies.

Statistic	Field Information	
	Size (acres)	# of EM Sites
Mean	28.17	1550.3
Std. Dev.	12.24	689.4
Skewness	0.00	1.82
Quantiles:		
Minimum	2.8	404
10%	10.7	735
25%	17.9	1254
Median (50%)	31.8	1429
75%	34.8	1795
90%	41.7	2487
Maximum	66.2	4998

Table 2.2 shows the number of 6- and 12-site soil sampling plans employed across these 77 field projects. As shown in table 2.2, the majority of fields ($n = 60$) were sampled using 6-site ESAP sampling plans. The remaining 17 fields were sampled using 12-site plans, although the soil sample data files for exactly four of these fields contain either 11 or 10 sampling locations, respectively. Overall, the validated soil sample database contains exactly $N = 6(60) + 10(1) + 11(3) + 12(13) = 559$ distinct sampling locations across 77 fields.

Table 2.2. Number of ESAP generated sampling sites for each of the 77 validated field studies.

Soil Sampling Design	Number of Fields	Notes
6-site design	60	
12-site design	17	3 fields were sampled at only 11 locations, 1 field at only 10 locations

Table 2.3 shows the number of sampling depths reported for each of the 77 field projects. Exactly 44 and 10 fields were sampled to depths of 90 cm (3 feet) and 120 cm (4 feet), respectively, in 30 cm (1 foot) depth increments. In the remaining 23 projects, soil sample data were collected from two depth increments. In 18 of these projects, the sampling depth increments were 30 cm (1 foot); the remaining five projects contained soil sample data collected using 60 cm (2 foot) depth increments. Note that in the validated soil sample database, the saturation percentage (SP), volumetric water content (θ_v), and log transformed salinity ($\ln(EC_e)$) readings in these five projects were interpolated to equivalent 30 cm depth increments using linear interpolation techniques. This interpolation was necessary to ensure that all validated soil sample data corresponded to aligned sample depth increments (i.e., 30 cm samples with depth midpoint positions at 0.15, 0.45, 0.75, and 1.05 m). Hence, in the validated soil sample database, there are exactly 77, 59, and 15 fields exhibiting 30 cm soil samples to depths of 60, 90, and 120 cm, respectively.

Table 2.3. Number of sampling depths for each of the 77 validated field studies.

Depth Increments	Number of Fields	Notes
2	23	5 of these 23 fields sampled in 2 foot (60 cm) depth increments
3	44	All 44 fields sampled in 1 foot (30 cm) depth increments
4	10	All 10 fields sampled in 1 foot (30 cm) depth increments

Figure 2.1 shows the physical locations (UTM coordinate grid, zone 11 North) of the 77 field surveys conducted across the Coachella Valley. As shown in figure 2.1, all 77 fields are located within a 30 km by 30 km area across the valley, with the majority of the fields located near the center of the valley. Overall, the coverage across the valley is fairly comprehensive, particularly given the non-random manner in which these fields were selected for study.

2.3 Field condition, Crop, and Soil type Meta-data Statistics

Table 2.4 displays the number of fields exhibiting soil sample data exhibiting adequate (> 70%), marginal (50% to 70%), and excessively dry (< 50%) average water content relative to field capacity (WCFC) levels, for the depth increment data originally reported in the soil sample data files. Note that 11 to 13 fields in the database appear to have been sampled under excessively dry soil conditions; at least five of these fields exhibited marked differences between their ordinary and dynamic DPPC correlation coefficients (see figure 2.2).

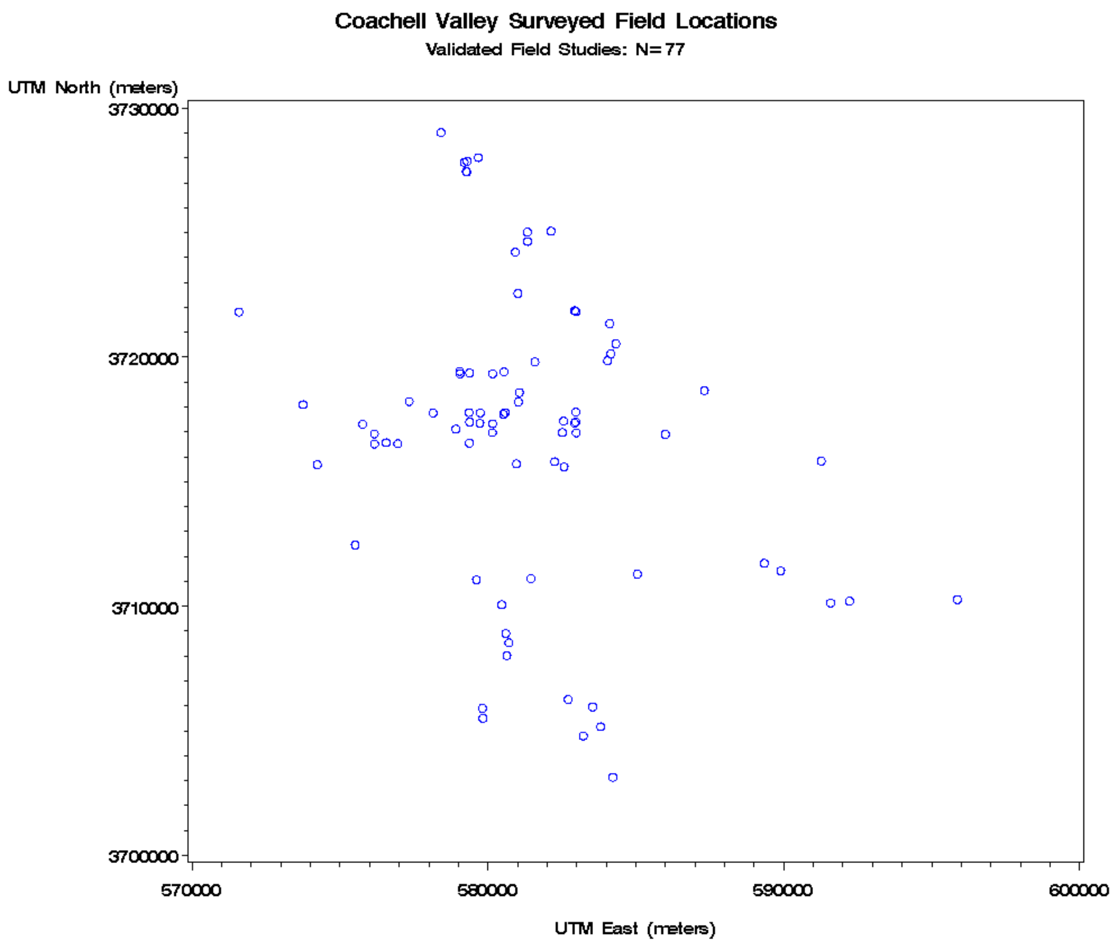

Figure 2.1. Physical locations of the 77 Coachella Valley field surveys.

Table 2.4. Calculated water content classes (relative to estimated field capacity) by sample depth increment.

Depth Increment	WCFC Class	Number of Fields
1st Increment	> 70%	37
	50% to 70%	28
	< 50%	12
2nd Increment	> 70%	50
	50% to 70%	14
	< 50%	13
3rd Increment (Note: 23 fields missing 3rd depth increment information)	> 70%	33
	50% to 70%	10
	< 50%	11

Table 2.5 shows the primary irrigation methods employed at the time of the initial surveys for each of the 77 field projects. Overall, 85% of the fields were irrigated using either drip or furrow methods. The vast majority of these drip and furrow irrigated fields were also supporting some type of vegetable crop (vegetable, herb, or fruit), as shown in table 2.6. Nearly all of the remaining fields essentially represented some type of orchard (citrus, dates, or grape vineyards). Only two fields exist in the database that can be classified both non-vegetable and non-orchard (one field supporting a Bermuda grass cover crop and a second field set aside as natural habitat).

Table 2.5. Primary irrigation method (used at the time of survey) for each of the 77 validated field studies.

Primary Irrigation Method	Number of Fields
Basin Flood	5
Drip (permanent & T-tape)	31
Furrow	35
Sprinkler	6

Finally, table 2.7 quantifies the amount of acreage associated with the various NRCS classified soil-types in the database. (In table 2.7, one surveyed field without any associated soil-type information has been excluded from the acreage calculations.)
As shown in table 2.7, there are Carsitas, Coachella, Gilman, Indio, Myoma, and Salton soil series included in the database, with the Gilman and Indio series representing the dominant soil types. It is worthwhile to note that all of these series except Salton are generally classified as either fine sands or fine sandy loams (Salton Sb is classified as a silty clay loam).

Table 2.6. Type of crops reported for each of the 77 valid field studies. Note: the total number of fields is > 77, since multiple fields supported either two or three crops grown in rotation, or two simultaneous crops.

Crop	Crop Type	Number of Fields
Artichokes	Vegetables	2
Bell Peppers	Vegetables	9
Bermuda grass	Cover crop	1
Broccoli	Vegetables	9
Cabbage	Vegetables	1
Carrots	Vegetables	2
Cauliflower	Vegetables	7
Celery	Vegetables	2
Chili Peppers	Vegetables (Herbs)	1
Cilantro	Vegetables (Herbs)	2
Citrus Trees	Orchard (Citrus)	1
Date Palms	Orchard (Dates)	6
Dill	Vegetables (Herbs)	5
Drum Stick	Vegetables	1
Eggplant	Vegetables	2
Fan Palms	Orchard (Palms)	2
Grapes	Orchard (Vineyard)	3
Green Beans	Vegetables	1
Habitat	Other	1
Herbs (general)	Vegetables (Herbs)	1
Lettuce (Leaf or Romaine)	Vegetables	14
Lemon Trees	Orchard (Citrus)	1
Melons	Vegetables (Fruit)	4
Okra	Vegetables	3
Peach Trees	Orchard	1
Rapani	Vegetables	1
Strawberries	Vegetables (Fruit)	5
Sweet Corn	Vegetables	5
Thyme	Vegetables (Herbs)	2
Vegetables (general)	Vegetables	10
Watermelon	Vegetables (Fruit)	1

Table 2.7. Total acreage of surveyed soil series associated with 76 of the 77 valid field studies.

Soil Series	Symbol	Acreage	% of Total Acreage
Carsitas	CdC	65.5	3.07
	CkB	7.7	0.36
Coachella	CpA	10.9	0.51
	CpB	18.1	0.85
	CrA	136.3	6.38
Gilman	GaB	11.5	0.54
	GbA	91.6	4.29
	GcA	548.7	25.69
	GdA	4.4	0.21
	GfA	125.4	5.87
Indio	Ip	5.2	0.24
	Ir	171.9	8.05
	It	728.0	34.09
Myoma	MaB	57.8	2.71
	McB	44.4	2.08
Salton	Sb	108.4	5.07

2.4 Soil EC_e and EM38 Meta-data Statistics

Table 2.8 shows the EC_e and collocated EM38 univariate summary statistics for the pooled dataset (corresponding to the 77 valid field studies). The majority of the soil salinity measurements are fairly low; note that the median salinity levels are < 3 dS/m across all four sampling depths and at least 25% of the readings in each depth are < 1.2 dS/m. However, approximately 10% of the measurements from each sampling depth are > 10 dS/m and the pooled salinity distributions are strongly right skewed (i.e., lognormally distributed). The EM38 signal readings exhibit similar characteristics.

The correlation coefficients between the log transformed EC_e and EM38 readings are all fairly high, as shown in table 2.9. (All coefficients are statistically significant below the 0.0001 level.) These results tentatively suggest that the EM38 data can be effectively used to predict the soil salinity levels, provided adequate models can be derived that suitably adjust for different secondary characteristics across fields.

Table 2.8. Soil salinity (EC_e) and collocated EM38 (EM_H, EM_V) summary statistics for the 77 valid field studies.

Statistic	Salinity Data (ECe, dS/m: by depth)				EM38 (mS/m)	
	0.15 m	0.45 m	0.75 m	1.05 m	EM_H	EM_V
N	559	559	450	106	559	559
Mean	5.06	4.77	4.47	4.63	73.42	95.73
Standard Deviation	9.05	8.81	6.65	5.65	91.90	107.94
Skewness	4.34	4.23	5.21	2.40	3.03	2.92
Quantiles:						
Minimum	0.12	0.10	0.10	0.19	8.50	10.88
5%	0.58	0.44	0.56	0.42	14.75	21.88
10%	0.72	0.56	0.71	0.72	17.50	27.88
25%	1.04	0.84	1.10	1.14	24.25	38.00
Median (50%)	2.11	1.80	2.12	2.90	39.38	57.88
75%	4.48	4.20	5.00	5.47	76.00	99.88
90%	12.60	11.62	10.65	9.60	179.25	223.13
95%	20.10	19.61	16.80	20.60	267.38	341.50
Maximum	66.70	72.60	83.50	27.40	658.88	718.25

Table 2.9. Correlation coefficients for $\ln(EC_e)$ v.s. $\ln(EM_H)$ and $\ln(EC_e)$ v.s. $\ln(EM_V)$; by depth (pooled data across all 77 valid field studies).

Correlation Statistic	Correlation Coefficients (by sample depth)			
	0.15 m	0.45 m	0.75 m	1.05 m
Corr[$\ln(EC_e)$, $\ln(EM_H)$]	0.668	0.801	0.751	0.802
Corr[$\ln(EC_e)$, $\ln(EM_V)$]	0.618	0.773	0.756	0.785

A more detailed assessment of the soil salinity / EM38 correlation structure can be determined by calculating the ordinary and dynamic DPPC correlation coefficients for each field. Summary statistics concerning these coefficients are presented in table 2.10. Recall that fields exhibiting dynamic DPPC coefficients < 0.5 were removed from the final data base; hence all 77 dynamic DPPC coefficients are obviously > 0.5. The average dynamic DPPC correlation value in the database is 0.858 and 75% of the fields produce correlation coefficients > 0.8; these results suggest that the majority of the survey data sets exhibit a high degree of internal consistency and reliability. The ordinary DPPC correlation coefficients are somewhat lower (average value = 0.777, 75% of the fields exhibit correlation coefficient values > 0.66). These generally lower ordinary DPPC coefficients suggest that the EM38 signal data in at least some of the 77 field surveys were moderately influenced by low field water content conditions.

Table 2.10. Summary statistics for the field specific ordinary and dynamic DPPC correlation coefficients.

Statistic	Correlation Coefficients	
	Ordinary DPPC	Dynamic DPPC
Mean	0.777	0.858
Std. Dev.	0.200	0.114
Skewness	-1.51	-1.14
Quantiles:		
Minimum	0.117	0.521
10%	0.515	0.653
25%	0.669	0.801
Median (50%)	0.866	0.894
75%	0.911	0.940
90%	0.964	0.977
Maximum	0.986	0.992

The DPPC correlation plot (figure 2.2) confirms this hypothesis; this plot shows that there are 7 fields which exhibit abnormally low ordinary DPPC correlation coefficients (i.e., < 0.5). Additionally, 10 fields exhibit dy-DPPC correlation coefficients that are at least 0.2 units greater than their corresponding ordinary DPPC coefficients. Hence, there appears to be a stronger than normal water content influence on the EM38 signal data in at least 10 of these 77 fields.

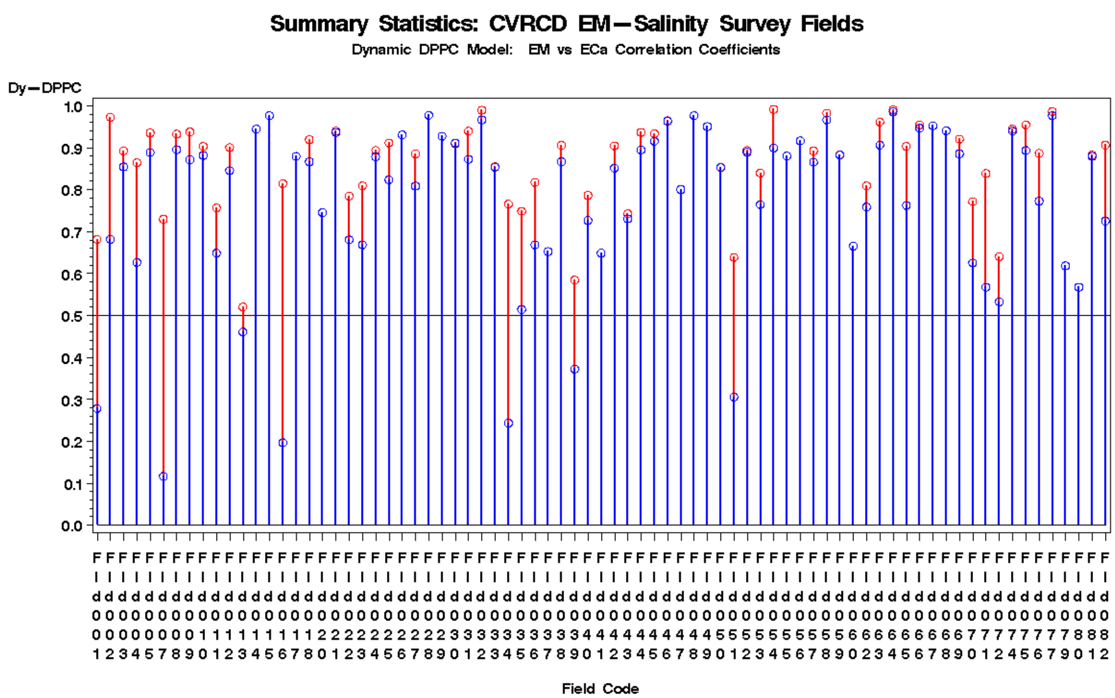

Figure 2.2. Ordinary (blue) and dynamic (red) DPPC correlation coefficients for the 77 validated field studies.

2.5 General Database Characteristics

Overall, the compiled database provides a good representation of EM38 survey situations encountered in the lighter textured soils across the Coachella Valley. As discussed in section 2.2, the median field size is the database is about 32 acres and the corresponding EM38 survey data contains roughly 1400+ survey positions. Additionally, in most fields soil samples were collected at six sites in 30 cm increments, down to either 0.6 or 0.9 m.

The database contains fields across a wide area of the valley and the vast majority of these fields were supporting some type of vegetable, herb, or fruit crop. Overall, 85% of the fields were irrigated using either drip or furrow methods and 95% of the soil types (soil series) can be classified as either fine sands or fine sandy loams. Not surprisingly, more that 50% of the soil samples exhibit fairly low salinity levels ($EC_e < 3$ dS/m), but about 10% of the samples exceed 10 dS/m.

The correlation levels between the log transformed soil salinity and log transformed EM38 signal data are fairly high ($r = 0.6$ to 0.8) and the field specific DPPC correlation coefficients suggest that the majority of the survey data sets exhibit a high degree of internal consistency and reliability. Although there does appear to be a stronger than normal water content influence on the EM38 signal data in at least 10 of these 77 fields, this result is probably consistent with typical surveying conditions encountered in the valley.

In summary, this compiled database of 77 fields should provide a robust and realistic data set to test the ANOCOVA modeling approach on. In turn, the estimated ANOCOVA model parameters should be applicable to Coachella Valley fields surveyed under similar soil and cropping conditions. More specifically, the estimated ANOCOVA model parameters should be applicable to nearly all specialty crops grown on lighter textured soils across the valley.

3. STATISTICAL METHODOLOGY

3.1 Model Definitions and Statistical Assumptions

Three types of linear models are considered here, for purposes of predicting the natural log transformed soil salinity values from the natural log transformed EM38 signal readings. For a soil salinity sample acquired from the jth sampling depth at the ith site within the kth field, these models are defined as follows:

The Field Specific Regression model (FSR model)

$$\ln(EC_{ijk}) = \beta_{0,jk} + \beta_{1,jk}\ln(EM_{V,ik}) + \beta_{2,jk}\ln(EM_{H,ik}) + \varepsilon_{ijk} \tag{3.1}$$

The Analysis of Covariance model (ANOCOVA model)

$$\ln(EC_{ijk}) = \beta_{0,jk} + \beta_{1,j}\ln(EM_{V,ik}) + \beta_{2,j}\ln(EM_{H,ik}) + \varepsilon_{ijk} \tag{3.2}$$

The Common Coefficient Regression model (CCR model)

$$\ln(EC_{ijk}) = \beta_{0,j} + \beta_{1,j}\ln(EM_{V,ik}) + \beta_{2,j}\ln(EM_{H,ik}) + \varepsilon_{ijk} \tag{3.3}$$

Note that all three of the coefficients in the FSR model (Eq. 3.1) change across fields and sampling depths; i.e., all of the regression model coefficients need to be re-estimated whenever a new field is surveyed. In contrast, the EM38 slope coefficients in the ANOCOVA model (Eq. 3.2) only change across sampling depths, but not across fields. Only the intercept coefficients change across fields and sampling depths in this second model. Finally, in the CCR model (Eq. 3.3), none of the coefficients change across fields; these coefficients can only change across sampling depths.

There are some well established theoretical reasons why we should expect the FSR model to represent the most accurate regression-based calibration equation (Lesch and Corwin, 2003). However, it is also well known that this equation can be difficult to accurately estimate when only a small number of calibration soil samples are available in a given field. Additionally, many of the field-specific effects on EM38 survey data such as seasonal changes in bulk soil temperature, bed-furrow geometry, surface roughness, and instrument placement height (during the survey process) are known to be approximately multiplicative. Therefore, on the log transformed scale these effects become additive constants, which in theory should only affect the intercept coefficient. Thus, there are both legitimate theoretical and statistical reasons to expect that the ANOCOVA model might actually perform better than the FSR model, particularly when only limited calibration data are available and the EM38 survey data is acquired over a short time-span (i.e., over a period of a few hours within any specific field).

Assuming that a total of M fields have been surveyed, for a specific sampling depth the FSR model requires $3M$ parameter estimates to produce soil salinity predictions (across all M fields). In contrast, for a specific sampling depth the ANOCOVA model requires only $M + 2$ parameter estimates (or just M estimates, if the EM38 slope coefficients have already been established). Of course, the CCR model requires just 3 parameter estimates, but in general we would not expect this model to be very accurate for the reasons mentioned above. Note that the CCR model has been included in this study primarily as a baseline reference model, rather than a formal prediction model per say.

In the following analyses, the residual errors associated with both the FSR and ANOCOVA models are assumed to be Normally distributed, independent across fields and spatially uncorrelated within a field (across different sites). These first two assumptions are typically quite reasonable, the third assumption can generally be met when model directed sampling strategies are used to select the calibration sample locations (such as the spatial response surface sampling strategy used by the ESAP-RSSD software program). Additionally, the errors are assumed to be correlated across sampling depths (at the same site) and the error variances are normally assumed to change across fields (and sometimes across sampling depths also). Under these residual error assumptions, either ordinary or mixed linear modeling techniques can be used to estimate the parameter coefficients in either the FSR or ANOCOVA models. Ordinary least squares (OLS) estimation techniques will always yield the best linear unbiased (BLU) parameter estimates for the FSR model, since unique coefficients are estimated for each field and sampling depth; note that the FSR model is equivalent to a multivariate multiple linear regression model (Johnson and Wichern, 1988). If the residual errors are assumed to exhibit a common variance component across different fields (for a specific sample depth), then OLS estimation will also yield the BLU parameter estimates for the ANOCOVA model (Milliken and Johnson, 2002; Searle, 1971). If the variance components change across fields, then a heterogeneous variance ANOCOVA model must be specified and estimated (typically using restricted maximum likelihood estimation) in order to derive the empirical BLU parameter estimates (McCulloch and Searle, 2001; Rao and Toutenburg, 1999). Note that in this latter case, the heterogeneous variance ANOCOVA model is essentially just a mixed linear analysis of covariance model, where the residual error variances are allowed to change across fields.

3.2 Model Assessment

In this study, our primary model assessment criteria is the mean square prediction error (MSPE). The MSPE is defined as the squared difference between the observed and jack-knifed (*a.k.a.* cross-validated) log salinity predictions; i.e.,

$$MSPE = \left(1/N_{ijk}\right)\sum_{i,j,k}\left(y_{ijk} - \hat{y}_{ijk,(-i)}\right)^2 \tag{3.4}$$

where $y_{ijk} = \ln(EC_{ijk})$, $\hat{y}_{ijk,(-i)}$ represents the model predicted $\ln(EC_{ijk})$ value where the i^{th} observed log salinity measurement has not been used to calibrate the regression equation, and N_{ijk} represents the total number of jack-knifed soil salinity samples. Jack-knifing is routinely used in many types of statistical modeling applications and can be readily computed for all linear models fit using OLS estimation techniques (Rao and Toutenburg, 1999; Myers, 1986). In regression-type models, jack-knifing is commonly done to assess a model's prediction accuracy and to select the "best" (i.e., most accurate) model from a set of competing prediction equations (Myers, 1986); note that a smaller

MSPE implies a more accurate model. Note also that the MSPE can be easily computed for multiple stratification variables, such as for specific sampling depths or individual fields, etc. In the analyses that follow, we have computed the MSPE estimates associated with Eqns. (3.1), (3.2) and (3.3) by both sampling depths and individual fields, in addition to global MSPE values.

Along with the computation of various MSPE estimates, we have also performed a detailed analysis of the ANOCOVA model residuals. This residual assessment analysis has been performed in order to verify that the previously discussed modeling assumptions are in fact reasonable; a more detailed discussion on this topic is presented in section 4.2.

4. RESULTS

4.1 Prediction accuracy (MSPE Statistics)

As discussed in section 2, a total of 77 distinct fields were included into the master database. All 77 fields contained soil salinity samples from the 0-30 cm and 30-60 cm sampling depths, 59 of these fields contained samples from the 60-90 cm depth, and only 15 fields contained samples from the 90-120 cm depth. Additionally, 60 of the 77 fields exhibited exactly 6 calibration sampling locations per field; the remaining 17 files exhibited 10 to 12 sites per field, respectively.

Table 4.1 displays the FSR, ANOCOVA, and CCR model summary statistics pertaining to the models for the 0-30, 30-60, and 60-90 cm sampling depths. Note that no model was fit to the 90-120 cm sample depth, due to the limited number of fields containing soil sample data at this depth. The R^2 and mean square error (MSE) estimates shown in table 4.1 for the FSR models represent composite statistics; i.e., composite estimates calculated by pooling all of the individual fields together. Additionally, the R^2 and MSE estimates correspond to the ANOCOVA models computed using OLS estimation; i.e., to ANOCOVA models that assume a homogeneous residual variance component across fields. As shown in table 4.1, the ANOCOVA model R^2 values are about 9-10% lower than the (composite) FSR model R^2 values and the ANOCOVA MSE estimates are about 0.05 to 0.06 units larger. In contrast, the CCR model R^2 values are noticeably lower (than both the ANOCOVA and FSR model R^2 values) and the CCR MSE estimates are clearly much larger.

The sample depth specific and overall average MSPE estimates associated with the FSR, ANOCOVA, and CCR models are shown in table 4.2. Unlike the model summary statistics (that essentially measure how well each model "fits" the sample data), these MSPE values provide a much more reliable estimate of the prediction accuracy associated with each model. All of the MSPE estimates associated with the ANOCOVA models are considerably smaller than either the FSR or CCR estimates. When compared specifically with the depth specific FSR estimates, the ANOCOVA MSPE is about 30% lower for the 0-30 cm sample depth, 29% lower for the 30-60 cm sample depth, and 41% lower for the 60-90 cm depth. Likewise the overall average ANOCOVA MSPE estimate is about 33% lower (than both the FSR and CCR MSPE estimates). These results indicate that the ANOCOVA models produce the most accurate log salinity predictions; i.e., the jack-knifed variance of the prediction errors associated with the ANOCOVA model is only about two-thirds as large (on average) as the jack-knifed variance of the FSR model prediction errors.

Table 4.1. FSR, ANOCOVA, and CCR model summary statistics.

Model	Statistic	Midpoint Sampling Depth		
		15 cm	45 cm	75 cm
FSR	R^2	0.886	0.890	0.837
	MSE	0.241	0.266	0.304
ANOCOVA	R^2	0.792	0.803	0.733
	MSE	0.302	0.326	0.349
CCR	R^2	0.460	0.642	0.578
	MSE	0.675	0.512	0.481

Table 4.2. Jack-knifed MSPE estimates for the FSR, ANOCOVA, and CCR models.

Model	Midpoint Sampling Depth			Pooled Average
	15 cm	45 cm	75 cm	
FSR	0.499	0.535	0.681	0.564
ANOCOVA	0.350	0.379	0.404	0.376
CCR	0.680	0.515	0.485	0.565

Table 4.3 shows the number of fields where the ANOCOVA models produced smaller jack-knifed MSPE estimates, in comparison to the FSR models. Overall, the jack-knifed salinity patterns in 58 out of 77 fields (75%) were more accurately predicted by the ANOCOVA models. For fields with only 6 sample sites, this percentage increased slightly (47 out of 60, or approximately 78%). For fields having 10 to 12 sample sites, this percentage was somewhat lower (11 out of 17, or approximately 65%). In general, as the number of calibration samples increases, we would expect a FSR model to out-perform the ANOCOVA model. However, these results suggest that the prediction accuracy in more than half of the fields associated with a normal (12-site) ESAP sampling plan can still be improved using the ANOCOVA modeling approach. Additionally, the prediction accuracy in about 80% of the fields associated with a reduced (6-site) plan were improved using the ANOCOVA modeling approach.

Table 4.3. Summary count statistics on field specific MSPE estimates for the FSR and ANOCOVA models.

Strata	$MSPE_{ANCV} < MSPE_{FSR}$	$MSPE_{ANCV} > MSPE_{FSR}$
All Fields (N=77)	58	19
Fields w/6 sample sites (N=60)	47	13
Fields w/10-12 sample sites (N=17)	11	6

Table 4.4 summarizes the distribution of the ANOCOVA model jack-knifed MSPE estimates (into four classes); these classes can be used to "grade" the reliability of the salinity predictions. Overall, approximately 82% of the fields (63 out of 77) exhibit either excellent (grade A), good (grade B), or fair (grade C) prediction reliability. MSPE estimates > 0.6 suggest that the salinity levels in a particular field are not well described by (i.e., strongly correlated with) the associated EM38 survey data; 14 of the 77 fields fall into this latter class. Figures 4.1 through 4.4 show the observed versus jack-knife predicted salinity measurements for the groups of fields exhibiting A, B, C, and U prediction accuracy grades, respectively.

Table 4.4. Distribution of ANOCOVA jack-knifed MSPE estimates (prediction accuracy statistics).

MSPE Range	Grade	Prediction Accuracy	Number of Fields	% of Total Sample Size	Correlation: Obs v.s. Prd Salinity
< 0.15	A	Excellent	16	21%	0.921
0.15 – 0.30	B	Good	23	30%	0.860
0.30 – 0.60	C	Fair	24	31%	0.795
> 0.60	U	Unacceptable	14	18%	0.693

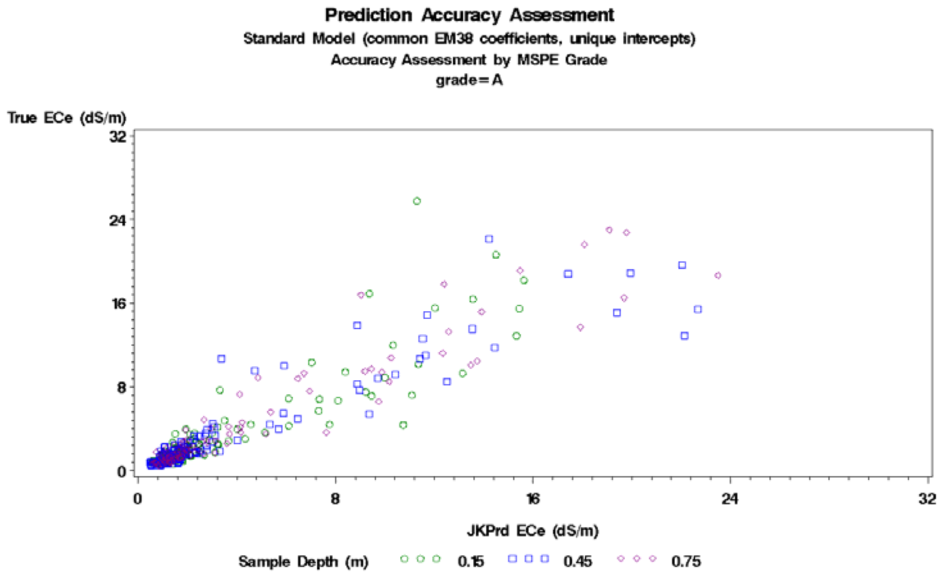

Figure 4.1. Observed versus jack-knife predicted soil salinity values (EC_e, dS/m), for fields with MSPE estimates < 0.15 (grade A fields).

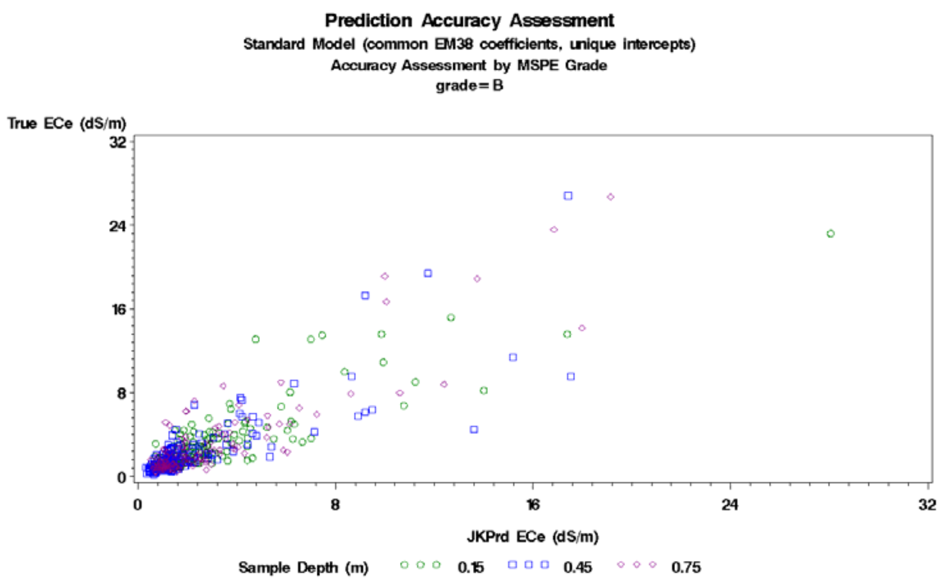

Figure 4.2. Observed versus jack-knife predicted soil salinity values (EC_e, dS/m), for fields with MSPE estimates > 0.15 and < 0.30 (grade B fields).

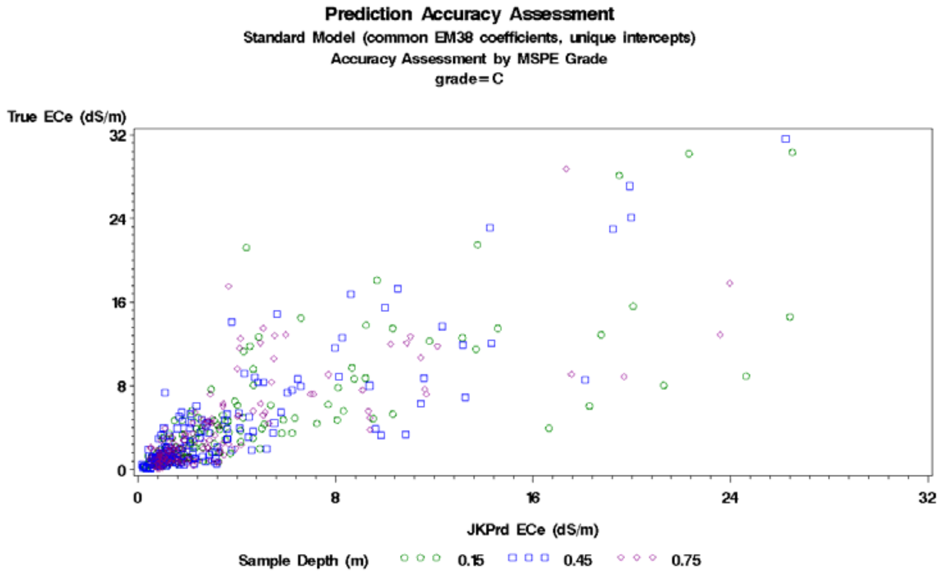

Figure 4.3. Observed versus jack-knifed predicted soil salinity values (EC_e, dS/m), for fields with MSPE estimates > 0.30 and < 0.60 (grade C fields).

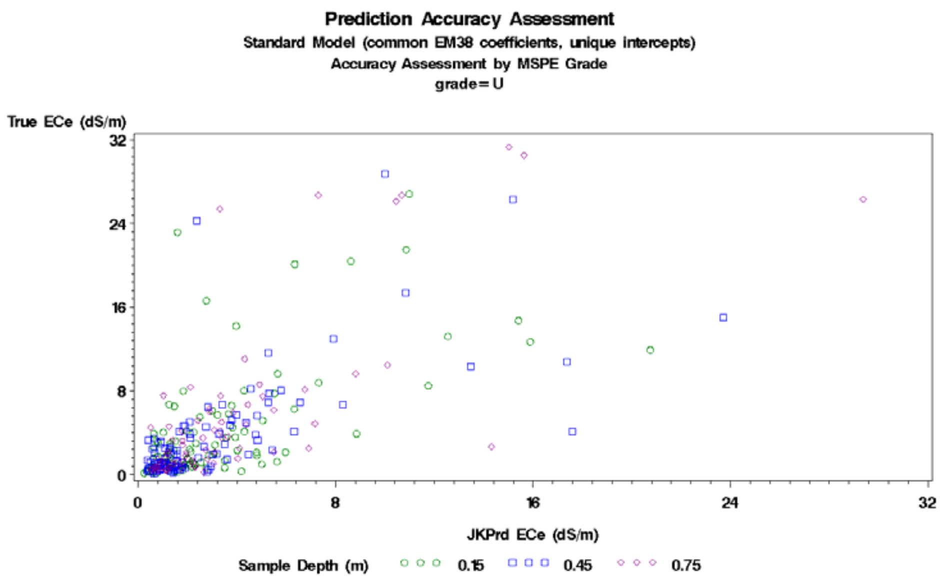

Figure 4.4. Observed versus jack-knifed predicted soil salinity values (EC_e, dS/m), for fields with MSPE estimates > 0.60 (grade U fields).

4.2 ANOCOVA model: Residual analysis and diagnostic statistics

As discussed in section 3.1, the residual errors associated with the ANOCOVA models are assumed to be Normally distributed, independent across fields and spatially uncorrelated within a field (across different sites). Additionally, when the ANOCOVA models are estimated using OLS techniques, we additionally assume that the residual variance does not change across fields. However, this last assumption often does not hold in practice, and thus should be carefully assessed.

A summary of the ANOCOVA model residual errors (i.e., a residual distribution assessment) is presented in table 4.5. The upper part of table 4.5 lists the global residual variance estimates for the three sampling depths, in addition to the pooled residual correlation matrix. All of the off-diagonal correlation coefficients are positive and statistically different from 0 ($p < 0.0001$), confirming that the ANOCOVA model residual errors associated with specific sampling locations are indeed correlated across sampling depths. (Note that this does not invalidate the OLS estimation technique, since unique ANOCOVA models have been fit to each depth.)

The center part of table 4.5 lists the test results and p-values for the Shapiro-Wilk Normality goodness-of-fit (GOF) test. The pooled set of ordinary residuals from the three ANOCOVA models clearly fail this test ($p < 0.0001$); the associated residual quantile plot (figure 4.5) suggests that the residual distribution is somewhat "heavy-tailed". Figure 4.6 displays the residual variance pattern across the 77 surveyed fields; this plot shows that the degree of residual variation is field dependent. However, after standardizing the pooled set of residuals by their individual field variance estimates, the new (variance standardized) residuals pass the Shapiro-Wilk GOF test ($p = 0.540$); note that the corresponding standardized residual quantile plot is shown in figure 4.7.

Overall, these residual diagnostic results suggest that the ANOCOVA model errors do follow a Normal distribution, but that the variance of the distribution changes across the 77 fields. The lower portion of table 4.5 shows the formal Chi-square test results for the non-constant residual variance hypothesis (i.e., the likelihood ratio tests associated with the mixed linear ANOCOVA models estimated using restricted maximum likelihood). The test results for all three ANOCOVA models are highly significant ($p < 0.0001$), implying that these heterogeneous variance ANOCOVA models should be used to produce our final set of $\ln(EM_V)$ and $\ln(EM_H)$ parameter coefficients.

Table 4.6 shows the calculated ANOCOVA model $\ln(EM_V)$ and $\ln(EM_H)$ parameter coefficients, for models estimated under both the homogeneous (common) and heterogeneous (field specific) variance assumptions; the estimated standard errors of the coefficients are shown in ( ). These standard errors confirm that the model coefficients associated with the heterogeneous variance (mixed linear) ANOCOVA models are more accurately estimated. These are the coefficients that should be used by the Coachella Valley RCD in all future survey operations where the ANOCOVA modeling approach is employed.

Table 4.5. ANOCOVA model residual errors; average depth correlation structure and pooled residual distribution assessments.

Midpoint Sample Depth	Residual Variance	Residual Correlation Matrix			
		Depth	0.15 cm	0.45 cm	0.75 cm
0.15 cm	0.302	0.15 cm	1.00	0.48	0.21
0.45 cm	0.326	0.45 cm	0.48	1.00	0.49
0.75 cm	0.349	0.75 cm	0.21	0.49	1.00

Shapiro-Wilks Normality Tests: polled residuals (N = 1568)		
Residuals	W-score	p-value
Ordinary residuals	0.9867	< 0.0001
Variance standardized residuals (by field)	0.9990	0.5400

Depth-specific χ^2 tests for non-constant residual variance across fields.			
Model / Sample Depth	χ^2-score	DF's	p-value
0.15 cm ANOCOVA model	189.9	76	< 0.0001
0.45 cm ANOCOVA model	144.4	76	< 0.0001
0.75 cm ANOCOVA model	139.6	58	< 0.0001

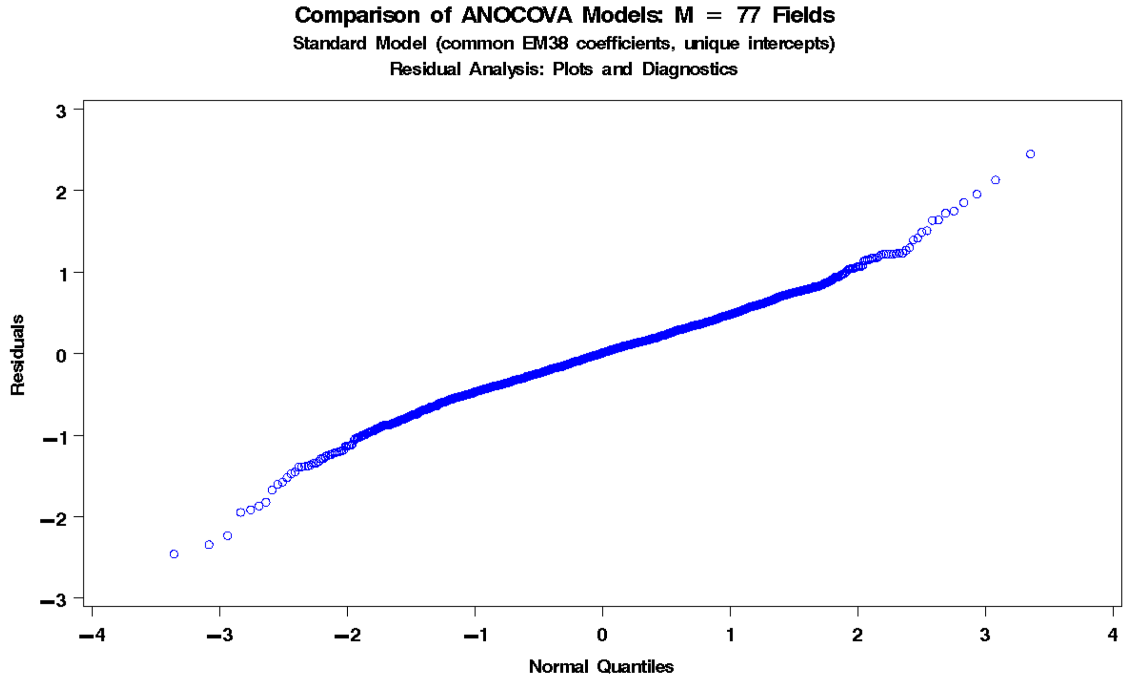

Figure 4.5. Residual QQ plot: ordinary ANOCOVA model residuals.

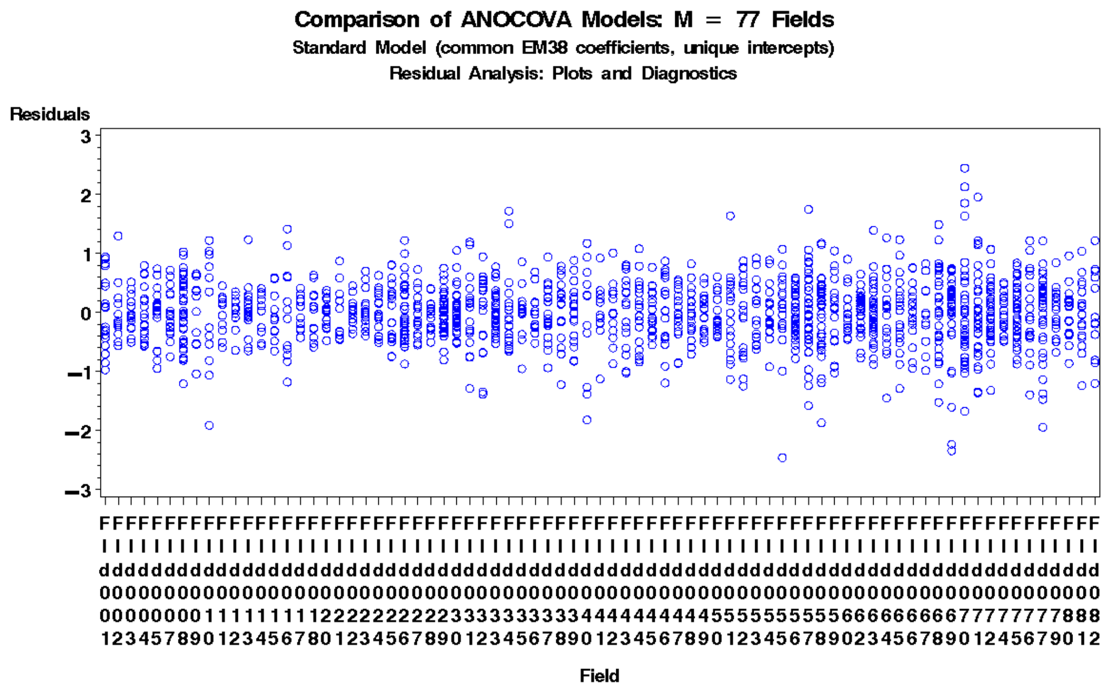

Figure 4.6. ANOCOVA residual variance plot, stratified by field.

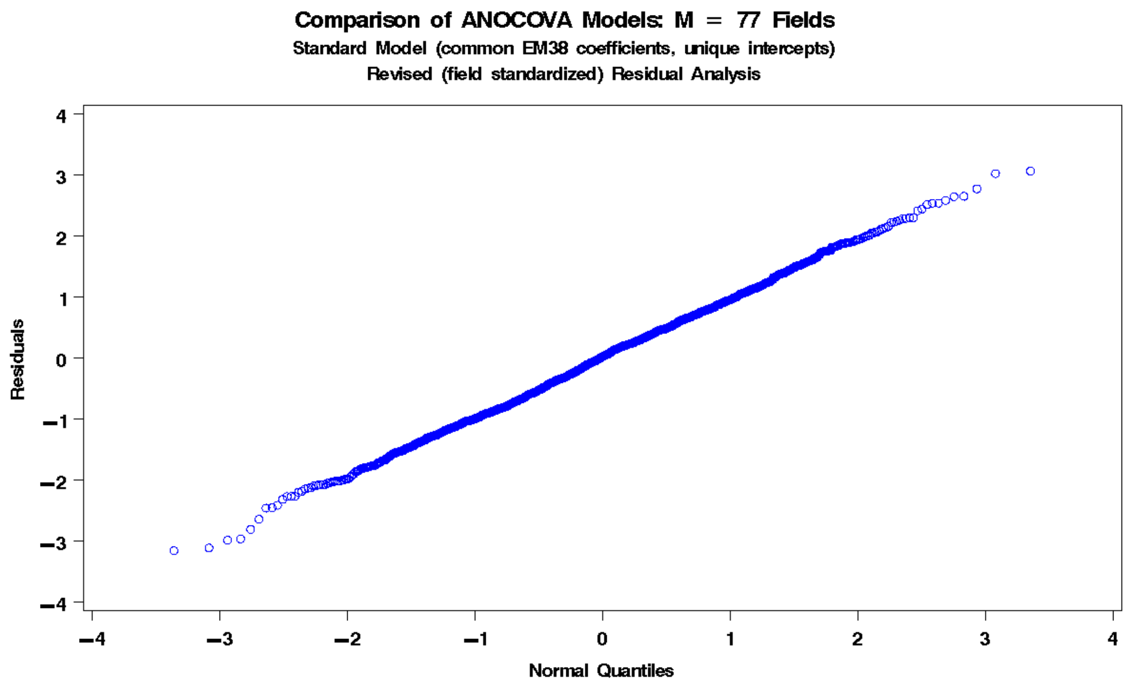

Figure 4.7. Residual QQ plot: variance (field) standardized ANOCOVA model residuals.

Table 4.6. Calculated ANOCOVA model $\ln(EM_V)$ and $\ln(EM_H)$ parameter coefficients, for models estimated under both the homogeneous (common) and heterogeneous (field specific) variance assumptions; estimated standard errors of the parameter coefficients shown in ( ).

Midpoint Sample Depth	Common variance assumption		Field specific variance assumption	
	$\ln(EM_V)$	$\ln(EM_H)$	$\ln(EM_V)$	$\ln(EM_H)$
0.15 cm	-0.696 (.170)	1.557 (.159)	-0.413 (.128)	1.178 (.117)
0.45 cm	0.114 (.177)	1.219 (.165)	0.150 (.127)	1.080 (.119)
0.75 cm	0.833 (.191)	0.317 (.178)	0.976 (.123)	0.129 (.117)

5. IMPLEMENTATION

Given the depth-specific EM38 parameter coefficients shown in Table 4.6, we can now use simple algebraic techniques to calculate an appropriate (depth-specific) intercept parameter estimate for any field surveyed using the ANOCOVA modeling approach. Section 5.1 below explicitly describes how these intercept parameter estimates should be calculated, along with a working example using one of the CVRCD fields included in the database. Section 5.2 then discusses how the ESAP-RSSD software program can be used to implement a sampling strategy that selects either four, five, or six sampling locations per field to facilitate the optimal estimation of these intercept estimates.

5.1 Calculation Details

As discussed in section 4, the ANOCOVA modeling approach can be used to make 30 cm depth predictions at the 0.15 m, 0.45 m, and 0.75 m sampling depths. The corresponding ANOCOVA model equation are

$$\ln\left(EC_{0.15,i}\right) = \hat{\beta}_{0.15} - 0.413\ln(EM_{V,i}) + 1.178\ln(EM_{H,i}) + \varepsilon_{0.15,i} \tag{5.1}$$

$$\ln\left(EC_{0.45,i}\right) = \hat{\beta}_{0.45} + 0.150\ln(EM_{V,i}) + 1.080\ln(EM_{H,i}) + \varepsilon_{0.45,i} \tag{5.2}$$

and

$$\ln\left(EC_{0.75,i}\right) = \hat{\beta}_{0.75} + 0.976\ln(EM_{V,i}) + 0.129\ln(EM_{H,i}) + \varepsilon_{0.75,i}\,, \tag{5.3}$$

where the subscript i = 1, 2, .. N refers to the i^{th} sampling and/or survey location in a particular field. Based on the available sample soil salinity data, we can use Eqns. (5.1), (5.2) and (5.3) to compute the intercept estimates and the corresponding pooled mean square prediction error (MSPE) estimate (which represents a measurement of how well the ANOCOVA model fits the data – i.e., see table 4.4 in section 4.1).

To calculate the intercept estimates, we first compute the differenced log salinity / EM readings for each sample depth as follows:

$$y_{0.15,i} = \ln\left(EC_{0.15,i}\right) + 0.413\ln(EM_{V,i}) - 1.178\ln(EM_{H,i}) \tag{5.4}$$

$$y_{0.45,i} = \ln\left(EC_{0.45,i}\right) - 0.150\ln(EM_{V,i}) - 1.080\ln(EM_{H,i}) \tag{5.5}$$

$$y_{0.75,i} = \ln\left(EC_{0.75,i}\right) - 0.976\ln(EM_{V,i}) - 0.129\ln(EM_{H,i}) \tag{5.6}$$

The intercept estimates are then simply the average values of these differenced estimates; i.e.,

$$\hat{\beta}_{0.15} = (1/n)\sum_{i=1}^{n} y_{0.15,i} \tag{5.7}$$

$$\hat{\beta}_{0.45} = (1/n)\sum_{i=1}^{n} y_{0.45,i} \tag{5.8}$$

$$\hat{\beta}_{0.75} = (1/n)\sum_{i=1}^{n} y_{0.75,i} \quad , \tag{5.9}$$

where n represents the number of sampling locations in the field. Once the intercepts have been obtained, the depth-specific MSPE estimates can then be calculated as

$$\hat{\sigma}^2_{0.15} = (1/(n-1))\sum_{i=1}^{n} (y_{0.15,i} - \hat{\beta}_{0.15,i})^2 \tag{5.10}$$

$$\hat{\sigma}^2_{0.45} = (1/(n-1))\sum_{i=1}^{n} (y_{0.45,i} - \hat{\beta}_{0.45,i})^2 \tag{5.11}$$

and

$$\hat{\sigma}^2_{0.75} = (1/(n-1))\sum_{i=1}^{n} (y_{0.75,i} - \hat{\beta}_{0.75,i})^2 \quad , \tag{5.12}$$

respectively. Finally, the pooled MSPE can be calculated by simply averaging the depth-specific MSPE estimates and then used to "grade" the overall prediction reliability, etc.

As can be seen from these derivations, the ANOCOVA parameter estimation technique is equivalent to simply calculating the means and variances of the differenced log salinity / EM data for each sample depth. Such calculations can be easily carried out using any spreadsheet program (such as Microsoft Excel), or even using a desktop calculator if necessary.

Table 5.1 shows a set of example calculations for Field #49 in the composite database. Columns three, four and five show the EC_e (dS/m), EM_V (mS/m), and EM_H (mS/m) data for the six sampling locations associated with this field, while column six shows the corresponding differenced data (computed from Eqns. 5.4, 5.5, and 5.6). The means and variances associated with these differenced readings are shown in columns seven and eight. Recall that the means represent the intercept estimates and the variances can be pooled (averaged) together to produce the field MSPE estimate. In this example, we find that $\hat{\beta}_{0.15} = -1.659$, $\hat{\beta}_{0.45} = -3.383$, and $\hat{\beta}_{0.75} = -2.993$, and the pooled MSPE is 0.127. As discussed in section 4.1, an MSPE < 0.15 implies excellent prediction accuracy, so in this example we would have high confidence in the corresponding ANOCOVA model salinity predictions.

Figure 5.1 shows the ESAP generated EM_V map for Field #49, while figure 5.2 shows the corresponding 0-90 cm bulk average EC_e map. The salinity map was generated by first computing the individual 0.15, 0.45, and 0.75 cm depth predictions using Eqns. 5.1 through 5.3 and then averaging these (back-transformed) values together. For this field, we find that the field-wide average 0-90 cm salinity level is 2.06 dS/m and that 90% of the individual predictions fall between 1.55 dS/m and 2.78 dS/m. Figure 5.2 shows that the majority of this field can be classified as non-saline, with some slightly saline areas located along the eastern half of the field and in the northwest corner.

Table 5.1. Example ANOCOVA model calculations (intercept estimates and MSPE) for Field #49 in the composite database.

Depth	Site	EC_e	EM_V	EM_H	Δ_i	Mean	Var
0.15 cm	137	0.91	24.38	18.13	-2.189	-1.659	0.1201
	581	2.99	40.00	30.13	-1.393		
	1178	2.25	29.00	25.00	-1.590		
	1451	2.72	30.88	26.38	-1.438		
	2100	1.42	20.63	21.00	-1.986		
	2379	4.16	38.75	38.38	-1.361		
0.45 cm	137	0.91	24.38	18.13	-3.703	-3.383	0.0438
	581	3.00	40.00	30.13	-3.133		
	1178	1.65	29.00	25.00	-3.481		
	1451	2.26	30.88	26.38	-3.234		
	2100	1.30	20.63	21.00	-3.480		
	2379	3.38	38.75	38.38	-3.270		
0.75 cm	137	1.68	24.38	18.13	-2.972	-2.993	0.2171
	581	1.70	40.00	30.13	-3.509		
	1178	1.48	29.00	25.00	-3.310		
	1451	3.89	30.88	26.38	-2.412		
	2100	1.05	20.63	21.00	-3.298		
	2379	4.87	38.75	38.38	-2.457		
Pooled MSPE = (0.1201 + 0.0438 + 0.2171) = 0.127 (Prediction Accuracy grade = Excellent)							

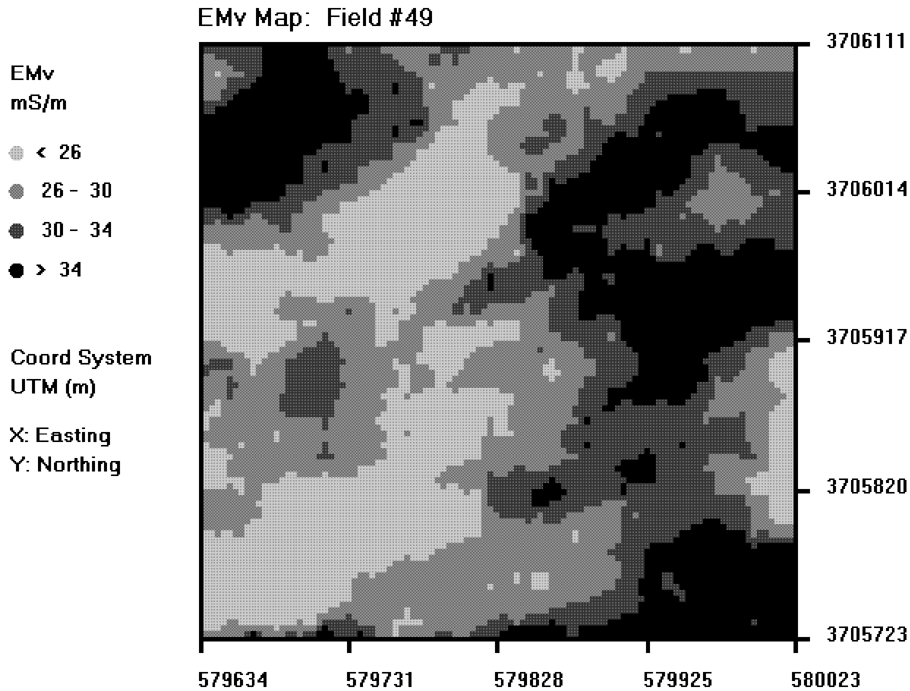

Figure 5.1. EMv map for Field #49.

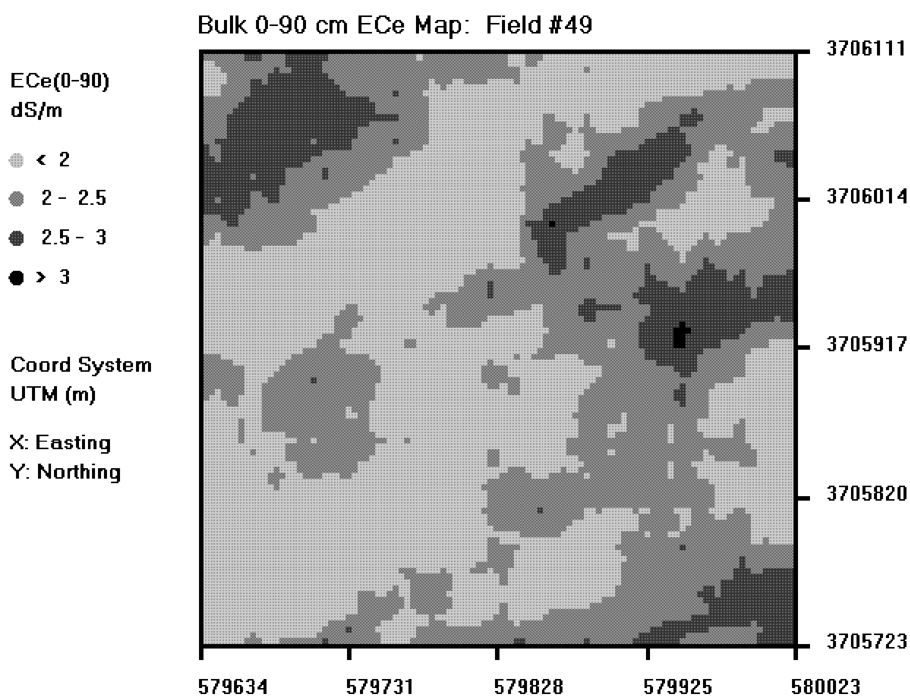

Figure 5.2. Bulk average (0-90 cm) ECe map for Field #49.

5.2 Generating ANOCOVA Sampling Designs using ESAP-RSSD Software

The ESAP-RSSD software program can generate sampling designs containing either n = 6, 12, or 20 sample locations, respectively (Lesch et al, 2000). Whenever the ANOCOVA modeling approach is adopted (for a given field), then four to six sites should be sufficient to produce reliable intercept estimates. (Note: acquiring < 4 sites in any field is not recommended.) In practice, a six site RSSD sampling design can be employed to generate a sampling plan containing four to six sites.

Every 6 site ESAP-RSSD sampling plan contains four sites that correspond to spatial response surface sampling (SRSS) locations, one additional center-point site (also a SRSS site, but less important) and one extra "spatial-support" site. These sites are clearly identified in the ESAP generated response surface design text file (created by the RSSD program whenever a sampling design is saved). The output shown in table 5.2 below shows the response surface design text file for Field #49. Note that the first listed sampling location (site 1178) represents the center-point site and the last location (site 1451) represents the spatial-support site, respectively.

Table 5.2. The response surface design text file for Field #49.

```
Full Path:     C:\US_Salinity_Lab\esap2\data\RanchoWN40\RN40rsd1.txt
Project:       RanchoWN40
File Name:     RN40rsd1.txt
Date & Time:   7/9/04 10:13:33 AM
Field Desc:    north 40 ac-prep for strawberries

Sample Size:   6   (Total Survey Size =  2516   Active Survey Size =  2130)
D-Factor Val: 0.90
Opt-Criteria: 1.24
Loop Count:    3

Parsing Algorithm Empolyed:  forced data stratification using Strata # 3

Target Information for SRS Sampling Design # 1

  Site ID      Design Levels   Ds1-STD  Ds2-STD   X-Coordinate   Y-Coordinate

   1178         0       0      -0.11     0.03      579840.5       3705794.4
   581          1.58    1.58    1.31     1.35      579924.7       3706049.8
   2100        -1.58   -1.58   -1.55    -1.62      579706         3705804.1
   2379         1.58   -1.58    1.92    -1.6       579656.4       3706033.4
   137         -1.58    1.58   -1.51     1.7       579997         3705828.3
   1451        support site     0.22     0.1       579802.3       3705941
```

To generate either a four, five, or six site sampling plan in ESAP, one should first use the RSSD program to generate a suitable 6-site design. If a five site plan is desired, then one can simply not collect any sample data at the support site location during the actual sampling process. (The support site represents the "statistically least important" sampling location in every 6-site ESAP sampling plan.) If a four site plan is desired, then one should drop out both the support site and the center-point site; i.e., the site exhibiting Design Level values of (0, 0) in table 5.2.

Figure 5.3 below shows the 6-site ESAP sampling plan for Field #49. As discussed above, a statistically optimal 5-site sampling plan would exclude site 1451. Likewise, a statistically optimal 4-site sampling plan would exclude sites 1178 and 1451. Hence, this same 6-site ESAP sampling design can be used to generate either a four, five, or six site sampling plan, respectively.

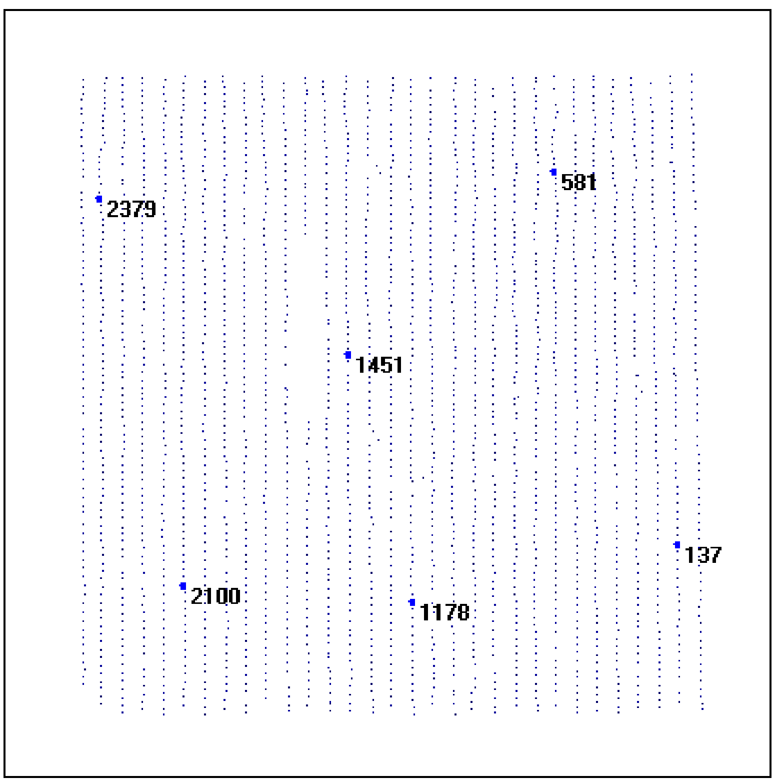

Figure 5.3. The ESAP-RSSD 6-site sampling plan for Field #49.

6.0 IRRIGATION & LEACHING EFFECTS ON FIELD AVERAGE SALINITY LEVELS

As discussed in section 2, this database of surveyed fields provides a robust and realistic data set for testing and validating the ANOCOVA modeling approach. However, this database also facilitates the testing of other salinity related hypotheses of interest to the Reclamation and Coachella Valley agricultural producers. One hypothesis of significant interest is if different irrigation and/or leaching techniques exhibit different degrees of success at controlling the field average salinity levels. More specifically, is there statistical evidence that one particular irrigation and/or leaching method is most effective at controlling (i.e., minimizing) these average salinity levels.

Overall, exactly 70 of the 77 fields in this database were supporting some type of vegetable, herb, or fruit crop when the EM surveys were conducted. Table 6.1 shows the frequency breakdown of both the irrigation and leaching techniques used in these 70 fields. Table 6.1 also shows the number of fields with and without tile lines (note that tile line information was unavailable for one field). As shown in this table, about 90% of these fields were irrigated using either drip or furrow methods. Likewise, slightly more than one half of these fields were sprinkler leached, and all but two of the remaining fields were either flood-leached or leached using a combination of flood and sprinkler techniques.

In section 6.1 we present a summary of the average soil salinity (EC_e, dS/m) and texture (SP, %) levels for these 70 fields. In section 6.2, we propose a suitable statistical model for analyzing what effect (if any) these irrigation and leaching methods have on the log transformed field average salinity levels. The results obtained from this modeling analysis are then discussed and interpreted in section 6.3.

Table 6.1. Frequency breakdown of irrigation and leaching techniques (and tile line status) used in the 70 fields supporting vegetable, herb, or fruit crops.

Irrigation Method	Frequency	Percent (%)
Drip	29	41.4
Furrow	35	50.0
Sprinkler	6	8.6
Leaching Method	Frequency	Percent (%)
Flood	19	27.1
Sprinkler	37	52.9
Flood & Sprinkler	12	47.1
Drip	2	2.9
Tile Lines	Frequency	Percent (%)
Present	55	78.6
Not Present	14	20.0
No Information Available	1	1.4

6.1 Field average EC_e and SP information

Table 6.2 shows the field average EC_e and SP univariate summary statistics for the 70 fields supporting vegetables, herbs, or fruit crops, where these averages have been calculated using all of the available (individual site) sample data in the database. In table 6.2, field average statistics are shown for each of the three sampling depths of interest; i.e., 0.15 m, 0.45 m, and 0.75 m. These salinity and texture readings appear to be consistent across the three sampling depths, although sample information from the 0.75 m depth is only available for 54 of the 70 fields. Additionally, the field average salinity levels appear to be reasonably similar to the (individual site) salinity levels shown in table 2.8.

Perhaps not surprisingly, the log transformed average salinity readings are moderately correlated with the field average SP levels (r = 0.53 to 0.59 across the three sample depths). It is well known that soil salinity levels tend to increase in heavier textured soils; note that the SP levels indirectly quantify the percentage of silt and clay in the soil. In general, the correlated field average SP levels should be included as a covariate when statistically analyzing for differential irrigation and/or leaching effects on field average (log) salinity levels (in order to adjust for the effects of different soil texture conditions across fields).

Table 6.2. Field average soil salinity (EC_e, dS/m) and texture (SP, %) summary statistics for the 70 fields supporting vegetable, herb, or fruit crops.

Statistic	0.15 m		0.45 m		0.75 m	
	EC_e	SP	EC_e	SP	EC_e	SP
N	70	70	70	70	54	54
Mean	5.03	38.11	4.37	38.53	4.07	38.64
Standard Deviation	7.69	7.80	6.36	8.08	4.45	9.21
Skewness	3.09	0.49	2.76	0.52	2.29	0.83
Quantiles:						
Minimum	0.31	24.0	0.27	24.3	0.80	24.1
10%	1.03	29.0	0.94	28.9	1.09	27.4
25%	1.30	32.8	1.19	33.1	1.27	32.6
Median (50%)	2.17	35.8	1.77	36.7	2.51	37.8
75%	4.35	44.6	3.68	45.0	4.50	44.0
90%	12.97	48.6	13.52	49.8	10.77	49.9
Maximum	41.65	56.8	34.60	61.7	22.25	68.8

6.2 Statistical modeling methodology

Analysis of variance modeling techniques can be used to formally test if different irrigation or leaching methods employed in the Coachella Valley influence the (log transformed) field average salinity levels. Taking $\ln(EC_{e,i})$ to represent the log transformed value of the field average salinity level in the ith field (for i = 1, 2, .. , 70), a first order analysis of variance model can be specified as

$$\ln(EC_{e,i}) = \mu + \theta_j + \delta_k + \beta_1(SP_i) + \beta_2(TL_i) + \varepsilon_i \qquad (6.1)$$

In Eq. (6.1), θ_j quantifies the three different irrigation techniques, δ_k quantifies the four different leaching methods, β_1 adjusts for differences in the field average SP levels, β_2 adjusts for tile line effects (TL = 1 if there are tile lines, 0 otherwise), and the error term (ε) is assumed to satisfy the usual Normality assumptions (Montgomery, 2001). Technically, Eq. (6.1) is actually another ANOCOVA model, but our purpose here is now statistical inference (rather than prediction) and the covariates in this model (the SP and tile line effects) essentially represent nuisance parameters.

Provided that the residual errors are assumed to be Normally distributed and spatially uncorrelated, standard OLS estimation techniques can be used to fit Eq. (6.1). Additionally, ordinary (ANOVA-type) F tests can be used to quantify the statistical significance of different irrigation and/or leaching techniques (Montgomery, 2001). In such an analysis, note that non-significant F test results would at least heuristically suggest that the different irrigation and leaching techniques do not differentially influence the field average salinity levels, etc.

6.3 Results and Discussion

Table 6.3 presents the basic model summary statistics associated with Eq. (6.1) for the 0.15 m, 0.45 m, and 0.75 m sample depths. As shown in this table, the specified ANOCOVA model was only able to explain 30% to 45% of the total variation in the (log transformed) field average salinity levels. None the less, the residual errors passed the Normality test for all three sampling depths and the residual quantile plots were devoid of any outliers (data not shown).

Table 6.3. Eqn. (6.1) ANOCOVA model summary statistics.

Model Statistics	Sampling Depth		
	0.15 m	0.45 m	0.75 m
R^2	0.354	0.300	0.457
CV (%)	84.1	101.5	66.1
Root MSE	0.851	0.889	0.674

Table 6.4. F test p-values for the irrigation and leaching effects, in addition to the tile line and SP covariate parameter estimates.

Effect	Sampling Depth		
	0.15 m	0.45 m	0.75 m
Irrigation Method	0.756	0.409	0.406
Leaching Method	0.119	0.952	0.324
Tile Lines	0.644	0.854	0.442
Field Average SP	< 0.0001	< 0.0001	< 0.0001

Table 6.4 shows the F test p-values associated with the irrigation and leaching effects, in addition to the tile line and field average SP covariate effects. As shown in table 6.4, neither the irrigation or leaching method effect is statistically significant in any of the three ANOCOVA models. These somewhat surprising results suggest that the (log transformed) average salinity levels in these fields are <u>not</u> differentially influenced by either the irrigation or leaching methods employed across the valley. Or equivalently, with respect to controlling the field average salinity level, it does not appear to matter which irrigation method (sprinkler, furrow, or drip) or leaching method (flood, sprinkler, or flood and sprinkler) is used on a particular field. Any one of these irrigation and leaching methods can be employed with equal effectiveness for controlling the soil salinity levels in a typical Coachella Valley vegetable field.

7. REFERENCES

Johnson, R.A., and D.W. Wichern. 1988. Applied Multivariate Statistical Analysis, 2nd Ed. Prentice Hall, Englewood Cliffs, New Jersey.

Lesch, S.M., and D.L. Corwin. 2003. Using the dual-pathway parallel conductance model to determine how different soil properties influence conductivity survey data. Agronomy J., 95: 365-379.

Lesch S.M., Rhoades, J.D., and D.L. Corwin. 2000. ESAP-95 Version 2.01R: User Manual and Tutorial Guide. Research report 146. USDA-ARS US Salinity Laboratory, Riverside, CA, USA.

McCulloch, C.E., and S.R. Searle. 2001 Generalized, Linear, and Mixed Models. John Wiley & Sons, New York, NY.

Milliken, G.A., and D.E. Johnson. 2002. Analysis of Messy Data. Volume III: Analysis of Covariance. Chapman & Hall / CRC Press, Boca Raton, FL.

Montgomery, D.C. 2001. Design and Analysis of Experiments, 5th Ed. John Wiley & Sons, New York, NY.

Myers, R.H. 1986. Classical and Modern Regression with Applications. Duxbury Press, Boston MA.

Rao, C.R., and H. Toutenburg. 1999. Linear Models: Least Squares and Alternatives, 2nd Ed. Springer, New York, NY.