Desarrollo de técnicas estadísticas basadas en computación intensiva para el análisis de medidas repetidas

Tesis doctoral realizada por:

José Antonio Sánchez Espigares

Dirigida por:

Jordi Ocaña Rebull

Esta tesis se presenta como compendio de ocho artículos:

1. ***Semiparametric Bootstrap in Linear Mixed Models with Misspecified Skewness and Kurtosis***

   Enviado a publicación JCR

2. ***Visualizing type II error in normality tests***

   American statistician, vol. 72, núm. 2, pàgs. 158-162 (2018)

3. ***Graphical comparison of normality tests for unimodal distribution data***

   Journal of statistical computation and simulation, vol. 89, núm. 1, pàgs. 145-154 (2019)

4. ***Mosaic Normality Test***

   Communication in Statistics—Theory and Methods (2020)

5. ***Transfer function and time series outlier analysis: modelling Soil salinity in loamy sand soil by including the influences of irrigation management and soil temperature***

   Irrigation and drainage, vol. 67, núm. 2, pàgs. 282-294 (2017)

6. ***An advanced process for evaluating a linear dielectric constant–bulk electrical conductivity model using a capacitance sensor in field conditions***

   Hydrological sciences journal, vol. 60, núm. 10, pàgs. 1828-1839 (2015)

7. ***Estimating pore water electrical conductivity of sandy soil from time domain reflectometry records using a time-varying dynamic linear model***

   Sensors, vol. 18, núm. 12, pàgs. 4403-4414 (2018)

8. ***Evaluating the variation of dissolved metals on a highway roadside using a generalized additive mixed model (GAMM)***

   Water, Air, and Soil Pollution, vol. 230, núm. 93 (2019)

A mis padres, a los que debo el ser como soy

AGRADECIMIENTOS

A lo largo del amplio periodo de desarrollo de este trabajo he tenido la fortuna de coincidir con muchos compañeros y amigos a los cuales desearía agradecer todas las enseñanzas y apoyos que me han permitido llegar a este punto.

En primer lugar, donde empezó todo hace ya mucho tiempo, en el departamento de Estadística de la Universitat de Barcelona. A continuación con los compañeros de la sección de Informática del departamento de Estadística e Investigación Operativa de la UPC y para acabar con los de la sección de TQG del mismo departamento. Debería incluir a todos y cada uno de los componentes de estos colectivos con los que he podido compartir momentos de docencia, investigación y diversión. La lista es muy larga pero querría mencionar algunos de ellos:

Agradezco a Jordi Ocaña, mi director, por la paciencia que ha tenido con mi vórtice de entropía.

También a Xavier Tort, tutor de mi tesis y que ya me daba por perdido.

Así mismo a Pere Grima que me facilitó la energía de activación para salir del estado de letargia.

En el recuerdo, a Manuel Martí-Recober, con el que aprendí mucho más que series temporales.

También deseo agradecer a Pilar Muñoz por haberme permitido trabajar y aprender con ella.

Para finalizar, agradezco a toda mi familia, hermanos, cuñadas y sobrinos, por soportar mi eterno "ya casi está acabada la tesis". A mi sobrina Ana que está haciendo el doctorado, espero que no siga mi ejemplo en cuanto a duración de la tesis.

ABSTRACT

This thesis includes contributions to statistical methodologies and practical applications in situations where data has been obtained as a result of repeated measures.

The first block includes a first article where different resampling methods in mixed linear models are evaluated by simulation against deviations from the assumptions of symmetry and Gaussian kurtosis of the variance components. One of these methods constitutes our proposal, designed and implemented in R.

The result of the previous simulations shows that the misspecifications induced by the simulation scenarios do not have the same properties if they affect random effects or residuals. Taking into account that both components of the variance tend to have different sample sizes, an exploration of the impact of the misspecification of the Gaussian distribution implemented in the simulation of scenarios is proposed, as a function of the sample size.

The second article of this block proposes a visualization of the type II error in goodness-of-fit test to the normal distribution, where the alternatives considered correspond to unimodal distributions with different asymmetry and kurtosis. In this article, a mosaic visualization is proposed to assess the capability to reject the hypothesis of normality of the test, taking into account the type of misspecification considered and the sample size.

The third article compares different normality tests based on the type II error represented in the previously designed mosaic. Considering as a set of alternative distributions those included in the unimodal parametric family included in the mosaic, the graphic representation obtained gives an idea of the ability of the tests to distinguish deviations from normality. It would be as an equivalent to the power curves obtained in tests with numerical parameter space.

The fourth article presents the application of the goodness-of-fit test to the distribution that represents each mosaic cell for a specific sample and the representation of the p-value of the test according to a gray scale. This last article aims to reflect the set of distributions of the family considered to be compatible with the sample evaluated, as a result of the concurrent application of the set of tests associated with the mosaic cells.

The second part of the thesis includes four publications resulting from the collaboration with Dr. Basem Aljoumani on applied topics of hydrology. The inclusion of these articles in this thesis is justified mainly by the statistical contributions to the analysis of hydrological data.

The fifth article focuses on the extension of ARIMA models in the context of prediction of water content in the field, including intervention analysis (scheduled irrigation), atypical detection (rainfall) and transfer function (temperature information and water content to predict salinity).

The sixth and seventh articles propose an extension of a classical deterministic model for the evaluation of salinity in the field to incorporate a stochastic component that allows the model to be adjusted to the data based on statistical criteria. The transformation of the model allows formulating it in a linear and Gaussian state space and applying the Kalman filter for its treatment. The difference between the two articles is based on the fact that in the first one the data is obtained in the field, while in the other one a laboratory experiment is performed.

The eighth article analyzes data on the contents of contaminating metals on a roadside near a highway based on a series of covariates. Linear mixed additive models is the technique applied in this context.

RESUMEN

Esta tesis incluye aportaciones a metodologías estadísticas y aplicaciones prácticas en situaciones donde los datos se han obtenido como resultado de medidas repetidas.

El primer bloque incluye un primer artículo donde se evalúan mediante simulación diferentes métodos de remuestreo en modelos lineales mixtos frente a desviaciones de los supuestos de simetría y curtosis gaussianas de las componentes de varianza. Uno de esos métodos constituye nuestra propuesta, diseñada e implementada en R.

El resultado de las simulaciones anteriores pone de manifiesto que las especificaciones erróneas inducidas por los escenarios de simulación no presentan las mismas propiedades si afectan a efectos aleatorios o a residuos. Teniendo en cuenta que ambas componentes de la varianza suelen presentar diferente tamaño de muestra, se plantea una exploración del impacto de la especificación errónea de la distribución gaussiana implementada en la simulación de escenarios, como función del tamaño de muestra.

El segundo artículo de este bloque propone una visualización del error de tipos II en test de bondad de ajuste a la normal, donde las alternativas consideradas corresponden a distribuciones unimodales con diferente asimetría y curtosis. En este artículo se propone una visualización en forma de mosaico para valorar la capacidad de rechazar la hipótesis de normalidad del test, atendiendo al tipo de especificación errónea considerada y el tamaño de la muestra.

El tercer artículo compara diferentes tests de normalidad en base al error de tipo II representado en el mosaico diseñado previamente. Considerando como conjunto de distribuciones alternativas las incluidas en la familia paramétrica unimodal que conforma el mosaico, la representación gráfica obtenida da una idea de la capacidad de los tests para distinguir las desviaciones de la normalidad. Sería como un equivalente a las curvas de potencias obtenidas en tests con espacio de parámetros numéricos.

El cuarto artículo plantea la aplicación del test de bondad de ajuste a la distribución que representa cada celda del mosaico para una muestra concreta y la representación del p-valor del test según una escala de grises. Este último artículo pretende reflejar el conjunto de distribuciones de la familia considerada que son compatibles con la muestra evaluada, como resultado de la aplicación concurrente del conjunto de tests asociado las celdas del mosaico.

La segunda parte de la tesis incluye cuatro publicaciones resultados de la colaboración con el Dr. Basem Aljoumani en temas aplicados de hidrología. La inclusión de estos artículos en esta tesis se justifica principalmente por las aportaciones estadísticas al análisis de datos de datos hidrológicos.

El quinto artículo se centra en la extensión de los modelos ARIMA en el contexto de la predicción del contenido de agua en el terreno, incluyendo análisis de intervención (riegos programados), detección de atípicos (lluvias puntuales) y función de transferencia (información de temperatura y contenido de agua para predecir la salinidad).

El sexto y séptimo artículos proponen una extensión de un modelo clásico determinista para la evaluación de la salinidad en el terreno para incorporar una componente estocástica que permita ajustar el modelo a los datos en base a criterios estadísticos. La transformación del modelo permite formularlo en espacio de estado lineal y gaussiano y aplicar el filtro de Kalman para su tratamiento. La diferencia entre ambos artículos se basa en que en el primero de ellos los datos son obtenidos en el terreno, mientras que en el otro se realiza un experimento en laboratorio.

El octavo artículo analiza datos de contenidos de metales contaminantes en los bordes de una autopista en base a una serie de covariables. Los modelos lineales aditivos mixtos es la técnica aplicada en este contexto.

Índice

Listado de Figuras

Lista de Tablas

1

Introducción

La estadística es una disciplina de ámbito claramente transversal a muchas áreas. Es una potente herramienta de análisis de datos que se aplica en diferentes campos como, por citar algunos, biología, medicina, economía, industria y en muchos otros. Este carácter instrumental permite que a la hora de realizar contribuciones en estadística se puedan considerar dos tipos de aportaciones: las centradas en desarrollar, implementar y evaluar mejoras en las metodologías estadísticas y las que corresponden a una aplicación de técnicas estadísticas para la resolución de problemas concretos.

El primer tipo de contribuciones permite extender las capacidades de la propia herramienta, dotándola de métodos robustos, innovadores y mejorados para análisis concretos. Las publicaciones de este tipo suelen incorporar el desarrollo teórico donde se enmarca la técnica y se completa con ejercicios de simulación y la aplicación casos reales, lo cual pretende habilitar el método para ser aplicado en situaciones similares a las presentadas.

En el segundo tipo de contribuciones se parte de un problema concreto que se presenta en un ámbito donde la estadística actúa como instrumento. La aplicación de la técnica estadística apropiada permite el análisis y la interpretación de los resultados de forma práctica. Las publicaciones aplicadas suelen incorporar la descripción de las técnicas estadísticas y sus especificidades en apartados que describen los "Materiales y Métodos".

En esta tesis se incluyen publicaciones de ambos tipos. El primer bloque contiene cuatro artículos en los que se pretende aportar extensiones metodológicas de carácter estadístico de forma genérica. El planteamiento original de esta parte se centra en mejorar las técnicas de inferencia en modelos mixtos como herramienta de análisis de medidas repetidas. Se proponen extensiones de tipo Bootstrap, que utilizan computación intensiva, para la obtención de intervalos de confianza y resolución de tests de hipótesis. Si bien existen múltiples contribuciones de este tipo en el contexto de los modelos lineales y lineales generalizados, donde sólo hay una componente aleatoria, la inclusión de efectos aleatorios en el predictor lineal implica la existencia de más de una componente de variabilidad. En este contexto, los métodos de remuestreo no son tan directos. El principal interés de disponer de técnicas alternativas a las clásicas, que suelen basarse en resultados asintóticos, es la posibilidad de analizar su rendimiento en situaciones donde no se cumplen de forma clara las condiciones de aplicación. Por ejemplo, si la muestra

es pequeña o la distribución de las componentes de variabilidad está incorrectamente especificada los resultados asintóticos pueden ser cuestionados. Por ello, en presencia de no normalidad de efectos aleatorios y residuos se plantean métodos de remuestreo alternativos en modelos lineales mixtos y se analiza la inferencia resultante, principalmente sobre los parámetros que describen las componentes de variabilidad.

El resto de las publicaciones de este bloque surgen del análisis de los resultados del primer artículo. En dicho artículo, la incorrecta especificación de la distribución de efectos aleatorios y residuos se concreta en modificaciones de su tercer y cuarto momento (asimetría y *Kurtosis*) respeto a la distribución Normal. En concreto, las simulaciones incluidas en el artículo utilizan como modelo teórico un ejemplo clásico y se sustituye la distribución Normal de efectos aleatorios y residuos por distribuciones con asimetría y Kurtosis diferente, Exponencial y Uniforme respectivamente. El efecto producido por estas desviaciones en los intervalos de confianza asintóticos da lugar a observar infrarecubrimiento si hay asimetría y sobrerecubrimento en presencia de baja kurtosis. Este comportamiento no se observa de igual manera en la inferencia para los parámetros de los efectos aleatorios y de los residuos. Por ello, tiene sentido plantearse el efecto del tamaño de muestra en los resultados. Para el modelo simulado, el número de valores para los efectos aleatorios es menor que el número de residuos.

El objetivo de los tres artículos se centra en usar una representación visual original para el análisis de bondad de ajuste ante especificaciones de una familia de distribuciones que incluyen cambios en la asimetría y *Kurtosis*. El segundo artículo presenta la representación gráfica correspondiente a un mosaico donde cada posición está asociada una distribución de la familia de distribuciones de Potencia Exponencial y Asimétrica (SEPD, *Skewed Exponential Power Distribution*). Esta familia incluye como casos particulares la distribución Normal (punto interior) y las distribuciones Exponencial y Uniforme (puntos frontera), entre otras. Se incluye la representación del error de tipo 2, obtenido mediante simulación a partir del tamaño de muestra, en forma de mancha sobre el gráfico. El tercer artículo incluye la posibilidad de dibujar sobre el mosaico descrito las curvas de nivel para el error de tipo 2 correspondiente a diferentes tests de Normalidad. Se incluyen un total de 9 tests diferentes de Normalidad sobre los que se analiza la potencia en base al tamaño muestral. Finalmente, el tercer artículo plantea, dada una muestra concreta y para cada distribución representada en el mosaico, el test de bondad de ajuste a dicha distribución. De esta manera, en este último artículo la mancha sobre el mosaico representa el conjunto de distribuciones de la familia SEPD con las que la muestra es compatible y que no permitirían rechazar la hipótesis nula de bondad de ajuste a aquella distribución. La conclusión más evidente establece que un incremento del tamaño de muestra suele implicar una reducción de la superficie representada en el mosaico por parte de las distribuciones compatibles con la muestra. La inclusión de la distribución Normal en este conjunto implica que no se rechace la hipótesis de normalidad para la muestra. Para tamaños de muestra pequeños, es frecuente que se produzca esta situación. Como conclusión de este análisis, aunque el

escenario de simulación sea diferente de la distribución Normal, en situaciones de tamaño de muestra pequeña las muestras generadas podrían ser consideradas generadas por esta distribución.

El segundo bloque se centra en la aplicación de técnicas estadísticas para la resolución de problemáticas concretas. Se incluyen 4 artículos que son el resultado de la colaboración con un grupo de investigación experimental del ámbito de la hidrología. La colaboración se inicia con el investigador del Parc Agrari del Baix Llobregat, Basem Aljoumani, el cual cursó la asignatura de Series Temporales del Máster de Estadística e Investigación Operativa. El objetivo era aplicar técnicas de predicción basadas en modelos ARIMA para el contenido de agua y la salinidad del terreno de cultivo, medidas de forma secuencial. A partir de las predicciones obtenidas se planteaba la configuración de riego automático que compensara la reducción de agua en el terreno e incremento de salinidad. El principal problema para aplicar la metodología estándar consiste en la presencia de eventos que alteran la evolución natural del fenómeno que se registra. En concreto, tanto los riegos que se realizan como los episodios de lluvia suponen una ruptura de la estructura de autocorrelación. Esta situación implica una reducción de la eficiencia en el proceso de estimación y predicción del modelo ARIMA clásico. Sin embargo, es posible extender los modelos ARIMA incluyendo el tratamiento de atípicos para eventos no programados y análisis de intervención para eventos programados. Además, la medición de las variables de interés (contenido de agua y salinidad) se realizan a diferentes profundidades, con lo que resulta natural introducir los modelos con función de transferencia que suponen una mejora de los procesos de predicción. La extensión de estos modelos fue presentada en la tesis doctoral del Dr. Aljoumani "Soil water management. Evaluation of infiltration systems, assessing water and salt content spatially and temporally in the Parc Agrari del Baix Llobregat area" (2012).

Respecto a la medición de la salinidad, aparece la problemática de que la señal medida no es directamente la salinidad del terreno, sino que el sensor registra una señal que se halla relacionada mediante un modelo físico teórico con la serie de interés (modelo de Hilhorst). En esta ecuación aparece un parámetro que no está determinado. Debido a que los terrenos de cultivo son heterogéneos en su composición y propiedades, el valor de este parámetro varía. Existen estudios que incluyen experimentación en laboratorio que permiten realizar una imputación aproximada para este parámetro con lo que la señal medida se puede transformar en la señal de interés. Una de las aportaciones realizadas y descritas en los tres primeros artículos de este bloque consiste en la utilización de modelos lineales dinámicos para obtener una imputación de este parámetro basada en los datos recogidos. La idea consiste en linealizar la ecuación del modelo de Hilhorst e incluir como variables de estado del modelo lineal dinámico tanto el parámetro como la serie de interés (salinidad). La estimación por máxima verosimilitud hace posible obtener los valores del parámetro específico y de los valores más verosímiles para la serie de salinidad. Los artículos incluidos presentan esta aproximación tanto con datos de campo como datos experimentales de laboratorio.

Finalmente, el último artículo incluido en la presente tesis supone una nueva línea de investigación, en colaboración con el departamento de Ecología y Conservación del Suelo de la Universidad Técnica de Berlín (TU Berlin). A partir de un diseño experimental, se dispone de información recogida de la presencia de metales en el suelo en puntos cercanos a una autopista en Alemania. Los datos recogidos no presentan la estructura de series temporales (los períodos entre observaciones no son constantes) y además, el interés del estudio es relacionar la cantidad de metales encuentro con variables ambientales. En este caso, se plantea la metodología de los modelos aditivos generalizados mixtos, donde los flexibiliza la relación de la respuesta con las variables explicativas para incorporar dependencias no paramétricas. Como las medidas se hacen el largo del tiempo en una serie de puntos concretos es necesario también incluir efectos aleatorios para las localizaciones, haciendo que los modelos que se utilizan deban tener la extensión a modelos mixtos para medidas repetidas.

1.1 Esquema de la tesis

Como se ha comentado, los artículos aparecen agrupados en dos bloques que corresponden a las dos partes en que se clasifica la presente tesis. La primera parte se centra en aportaciones de tipo metodológico en estadística. Concretamente para el análisis de medidas repetidas mediante modelos mixtos, se proponen y analizan extensiones de las técnicas inferenciales clásicas usando métodos Bootstrap. El primer artículo (capítulo 2) describe esta aproximación mediante la aplicación, entre otras, de la técnica *Wild Bootstrap* para el caso de remuestreo de los componentes de la varianza (*Residual Resampling*). Las siguientes publicaciones (capítulos 3 a 5) profundizan en el análisis de la potencia de tests de bondad de ajuste cuando aplicamos desviaciones de normalidad a la hora de realizar las simulaciones para evaluar métodos inferenciales como los descritos en la primera publicación.

En la segunda parte de la tesis se ilustra la aplicación de técnicas estadísticas para analizar medidas repetidas en el ámbito de la hidrología. Estas publicaciones son el resultado de la colaboración con un equipo del Departamento de Ingeniería Alimentaría y Biotecnología (UPC) y del Departamento de Ecología y Conservación del suelo (TU Berlin). La primera publicación de esta parte (capítulo 6) ilustra la aplicación de técnicas avanzadas de series temporales para analizar mediciones repetidas de contenido de agua y salinidad en el suelo. Las dos publicaciones siguientes profundizan en la determinación de la salinidad mediante el uso de modelos lineales dinámicos y el filtro de Kalman con datos de campo (capítulo 7) y con datos de experimentación en laboratorio (capítulo 8). Finalmente, la última publicación (capítulo 9) describe la aplicación de modelos GAMM para datos longitudinales. En este caso consiste en el análisis del contenido de metales pesados en zonas próximas a autopistas, en función de factores climatológicos y medioambientales.

El último capítulo (capítulo 10) sumariza las aportaciones presentadas y futuras extensiones del trabajo realizado.

Parte I. Aportaciones metodológicas a la estadística

- *Capítulo 2:*

  Semiparametric Bootstrap in Linear Mixed Models with Misspecified Skewness and Kurtosis
  Enviado a publicación JCR

- *Capítulo 3:*

  Visualizing type II error in normality tests
  American statistician, vol. 72, núm. 2, pàgs. 158-162 (2018)

  STATISTICS & PROBABILITY (JCR Q1)

  DOI: 10.1080/00031305.2016.1278035

- *Capítulo 4:*

  Graphical comparison of normality tests for unimodal distribution data
  Journal of statistical computation and simulation, vol. 89, núm. 1, pàgs. 145-154 (2019)

  STATISTICS & PROBABILITY (JCR Q3), COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS (JCR Q4)

  DOI: 10.1080/00949655.2018.1539085

- *Capítulo 5:*

  Mosaic Normality Test
  Communication in Statistics—Theory and Methods (2020)

  STATISTICS & PROBABILITY (JCR Q4)

  DOI: 10.1080/03610926.2020.1734828

Parte II. Aplicaciones de técnicas estadísticas en hidrología

- *Capítulo 6:*

  Transfer function and time series outlier analysis: modelling Soil salinity in loamy sand soil by including the influences of irrigation management and soil temperature
  Irrigation and drainage, vol. 67, núm. 2, pàgs. 282-294 (2017)

  AGRONOMY (JCR Q3), WATER RESOURCES (JCR Q4)

  DOI: 10.1002/ird.2187

- *Capítulo 7:*

  An advanced process for evaluating a linear dielectric constant–bulk electrical conductivity model using a capacitance sensor in field conditions
  Hydrological sciences journal, vol. 60, núm. 10, pàgs. 1828-1839 (2015)

  WATER RESOURCES (JCR Q1)

  DOI: 10.1080/02626667.2014.932053

- *Capítulo 8:*

Estimating pore water electrical conductivity of sandy soil from time domain reflectometry records using a time-varying dynamic linear model

Sensors, vol. 18, núm. 12, pàgs. 4403-4414 (2018)

CHEMISTRY ANALYTICAL (JCR Q2), ELECTROCHEMISTRY (JCR Q2), INSTRUMENTS & INSTRUMENTATION (JCR Q1)

- *Capítulo 9:*
Evaluating the variation of dissolved metals on a highway roadside using a generalized additive mixed model (GAMM)

Water, Air, and Soil Pollution, vol. 230, núm. 93 (2019)

ENVIRONMENTAL SCIENCES (JCR Q3), METEOROLOGY & ATMOSPHERIC SCIENCES (JCR Q3), WATER RESOURCES (JCR Q3)

Parte I

Aportaciones metodológicas a la estadística

2. *Semiparametric Bootstrap in Linear Mixed Models with Misspecified Skewness and Kurtosis*

José A. Sánchez-Espigares[1], Jordi Ocaña[2]

[1]Department of Statistics and Operations Research
Universitat Politècnica de Catalunya-BarcelonaTech, Barcelona, Spain
[2]Departament of Genetics, Microbiology and Statistics, University of Barcelona, Spain

ABSTRACT

Standard inferential procedures for Linear Mixed Models (namely, asymptotic inference and parametric bootstrap) rely on normality assumptions for random effects estimates. Even if their distributions' first and second moments are properly specified, skewness and kurtosis misspecification can lead to unsuitable inferences, e.g., insufficient or excessive coverage of confidence intervals. Several forms of semi-parametric bootstrap (adjusted and Wild bootstrap, among them) have been proposed to improve the estimates. This paper shows, using simulation, that when the misspecification tends to be exponential-like (asymmetry, high kurtosis) the Wild bootstrap over-performs the adjusted; when the misspecification tends to be uniform-like (symmetry, low kurtosis) the adjusted semi-parametric bootstrap should be preferred over the Wild. It is also shown that when the two sources of variability (random effects and residual variance) have contrasting ways of misspecification, the choice between Wild and adjusted is not clear and involves a trade-off, but any of the two options are better than the parametric bootstrap with the normality assumption.

KEYWORDS: linear mixed model, parametric bootstrap, semi-parametric bootstrap, simulation

2.1.   *Introduction*

It is well known that the violation of statistical assumptions is not always assessed in statistical practice. A common example of this apparently widespread phenomenon occurs when confidence intervals are assumed to have the expected coverage even when key assumptions for their computation do not hold. Typically, assumptions relate to the probability distributions of some terms in a model.

Dubious actual coverages due to unfulfilled distributional assumptions can affect both asymptotically justified confidence intervals and confidence intervals obtained using parametric bootstrap.

The versatility of Linear Mixed Models (LMMs) to analyze various kinds of data comes from the fact that they allow more random terms beyond the single residual error found in Linear Models (LMs). Since each random term involves the specification of a distribution, LMMs pose particular challenges concerning how to deal with misspecification.

LMMs strongly rely on normality assumptions for the estimation of the covariance matrix of random effects and the residual variance (Schützenmeister and Piepho 2011; Laird and Ware 1982); such assumptions are mostly made for the convenience of getting an analytic treatment of the marginal density of the response variable.

The extent to which the normality assumption can lead to wrong conclusions when one or more than one of the sources of variability (random effects or residual errors) are misspecified, largely depends on the study objectives. For instance, it is well established that the maximum-likelihood estimation of fixed effects is robust to random-effects misspecification (Butler and Louis 1992; Agresti, Caffo, and Ohman-Strickland 2004; Charles E McCulloch and Neuhaus 2011; Geert Verbeke and Lesaffre 1997; D. Zhang and Davidian 2001); in most situations (except, for instance, when the residual variance depends on time-varying covariates), they are robust not only to non-normality but also to heteroscedasticity (Jacqmin-Gadda et al. 2007). Similar results have been found for fixed effects estimates obtained by Generalised Estimating Equations (GEE) (Liang and Zeger 1986) and, at least asymptotically, for robust sandwich variance estimators (Royall 1986). Since research objectives frequently focus on fixed effects estimates and random effects are only included to improve the inference on them, misspecification is not necessarily problematic in all cases.

Furthermore, asymptotic consistency results are well established for maximum likelihood estimators not only of fixed effects but also of variance components, even when their distribution is not normal: under general regularity conditions, the maximum likelihood estimates of fixed effects, random effects and residual variance are strongly consistent estimators when the number of observations tends to infinity, as shown in (Geert Verbeke and Lesaffre 1997).

When sample sizes are not large enough, misspecification can have several consequences on the estimation of random effects (Agresti, Caffo, and Ohman-Strickland 2004). Empirical Bayes estimates of random effects have been shown to be sensitive to non-normality (Ghidey, Lesaffre, and Verbeke 2010). It is also well established that the normal model can be inefficient in the case of binary or mixture random-effects distributions (Agresti, Caffo, and Ohman-Strickland 2004). Bimodal distribution of the random effects may indicate that subject-specific covariates have not been taken into account. McCulloch and Neuhaus (Charles E McCulloch and Neuhaus 2011) further study the impact of the shape of random-

effects distribution on the estimation of random intercept coefficients, a topic closely related to the present paper, which focus on asymmetry and skewness misspecification in the case unimodal distributions.

Such problems are relevant in practice because LMMs are widely used for the analysis of longitudinal and repeated measurements data, for instance in growth curve models including random effects for intercept and slope (Nguyen and McLachlan 2016; C. E. McCulloch and Searle 2001; J. C. Pinheiro and Bates 2000; G. Verbeke and Molenberghs 2000). In behavioral or medical research, repeated measures and split-plot designs are frequently used because they require less subjects (Davis 2002; Hedeker and Gibbons 2006; Howell 2013; Stevens 2012; Friedrich, Konietschke, and Pauly 2016). Indeed, the cost considerations or subject availability reasons behind the use of this family of clustered designs do not play in favor of the widespread reliance on asymptotic results, either in the context of multivariate ANOVA methods, repeated measures ANOVA, or LMMs (Davis 2002; Hedeker and Gibbons 2006; Johnson and Wichern 2007). The required parametric assumptions (which are crucial when asymptotic considerations are not reliable) are not always met for this kind of data (Kherad-Pajouh and Renaud 2014; Konietschke et al. 2015; Suo et al. 2013; Xu and Cui 2008).

Several practical applications are exposed to misspecification problems. Again in behavioral and medical research, measurements of interest can be assessed by differents observers or diagnostic methods; such measurements can involve nominal, ordinal, count or continuous scales that not always are going to satisfy stringent distributional assumptions. But nevertheless it is important to assess the agreement among observers or the reproducibility of replicated measures of a single method (Tsai and Lin 2017). The Intraclass Correlation Coefficient (ICC), the proportion of the total variance due to the between-subjects variance, can be used for assessing the agreement between different observers taking quantitative measures (Bartko 1966; Shrout and Fleiss 1979; Tsai 2015). ICC can be analyzed via LMMs with serial correlation for inter-observer, intra-observer, and absolute agreement, with observer and time as random effects (Vangeneugden et al. 2005; Chen and Barnhart 2013).

The Concordance Correlation Coefficient (CCC), proposed by Lin (L. Lin 1989), was shown to be equivalent with the ICC by Carrasco and Jover (Carrasco and Jover 2004) for a two-way linear mixed model without interaction; both can be estimated through variance components (in addition to other methods such as GEE and U-statistics). When repeated measurements are assessed from different observers over time, a longitudinal version can be obtained by similar methods (Carrasco, King, and Chinchilli 2009).

Another example of relevant application of LMM sensitive to misspecification is the prediction of vectors of small area quantities based on a multivariate Fay-Herriot model (González-Manteiga et al. 2008). In this case, an empirical best linear unbiased predictor (EBLUP) of the target vector can obtained either analytically and using bootstrap. Small sample properties of bootstrap estimators have been

compared to the analytical approximation under non-normality, showing that empirical mean square errors (MSEs) may increase considerably (González-Manteiga et al. 2008).

The diagnosis of misspecification problems in LMM can be done with generalized weighted normal plots with weights depending on the sampling variances of the estimated random effects (Lange and Ryan 1989), even though Verbeke and Molenberghs found that these plots are sensitive to the choice of covariates and distributional assumptions on random effects or residual variance (G. Verbeke and Molenberghs 2000). Pinheiro and Bates (J. C. Pinheiro and Bates 2000) suggest to assess the normality of residual errors and random effects with quantile-quantile plots.

Proposed solutions to the problems of misspecification in LMM include analytic methods using t-distributed random-effects and residuals (José C. Pinheiro, Liu, and Wu 2001; Song, Zhang, and Qu 2007), or using the skew normal distribution (Arellano-Valle, Bolfarine, and Lachos 2005; Lachos, Ghosh, and Arellano-Valle 2010; Ho and Lin 2010). Non-parametric maximum likelihood (NPML) densities (Laird 1978) can be shown to provide more efficient estimates of random-effects parameters compared to parametric inference when the true densities are not as assumed (Agresti, Caffo, and Ohman-Strickland 2004). Kernel density methods for mixture modeling with symmetric components are proposed in (Chee and Wang 2013).

Bootstrap methods have been also considered (Field and Welsh 2007; O'Shaughnessy and Welsh 2018), and they are the main focus in the present paper. Field et al. (Field, Pang, and Welsh 2008) considered semiparametric bootstraps for unbalanced clustered data. Several forms of generalized bootstrap techniques have been obtained applying bootstrap weights into GEE (Chatterjee and Bose 2005; Field, Pang, and Welsh 2010). These methods have been used in the analysis of clustered data with multiple levels of random effects (Field, Pang, and Welsh 2010), both in balanced and unbalanced cases (Samanta and Welsh 2013). O'Shaughnessy and Welsh (O'Shaughnessy and Welsh 2018) use bootstrap inference for the correlation between random effects under non-normality.

This paper also considers bootstrap methods in the case of misspecification of LMMs. Misspecification is understood here in terms of the shape of the distribution of random effects and residuals, namely in terms of their asymmetry and skewness.

In section 2 we introduce the notation together with the description of the model and its hypotheses. We discuss asymptotic inference for the model's parameters and we summarize three distinct forms of bootstrap resampling: parametric bootstrap, adjusted semi-parametric bootstrap, and Wild bootstrap. In section 3 we introduce a simulation experiment aimed at showing how the parametric bootstrap fails in the case of misspecification and in which scenarios the adjusted and Wild bootstraps are preferable. Simulation results are shown in section 4 and the corresponding appendix, and discussed in section 5.

2.2.   Inference for Linear mixed models

2.2.1. Model and hypotheses

The general specification of LMMs using Laird and Ware's (Laird and Ware 1982) notation (with minor variabilitys) is:

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i$$
$$\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{D})$$
$$\boldsymbol{\varepsilon}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_i)$$

where i ranges from 1 to M experimental units. In contrast to classical Linear Models, each $\mathbf{Y}_i$ is a vector of n_i random variables:

$$\mathbf{Y}_i = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{in_i} \end{pmatrix}$$

For observations $\mathbf{Y}_i = \mathbf{y}_i$ we use the notation $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})^T$.

Fixed factors have a known $n_i \times p$ design matrix $\mathbf{X}_i$ and an unknown parameter vector $\boldsymbol{\beta}$ of dimension $p \times 1$ common to all experimental units:

$$\mathbf{X}_i = \begin{pmatrix} X_{i1}^{(1)} & \dots & X_{i1}^{(p)} \\ & & \vdots \\ X_{in_i}^{(1)} & \dots & X_{in_i}^{(p)} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$$

Random factors have a known $n_i \times k$ design matrix and an unknown parameter vector $\mathbf{b}_i$ of dimension $k \times 1$ assumed to be normally distributed around the zero vector:

$$\mathbf{Z}_i = \begin{pmatrix} Z_{i1}^{(1)} & \dots & Z_{i1}^{(k)} \\ & & \vdots \\ Z_{in_i}^{(1)} & \dots & Z_{in_i}^{(k)} \end{pmatrix}, \quad \mathbf{b}_i = \begin{pmatrix} b_{i1} \\ \vdots \\ b_{ik} \end{pmatrix} \sim \mathcal{N}(\mathbf{0}, \mathbf{D})$$

Here $\mathbf{D}$ is a $k \times k$ positive semi-definite (thus symmetric) matrix to be interpreted as the variance-covariance matrix

$$\mathbf{D} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1k} \\ \sigma_{12} & \sigma_2^2 & \dots & \sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1k} & \sigma_{2k} & \dots & \sigma_k^2 \end{pmatrix}$$

Notice that $\mathbf{D}$ does not depend on the individual i, the subscript in $\mathbf{b}_i$ (just like in $\boldsymbol{\varepsilon}_i$) meaning that each individual has a particular realization of the random variable. Accordingly, in the random part of the model only $\mathbf{Z}_i$ is specific to the individual.

Each experimental unit has a $n_i \times 1$ residuals vector also taken to be normally distributed around the zero vector:

$$\boldsymbol{\varepsilon}_i = \begin{pmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{in_i} \end{pmatrix} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_i)$$

In this paper we shall assume that $\mathbf{R}_i = \sigma^2 \mathbf{I}_{n_i}$, where $\mathbf{I}_{n_i}$ is the $n_i \times n_i$ identity matrix. The matrix $\sigma^2 \mathbf{I}_{n_i}$ only depends on the individual i through its dimension, and σ^2 is common to all the experimental units.

According to the last remark, every pair of residuals ε_{it} and $\varepsilon_{it'}$ of any individual i are uncorrelated. In addition, for any two individuals i and j, any pair of errors ε_{it} and $\varepsilon_{jt'}$ are independent. Any two b_l and $b_{l'}$ in $\mathbf{b}$ are tipically correlated, but any b_l and ε_{it} are independent.

Let $\boldsymbol{\theta}$ be a $k \times 1$ vector of the variance and covariance parameters in $\mathbf{D}$, thus

$$\boldsymbol{\theta} = (\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2, \sigma_{12}, \sigma_{13}, \dots, \sigma_{k(k-1)})^T$$

We denote by $\boldsymbol{\zeta}$ the vector of the parameters for the whole model, which include $\boldsymbol{\beta}$ (fixed effects) and $\Theta = (\boldsymbol{\theta}, \sigma^2)$ (variance components), so

$$\boldsymbol{\zeta} = (\beta_1, \dots, \beta_p, \sigma_1^2, \sigma_2^2, \dots, \sigma_k^2, \sigma_{12}, \sigma_{13}, \dots, \sigma_{k(k-1)}, \sigma^2)^T$$

When appropriate, standard deviations and correlations will be used instead of variances and covariances. For instance, in the particular case of $p = k = 2$ and

$$\mathbf{D} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$$

we have $\boldsymbol{\theta} = (\sigma_1^2, \sigma_2^2, \sigma_{12})^T$ but we may use $\boldsymbol{\zeta} = (\beta_1, \beta_2, \sigma_1, \sigma_2, \rho, \sigma)^T$, with $\rho = \frac{\sigma_{12}}{\sigma_1 \sigma_2}$, as the parametrization for the whole model.

2.2.2. Point estimation from data

Once a realization $\mathbf{Y}_i = \mathbf{y}_i$, with $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})^T$, has been obtained for each $i = 1, \dots, M$, the observed response variable is

$$\mathbf{y} = (y_{11}, \dots, y_{1n_1}, y_{21}, \dots, y_{2n_2}, \dots, y_{M1}, \dots, y_{Mn_M})^T$$

of dimension $N \times 1$, with $N = \sum_{i=1}^{M} n_i$. Let $\mathbf{X}$ be the $N \times p$ matrix

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_M \end{pmatrix}$$

and let $\mathbf{Z}$ be the $N \times kM$ matrix

$$\mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{Z}_M \end{pmatrix}$$

Assuming the model [lmmodel]–[mod_resid_dist], from $\mathbf{y}$, $\mathbf{X}$ and $\mathbf{Z}$ one can get a point estimate

$$\hat{\boldsymbol{\zeta}} = (\hat{\beta}_1, \dots, \hat{\beta}_p, \hat{\sigma}_1^2, \hat{\sigma}_2^2, \dots, \hat{\sigma}_k^2, \hat{\sigma}_{12}, \hat{\sigma}_{13}, \dots, \hat{\sigma}_{k(k-1)}, \hat{\sigma}^2)^T$$

of the model parameters.

As Pinheiro and Bates (J. C. Pinheiro and Bates 2000) recall, technically the random effects $\mathbf{b}_i$ are not parameters for the statistical model, but sometimes it is convenient to get their predicted values given the observed $\mathbf{y}$. The conditional modes of the random effects, evaluated at the conditional estimate of $\boldsymbol{\beta}$, are the Best Linear Unbiased Predictors (BLUP) of the $\mathbf{b}_i$, $i = 1, \dots, M$. Accordingly, as usual in the literature, $\hat{\mathbf{b}}_i = (\hat{b}_{i1}, \dots, \hat{b}_{ik})^T$ denotes the prediction of $\mathbf{b}_i$ and $\hat{\mathbf{b}}$ is the $kM \times 1$ vector

$$\hat{\mathbf{b}} = (\hat{b}_{11}, \dots, \hat{b}_{1k}, \hat{b}_{21}, \dots, \hat{b}_{2k}, \dots, \hat{b}_{M1}, \dots, \hat{b}_{Mk})^T$$

In a similar abuse of notation, given an observation $\mathbf{Y}_i = \mathbf{y}_i$, with $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})^T$, the conditional residual is defined as $\hat{\boldsymbol{\varepsilon}}_i = \mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}} - \mathbf{Z}_i \hat{\mathbf{b}}_i$, so we have $\hat{\boldsymbol{\varepsilon}}_i = (\hat{\varepsilon}_{i1}, \dots, \hat{\varepsilon}_{in_i})^T$. Let $\hat{\boldsymbol{\varepsilon}}$ be the $N \times 1$ vector

$$\hat{\boldsymbol{\varepsilon}} = (\hat{\varepsilon}_{11}, \dots, \hat{\varepsilon}_{1n_1}, \hat{\varepsilon}_{21}, \dots, \hat{\varepsilon}_{2n_2}, \dots, \hat{\varepsilon}_{M1}, \dots, \hat{\varepsilon}_{Mn_M})^T$$

It can be checked that

$$\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{b}} + \hat{\boldsymbol{\varepsilon}}$$

2.2.3. Asymptotic inference

Throughout this manuscript, inference shall mean confidence interval estimation of $\boldsymbol{\zeta}$.

Asymptotic confidence intervals are based on Wald's test, since the standard error of the maximum likelihood (ML) estimates is computed from the inverse of Fisher's information matrix $\mathcal{J}$, since

$$(\hat{\boldsymbol{\beta}}, \hat{\mathbf{D}}, \hat{\sigma}^2) \xrightarrow{d} \mathcal{N}((\boldsymbol{\beta}, \mathbf{D}, \boldsymbol{\sigma}), \mathcal{J}^{-1}(\boldsymbol{\beta}, \mathbf{D}, \sigma^2))$$

where $\xrightarrow{d}$ means convergence in distribution. Though asymptotically efficient, ML estimates of $\mathbf{D}$ and σ^2 are biased for finite samples. More precisely, σ^2 and the diagonal of $\mathbf{D}$ are underestimated ($\hat{\sigma}^2$ and $\hat{\mathbf{D}}$ are downward biased) because of the loss of degrees of freedom due to the estimation of $\boldsymbol{\beta}$ at the same time.

Restricted maximum likelihood (REML) gets unbiased estimates of $\mathbf{D}$ and σ^2 by decoupling the estimation of variance components from the estimation of $\boldsymbol{\beta}$. That is, it maximizes a modified version of

the likelihood function which does not contain information about the fixed effects but only about the marginal residuals $\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}} = \mathbf{Z}_i\widehat{\mathbf{b}}_i + \widehat{\boldsymbol{\varepsilon}}_i$, so the necessary degrees of freedom are regained for the estimation of the variance components.

2.2.4. Bootstrap

In resampling, it turns out to be particularly relevant that $\mathbf{D}$ can be decomposed as $\mathbf{D} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^T$, where $\boldsymbol{\Lambda}$ is a lower triangular matrix with positive diagonal. Such decomposition (the Cholesky decomposition) satisfies

$$\boldsymbol{\Lambda}^{-1}\mathbf{b}_i = \mathbf{u}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_k)$$

where $\mathbf{I}_k$ is the $k \times k$ identity matrix. Therefore, $\mathbf{u}_i$ is a $k \times 1$ vector of uncorrelated parameters from which we can retrieve the original correlated parameters, since $\mathbf{b}_i = \boldsymbol{\Lambda}\mathbf{u}_i$. The model [lmmodel] can thus be written as $\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\Lambda}\mathbf{u}_i + \boldsymbol{\varepsilon}_i$ with $\mathbf{u}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_k)$ and $\boldsymbol{\varepsilon}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_i)$. Using this setting, it is possible to resample from the estimations $\widehat{\mathbf{u}}_i$ from the data, which makes more sense than resampling from the presumably correlated $\widehat{\mathbf{b}}_i$.

When a realization $\mathbf{Y}_i = \mathbf{y}_i$ is obtained for each one of the N individuals, we denote by $\underline{\mathbf{y}}$ the vector of all N observations $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})^T$. Since each $\mathbf{X}_i$ and $\mathbf{Z}_i$ are assumed to be known, $\underline{\mathbf{y}}$ can be used to compute estimates $\widehat{\mathbf{b}}_i$ (for each individual) and $\widehat{\boldsymbol{\beta}}$, thus observed residuals $\widehat{\boldsymbol{\varepsilon}}_i$ can be obtained for each individual.

Bootstrap methods are based on generating new observations either using an underlying model (parametric bootstrap) or the empirical distribution function of the parameters (semi-parametric bootstrap). Both sorts of bootstrap are based on so-called "residual resampling"—or, more generally, in a terminology adapted to LMMs, variance components resampling.

The strictly non-parametric bootstrap based on "case resampling" is not suitable for LMMs because either experimental units should have to be resampled with all their measures or measures should have to be resampled in disregard of the experimental units; in either case the design matrix would not be respected. In this manuscript, non-parametric bootstrap is not taken into account, since we assume that design matrices are determined.

Given any bootstrap procedure, resamples $\mathbf{b}_i^*$ and $\boldsymbol{\varepsilon}_i^*$ are obtained. The number of resamples is typically denoted by B. Details on the three ways considered here to obtain such resamples (parametric, adjusted semi-parametric, and semi-parametric Wild bootstrap) are explained below. Each bootstrap sample has the same size $N = \sum_{i=1}^{M} n_i$ as the original sample, and $\mathbf{b}_i^*$ and $\boldsymbol{\varepsilon}_i^*$ have the same dimension as $\widehat{\mathbf{b}}_i$ and $\widehat{\boldsymbol{\varepsilon}}_i$, respectively.

There are three kinds of bootstrap confidence intervals which can be computed with the resamples: normal, basic, and percentile (Davison and Hinkley 1997). In this study, percentile confidence intervals $(\theta^*_{(\alpha/2)}, \theta^*_{(1-\alpha/2)})$ are used, which simply consist of taking the $\alpha/2$-th percentile of the B estimations of the parameter θ as lower limit, and the $1 - \alpha/2$-th percentile as upper limit.

The assessment of confidence interval estimations can be done via simulation. If the simulation is repeated nsim times, so that B resamples of sample size $N = \sum_{i=1}^{M} n_i$ are obtained nsim times, then we obtain nsim percentile confidence intervals. The empirical coverage is defined as the proportion of the nsim intervals which effectively include the value of the parameter used in their respective simulation instances. Below [above] rates can be defined as the proportion of the nsim intervals for which the parameter value used in the simulation instances falls below [above] the lower [upper] limit.

The simplest way to obtain resamples is the parametric bootstrap, which assumes the family of distributions of random effects and residuals to be known, and just simulates resamples according to such families. That is, if the model [lmmodel]–[mod_resid_dist] holds, $\mathbf{b}_i$ and $\boldsymbol{\varepsilon}_i$ are random variables assumed to be normal with zero means and covariance matrices $\mathbf{D}$ and $\sigma^2 \mathbf{I}_{n_i}$, respectively. When data is available, this model can be fitted, yielding estimates $\widehat{\mathbf{D}}$ and $\widehat{\sigma^2}$. This makes it possible to compute an arbitrary number B of resamples $\mathbf{b}_i^*$ and $\boldsymbol{\varepsilon}_i^*$ just simulating from

$$\mathbf{b}_i^* \sim \mathcal{N}(\mathbf{0}, \widehat{\mathbf{D}})$$
$$\boldsymbol{\varepsilon}_i^* \sim \mathcal{N}(\mathbf{0}, \widehat{\sigma}^2 \mathbf{I}_{n_i})$$

so obtaining fitted values

$$\widehat{\mathbf{y}}_i^* = X_i\,\widehat{\boldsymbol{\beta}} + \mathbf{Z}_i\mathbf{b}_i^* + \boldsymbol{\varepsilon}_i^*$$

A resample is a vector $(\widehat{\mathbf{y}}_1^*, \dots, \widehat{\mathbf{y}}_N^*)^T$.

2.2.5. Adjusted semi-parametric bootstrap

If the model [lmmodel]–[mod_resid_dist] cannot be assumed, the empirical distribution of the parameters can be used:

$$\mathbf{b}_i^* \sim F_n(\widehat{\mathbf{b}}_i)$$
$$\boldsymbol{\varepsilon}_i^* \sim F_n(\widehat{\boldsymbol{\varepsilon}}_i)$$

Estimates $\widehat{\mathbf{D}}$ and $\widehat{\sigma}^2$ can be shown to be biased. In order to avoid the bias, the following adjustment can be used (Davison and Hinkley 1997):

$$\mathbf{b}_i^* \sim F_n(\widehat{\mathbf{D}} \frac{\widehat{\mathbf{b}}_i}{s_{\mathbf{b}_i}})$$

$$\boldsymbol{\varepsilon}_i^* \sim F_n(\widehat{\sigma} \frac{e_i^*}{s_{e_i^*}})$$

where $e_i = \widehat{\boldsymbol{\varepsilon}}_i$.

2.2.6. Semi-parametric bootstrap: Wild bootstrap

In the Wild boostrap each resampled residual is randomly multiplied by 1 or -1. That is, resampled residuals are $e'_i = w_i e_i$ where w_i is the Rademacher distribution (Davison and Hinkley 1997).:

$$w_i = \begin{cases} 1 & \text{with probability } 0.5 \\ -1 & \text{with probability } 0.5 \end{cases}$$

The Wild bootstrap was proposed by Davidson and Flachaire (Davidson and Flachaire 2008) and, even though it may increase variability and impose symmetry on the empirical distribution, it can capture other features of the underlying density. An alternative approach can be obtained using Mammen's two-point distribution (Mammen 1993) (Mammen 1993):

$$w_i = \begin{cases} -(\sqrt{5}-1)/2 & \text{with probability } (\sqrt{5}+1)/(2\sqrt{5}) \\ (\sqrt{5}+1)/2 & \text{with probability } (\sqrt{5}-1)/(2\sqrt{5}) \end{cases}$$

Both the Rademacher and Mammen distributions fix the third and fourth moments, so can be suitable to some kinds of misspecifications of skewness and kurtosis, but not others.

2.3.   Simulation Study

2.3.1. Simulation settings

Three distributions are used in the simulation study for both the random effects and the residuals: normal (symmetric, reference kurtosis), uniform (symmetric, platykurtic), and exponential (assymetric, leptokurtic).

Depending of the distribution of each source of variability (random effects or r.e. - residuals or res.), we have the following 9 scenarios:

- Normal r.e. - Normal res.

- Uniform r.e. - Normal res.

- Exponential r.e. - Normal res.

- Uniform r.e. - Uniform res.

- Exponential r.e. - Exponential res.

- Exponential r.e. - Uniform res.

- Uniform r.e. - Exponential res.

In order to make the simulation, the model from the Oxboys data is assumed. It is a dataset about the height growth of boys in Oxford, initially studied in (Harrison and Brush 1990) and then described in (Goldstein 2011); it is included (J. Pinheiro et al. 2019). The following variables are used in this paper:

- *height* : height of the boy (cm).

- *age* : standardized age (dimensionless).

- *Subject* : a unique identifier for each boy in the experiment.

The response variable $\mathbf{y}_i$ is the height of $M = 26$ boys in each of $n_i = 9$ measures, thus $N = 234$. The matrix $\mathbf{Z}_i = \mathbf{X}_i$ has dimension 26×2, corresponding to the variables Intercept and Age. The parameters for fixed effects are β_0 (intercept) and β_A (age), and we have variance components $(\sigma_0, \sigma_A, \rho, \sigma)$, where ρ is the correlation between (random) intercept and age. The theoretical model is on the adjusted model for the Oxboys data, taken here to be a suitable model for simulation:

- $\beta_0 = 149.37175$

- $\beta_A = 6.52547$

- $\sigma_0 = 8.081077$

- $\sigma_A = 1.680717$

- $\rho = 0.641$

- $\sigma = 0.659889$

The simulation consists of nsim = 30000 simulation replicates for each scenario. In each simulation replicate B = 200 resamples have been drawn.

2.4.   Results and discussion

The Appendix shows the simulation results for three confidences (90%, 95% and 99%) and the nine scenarios. In each one of the 27 settings, the results from using parametric, semi-parametric and Wild bootstrap, together with the reference result of bootstrapping from the true distribution (known in the simulation) are shown.

In each case we provide the following information for each one of the six parameters $(\beta_0, \beta_A, \sigma_0, \sigma_A, \rho, \sigma)$:

- *Below rate*. Proportion of confidence intervals in which the true parameter is below the lower limit.

- *Cover rate*. Proportion of confidence intervals in which the true parameter falls within the limits.

- *Above rate*. Proportion of confidence intervals in which the true parameter is above the upper limit.

- *Mean width*. Average of the confidence intervals' widths.

- *Width SD*. Standard deviation of the confidence intervals' widths.

Among the obtained results, the most expected one it that the fixed effects estimation is robust to misspecification: misspecification in random effects and/or residuals does not corrupt the fixed effects estimates.

On the other hand, random effects and residuals show a relatively independent behavior. If random effects are normal but residuals are misspecified, the random effects' estimates do not seem to be wrong in any obvious way. Likewise, if residuals are normal but random effects misspecified, the residuals' estimates seem to be correct. Even if one source is misspecified in some way (say, towards the uniform distribution) and the other source in the contrary way (towards the exponential), their behavior does not interact in any clear way.

Another conclusion mostly independent of the estimation method is that when a source of variability has an exponential behavior, the confidence interval for its standard deviation presents clear under-coverage: for a 95% confidence, it ranges from about 70% in the worst case (parametric bootstrap) to at most 90% for the best option (the Wild bootstrap), through the 85% in the adjusted semi-parametric bootstrap. In all these cases the Above Rate is clearly larger than the Below Rate, so the fact that the true parameter tends to fall above the upper limit accounts for most of the reduced coverage. The under-coverage for the correlation between random effects does not seem to be as pronounced as in the case of the standard deviations.

Contrariwise, when a source of variability has a uniform behavior, the confidence intervals for its parameters (including the correlation between random effects) presents over-coverage. Of course, the over-coverage is upper bounded, so not strictly comparable with the magnitude of the under-coverage in the exponential case, but for a 95% confidence it can rise as high as 99.0 - 99.5% (standard deviation of the residual error in the parametric and Wild bootstraps).

Of course, another expected result is that, neither the Wild nor the semi-parametric bootstrap outperform the parametric bootstrap if the model is correctly specified. In the exponential-exponential scenario (asymmetry, higher kurtosis), the Wild bootstrap seems to be preferable over the adjusted (and of course over the parametric bootstrap); this is also the case when only one source is misspecified towards the exponential. A possible reason is that the Wild bootstrap introduces bias, which (by the asymmetry of the exponential) may "artificially" increase the coverage. As noted, the exponential

distribution essentially generates under-coverage, so anything increasing the coverage appears as better-behaved.

In the uniform case less or no bias is generated, so the Wild bootstrap does not take advantage of an artificial impact on coverage, and in the uniform-uniform scenario (symmetry, lower kurtosis), the semi-parametric bootstrap seems to be preferable. It is also preferable if only one source is misspecified towards the uniform. If both sources are misspecified, but in different ways (one towards the uniform distribution and the other towards the exponential), the results are not conclusive: when choosing between semi-parametric and Wild bootstrap there is a tradeoff between getting a better estimation of the random effects or a better estimate of the residual variance. But, crucially, both the semi-parametric and the Wild bootstrap are better than the parametric bootstrap (which tends to be the most prevalent choice). In fact, the parametric bootstrap never outperforms the Wild nor the semi-parametric bootstraps in the presence of any misspecified source of variability.

2.5.   Conclusions

Taking the results into account, a heuristic for bootstrap method selection based on sample estimates of skewness and kurtosis could be easily designed and implemented. The parametric bootstrap should be preferable only under normality. Further research can improve the criteria proposed for each of the considered scenarios, in particular resolving what is the best choice if random effects and residuals have different kinds of misspecification (namely, exponential-like and uniform-like).

Of course, the methodology adopted in this study has its limitations. Crucially, it has been assumed that the same bootstrap method is performed in random effects and residuals; allowing separate methods presumably would improve the results and the possible heuristics. On the other hand, only one simulation study (assuming a particular model) has been carried out; therefore, other findings could possibly emerge in models with other characteristics. For instance, if in the characterization of residuals $\boldsymbol{\varepsilon}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_i)$ the assumption that $\mathbf{R}_i = \sigma^2 \mathbf{I}_{n_i}$ is relaxed, some conclusions could conceivably change.

The main line of future research lays on the development of efficient misspecification diagnostic methods with sample size considerations in mind. Small sample sizes may impact on the ability to identify misspecification; e.g., small samples of uniform or exponential populations might appear as acceptable under the normal model. This is especially relevant in the case of LMMs because the number of observations available for the estimation of variance components differs: samples of random effects and residuals for a given dataset have different sizes (necessarily larger for the random effects). Therefore, it is desirable to be able (1) to assess the "distance" between distributions with the same mean and variance but different skewness and kurtosis, and (2) to understand how these distinct distributions can be correctly identified as different depending on the sample size.

The simulation shown in this paper has been carried out using the R library merboot, based on the package lme4 and still under development and documentation. The library already implements all the bootstrap methods used here, and the simulation engine, but future work may include modifications based on the experience and findings of the presented simulations.

Additional future work can include research on (and implementation of) other bootstrap methods, such as a version of the Wild bootstrap using a kernel density estimation to draw resamples, or quantiles semi-parametric bootstrap. The generalization of the current research to Generalized Linear Mixed Models (GLMMs) could also be explored.

2.6. *References*

1. Agresti, Alan, Brian Caffo, and Pamela Ohman-Strickland. 2004. "Examples in Which Misspecification of a Random Effects Distribution Reduces Efficiency, and Possible Remedies." *Computational Statistics & Data Analysis* 47: 639–53.

2. Arellano-Valle, Reinaldo Boris, Heleno Bolfarine, and Victor H. Lachos. 2005. "Skew-Normal Linear Mixed Models." *J. Data Sci.* 3: 415–38.

3. Bartko, John. 1966. "The Intraclass Correlation Coefficient as a Measure of Reliability." *Psychological Reports* 19 (September): 3–11. doi:10.2466/pr0.1966.19.1.3.

4. Butler, S. M., and T. A. Louis. 1992. "Random Effects Models with Non-Parametric Priors." *Stat. Med.* 11: 1981–2000.

5. Carrasco, Josep, and Lluis Jover. 2004. "Estimating the Generalized Concordance Correlation Coefficient Through Variance Components." *Biometrics* 59 (January): 849–58. doi:10.1111/j.0006-341X.2003.00099.x.

6. Carrasco, Josep, Tonya King, and Vernon Chinchilli. 2009. "The Concordance Correlation Coefficient for Repeated Measures Estimated by Variance Components." *Journal of Biopharmaceutical Statistics* 19 (February): 90–105. doi:10.1080/10543400802527890.

7. Chatterjee, Snigdhansu, and Arup Bose. 2005. "Generalized Bootstrap for Estimating Equations." *Ann. Statist.* 33: 414–36.

8. Chee, Chew-Seng, and Yong Wang. 2013. "Estimation of Finite Mixtures with Symmetric Components." *Statistics and Computing* 23 (2): 233–49. doi:10.1007/s11222-011-9305-5.

9. Chen, Chia-Cheng, and Huiman X. Barnhart. 2013. "Assessing Agreement with Intraclass Correlation Coefficient and Concordance Correlation Coefficient for Data with Repeated Measures." *Comput. Stat. Data Anal.* 60 (April). Amsterdam, The Netherlands, The Netherlands: Elsevier Science Publishers B. V.: 132–45. doi:10.1016/j.csda.2012.11.004.

10. Davidson, Russell, and Emmanuel Flachaire. 2008. "The Wild Bootstrap, Tamed at Last." *Journal of Econometrics* 146 (1): 162–69.

11. Davis, C. S. 2002. *Statistical Methods for the Analysis of Repeated Measurements*. Springer Science & Business Media.

12. Davison, A. C., and D. V. Hinkley. 1997. *Bootstrap Methods and Their Application*. Cambridge University Press.

13. Field, Chris, and Alan Welsh. 2007. "Bootstrapping Clustered Data." *J. R. Stat. Soc. Ser. B Stat. Methodol.* 69: 369–90.

14. Field, Chris, Zhen Pang, and Alan Welsh. 2008. "Bootstrapping Data with Multiple Levels of Variation." *Canad. J. Statist.* 36: 521–39.

15. ———. 2010. "Bootstrapping Robust Estimates for Clustered Data." *J. Amer. Statist. Assoc.* 105: 1606–16.

16. Friedrich, Sarah, Frank Konietschke, and Markus Pauly. 2016. "A Wild Bootstrap Approach for Nonparametric Repeated Measurements." *Computational Statistics & Data Analysis* 113 (July). doi:10.1016/j.csda.2016.06.016.

17. Ghidey, Wendimagegn, Emmanuel Lesaffre, and Geert Verbeke. 2010. "A Comparison of Methods for Estimating the Random Effects Distribution of a Linear Mixed Model." *Statistical Methods in Medical Research* 19: 575–600.

18. Goldstein, Harvey. 2011. *Multilevel Statistical Models*. 4th ed. Wiley.

19. González-Manteiga, Wenceslao, María José Lombardía, Isabel Molina, Domingo Morales, and Laureano Santamaría. 2008. "Analytic and Bootstrap Approximations of Prediction Errors Under a Multivariate Fay-Herriot Model." *Computational Statistics & Data Analysis* 52: 5242–52.

20. Harrison, G.A., and G. Brush. 1990. "On Correlations Between Adjacent Velocities and Accelerations in Longitudinal Growth Data." *Annals of Human Biology* 17 (1).

21. Hedeker, D., and R.D. Gibbons. 2006. *Longitudinal Data Analysis*. Wiley Series in Probability and Statistics 451. John Wiley & Sons.

22. Ho, Hsiu J., and Tsung-I Lin. 2010. "Robust Linear Mixed Models Using the Skew T Distribution with Application to Schizophrenia Data." *Biometrical Journal. Biometrische Zeitschrift* 52 (4): 449–69.

23. Howell, D. 2013. *Fundamental Statistics for the Behavioral Sciences*. Cengage Learning.

24. Jacqmin-Gadda, Hélène, Solenne Sibillot, Cécile Proust-Lima, Jean-Michel Molina, and Rodolphe Thiébaut. 2007. "Robustness of the Linear Mixed Model to Misspecified Error Distribution." *Computational Statistics & Data Analysis* 51: 5142–54.

25. Johnson, R. A., and D. W. Wichern. 2007. *Applied Multivariate Statistical Analysis*. Pearson.

26. Kherad-Pajouh, Sara, and Olivier Renaud. 2014. "A General Permutation Approach for Analyzing Repeated Measures Anova and Mixed-Model Designs." *Statistical Papers* 56 (August): 1–21. doi:10.1007/s00362-014-0617-3.

27. Konietschke, Frank, Arne Bathke, Solomon Harrar, and Markus Pauly. 2015. "Parametric and Nonparametric Bootstrap Methods for General Manova." *Journal of Multivariate Analysis* 140 (May). doi:10.1016/j.jmva.2015.05.001.

28. Lachos, Victor H., Pulak Ghosh, and Reinaldo Boris Arellano-Valle. 2010. "Likelihood Based Inference for Skew-Normal Independent Linear Mixed Models." *Statist. Sinica* 20: 303–22.

29. Laird, Nan M. 1978. "Nonparametric Maximum Likelihood Estimation of a Mixing Distribution." *Journal of the American Statistical Association* 73: 805–11.

30. Laird, Nan M., and James H. Ware. 1982. "Random-Effects Models for Longitudinal Data." *Biometrics* 38 (4): 963–74.

31. Lange, Nicholas, and Louise Ryan. 1989. "Assessing Normality in Random Effects Models." *Ann. Statist.* 17 (2). The Institute of Mathematical Statistics: 624–42. doi:10.1214/aos/1176347130.

32. Liang, Kung Yee, and Scott Zeger. 1986. "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika* 73: 13–22.

33. Lin, Lawrence. 1989. "A Concordance Correlation-Coefficient to Evaluate Reproducibility." *Biometrics* 45 (April): 255–68. doi:10.2307/2532051.

34. Mammen, Enno. 1993. "Bootstrap and Wild Bootstrap for High Dimensional Linear Models." *Ann. Statist.* 21 (1). The Institute of Mathematical Statistics: 255–85. doi:10.1214/aos/1176349025.

35. McCulloch, C. E., and S. R. Searle. 2001. *Generalized, Linear, and Mixed Models*. Wiley.

36. McCulloch, Charles E, and John M Neuhaus. 2011. "Misspecifying the Shape of a Random Effects Distribution: Why Getting It Wrong May Not Matter." *Stat. Sci.* 26: 388–402.

37. Nguyen, Hien Duy, and Geoffrey J. McLachlan. 2016. "Linear Mixed Models with Marginally Symmetric Nonparametric Random Effects." *Computational Statistics & Data Analysis* 103: 151–69.

38. O'Shaughnessy, P. Y., and Alan H. Welsh. 2018. "Bootstrapping Longitudinal Data with Multiple Levels of Variation." *Computational Statistics & Data Analysis* 124: 117–31.

39. Pinheiro, J. C., and D. M. Bates. 2000. *Mixed-Effects Models in S and S-Plus*. Springer.

40. Pinheiro, José C., Chuanhai Liu, and Ying Nian Wu. 2001. "Efficient Algorithms for Robust Estimation in Linear Mixed-Effects Models Using the Multivariate T Distribution." *J. Comput. Graph. Statist.* 10: 249–76.

41. Pinheiro, Jose, Douglas Bates, Saikat DebRoy, Deepayan Sarkar, and R Core Team. 2019. *nlme: Linear and Nonlinear Mixed Effects Models*. https://CRAN.R-project.org/package=nlme.

42. Royall, Richard M. 1986. "Model Robust Confidence Intervals Using Maximum Likelihood Estimators." *Internat. Statist. Rev.* 2: 221–26.

43. Samanta, Mayukh, and Alan H. Welsh. 2013. "Bootstrapping for Highly Unbalanced Clustered Data." *Computational Statistics & Data Analysis* 59: 70–81.

44. Schützenmeister, A., and H.-P. Piepho. 2011. "Residual Analysis of Linear Mixed Models Using a Simulation Approach." *Computational Statistics and Data Analysis* 56: 1405–16. doi:10.1016/j.csda.2011.11.006.

45. Shrout, P. E., and J. L. Fleiss. 1979. "Intraclass Correlations: Uses in Assessing Rater Reliability." *Psychol Bull* 86 (2): 420–38.

46. Song, P. X.-K., P. Zhang, and A. Qu. 2007. "Maximum Likelihood Inference in Robust Linear Mixed-Effects Models Using Multivariate T Distributions." *Statist. Sinica* 17: 929–43.

47. Stevens, J. P. 2012. *Applied Multivariate Statistics for the Social Sciences*. Routledge.

48. Suo, Chen, Timothea Toulopoulou, Elvira Bramon, Muriel Walshe, Marco Picchioni, Robin Murray, and J Ott. 2013. "Analysis of Multiple Phenotypes in Genome-Wide Genetic Mapping Studies." *BMC Bioinformatics* 14 (May).

49. Tsai, Miao-Yu. 2015. "Comparison of Concordance Correlation Coefficient via Variance Components, Generalized Estimating Equations and Weighted Approaches with Model Selection." *Computational Statistics & Data Analysis* 82 (February): 47–58. doi:10.1016/j.csda.2014.08.005.

50. Tsai, Miao-Yu, and Chaochun Lin. 2017. "Concordance Correlation Coefficients Estimated by Variance Components for Longitudinal Normal and Poisson Data." *Computational Statistics & Data Analysis* 121 (December). doi:10.1016/j.csda.2017.12.003.

51. Vangeneugden, Tony, Annouschka Laenen, Helena Geys, Didier Renard, and Geert Molenberghs. 2005. "Applying Concepts of Generalizability Theory on Clinical Trial Data to Investigate Sources of Variation and Their Impact on Reliability." *Biometrics* 61 (1): 295–304. https://ideas.repec.org/a/bla/biomet/v61y2005i1p295-304.html.

52. Verbeke, G., and G. Molenberghs. 2000. *Linear Mixed Models for Longitudinal Data*. Springer.

53. Verbeke, Geert, and Emmanuel Lesaffre. 1997. "The Effect of Misspecifying the Random-Effects Distribution in Linear Mixed Models for Longitudinal Data." *Computational Statistics & Data Analysis* 23: 541–56.

54. Xu, Jin, and Xinping Cui. 2008. "Robustified Manova with Applications in Detecting Differentially Expressed Genes from Oligonucleotide Arrays." *Bioinformatics (Oxford, England)* 24 (May): 1056–62. doi:10.1093/bioinformatics/btn053.

55. Zhang, Daowen, and Marie Davidian. 2001. "Linear Mixed Models with Flexible Distributions of Random Effects for Longitudinal Data." *Biometrics* 57 (3): 795–802.

2.7. Annex

Normal-Normal scenario, 90% confidence.

Method	Statistic	β_{int}	β_{age}	σ_{int}	σ_{age}	ρ	σ_{resid}
Parametric bootstrap	Below rate	0.0610	0.0576	0.0134	0.0154	0.0652	0.0372
	Cover rate	0.8804	0.8880	0.8818	0.8844	0.8972	0.8960
	Above rate	0.0586	0.0544	0.1048	0.1002	0.0376	0.0668
	Mean width	5.2064	1.1055	3.7449	0.8142	0.4151	0.1149
	Width SD	0.8092	0.1701	0.5776	0.1211	0.1047	0.0094
Semi-parametric bootstrap	Below rate	0.0566	0.0582	0.0188	0.0158	0.0714	0.0360
	Cover rate	0.8832	0.8856	0.8470	0.8656	0.8948	0.8950
	Above rate	0.0602	0.0562	0.1342	0.1186	0.0338	0.0690
	Mean width	5.1587	1.0951	3.5198	0.7864	0.4098	0.1146
	Width SD	0.7919	0.1706	0.7101	0.1259	0.1080	0.0116
Wild bootstrap	Below rate	0.0536	0.0560	0.0074	0.0094	0.0762	0.0232
	Cover rate	0.8896	0.8866	0.9260	0.8850	0.8916	0.9462
	Above rate	0.0568	0.0574	0.0666	0.1056	0.0322	0.0306
	Mean width	5.2607	1.0918	4.3458	0.8639	0.4206	0.1376
	Width SD	0.8144	0.1720	0.7722	0.1442	0.1079	0.0130
True Distribution	Below rate	0.0570	0.0556	0.0134	0.0156	0.0662	0.0358
	Cover rate	0.8842	0.8906	0.8830	0.8798	0.8986	0.9004
	Above rate	0.0588	0.0538	0.1036	0.1046	0.0352	0.0638
	Mean width	5.2078	1.1060	3.7379	0.8120	0.4151	0.1148
	Width SD	0.8080	0.1706	0.5652	0.1206	0.1045	0.0092

Uniform-Uniform scenario, 90% confidence.

Method	Statistic	β_{int}	β_{age}	σ_{int}	σ_{age}	ρ	σ_{resid}
Parametric bootstrap	Below rate	0.0560	0.0576	0.0004	0.0078	0.0448	0.0064
	Cover rate	0.8874	0.8842	0.9694	0.9226	0.9302	0.9726
	Above rate	0.0566	0.0582	0.0302	0.0696	0.0250	0.0210
	Mean width	5.2695	1.1152	3.7822	0.8183	0.4152	0.1148
	Width SD	0.5972	0.1549	0.4216	0.1090	0.0973	0.0084
Semi-parametric bootstrap	Below rate	0.0578	0.0596	0.0148	0.0144	0.0774	0.0172
	Cover rate	0.8824	0.8838	0.9044	0.8836	0.8946	0.9392
	Above rate	0.0598	0.0566	0.0808	0.1020	0.0280	0.0436
	Mean width	5.2115	1.1030	2.6893	0.7087	0.3776	0.0956
	Width SD	0.5869	0.1526	0.3277	0.0949	0.1035	0.0069
Wild bootstrap	Below rate	0.0508	0.0584	0.0030	0.0064	0.0766	0.0082
	Cover rate	0.8966	0.8822	0.9752	0.9084	0.8994	0.9806
	Above rate	0.0526	0.0594	0.0218	0.0852	0.0240	0.0112
	Mean width	5.3171	1.1005	3.7132	0.8025	0.3972	0.1200
	Width SD	0.5894	0.1549	0.3350	0.1065	0.0991	0.0083
True Distribution	Below rate	0.0578	0.0556	0.0162	0.0170	0.0840	0.0342
	Cover rate	0.8842	0.8862	0.8886	0.8798	0.8862	0.8960
	Above rate	0.0580	0.0582	0.0952	0.1032	0.0298	0.0698
	Mean width	5.2687	1.1140	2.5018	0.6962	0.3700	0.0835
	Width SD	0.5962	0.1555	0.2805	0.0969	0.1031	0.0060

Exponential-Exponential scenario, 90% confidence.

Method	Statistic	β_{int}	β_{age}	σ_{int}	σ_{age}	ρ	σ_{resid}
Parametric bootstrap	Below rate	0.0290	0.0320	0.0968	0.0682	0.1512	0.1438
	Cover rate	0.8564	0.8728	0.6064	0.7154	0.7604	0.6352
	Above rate	0.1146	0.0952	0.2968	0.2164	0.0884	0.2210
	Mean width	5.0712	1.0948	3.6455	0.8031	0.4124	0.1144
	Width SD	1.3893	0.2411	1.0041	0.1670	0.1351	0.0129
Semi-parametric bootstrap	Below rate	0.0348	0.0386	0.0224	0.0194	0.1022	0.0360
	Cover rate	0.8544	0.8714	0.7096	0.7800	0.8354	0.8124
	Above rate	0.1108	0.0900	0.2680	0.2006	0.0624	0.1516
	Mean width	5.0133	1.0820	4.9466	0.9662	0.4713	0.1715
	Width SD	1.3650	0.2384	2.0690	0.3105	0.1352	0.0460
Wild bootstrap	Below rate	0.0328	0.0416	0.0528	0.0182	0.1288	0.0404
	Cover rate	0.8666	0.8684	0.7310	0.7928	0.7926	0.8564
	Above rate	0.1006	0.0900	0.2162	0.1890	0.0786	0.1032
	Mean width	5.1015	1.0774	4.9715	1.0142	0.4440	0.1899
	Width SD	1.3687	0.2365	1.8458	0.3090	0.1269	0.0442
True Distribution	Below rate	0.0388	0.0388	0.0154	0.0132	0.0548	0.0314
	Cover rate	0.8592	0.8766	0.8492	0.8762	0.8964	0.8940
	Above rate	0.1020	0.0846	0.1354	0.1106	0.0488	0.0746
	Mean width	5.0442	1.0918	6.5112	1.1965	0.5291	0.2063
	Width SD	1.3912	0.2398	1.8057	0.2736	0.1193	0.0239

Visualizing Type II Error in Normality Tests

José A. Sánchez-Espigares, Pere Grima, and Lluís Marco-Almagro

Department of Statistics and Operational Research, Universitat Politècnica de Catalunya - Barcelona Tech, Barcelona, Spain

ABSTRACT

A skewed exponential power distribution, with parameters defining kurtosis and skewness, is introduced as a way to visualize Type II error in normality tests. By varying these parameters a mosaic of distributions is built, ranging from double exponential to uniform or from positive to negative exponential; the normal distribution is a particular case located in the center of the mosaic. Using a sequential color scheme, a different color is assigned to each distribution in the mosaic depending on the probability of committing a Type II error. This graph gives a visual representation of the power of the performed test. This way of representing results facilitates the comparison of the power of various tests and the influence of sample size. A script to perform this graphical representation, programmed in the R statistical software, is available online as supplementary material.

1. Introduction

Generally in hypothesis testing, what we want to verify has the burden of proof. Nevertheless, this is not the case in normality tests, where what we want to prove (the distribution is normal) is assumed as true. Moreover, not having an alternative distribution, the discussion on Type II error (considering the distribution is normal when it is not) is frequently skipped or at least minimized in general statistics courses, although this Type II error can be high.

This problem has been widely discussed in the statistical literature, and several studies on the power of existing tests have been published. For example, Farrell and Rogers-Stewart (2006), Yacini and Yolacan (2007), and Yap and Sim (2011) compared the power of different normality tests against a set of alternative distributions. Results are shown in tables or power curves based on sample size, and a discussion of the most appropriate test depending on the kind of deviation from normality for each of the alternative distributions is included.

Other articles emphasize the interest of having a visual representation of the test, facilitating the observation of departures from normality. This happens with the Kolmogorov–Smirnov test (KS), which is easy to understand and visualize, and still remains common in both textbooks and software packages (although its power is worse than that of other normality tests). In this regard, Rosenkrantz (2000) derived bounds for the theoretical quantile function suggesting a test that has the same visual representation advantages of the KS test. In a similar manner, Aldor-Noiman et al. (2013) proposed a new method that provides simultaneous confidence bands for a normal quantile-quantile plot that are narrower in the extremes than those associated with the KS test, so higher powers are obtained, keeping the visualization benefit.

This article presents a visual and easy to understand way to show Type II error in normality tests. A graph shows a set of alternative distributions in a grid—a kind of mosaic of distributions—with the normal distribution in the center, and changing asymmetry as we move left and right, and changing kurtosis as we move up and down. A wide range of alternatives is then visible, and it is possible to test the hypothesis of normality for samples generated from each distribution. A sequential color scheme is then used to represent the probability of declaring normality, giving a visual representation of the power of the normality test used.

The following sections develop in detail the construction of the graph, together with an example showing its practical application.

2. Mosaic of Distributions

Zhu and Zinde-Walsh (2009) showed (as a particular case of a more general distribution) a version of the skewed exponential power distribution (SEPD) that can be adapted to our needs:

$$
f\left(x \mid \mu^*, \sigma^*, \alpha, p\right)
$$
$$
= \begin{cases} \frac{1}{\sigma^*} K(p) \exp\left(-\frac{1}{p}\left|\frac{x-\mu^*}{2\alpha\sigma^*}\right|^p\right) & \text{if } x \leq \mu^* \\ \frac{1}{\sigma^*} K(p) \exp\left(-\frac{1}{p}\left|\frac{x-\mu^*}{2(1-\alpha)\sigma^*}\right|^p\right) & \text{if } x > \mu^* \end{cases}. \quad (1)
$$

We say that $X \sim \text{SEPD}\ (\mu^*, \sigma^*, \alpha, p)$, where μ^* and σ^* are, respectively, the location and scale parameters, and correspond to the mean and standard deviation in the case of the Normal distribution, $p \geq 0$ is the parameter related to kurtosis, $\alpha \in [0, 1]$ is related to skewness, and $K(p)$ is the normalization constant, $K\ (p) = 1/[2p^{1/p}\Gamma(1 + 1/p)]$. Values $p = 2$ and $\alpha = 0.5$ correspond to the Normal distribution.

3. Visualizing Type II Error in Normality Tests

José A. Sánchez-Espigares, Pere Grima, Lluís Marco-Almagro

Department of Statistics and Operations Research
Universitat Politècnica de Catalunya-BarcelonaTech, Barcelona, Spain

ABSTRACT

A Skewed Exponential Power Distribution, with parameters defining kurtosis and skewness, is introduced as a way to visualize Type II error in normality tests. By varying these parameters a mosaic of distributions is built, ranging from double exponential to uniform or from positive to negative exponential; the normal distribution is a particular case located in the center of the mosaic. Using a sequential color scheme, a different color is assigned to each distribution in the mosaic depending on the probability of committing a Type II error. This graph gives a visual representation of the power of the performed test. This way of representing results facilitates the comparison of the power of various tests and the influence of sample size. A script to perform this graphical representation, programmed in the R statistical software, is available online as supplementary material.

KEYWORDS: Normality test, Type II error, Visualizing information, Teaching statistics, SEPD Distributions

3.1.   Introduction

Generally in hypothesis testing, what we want to verify has the burden of proof. Nevertheless, this is not the case in normality tests, where what we want to prove (the distribution is normal) is assumed as true. Moreover, not having an alternative distribution, the discussion on Type II error (considering the distribution is normal when it is not) is frequently skipped or at least minimized in general statistics courses, although this Type II error can be high.

This problem has been widely discussed in the statistical literature, and several studies on the power of existing tests have been published. For example, Farrell and Rogers-Stewart (2006), Yacini and Yolacan (2007) and Yap and Sim (2011) compare the power of different normality tests against a set of alternative distributions. Results are shown in tables or power curves based on sample size, and a discussion of the most appropriate test depending on the kind of deviation from normality for each of the alternative distributions is included.

Other papers emphasize the interest of having a visual representation of the test, facilitating the observation of departures from normality. This happens with the Kolmogorov-Smirnov test (KS), which is easy to understand and visualize, and still remains common in both textbooks and software packages (although its power is worse than that of other normality tests). In this regard, Rosenkrantz (2000) derives bounds for the theoretical quantile function suggesting a test that has the same visual representation advantages of the KS test. In a similar manner, Aldor-Noiman et al. (2013) propose a new method which provides simultaneous confidence bands for a normal quantile-quantile plot that are narrower in the extremes than those associated with the KS test, so higher powers are obtained, keeping the visualization benefit.

This paper presents a visual and easy to understand way to show Type II error in normality tests. A graph shows a set of alternative distributions in a grid – a kind of mosaic of distributions – with the normal distribution in the center, and changing asymmetry as we move left and right, and changing kurtosis as we move up and down. A wide range of alternatives is then visible, and it is possible to test the hypothesis of normality for samples generated from each distribution. A sequential color scheme is then used to represent the probability of declaring normality, giving a visual representation of the power of the normality test used.

The following sections develop in detail the construction of the graph, together with an example showing its practical application.

3.2.   *Mosaic of Distributions*

Zhu and Zinde-Walsh (2009) show (as a particular case of a more general distribution) a version of the Skewed Exponential Power Distribution (SEPD) that can be adapted to our needs:

$$f(x \mid \mu^*, \sigma^*, \alpha, p) = \begin{cases} \dfrac{1}{\sigma^*} K(p) \exp\left(-\dfrac{1}{p}\left|\dfrac{x - \mu^*}{2\alpha\sigma^*}\right|^p\right) & \text{if } x \leq \mu^* \\ \dfrac{1}{\sigma^*} K(p) \exp\left(-\dfrac{1}{p}\left|\dfrac{x - \mu^*}{2(1-\alpha)\sigma^*}\right|^p\right) & \text{if } x > \mu^* \end{cases} \tag{1}$$

We say that $X \sim SEPD\ (\mu^*, \sigma^*, \alpha,\ p)$, where μ^* and σ^* are respectively the location and scale parameters, and correspond to the mean and standard deviation in the case of the Normal distribution, $p \geq 0$ is the parameter related to kurtosis, $\alpha \in [0, 1]$ is related to skewness and $K(p)$ is the normalization constant, $K(p) = 1/[2p^{1/p}\Gamma(1 + 1/p)]$. Values $p = 2$ and $\alpha = 0.5$ correspond to the Normal distribution.

The kurtosis changes when moving p and keeping all other parameters constant. Analogously, the skewness changes when varying α (Figure 1).
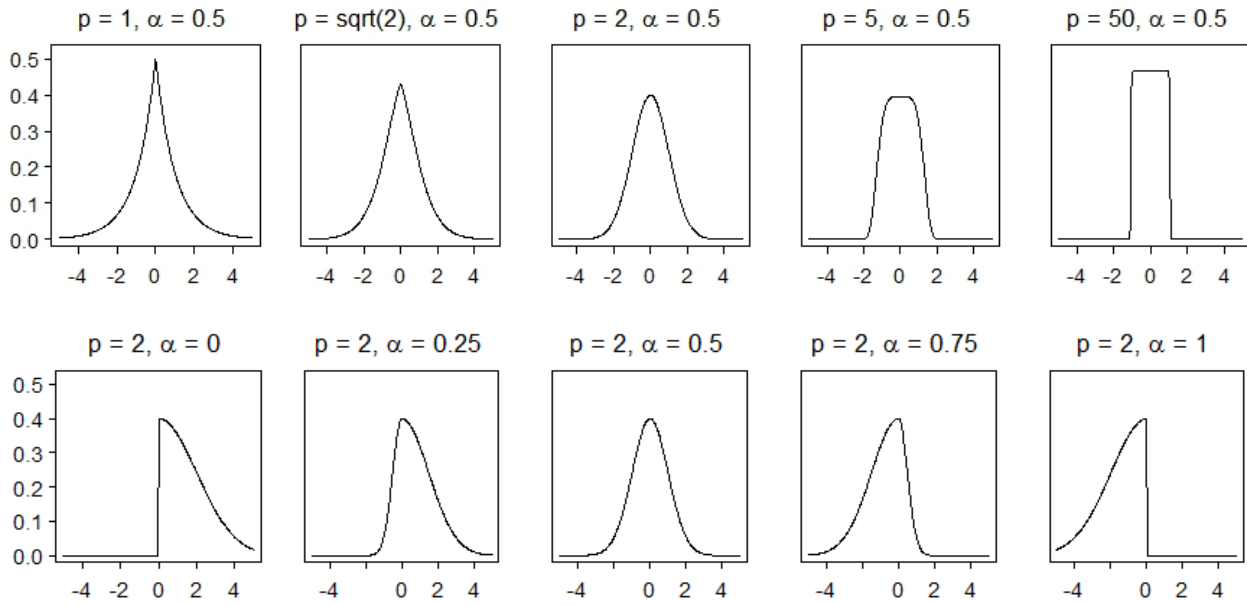
Figure 3.1: pdf of SEPD with $\mu^* = 0$ and $\sigma^* = 1$, with different values of p keeping $\alpha = 0.5$ (upper row) and different values of α keeping $p = 2$ (lower row)

As we want all distributions to have the same values of μ and σ, consider $X_z \sim SEPD\,(0; 1; \alpha; p)$. From the expressions obtained by Zhu and Zinde-Walsh (2009) for $E(X)$ and $V(X)$ follows that:

$$E(X_z) = \frac{1}{K(p)}\left[(1 - 2\alpha)\frac{p\Gamma(2/p)}{\Gamma^2(1/p)}\right] \tag{2}$$

$$V(X_z) = \frac{1}{[K(p)]^2}\left[(3\alpha^2 - 3\alpha + 1)\frac{p^2\Gamma(3/p)}{\Gamma^3(1/p)}\right]$$
$$- \frac{1}{[K(p)]^2}\left[(1 - 2\alpha)\frac{p\Gamma(2/p)}{\Gamma^2(1/p)}\right]^2 \tag{3}$$

We write $E(X_z) = A$ and $V(X_z) = B^2$ to lighten the notation and define:

$$Y \equiv g(X_Z) = \mu + \sigma\left(\frac{X_z - A}{B}\right) \tag{4}$$

One can check that $E(Y) = \mu$ and $V(Y) = \sigma^2$. In order to define the probability density function (pdf) of Y, a change of variable is applied:

$$f_Y(y) = f_X[g^{-1}(y)]\left|\frac{d}{dy}g^{-1}(y)\right|$$

with:

$$f_X(x_z|0,1,p,\alpha) = \begin{cases} K(p)\exp\left(-\dfrac{1}{p}\left|\dfrac{x_z}{2\alpha}\right|^p\right) & \text{if } x \le 0 \\[3mm] K(p)\exp\left(-\dfrac{1}{p}\left|\dfrac{x_z}{2(1-\alpha)}\right|^p\right) & \text{if } x > 0 \end{cases}$$

Isolating X_z in (4) we obtain:

$$g^{-1}(y) = \frac{y-\mu}{\sigma}B + A$$

Therefore, we have:

$$f(y) = \begin{cases} \dfrac{B}{\sigma}K(p)\exp\left(-\dfrac{1}{p}\left|\dfrac{A+B\left(\frac{y-\mu}{\sigma}\right)}{2\alpha}\right|^p\right) & \text{if } y \le \mu - \dfrac{A}{B}\sigma \\[5mm] \dfrac{B}{\sigma}K(p)\exp\left(-\dfrac{1}{p}\left|\dfrac{A+B\left(\frac{y-\mu}{\sigma}\right)}{2(1-\alpha)}\right|^p\right) & \text{if } y > \mu - \dfrac{A}{B}\sigma \end{cases} \tag{5}$$

In order to draw the mosaic of distributions, values of p between 1 (double exponential) and 50 (almost an uniform distribution) are used. The normal distribution ($p = 2$) is placed in the center. Each value of p is equal to the previous one raised to a certain value j, except for the second, which will always be equal to $2^{1/j^{\left(\frac{m}{2}-1.5\right)}}$, where m is the number of distributions on each side of the mosaic. For instance, for a 7x7 mosaic, these values will be: 1, $2^{1/j^2}$, $2^{1/j}$, 2, 2^j, 2^{j^2}, 2^{j^3}. As the highest value of p is equal to 50, in this case of $m = 7$ it can be deduced that $j = \sqrt[3]{\dfrac{\log 50}{\log 2}}$. In order to have the Normal distribution in the center, m must be an odd number; therefore, in general, $j = \sqrt[\frac{m}{2}-0.5]{\dfrac{\log 50}{\log 2}}$.

For α we take equidistant values between 0 and 1 (inclusive). Therefore, if we have n values, the i-th position will be $\alpha = \dfrac{i-1}{n-1}$

With these criteria it is possible to obtain mosaics of distributions that vary in a reasonably equidistant way between the negative ($\alpha = 0$) and positive ($\alpha = 1$) exponential distributions, and from the double exponential ($p = 1$) to an almost uniform distribution ($p = 50$). Figures 2 and 3 show 7x7 and 25x25 mosaics made with the statistical software R.
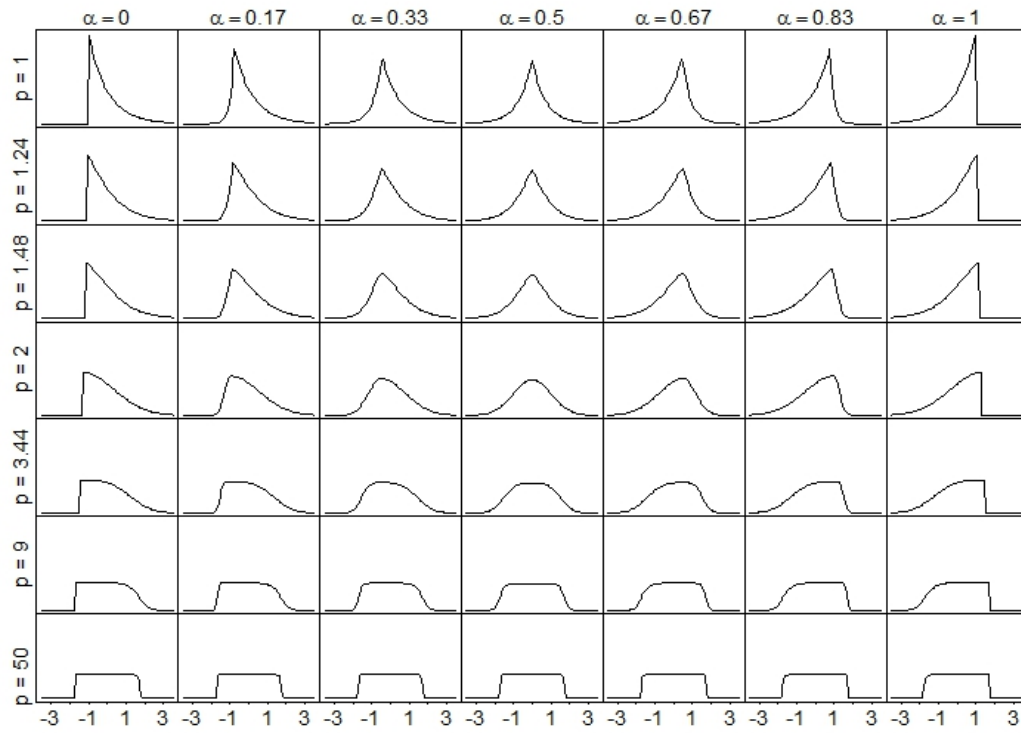
Figure 3.2: Mosaic of distributions (7x7) with $\mu = 0$, $\sigma = 1$
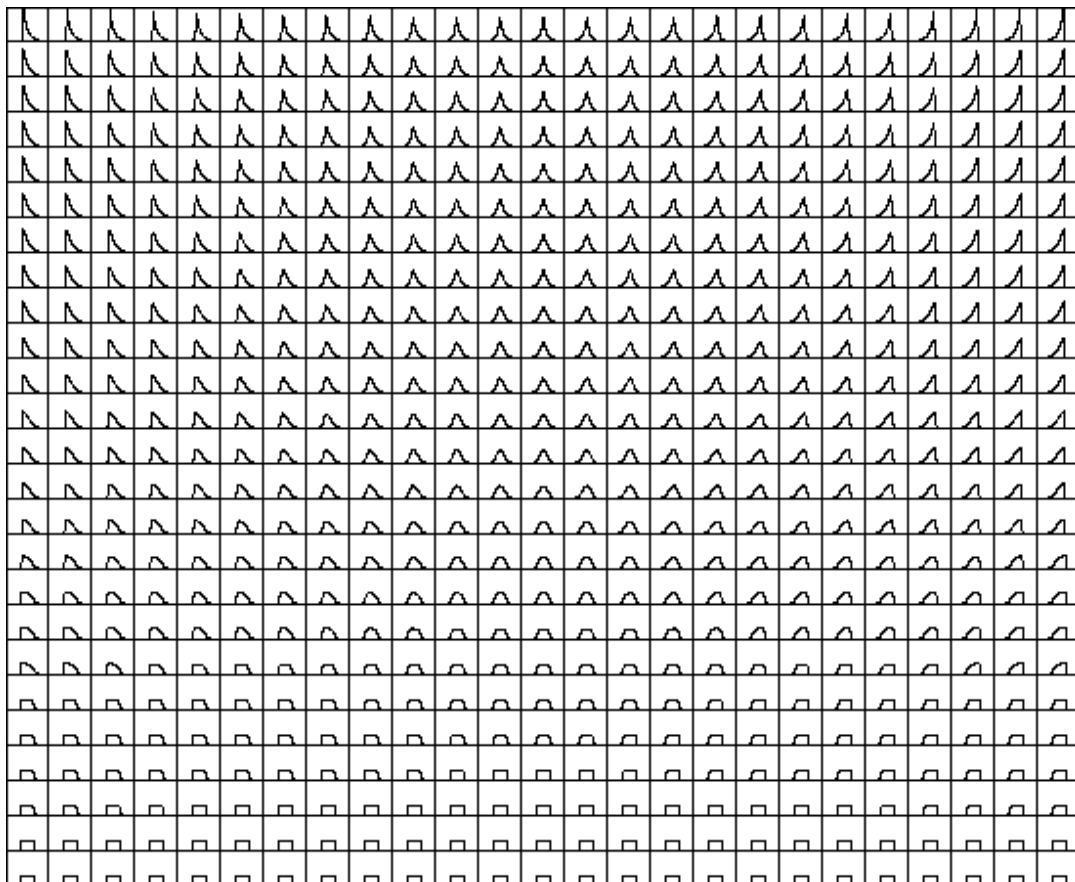
Figure 3.3: Mosaic of distributions (25x25)

3.3. *Visualizing the Probability of Declaring Normality*

Given a sample of size n, N samples of each of the distributions in the mosaic are generated. The values of μ and σ do not affect the appearance of the mosaic as the distributions do not have a scale in the axis. The random number generation for each distribution is done by a procedure analogous to that used by Zhu and Zinde-Walsh (2009). Values following $Y \sim SEPD$ $(0; 1; \alpha; p)$ are generated from a random number U from a uniform distribution $U(0,1)$, and another random number W from a gamma distribution with shape parameter $k = 1/p$ and scale parameter $\theta = 1$, in the form:

$$Y = \begin{cases} \dfrac{-\alpha W^{1/p}}{\Gamma\left(1 + \dfrac{1}{p}\right)} & \text{if } U < \alpha \\[4mm] \dfrac{(1-\alpha)W^{1/p}}{\Gamma\left(1 + \dfrac{1}{p}\right)} & \text{if } U \geq \alpha \end{cases}$$

A normality test is performed for each of the N samples of each distribution, and each p-value is recorded. Given a certain level of significance, we assign a value to each distribution equal to the proportion of samples (out of N) for which the assumption of normality is not rejected.

Figure 4 shows 7x7 mosaics with the probability (for each distribution) of declaring normality when using the Shapiro-Wilk test. These values are based on simulation of N = 10000 samples for each distribution. A level of significance of 5% is used. The calculations were performed with the R statistical software, using the 'nortest' package (Gross, 2013). Obviously, the power of the test increases with the sample size n.

A sequential gray scale is used to display results more clearly. From lighter to darker colors the ranges are 0-0.05, 0.05-0.25, 0.25-0.50 and 0.50-1.

Figure 5 shows the results obtained using the following normality tests: Kolmogorov-Smirnov-Lilliefords, Anderson-Darling and Shapiro-Wilk, with samples of size $n = 25$, 50 and 100. Dark areas in the mosaics are basically symmetric about the vertical axis crossing the center of the mosaic, due to the fact that distribution shapes are also symmetric about this axis. At the lower lines of the mosaic the distribution is almost uniform, so it changes scarcely from left to right. Therefore, the test either declares or not normality for all distributions in that line. On the contrary, skewness is relevant in the upper lines of the mosaic, with very asymmetric distributions on the left and the right parts of the graph, so the test only declares normality, at most, in the central distributions. These behaviors give the candle flare shape that can be seen in the graphs.
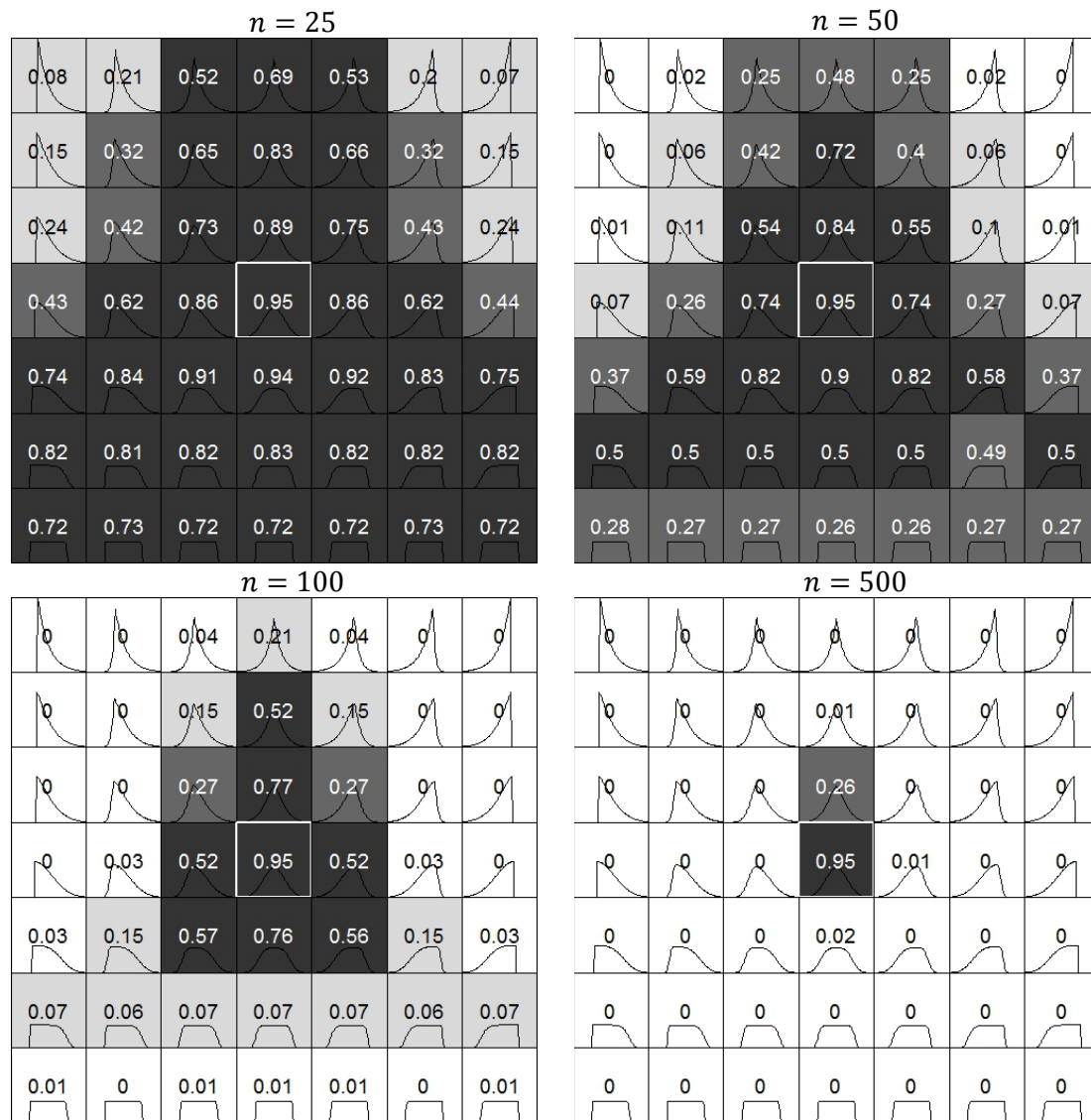
Figure 3.4: Mosaic of distributions including the proportion of times that the hypothesis of normality is not rejected (with 0.05 as significance level) using the Shapiro-Wilk test for the sample sizes shown

The smaller the dark area in the graph, the more powerful the test is. Obviously, when the sample size is increased, the dark area becomes smaller. Looking at Figure 5 we can also conclude that Anderson-Darling test works better than KS-Lillefords test, and Shapiro-Wilk test better than Anderson-Darling. The ideal situation –unattainable when working with samples – would be having only dark black in the center, where the normal distribution is located.

	KS-Lillefords	Anderson-Darling	Shapiro-Wilk
n= 25			
n= 50			
n= 100			
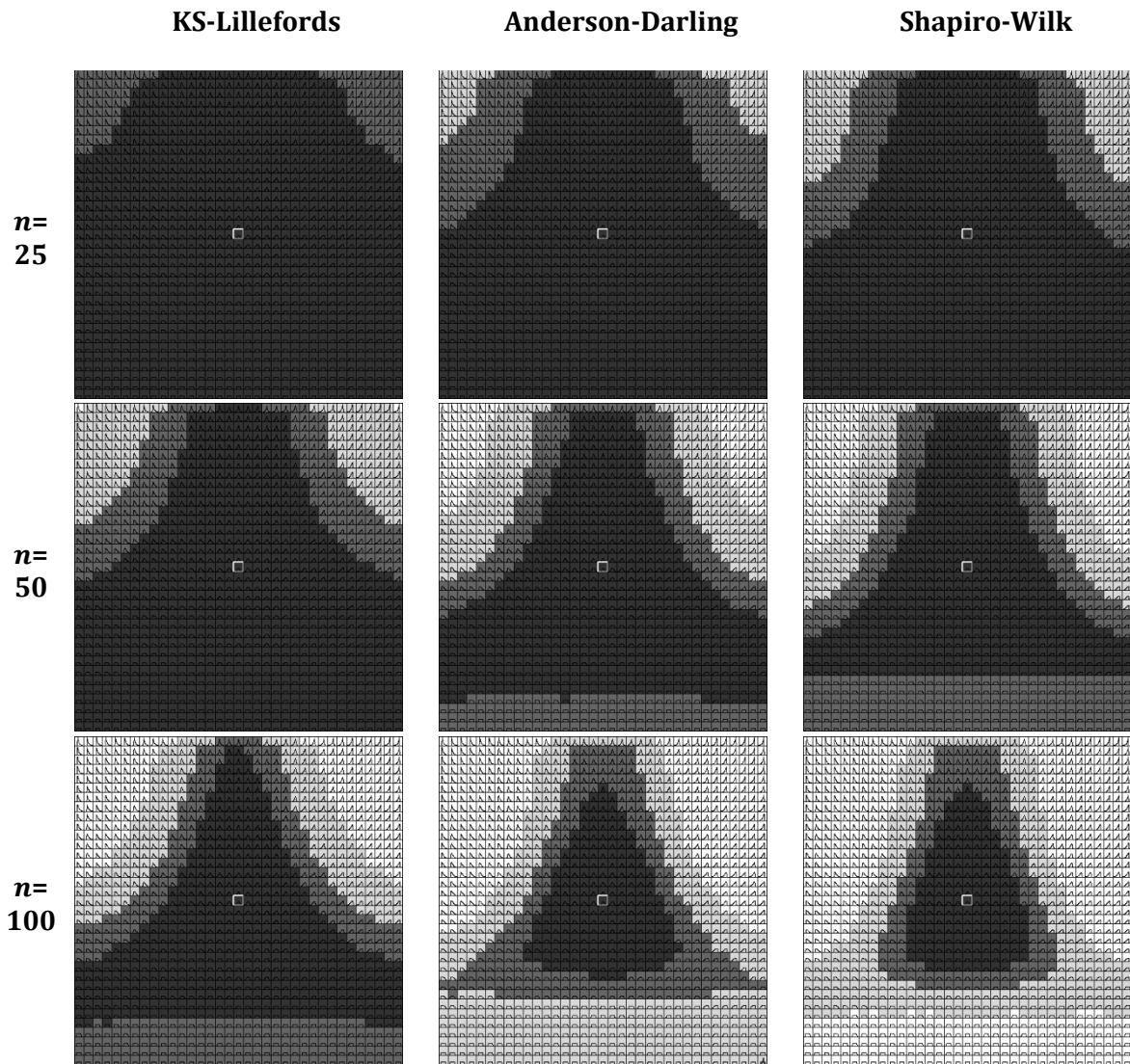
Figure 3.5: 35x35 mosaic of distributions for comparing the results using three different normality tests with three sample sizes.

In our experience, the existence and importance of Type II error in normality tests is often neglected both in academic and business settings. The graphs shown in this paper allow the visualization of Type II error and, in this way, help in understanding the limitations of these kind of tests. Also, the mosaic facilitates the comparison among normality tests and the impact of sample size in a visual manner.

Supplementary Materials

An R script named "Visualizing Type II Error.R" is attached as supplementary material. The script allows graphical representations as those shown in this paper. Instructions on how to use the script are contained in the script code.

3.4. *References*

1. Aldor-Noiman, S., Brown, L.D., Buja, A., Rolke, W. and Stine, R.A. (2013), "The Power to See: A New Graphical Test of Normality," The American Statistician, 67, 249-260.

2. Farrell, P.J., Rogers-Stewart, K. (2006), "Comprehensive study of tests for normality and symmetry: extending the Spiegelhalter test," Journal of Statistical Computation and Simulation, 76, 803-816.

3. Gross, J. and bug fixes by Ligges, U. (2012). nortest: Tests for Normality. R package version 1.0-2. http://CRAN.R-project.org/package=nortest

4. Rosenkrantz, W.A. (2000), "Confidence Bands for Quantile Functions: A Parametric and Graphic Alternative for Testing Goodness of Fit," The American Statistician, 54, 185-190.

5. Yacini, B., Yolacan, S. (2007), "A comparison of various tests of normality," Journal of Statistical Computation and Simulation, 77, 175-183.

6. Yap, B.W., Sim, C.H. (2011), "Comparisons of various types of normality tests," Journal of Statistical Computation and Simulation, 81, 2141-2155.

7. Zhu, D. and Zinde-Walsh, V. (2009), "Properties and estimation of asymmetric exponential power distribution," Journal of Econometrics, 148, 89-99

3.6.   *Supplementary Materials*

VisualizingTypeIIError.R

```r
###########################################################
# Supplementary material that accompanies the article:  #
#    Visualizing Type II Error in Normality Tests       #
###########################################################


# You can change the values of the following parameters:
siz=50      # sample size
nsig=0.05   # significance level
n=7         # mosaic size (nxn), with 1<n<50. If n is even, then n=n+1
nsim=1000   # number of simulations
test="SW"   # Type of test, AD=Anderson-Darling,SW=Shapiro-Wilks,KS=Kolmogorov-Smirnov
label=T     # T= type II error is shown in each cell; F= type II error is not shown
Opt=0       # 0: Standard; 1: Curve always black; 2: Only mosaic (without test)
###########################################################

library(nortest)
####Random generation for SEPD
rsepd=function(n,mu=0,sigma=1,alpha=0.5,p=2){
  kp=1/(2*gamma(1+1/p)*p^(1/p))
  u=runif(n)
  w=rgamma(n,shape=1/p,scale=1)
  y=ifelse(u>alpha,1-alpha,-alpha)*(w^(1/p))/gamma(1+1/p)
  y=y/kp
  A=2*p^(1/p)*((1-alpha)^2-alpha^2)*gamma(2/p)/gamma(1/p)
  A2=(2*p^(1/p))^2*((1-alpha)^3+alpha^3)*gamma(3/p)/gamma(1/p)
  B=sqrt(A2-A^2)
  mu+sigma*(y-A)/B
}


####Density function for SEPD
dsepd=function(x,mu=0,sigma=1,alpha=0.5,p=2){
  A=2*p^(1/p)*((1-alpha)^2-alpha^2)*gamma(2/p)/gamma(1/p)
  A2=(2*p^(1/p))^2*((1-alpha)^3+alpha^3)*gamma(3/p)/gamma(1/p)
  B=sqrt(A2-A^2)
  y=A+B*(x-mu)/sigma
```

```r
  kp=1/(gamma(1+1/p)*(2*p^(1/p)))
  kp*exp(-(abs(y/(2*ifelse(x<mu-sigma*A/B,alpha,1-alpha)))^p)/p)*B/sigma
}


###Monte Carlo p-value for godness-of-fit statistics
et2_MC=function(siz,nsim,mu,sigma,alpha,p,nsig,test){
  ysim=matrix(rsepd(siz*nsim,mu=mu,sigma=sigma,alpha=alpha,p=p),nc=siz)
  prova=switch(test,
        AD = function(el) ad.test(el)$p.value,
        SW = function(el) shapiro.test(el)$p.value,
        KS = function(el) lillie.test(el)$p.value)
  res=apply(ysim,1,prova)
  length(res[res>nsig])/nsim
}


#### Mosaic function
mosaic=function(siz,n,nsim=499,nsig=0.05,test="AD",label=F,Opt=0){
  if (n<=1 | n>=50) {
    stop("n must be a value between 2 and 49\n")
  }
  mu=0
  sigma=1
  if (n%%2==0) n=n+1
  cex1=c(1.2,1,0.8,0.7,0.6)
  old.par = par(mfrow = c(n,n), xaxt="n", yaxt="n", mar=c(0,0,0,0), cex=cex1[findInte
rval(n,c(1,5,10,15,25,9999))], xaxs="i", yaxs="i")
  s=3.5
  ymax=c(0,1/sigma)
  j=(log10(50)/log10(2))^(1/(n/2-0.5))
  alfa=(0:(n-1))/(n-1)
  alfa[alfa==0]=1e-12
  alfa[alfa==1]=1-1e-12
  p=c(1,2,50)
  if (n>3) p=c(1,2^(1/(j^(((n-3)/2):1))),2,2^(j^((1:(n/2-0.5)))))
  graf1=function(param,label){
    x=seq(from=mu-s*sigma,to=mu+s*sigma,length.out=101)
    plot(0,0,xlim=c(mu-s*sigma,mu+s*sigma),ylim=ymax,type="n",axes=F)
    pvalue=round(et2_MC(siz=siz,nsim=nsim,mu=mu,sigma=sigma,alpha=param[1],p=param[2]
,nsig=nsig,test=test),2)
    id=findInterval(pvalue,c(0,0.05,0.25,0.5,1))
    colo=gray(c(0.9,0.85,0.4,0.2,0))[id]
```

```r
    if(id>1) rect(mu-s*sigma,0,mu+s*sigma,ymax[2],col=colo)
    coline=switch(Opt+1, ifelse(id>2,"white","black"),"black","black")
    curve(dsepd(x,mu=mu,sigma=sigma,alpha=param[1],p=param[2]),col=coline,add=T)
    if (param[1]==0.5 & param[2]==2) box(lwd=2,col="white") else box()
    if (label) text(mu,ymax[2]*0.5,round(pvalue,2),col=ifelse(id>2,"white","black"))
  }
  graf2=function(param,label){
    x=seq(from=mu-s*sigma,to=mu+s*sigma,length.out=101)
    plot(0,0,xlim=c(mu-s*sigma,mu+s*sigma),ylim=ymax,type="n",axes=F)
    curve(dsepd(x,mu=mu,sigma=sigma,alpha=param[1],p=param[2]),col="black",add=T)
    box()
  }
  if (Opt==2) graf=graf2 else graf=graf1
  apply(expand.grid(alfa,p),1,graf,label=label)
  par(old.par)
}
mosaic(siz,n,nsim,nsig,test,label,Opt)
```

Check for updates

Graphical comparison of normality tests for unimodal distribution data

José A. Sánchez-Espigares, Pere Grima and Lluís Marco-Almagro

Department of Statistics and Operations Research, Universitat Politècnica de Catalunya-BarcelonaTech, Barcelona, Spain

ABSTRACT

A methodology is proposed to compare the power of normality tests with a wide variety of alternative unimodal distributions. It is based on the representation of a distribution mosaic in which kurtosis varies vertically and skewness horizontally. The mosaic includes distributions such as exponential, Laplace or uniform, with normal occupying the centre. Simulation is used to determine the probability of a sample from each distribution in the mosaic being accepted as normal. We demonstrate our proposal by applying it to the analysis and comparison of some of the most well-known tests.

1. Introduction

There are a wide variety of normality tests, from the classic Kolmogorov–Smirnov to the widely used Shapiro–Wilk and Anderson–Darling tests. Books are dedicated exclusively to normality tests, such as Thode [1], who describes dozens of them. The problem is not yet closed, as new tests and modifications of existing ones continue to emerge (see, for example, Desgagné and de Micheaux [2]).

The fact that so many exist surely indicates that not one is better than all the others in all circumstances. The comparison of normality tests has been addressed in articles such as Farell and Rogers-Stewart [3], where 14 test types are compared to 48 possible alternative distributions and the results are presented in a table with the power of the test indicated against each alternative for a significance level of 0.10 and a sample size of $n = 20$. Yacini and Yolocan [4] compare 12 tests against 5 alternative distributions and present the results also in a table where the power of the test is indicated according to the alternative distribution for $\alpha = 0.05$ and sample sizes of $n = 20, 30, 40$ and 50. Romão et al. [5] present an exhaustive study in which they describe and analyze the performance of 33 normality tests against data from a wide variety of distributions, presenting the power obtained in tables but also with line charts in which the horizontal axis indicates which test was performed. In a similar manner, Yap and Sim [6] compare 8 types of tests against 9 alternative distributions, presenting the results in tables as well as graphs, where for each alternative distribution the power curves are presented according to the size of the sample.

CONTACT Pere Grima ✉ pere.grima@upc.edu 🖃 Department of Statistics and Operations Research, Universitat Politècnica de Catalunya-BarcelonaTech, Barcelona 08028, Spain

4. Graphical Comparison of Normality Tests for Unimodal Distribution Data

José A. Sánchez-Espigares, Pere Grima, Lluís Marco-Almagro

Department of Statistics and Operations Research
Universitat Politècnica de Catalunya-BarcelonaTech, Barcelona, Spain

ABSTRACT

A methodology is proposed to compare the power of normality tests with a wide variety of alternative unimodal distributions. It is based on the representation of a distribution mosaic in which kurtosis varies vertically and skewness horizontally. The mosaic includes distributions such as exponential, Laplace or uniform, with normal occupying the center. Simulation is used to determine the probability of a sample from each distribution in the mosaic being accepted as normal. We demonstrate our proposal by applying it to the analysis and comparison of some of the most well-known tests.

KEYWORDS: Goodness-of-Fit, Test Power, Graphical Techniques, Comparing Normality Tests, Anderson-Darling, Shapiro-Wilk

4.1.   Introduction

There are a wide variety of normality tests, from the classic Kolmogorov-Smirnov to the widely used Shapiro-Wilk and Anderson-Darling tests. Books are dedicated exclusively to normality tests, such as Thode [1], who describes dozens of them. The problem is not yet closed, as new tests and modifications of existing ones continue to emerge (see, for example, Desgagné and de Micheaux [2]).

The fact that so many exist surely indicates that not one is better than all the others in all circumstances. The comparison of normality tests has been addressed in articles such as Farell and Rogers-Stewart [3], where 14 test types are compared to 48 possible alternative distributions and the results are presented in a table with the power of the test indicated against each alternative for a significance level of 0.10 and a sample size of n=20. Yacini and Yolocan [4] compare 12 tests against 5 alternative distributions and present the results also in a table where the power of the test is indicated according to

the alternative distribution for α=0.05 and sample sizes of n=20, 30, 40 and 50. Romão *et al.* [5] present an exhaustive study in which they describe and analyze the performance of 33 normality tests against data from a wide variety of distributions, presenting the power obtained in tables but also with line charts in which the horizontal axis indicates which test was performed. In a similar manner, Yap and Sim [6] compare 8 types of tests against 9 alternative distributions, presenting the results in tables as well as graphs, where for each alternative distribution the power curves are presented according to the size of the sample.

Our work proposes a methodology for comparing the power of normality tests to a wide variety of alternative unimodal distributions in a highly visual manner and with a single graph. It is based on the representation suggested by Sánchez-Espigares *et al.* [7], who builds a distribution mosaic in which the kurtosis varies vertically and the skewness horizontally. The mosaic includes distributions such as exponential, Laplace and uniform, with normal occupying the center. Simulation is used to determine the probability of a sample from each of the distributions in the mosaic being accepted as normal.

The next section describes the tests to be compared. Next, we describe the characteristics of the distribution mosaic, how the power of the tests is represented in the mosaic and, finally, the studied tests are analyzed and compared.

4.2.   Test Selection

To demonstrate the possibilities of the proposed procedure and also compare some of the most well-known tests, we have chosen three from each of the strategies in which normality tests can be grouped: regression tests, tests based on the empirical distribution function (EDF) and tests based on moments.

Regression tests are based on the fact that the distribution function $F(x)$ of a random variable $X \sim N(\mu, \sigma)$ is a straight line when represented on a normal probability plot (Q-Q plot). Therefore, given an ordered sample of values $x_{(1)} \cdots x_{(n)}$ with $F\left(x_{(i)}\right), \cdots F\left(x_{(n)}\right)$ values of their distribution function, points $\left(x_{(i)}, F(x_{(i)})\right)$ should align approximately according to a straight line in a Q-Q plot, and any departure from that alignment indicates the data's lack of normality. From among the tests based on this idea, we have selected:

- Shapiro-Wilk (SW) is probably the best known and most often used (a detailed description can be found, for instance, in Thode [1]). The original version [8] has some computational limitations, especially for large sample sizes. Royston [9] suggested a transformation of the original statistic that allows it to be applied to sample sizes of up to $n = 2000$ without any demand for great computational resources.

- Shapiro-France (SF) (see Thode [1]) is a variant of SW. When it appeared in 1972 [10], its main advantage was the demand for fewer computational resources than the original SW. This advantage has ceased to be of interest, especially after Royston's contributions to the SW test; but it continues to be among the most representative of this group of tests that are based on correlation.

- Filliben [11] uses the correlation between the sample order statistics and the estimated median values of the theoretical order statistics. Its main advantage is that the calculations are very easy because there is no need to calculate the expected values of the normal order statistics.

The test statistic in tests based on the EDF is a measure of the discrepancy between the EDF and the theoretical distribution function. The most typical is that of Kolmogorov-Smirnov (KS), which uses the maximum distance – in absolute value – between both distributions. The tests of this type that we analyze here are:

- Lilliefors. This uses the same test statistic as that of KS, but with a different reference distribution, due to the fact that the KS test requires knowledge of the population parameters, while Lilliefors bases its estimation on the sample. Lilliefors deduced the critical values through simulation, but analytical methods for determining them have also been published [12].

- Cramer-von Mises (CvM). The test statistic is determined from the discrepancy between the theoretical distribution function F and the empirical function F_n accumulated throughout all the variation space of x. It is specified in a relatively simple formula (see, for example, [13]). The critical values depend on the size of the sample and the number of known population parameters.

- Anderson-Darling (AD). This is surely the most commonly used of this group. It is similar to the CvM but gives more weight to the discrepancy in the tails of the distribution (see, for example, [13]).

Finally, we have the group that uses a test statistic based on the difference between, on the one hand, the kurtosis and the skewness of the data (third and fourth moment) and, on the other, their theoretical values. For this group, we have selected:

- D'Agostino-Pearson K^2 (DA). The test statistic is a function of the kurtosis and skewness of the sample. It follows a Chi-square distribution with 2 degrees of freedom if the hypothesis of normality is true [14].

- Jarque-Bera (JB). Conceptually similar to the one above. The test statistic is also calculated from the kurtosis and skewness of the sample. Thus, it is also distributed as a Chi-square with 2 degrees of freedom; but when $n < 2000$, the p-value is determined by simulation [15].

- Adjusted Jarque-Bera (AJB). It uses a new test statistic computed from the first four moments about the origin. The p-values are determined by simulation [16].

To apply these tests and analyze their performance, we have used functions that have already been developed and implemented in R statistical software packages [17]. Table 1 indicates which package and function were used for each test.

Table 4.1 Tests analyzed and the R packages and functions that were used to apply them.

Test	Package	Function
Shapiro-Wilk	`stats`, R Core Team [17]	`shapiro.test(x)`
Shapiro-Francia	`nortest`, Gross and Ligges [18]	`sf.test(x)`
Filliben's	`ppcc`, Pohlert [19]	`ppccTest(x, "qnorm")`
Lilliefords	`nortest`, Gross and Ligges [18]	`lillie.test(x)`
Cramer-von Mises	`nortest`, Gross and Ligges [18]	`cvm.test(x)`
Anderson-Darling	`nortest`, Gross and Ligges [18]	`ad.test(x)`
D'Agostino-Pearson K_s^2	`fBasics`, Wuertz et al. [20]	`dagoTest(x)`
Jarque-Bera	`normtest`, Gavrilov and Pusev [21]	`jb.norm.test(x)`
Adjusted Jarque-Bera	`normtest`, Gavrilov and Pusev [21]	`ajb.norm.test(x)`

4.3.   *Distribution mosaic. Representation of the power of a test*

Based on a Skewed Exponential Power Distribution (SEPD) used by Zhu and Zinde-Walsh [22], Sánchez-Espigares *et al.* [7] consider the probability density function that is used to create the mosaic distributions. This function is characterized by the mean and variance of the variable considered and also a third parameter, p, which is related to kurtosis and varies between 1 (double exponential distribution) and 50 (practically a uniform distribution). It also employs a fourth parameter, α, which is related to the asymmetry that varies between 0 (very asymmetric distribution with tail to the right) and 1 (with tail to the left). The values $p = 2$ and $\alpha = 0.5$ correspond to a normal distribution.

We want the number of distributions on each side of the mosaic to be odd so that the normal distribution remains exactly in the center. It is easily deduced that for α to vary between 0 and 1 in equidistant intervals and for $\alpha = 0.5$ to remain in the center, it is sufficient that the i-th position has the value $\alpha = \frac{i-1}{m-1}$, with m being the number of distributions on each side of the mosaic. Regarding the values of p, their determination is not so immediate. Each value must be equal to the previous one raised to a power of $j = \sqrt[\frac{m}{2}-0.5]{\frac{\log 50}{\log 2}}$, except for the second one, which always equals $2^{1/j^{\left(\frac{m}{2}-1.5\right)}}$ [7]. For example, if the mosaic is of size 11x11 ($m = 11$), we have $j = 1.4136$ and the values of α and p will be:

i	1	2	3	4	5	6	7	8	9	10	11
α	0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1
p	1	.19	.28	.41	.63	2	2.66	3.99	7.08	15.92	50

Figure 1 shows an 11x11 mosaic. The values of μ and σ are the same for all distributions and do not affect their shape but only the scale of the axes, which we do not consider here. Each distribution corresponds to the values of α and p that are indicated. In [**¡Error! Marcador no definido.**], the R code is included to create mosaics as large as 49x49, although it can easily be changed to obtain larger mosaics.
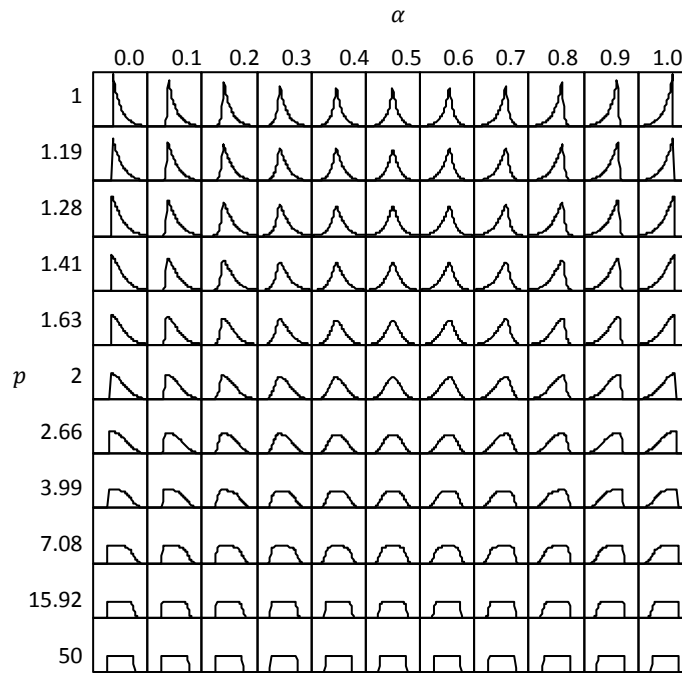
Figure 4.1: Mosaic of 11x11 distributions with the p and α values that correspond to each one

From each of the distributions that appear in the mosaic, a random sample of size n can be obtained and contrasted against the normal distribution by means of the test whose power we want to analyze. We chose, for example, the Anderson-Darling test and generate 10000 samples of size n=100 from each of the distributions that appear in the mosaic of Figure 1. We consider that the hypothesis of normality is not rejected if the p-value obtained is greater than 0.05; and we annotate on each distribution the proportion of times that the hypothesis of normality would not be rejected with samples from that distribution. The values obtained are indicated in Figure 2. The area is outlined for the distributions with values of this proportion greater than 0.5. In this figure, it can be observed that, if the population from which the sample comes is exponential, the probability of not rejecting the hypothesis of normality is practically null with a sample size of n=100 when applying the AD test. The probability of not rejecting is approximately 18% if the sample comes from a Laplace distribution and around 5-6% if it comes from a uniform distribution.

0	0	0	0.02	0.1	0.18	0.1	0.02	0	0	0
0	0	0	0.06	0.27	0.45	0.28	0.05	0	0	0
0	0	0.01	0.09	0.37	0.57	0.37	0.09	0.01	0	0
0	0	0.02	0.15	0.5	0.73	0.51	0.15	0.02	0	0
0	0	0.04	0.26	0.68	0.89	0.68	0.26	0.04	0	0
0.01	0.02	0.12	0.44	0.82	0.95	0.82	0.43	0.11	0.02	0.0
0.05	0.09	0.28	0.58	0.82	0.91	0.83	0.57	0.28	0.1	0.0
0.19	0.27	0.38	0.52	0.61	0.65	0.61	0.51	0.37	0.26	0.1
0.23	0.24	0.24	0.25	0.25	0.26	0.26	0.25	0.24	0.23	0.2
0.11	0.1	0.1	0.09	0.09	0.09	0.09	0.09	0.1	0.1	0.1
0.06	0.05	0.05	0.06	0.06	0.06	0.05	0.06	0.06	0.06	0.0
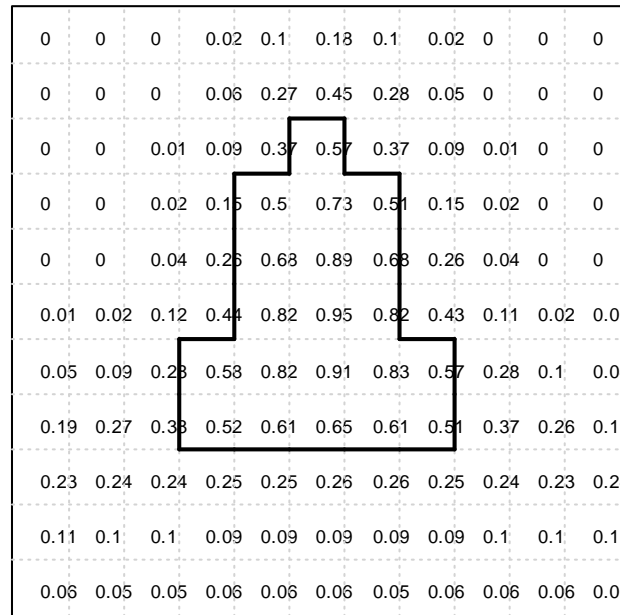
Figure 4.2: The box corresponding to each distribution indicates the proportion of times that the hypothesis of normality is not rejected when applying the Anderson-Darling test to samples of size n=100. The area where this proportion is greater than 0.5 is outlined.

Naturally, a larger mosaic can be constructed. Figure 3 shows one with 101 distributions on each side, with curves outlining the distributions in which normality is not rejected for the proportion of times indicated. The thickest line corresponds to the proportion p=0.5. The appearance of these curves can also be compared when the sample size is varied.
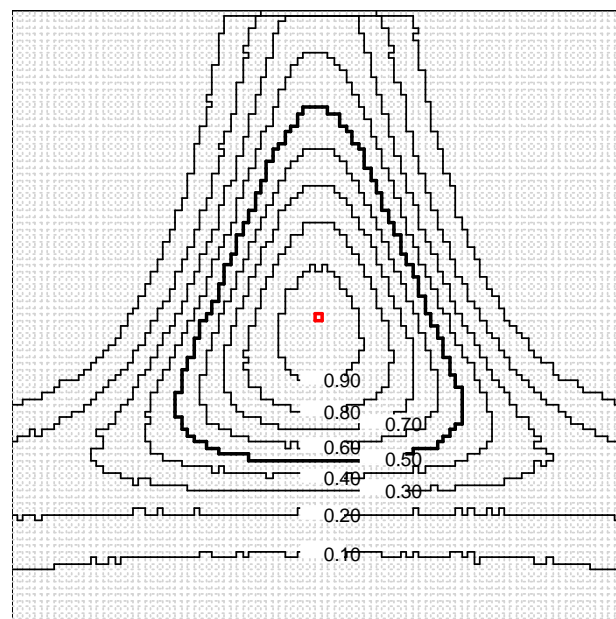
Figure 4.3: Curves on a 101x101 mosaic that indicate the proportion of times that the hypothesis of normality is not rejected when applying the Anderson-Darling test to samples of size n=100.

Figure 4 shows the curves that delimit the areas in which normality is not rejected with a probability of 50%, depending on the sample size, which is indicated on the curve itself. The curve that corresponds to $n = 10$ does not appear in the figure because any of the mosaic distributions for that sample size would be accepted with a probability greater than 50%. If the distribution from which the data come is uniform, a sample of $n = 20$ observations will also result in a greater than 50% probability of not rejecting the hypothesis of normality.
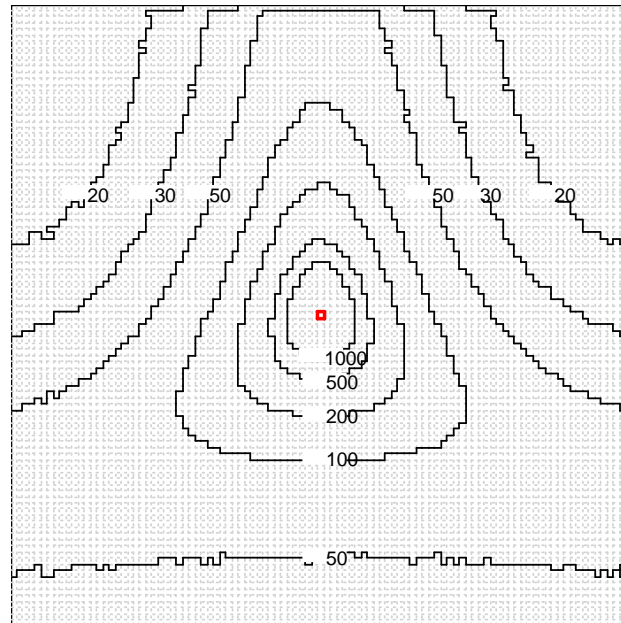
Figure 4.4: Curves that delimit the distributions for which normality is not rejected more than 50% of the time with the Anderson-Darling test for the sample sizes indicated

4.4.   Comparison of tests

As an example of the possibilities of our method, 10000 samples of $n = 20$, 50 and 100 observations have been generated from each of the distributions that appear in the mosaic. The figures indicate the curves that enclose the distributions for which the hypothesis of normality is not rejected more than 50% of the time. The smaller the surface this curve encloses, that is, the fewer distributions it includes, the better the test is. This is because it is then more likely to reject the hypothesis of normality of a greater number of distributions that are not actually normal (those that are outside the curve).

Looking at the distributions in group 1 (Figure 5), based on correlation and regression measures, we observe that the Shapiro-France and Filliben tests have practically identical performance. With sample sizes of $n = 20$, the curve corresponding to the Shapiro-Wilk test is not very different either. For $n = 50$, and much more clearly for $n = 100$, the Shapiro-Wilk test shows greater power with regard to low

kurtosis distributions, since its curve moves away from the lower zone. However, it extends somewhat further into the zone of symmetrical distributions with high kurtosis (towards the Laplace distribution). Overall, and taking the area enclosed by the curves as a measure of the performance of the test, we can say that SW produces the best performance.

Regarding the distributions of group 2 (Figure 6), their ranking is clear. The one that performs best for the 3 sample sizes considered is AD. It is noteworthy that if the Lilliefors test is applied with sample size $n = 20$, the probability of rejecting the null hypothesis for data coming from a uniform distribution is greater than 50%, and only slightly lower if they come from an exponential distribution.

In group 3 (Figure 7), the d'Agostino test clearly performs better when the sample size is n=100, whereas Jarque-Bera test performs slighthly better for n=20. The performance of both is very similar when n=50. Keeping in mind that we generally work with small samples, Jarque-Bera test is best for us, although it is a debatable point since it depends on the sample size. Figure 8 compares those that have been considered the best from each group, and in this case it is very clear that for any of the sample sizes considered the test that performs best is Shapiro-Wilk.

Interested readers can find in GitHub: https://github.com/PereGrima/Graphical-Comparison-Normality-Tests an R Markdown file (TestsComparison.Rmd) with the explanations and the R code to draw the curves that allow comparing the Lilliefords, Cramer-Von Misses and Anderson-Darling tests with n = 100 (Figure 6, right). The probabilities of rejecting the hypothesis of normality are calculated separately since the computation time is long. The file (CalculationTypeIIError.Rmd) contains the explanations and the code to calculate these probabilities. We also include three txt files with data, one for each test considered, so that you will be able to quickly see the curves drawn by the R program.
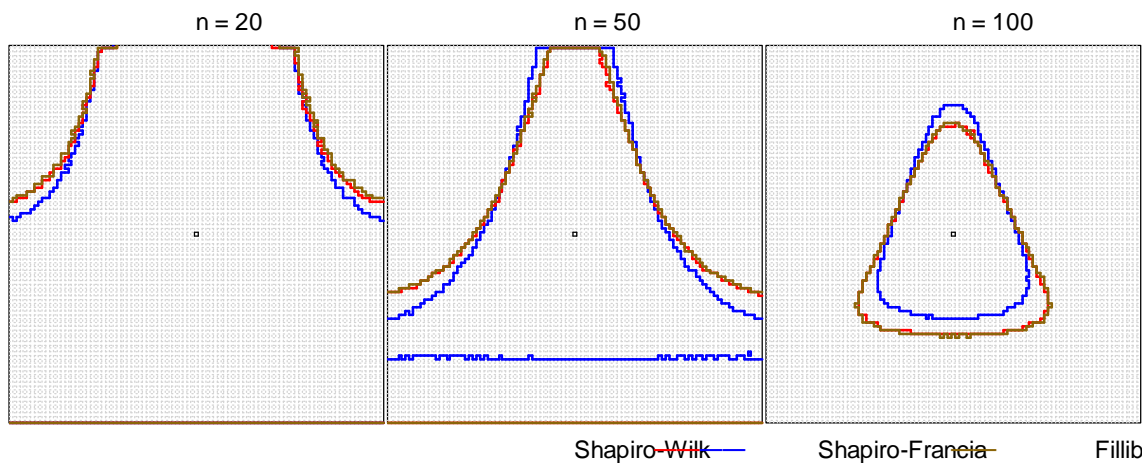
Figure 4.5: Tests based on correlation measures. If the sample comes from one of the distributions enclosed by the curve, the probability of not rejecting the hypothesis of normality is greater than 50%
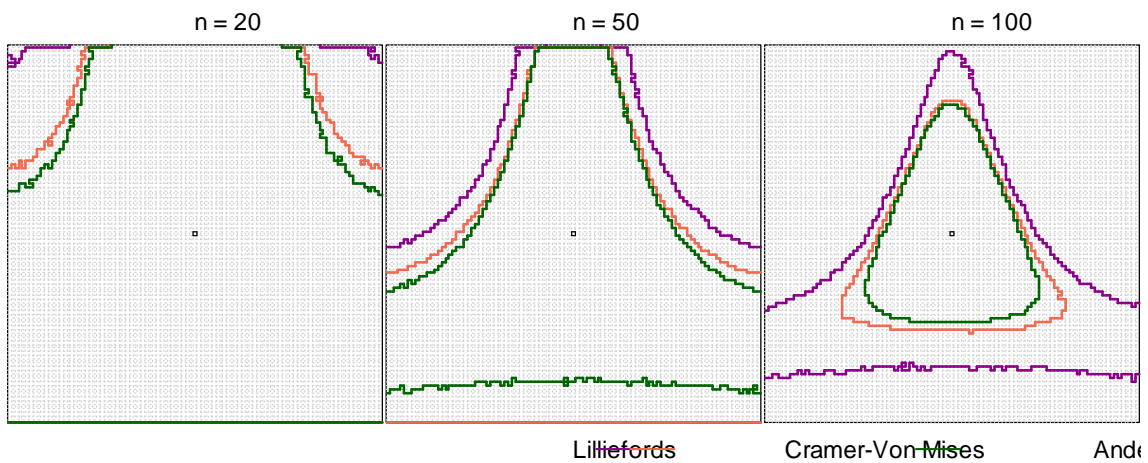
Figure 4.6: Tests based on EDF. If the sample comes from one of the distributions enclosed by the curve, the probability of not rejecting the hypothesis of normality is greater than 50%.
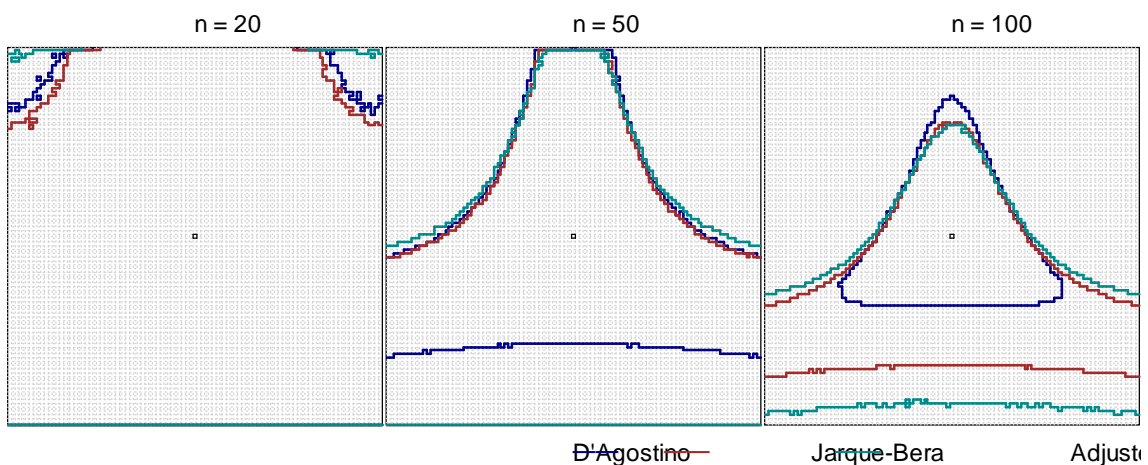
Figure 4.7: Tests based on moments. If the sample comes from one of the distributions enclosed by the curve, the probability of not rejecting the hypothesis of normality is greater than 50%.

n = 20

n = 50

n = 100

Shapiro-Wilk

Anderson-Darling

Jarque-Bera
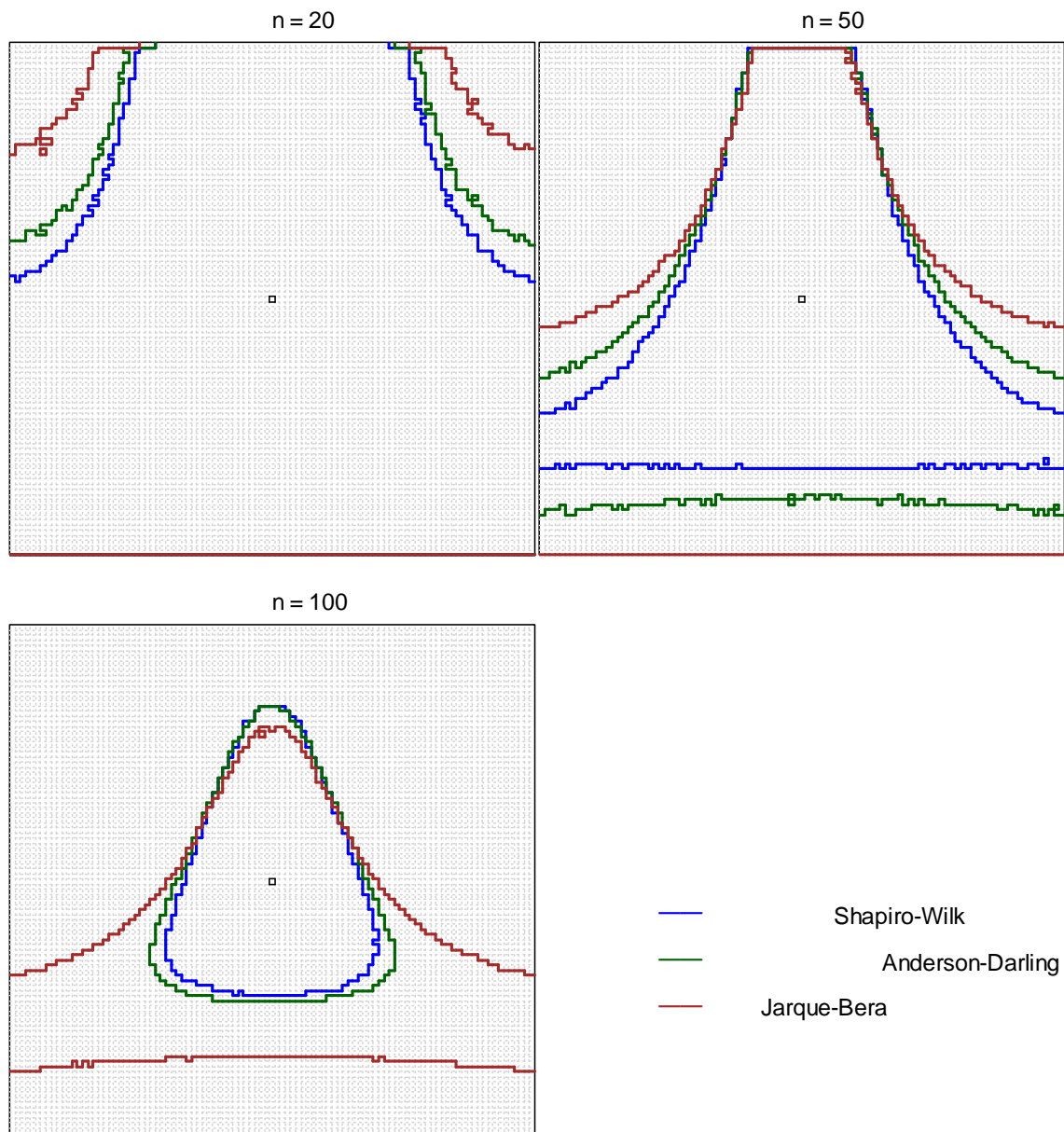
Figure 4.8: Comparison of the tests considered best for each group.

4.5.   Final remarks

The proposed method allows visualizing the power of a normality test in comparison to a wide range of unimodal distributions. This procedure is especially useful for graphically comparing the power of different tests as well as the influence of the sample size.

This way of visualizing the results allows us to consider the difference between a statistically significant difference and an important difference in the context of normality tests. When a statistical method is valid only under the hypothesis of normality of the data, it is worth asking what deviation from normality is tolerable. For example, if the data belongs to one of the distributions that are next to the normal in a mosaic of 101x101 distributions, would the method be good? The answer is surely yes, but at

what distance from normal would that no longer be true? And if we know the answer to this question, then what sample size is necessary for a high probability of rejecting the normality hypothesis if the data come from that distribution? Furthermore, what test performs best for that objective?

The proposed graphics naturally give rise to questions of this type, and they also allow a clearer view of the possibilities and limitations of normality tests.

4.6.   References

1. Thode, HC Jr. Testing for Normality. Marcel Dekker, Inc. New York. 2002

2. Desgagné, A. and de Micheaux, P.L. (2017) A powerful and interpretable alternative to the Jarque–Bera test of normality based on 2nd-power skewness and kurtosis, using the Rao's score test on the APD family, Journal of Applied Statistics, DOI: 10.1080/02664763.2017.1415311.

3. Farell, PJ and Rogers-Stewart, K (2006). Comprehensive study of tests for normality and symmetry: extending the Spiegelhalter test. Journal of Statistical Computation and Simulation. 76:9, 803-816

4. Yazici, B and Yolacan, S. (2007). A comparison of various tests of normality. Journal of Statistical Computation and Simulation. Vol. 77, No. 2, February 2007, 175–183

5. Romão, X., Delgado, R., Costa, A. (2010). An empirical power comparison of univariate goodness-of-fit tests for normality. Journal of Statistical Computation and Simulation. 80:5, 545-591.

6. Yap, BW and Sim, CH (2011). Comparisons of various types of normality tests. Vol. 81, No. 12, December 2011, 2141–2155

7. Sánchez-Espigares, JA, Grima, P and Marco-Almagro, Ll. (2017): Visualizing Type II Error in Normality Tests, The American Statistician, DOI: 10.1080/00031305.2016.1278035

8. Shapiro, S.S and Wilk, M.B. (1965): An Analysis of Variance Test for Normality (Complete Samples). Biometrika. Vol. 52, No. 3/4 (Dec., 1965), pp. 591-611

9. Royston, P. (1991): Approximating the Shapiro-Wilk W-test for non-normality. Statistics and Computing (1992) 2, 117-119.

10. Shapiro, S.S. and Francia, R.S. "An approximate analysis of variance test for normality", Journal of the American Statistical Association 67 (1972) 215–216.

11. Filliben, J.J. (1975): The Probability Plot Correlation Coefficient Test for Normality. Technometrics, Vol. 17, No. 1, pp.: 111-116.

12. Dallal, G.E. and Wilkinson, L. (1986): An analytic approximation to the distribution of Lilliefors' test for normality. The American Statistician, 40, 294–296.

13. Stephens, M.A. (1986): Tests based on EDF statistics. In: D'Agostino, R.B. and Stephens, M.A., eds.: Goodness-of-Fit Techniques. Marcel Dekker, New York.

14. D'Agostino, Ralph B.; Albert Belanger; Ralph B. D'Agostino, Jr (1990). "A suggestion for using powerful and informative tests of normality" (PDF). The American Statistician. 44 (4): 316–321. doi:10.2307/2684359.

15. Jarque, C.M. and Bera, A.K. (1987): A Test for Normality of Observations and Regression Residuals. International Statistical Review, 55, 2, pp. 163-172.

16. Urzua, C. M. (1996): On the correct use of omnibus tests for normality. — Economics Letters, vol. 53, pp. 247–251.

17. R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

18. Juergen Gross and Uwe Ligges (2015). nortest: Tests for Normality. R package version 1.0-4. https://CRAN.R-project.org/package=nortest

19. Thorsten Pohlert (2017). ppcc: Probability Plot Correlation Coefficient Test. R package version 1.0. https://CRAN.R-project.org/package=ppcc

20. Diethelm Wuertz, Tobias Setz and Yohan Chalabi (2017). fBasics: Rmetrics - Markets and Basic Statistics. R package version 3042.89. https://CRAN.R-project.org/package=fBasics

21. Ilya Gavrilov and Ruslan Pusev (2014). normtest: Tests for Normality. R package version 1.1. https://CRAN.R-project.org/package=normtest

22. Zhu, D., and Zinde-Walsh, V. (2009), "Properties and Estimation of Asymmetric Exponential Power Distribution," Journal of Econometrics, 148, 89–99.

4.8.   *Supplementary Materials*

CalculationTypeIIError.Rmd

Calculation of Type II Error for a given test and sample size

You can set the sample size in the n variable. The variable m is the dimension of the mosaic (i.e. the number of squares in each row and column). You can change the number of simulation iterations to calculate the type II error in each cell by changing nsim. By default, the significance level to solve the test, siglevel, is set to 0.05

We use here a mosaic size m = 11. With this size it takes about 4.25 minutes to produce the file (using a computer with an Intel i7 @ 3.50GHz processor). With m = 101, the size used for files like the ones attached, the time is about 6 hours.

```
n = 100          # Sample size
m = 11           # n: mosaic size (nxn)
nsim = 10000     # nsim: number of simulations in each cell
siglevel = 0.05 # siglevel: significance level of the normality test
```

There are several possibilities for the test to use in the calculations. Some options with their implementation are listed below.

 # Lillefords Test library(nortest) nomtest = "LI" test = function(x) lillie.test(x)$p.value

 # Cramer-Von Mises Test library(nortest) nomtest = "CVM" test = function(x) cvm.test(x)$p.value

 # Shapiro-Wilks library(fBasics) nomtest = "SW" test = function(x) shapiroTest(x)@test$p.value

#Kolmogorov-Smirnov Test library(fBasics) nomtest = "KS" test = function(x) ksnormTest(x)@test$p.value [1]

 # Jarque-Bera Test library(fBasics) nomtest = "JB" test = function(x) jarqueberaTest(x)@test$p.value

 # Chi-Square Test library(fBasics) nomtest = "CSQ" test = function(x) pchiTest(x)@test$p.value [1]

 # D'Agostino-Pearson Test library(fBasics) nomtest = "DP" test = function(x) dagoTest(x)@test$p.value [1]

 # Your own test You only have to implement a function named *test* that calculates the p-value given the sample. You also need to specify an acronym *nomtest*, to identifie this new test. *nomtest = "NewTest"*… (identifier of the test, prefix of the file) *test = function(x) …* (function implementing the test, given a sample must return the p-value from the normality test)

Example with the Anderson-Darling Test

Calculation of the type II errors for each cell in the mosaic. The values are saved in a file with the same structure as the mosaic. Each value represents the percentatge of times in the simulation for that cell in which the null hypothesis has not been refused (i.e. type II error)

```r
library(nortest)
nomtest="AD"
test=function(x) ad.test(x)$p.value
####Random generation for SEPD
rsepd=function(n,mu=0,sigma=1,alpha=0.5,p=2){
  kp=1/(2*gamma(1+1/p)*p^(1/p))
  u=runif(n)
  w=rgamma(n,shape=1/p,scale=1)
  y=ifelse(u>alpha,1-alpha,-alpha)*(w^(1/p))/gamma(1+1/p)
  y=y/kp
  A=2*p^(1/p)*((1-alpha)^2-alpha^2)*gamma(2/p)/gamma(1/p)
  A2=(2*p^(1/p))^2*((1-alpha)^3+alpha^3)*gamma(3/p)/gamma(1/p)
  B=sqrt(A2-A^2)
  mu+sigma*(y-A)/B
}

alpha=(0:(m-1))/(m-1)
j=(log10(50)/log10(2))^(1/(m/2-0.5))
p=c(1,2^(1/(j^(((m-3)/2):1))),2,2^(j^((1:(m/2-0.5)))))

ftypeIIerror=function(i){
  ysim=matrix(rsepd(n*nsim,mu=0,sigma=1,alpha=alpha[i[1]],p=p[i[2]]),nc=n)
  setTxtProgressBar(pb, ((i[2]-1)*m+i[1])/(m*m))
  mean(apply(ysim,1,function(el) test(el)>siglevel))
}

pb <- txtProgressBar(style=3)
typeIIerror=matrix(apply(expand.grid(1:m,1:m),1,ftypeIIerror),nc=m,byrow=T)
write.table(typeIIerror, paste0(nomtest,"_m",m,"_n",n,".txt"),row.names=F,col.names=F
)
```

TestsComparison.Rmd

Drawing the curves for the tests comparison

In general, type II errors for each distribution must be previously calculated and saved in a file with the name:

[Test acronym]_m[mosaic size]_n[sample size].txt

This is not included in this program because the computing time is very long. Here we use three files previously created that are included in the supplemental online material:

- **LI_m101_n100.txt** (Lillieford test in 101x101 mosaic and for samples of size n = 100)

- **CVM_m101_n100.txt** (Cramer- Vom Misses test in 101x101 mosaic and for samples of size n = 100)

- **AD_m101_n100.txt** (Anderson-Darling test in 101x101 mosaic and for samples of size n = 100)

These files or other similar for other tests, mosaic size or sample size can be created with the code included in the file "Calculing type II errors.html", which can also be found in the supplemental online material.

The mosaic size (*m*) and the sample size (*n*) are needed to identify these files.

```
m = 101     # Mosaic size
n = 100     # Sample size
```

Also, it is necessary to indicate which tests are to be represented. The acronym of the test is included in the array *tests*. The description to be included in the legend must be included in the *legends* array. The color used for this test is specified in the *co* array. Any of these values can be changed for other tests.

```
tests=c("LI","CVM","AD")
legends=c("Lillefords","Cramer-Von Mises","Anderson-Darling")
co=c("magenta4","coral2","darkgreen")
```

The variable *typeIIerrorlevel* indicates the value of the type II error to draw in the mosaic. Cells inside this line have a type II error greater than this value. In our paper we use *typeIIerrorlevel=0.5*, but this value can be changed if desired.

```
typeIIerrorlevel=0.5
```

Below is the rest of the code. This part should not be changed.

Copying all the code lines and executing them as an R script will produce the figure at the end. Do not forget to correctly specify the directory where the *.txt* files are located

```r
par(mar=c(3,0.1,3,0.1), oma=c(0,0,0,0), mfrow=c(1,1))
plot (NULL, xlim=c(0.5,m+0.5), ylim=c(0.5,m+0.5), yaxs="i", xaxs="i", xaxt='n', yaxt=
'n', xlab="",ylab="")
grid(m,m,lwd=1)
title(paste0('n = ',n), line = 1, cex.main=1.8, font.main=1)
legend("topleft",legend=legends,col=co,lty=1,lwd=2)


    typeIIerrorCurve=function(typeIIerror,co,typeIIerrorlevel,m){
      for (i in 1:m) {
        for (j in 1:m) {
          if (i<m) {
            if ((typeIIerror[i,j] < typeIIerrorlevel) & (typeIIerror[i+1,j] >= typ
    eIIerrorlevel)) segments(j-0.5, m+1-i-0.5, j+0.5, m+1-i-0.5, lty=1, col=co, lw
    d=2)
            if ((typeIIerror[i,j] >= typeIIerrorlevel) & (typeIIerror[i+1,j] < typ
    eIIerrorlevel)) segments(j-0.5, m+1-i-0.5, j+0.5, m+1-i-0.5, lty=1, col=co, lw
    d=2)
          }
          if (j<m) {
            if (typeIIerror[i,j] < typeIIerrorlevel & typeIIerror[i,j+1] >= typeII
    errorlevel) segments(j+0.5,m+1-i-0.5,j+0.5,m+1-i+0.5, lty=1, col=co, lwd=2)
            if (typeIIerror[i,j] >= typeIIerrorlevel & typeIIerror[i,j+1] < typeII
    errorlevel) segments(j+0.5,m+1-i+0.5,j+0.5,m+1-i-0.5, lty=1, col=co, lwd=2)
          }
        }
      }
      segments(m/2, m/2, m/2+1, m/2,col="black", lty="solid", lwd=1)
      segments(m/2, m/2, m/2, m/2+1,col="black", lty="solid", lwd=1)
      segments(m/2, m/2+1, m/2+1, m/2+1,col="black", lty="solid", lwd=1)
      segments(m/2+1, m/2, m/2+1, m/2+1,col="black", lty="solid", lwd=1)
    }
    for (i in 1:length(tests)){
        typeIIerror <- read.table(paste0(tests[i],"_m",m,"_n",n,".txt"), header=FA
    LSE)
        typeIIerrorCurve(typeIIerror,co[i],typeIIerrorlevel,m)
    }
```

Check for updates

Mosaic normality test

José A. Sánchez-Espigares 🆔, Pere Grima 🆔, and Lluís Marco-Almagro 🆔

Department of Statistics and Operations Research, Universitat Politècnica de Catalunya-BarcelonaTech, Barcelona, Spain

ABSTRACT

A procedure is proposed here for jointly visualizing the compatibility of a sample with a family of Skewed Exponential Power Distributions, of which the distributions known as Normal, Exponential, Laplace and Uniform are particular cases. The procedure involves constructing a mosaic that contains these distributions in such a way that the asymmetry varies from left to right and the kurtosis varies from top to bottom. The null hypothesis that the sample belongs to each of the mosaic distributions is tested, with the corresponding box for each distribution being shaded in a gray scale according to the p-value obtained. The location and shape of the shaded area that appears on the mosaic facilitates not only identifying which distributions are compatible with the sample, but also assessing the power of the test performed. All the parameters that define the distribution are considered to be known. The problem remains open for expanding the procedure to cases in which these parameters are not known but must be estimated from the sample. As additional material, a code written in R that carries out the proposed test is included.

1. Introduction

There are a great variety of tests and procedures to check the normality of the data, so many that it may seem unnecessary to create one more. Thode (2002) lists an exhaustive series of tests and procedures that can be used for assessing normality. These procedures can be divided between those that use graphic techniques and those that use analytical techniques. Graphical methods such as the representation of values in Normal Probability Plot (NPP) allow visualizing the degree of discrepancy compared to the theoretical distribution together with the specific values that stand apart, but they do not offer an objective measure of that discrepancy. Analytical tests do measure this discrepancy, but it is advisable to complement them with some type of graphic representation that allows observing possible anomalous values that could affect the test result.

The analytical methods use different criteria to assess the discrepancy between the available data and the theoretical distribution. Regardless of which criterion is used, they are all based on the same reasoning schema:

5. Mosaic Normality Test

José A. Sánchez-Espigares, Pere Grima, Lluís Marco-Almagro

Department of Statistics and Operations Research
Universitat Politècnica de Catalunya-BarcelonaTech, Barcelona, Spain

ABSTRACT

A procedure is proposed here for jointly visualizing the compatibility of a sample with a family of Skewed Exponential Power Distributions, of which the distributions known as Normal, Exponential, Laplace and Uniform are particular cases. The procedure involves constructing a mosaic that contains these distributions in such a way that the asymmetry varies from left to right and the kurtosis varies from top to bottom. The null hypothesis that the sample belongs to each of the mosaic distributions is tested, with the corresponding box for each distribution being shaded in a gray scale according to the p-value obtained. The location and shape of the shaded area that appears on the mosaic facilitates not only identifying which distributions are compatible with the sample, but also assessing the power of the test performed. All the parameters that define the distribution are considered to be known. The problem remains open for expanding the procedure to cases in which these parameters are not known but must be estimated from the sample. As additional material, a code written in R that carries out the proposed test is included.

KEYWORDS: Normality tests, Goodness-of-Fit Methods, Power test, Graphical Methods

5.1.   Introduction

There are a great variety of tests and procedures to check the normality of the data, so many that it may seem unnecessary to create one more. Thode (2002) lists an exhaustive series of tests and procedures that can be used for assessing normality. These procedures can be divided between those that use graphic techniques and those that use analytical techniques. Graphical methods such as the representation of values in Normal Probability Plot (NPP) allow visualizing the degree of discrepancy compared to the theoretical distribution together with the specific values that stand apart, but they do not offer an objective measure of that discrepancy. Analytical tests do measure this discrepancy, but it is advisable to complement them with some type of graphic representation that allows observing possible anomalous values that could affect the test result.

The analytical methods use different criteria to assess the discrepancy between the available data and the theoretical distribution. Regardless of which criterion is used, they are all based on the same reasoning schema:

1) The null hypothesis states that the data do indeed come from a Normal distribution.

2) A measure of the discrepancy between the data and the theoretical distribution is calculated. This measure can be gathered in various ways, such as: the regression and correlation techniques of Shapiro-Wilk and Shapiro-Francia; the difference between the theoretical and empirical distribution, as in the techniques used by Kolmogorv-Smirnov and Anderson-Darling; and estimating moments, like the methods of D'Agostino and Jarque-Bera.

3) The distribution of this discrepancy measure is known when the data actually come from a normal distribution. This distribution is called the "reference distribution" and may have been obtained either analytically or by simulation.

4) The discrepancy measure is compared with its reference distribution and the p-value p is determined. The hypothesis of normality is rejected if $p < \alpha$, with α being the level of significance chosen for the test.

This is actually the typical schema of hypothesis testing. However, even though general practice places the burden of proof on demonstrating a particular matter of interest (for example, that a certain drug is more effective than a placebo), testing normality assumes that the null hypothesis is true by default. This leads to paradoxes such as: the less data one has, the better it is to pass the test and be able to work with the normality assumption.

On the other hand, when a hypothesis test is carried out, it is usual to test a value that will serve as the parameter of the null hypothesis, such that by varying this value the probability of type II error can be continuously calculated and, thus, the power curve of the test can be drawn. However, this is not possible when testing normality, since the null hypothesis does not refer to the value of a parameter but to a type of distribution. Some works – such as those by Yacini and Yolacan (2007), Romão *et al*. (2010) and Yap and Sim (2011) – compute the probability of type II error for different normality tests using simulated data from different alternative distributions. This information helps in choosing the most convenient test in each case; although, when the test is performed, it is not easy to visualize what its power curve would be. This difficulty can lead to the probability of type II error going unnoticed and its value being much greater than what the analyst would assume.

5.1.1. Visualizing type II error in Normality Tests

Sánchez-Espigares *et al*. (2018) propose a procedure for visualizing this type II error probability in a normality test. It is based on representing a mosaic comprised of $m \times m$ boxes, each box having a different distribution. Based on the Skewed Exponential Power Distribution (SEPD) used by Zhu and Zinde-Walsh (2009), they propose the following probability density function to create the distributions of the mosaic:

$$f(y) = \begin{cases} \dfrac{B}{\sigma} K(p) \exp\left(-\dfrac{1}{p}\left|\dfrac{A + B\left(\frac{y-\mu}{\sigma}\right)}{2\alpha}\right|^p\right) & \text{if } y \le \mu - \dfrac{A}{B}\sigma\,, \\[3em] \dfrac{B}{\sigma} K(p) \exp\left(-\dfrac{1}{p}\left|\dfrac{A + B\left(\frac{y-\mu}{\sigma}\right)}{2(1-\alpha)}\right|^p\right) & \text{if } y > \mu - \dfrac{A}{B}\sigma\,, \end{cases}$$

where:

$$K(p) = 1/\left[2p^{1/p}\Gamma(1 + 1/p)\right],$$

$$A = \frac{1}{K(p)}\left[(1 - 2\alpha)\frac{p\Gamma(2/p)}{\Gamma^2(1/p)}\right],$$

$$B = \sqrt{\frac{1}{[K(p)]^2}\left[(3\alpha^2 - 3\alpha + 1)\frac{p^2\Gamma(3/p)}{\Gamma^3(1/p)}\right] - \frac{1}{[K(p)]^2}\left[(1 - 2\alpha)\frac{p\Gamma(2/p)}{\Gamma^2(1/p)}\right]^2}.$$

This function is characterized by the mean (μ) and the variance (σ^2) of the variable under consideration, as well as two more parameters: p, which is related to kurtosis and can vary between 1 (maximum kurtosis) and infinity (uniform distribution); and α, which is related to asymmetry and varies between 0 (very asymmetric distribution with a tail to the right) and 1 (with a tail to the left). The values $p = 2$ and $\alpha = 0.5$ correspond to the Normal distribution.

For constructing the mosaic, the value of p is varied between 1 and 50 because it is necessary that the maximum value is finite; and when $p = 50$, the distribution is already practically uniform. The number of distributions on each side of the mosaic should be an odd number so that the Normal distribution remains directly in the center. In order for the values of α to cover their entire range of variation, the i-th value of α must be $\alpha_i = \frac{i-1}{m-1}$, where m is the number of distributions on each side of the mosaic.

Regarding the values of p, we begin with $p = 1$; and then each value is equal to the previous one raised to the power of $j = \sqrt[\frac{m}{2}-0.5]{\dfrac{\log 50}{\log 2}}$, except for the second one, which is always equal to $2^{1/j^{\left(\frac{m}{2}-1.5\right)}}$. For example, if $m = 9$ (9×9 mosaic), the values of α and p are those indicated in Figure 1, which also shows the shape that each of the distributions will have. The first row includes the exponential distribution ($p =$

1, $\alpha = 0$) and also the Laplace distribution ($p = 1$, $\alpha = 0.5$). The Normal distribution has been boxed in the center. In the last row, the distributions are practically uniform.
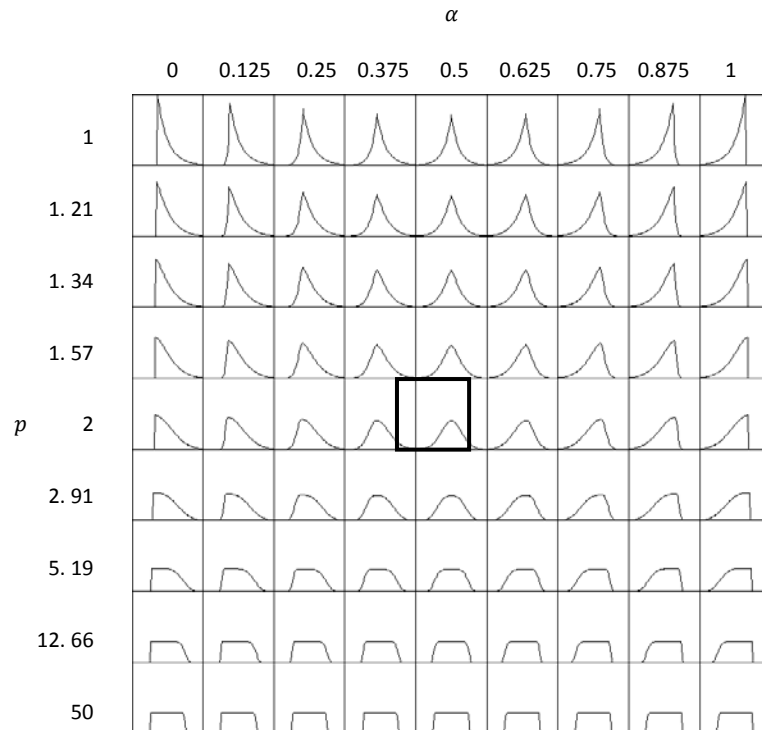
Figure 5.1: Mosaic of 9x9 distributions, with values of p (vertical) and α (horizontal).

Each distribution in the mosaic is used to generate k random samples of size n, and the normality of each of them is tested. Based on a level of significance that has been previously established (normally equal to 0.05), we calculate the proportion of times that the hypothesis of normality in each of the distributions is not rejected. Finally, each mosaic box (distribution) is shaded in a scale of gray according to the value of that proportion.

For example, upon generating $k = 1000$ samples of size $n = 50$ for each of the distributions that appear in the 9x9 mosaic in Figure 2, the number that appears in each box indicates the proportion of times that the null hypothesis of normality has not been rejected (using the Kolmogorov test). It can be seen that this proportion drops below 5% only for the exponential distributions in the upper corners, showing that the power of the test is low even with a sample that we could consider large (50 observations). A sequential gray scale is used to display the results more clearly. From lighter to darker shades, the ranges are 0-0.05, 0.05-0.25, 0.25-0.50 and 0.50-1.
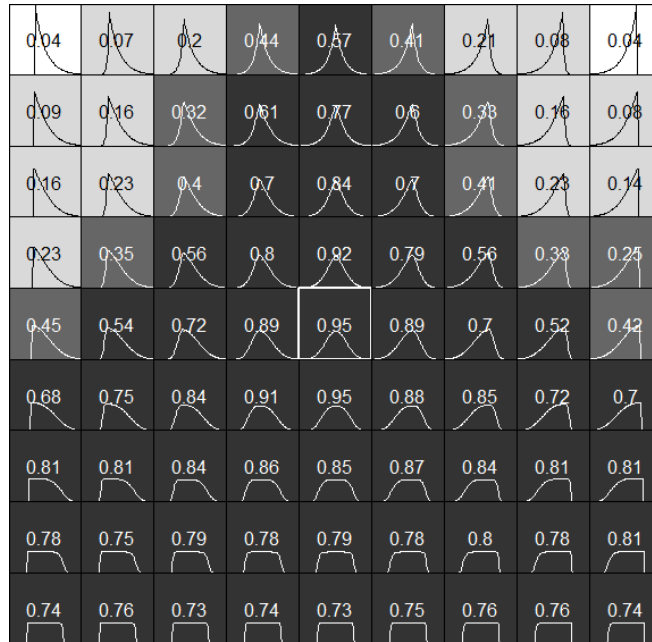
Figure 5.2: Each box shows the proportion of times that the hypothesis of normality is not rejected (α=0.05) when using the Kolmogorov test on samples of size n=50 coming from the given distributions.

5.1.2. Our proposal

We propose using the mosaic of distributions not only to visualize the probability of type II error but also to provide a new approach to normality tests. To test the normality of a sample, we propose testing if it fits in each of the distributions that appear in the mosaic and giving a dark color to those boxes (distributions) in which the fitting hypothesis cannot be rejected. In this way, a dark area on the mosaic covers all the distributions that are compatible with the available sample. If the box corresponding to the Normal distribution is within the dark area, the normality hypothesis cannot be rejected. This procedure can be easily incorporated into statistical software packages and has the additional advantage to provide an idea about the power of the test, which is impossible to observe when using conventional tests. Indeed, it is different for a sample to pass the normality test with a small or a large dark area on the mosaic. Clearly, in the first case there are few possible alternatives to normality, while in the second there are many alternatives and with very different characteristics.

In the next sections we discuss some of the goodness of fit measures that can be used in our test. We describe the possibilities of using the graphic representation that we propose as well as its implementation with the statistical software R (R Core Team 2017). After that, we present two case studies in which the advantages of our suggested methodology will be apparent. Finally, we summarize the most relevant aspects of the proposed procedure, highlight its limitations, and suggest some possible improvements.

5.2.   *Goodness of fit criteria*

We need goodness-of-fit tests that can be applied to any of the distributions that appear in the mosaic. For instance, even though the Shapiro-Wilk test gives excellent results for testing normality, it would not be useful for the remaining distributions. We can use tests based on measuring the discrepancy between the empirical distribution function of the sample, $F_n(x)$, and the theoretical distribution function that is proposed as the null hypothesis, $F(x)$. The values of $F(x)$ can be considered values of a uniform distribution $U(0,1)$ regardless of the original distribution of X, as long as it is continuous. Therefore, whatever the distribution in question, we measure the difference between the values of the empirical distribution of the sample distribution being tested and another sample that is always a random sample of a distribution $U(0,1)$.

In turn, the criterion of Stephens (1986) establishes that adjustment tests based on the Empirical Distribution Function (EDF) can be divided into two classes: the supremum class and the quadratic class. The most typical of each group are the Kolmogorov-Smirnov test (supremum) and the Anderson-Darling test (quadratic). These are the ones we will use. Below, we describe the general characteristics of each.

5.2.1. Kolmogorov-Smirnov

Given n data points $X_1, X_2, \cdots, X_n$, ordered from the smallest to largest value, the EDF is defined as $F_n(X_i) = k(i)/n$, where $k(i)$ is the number of points less than X_i. This is a step function that increases the value of each ordered data point by $1/n$. The Kolmogorov-Smirnov test statistic is defined as:

$$D_n = \max_{1 \le i \le n} \left(F(X_i) - \frac{i-1}{n}, . \frac{i}{n} - F(X_i) \right)$$

Note that D_n is the maximum distance between both distributions. Let us take, for example, a sample of 15 random values of a distribution $N(0,1)$ generated with R (seed=1). Figure 3 shows its EDF (stepped form) and the theoretical cumulative distribution (solid line), and highlights the value of the test statistic D_n.
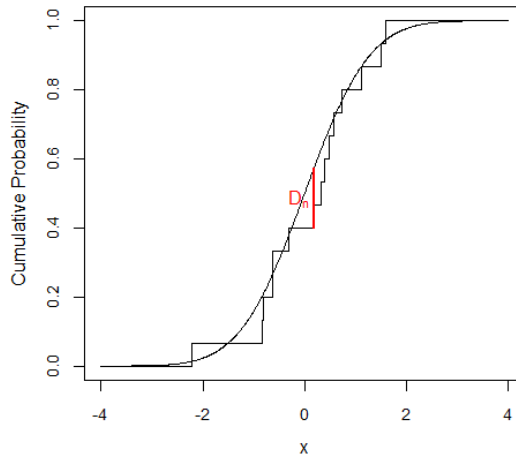
Figure 5.3: Kolmogorov-Smirnov statistic test (D_n): Maximum difference between theoretical and empirical distribution function

As we have said, the distribution of D_n does not depend on the distribution F, but it does depend on n. Using the same form suggested by Kolmogorov (see, for example, Gibbons and Chakraborti 2003), its asymptotic expression is:

$$\lim_{n \to \infty} P\big(D_n \sqrt{n} \le x\big) = 1 - 2 \sum_{i=1}^{\infty} (-1)^{i-1} e^{-2i^2 x^2}$$

Marsaglia *et al*. (2003) detail a procedure for determining p-values with a precision of at least 13 digits for values of n, from 2 up to at least 16000. This procedure is used in the `ks.test` function, which is included in the `stats` package (R Core Team 2017) of the statistical software R.

5.2.2. Anderson-Darling test

This is based on the Cramer-von Mises test, which quantifies the difference between F and F_n through the expression:

$$Q = n \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 dF(x)$$

Anderson and Darling (1954) added the weighting factor $\frac{1}{[F(x)(1-F(x)]}$, which gives more weight to the values located at the extremes of the sample. In this way, the test is more powerful and, in addition, the critical values do not depend on n for $n > 5$. The discrepancy measure, including the weighting factor, can be expressed as (see, for example, Marsaglia and Marsaglia 2004):

$$A^2 = -n - \sum_{i=1}^{n} \frac{2i-1}{n} \big[\ln\big(F(X_i)\big) + \ln\big(1 - F(X_{n+1-i})\big)\big]$$

where $X_1 < \cdots < X_n$ are the values of the ordered sample. As in the Kolmogorov-Smirnov test, the values of $F(X_i)$ can be understood as a random sample of a distribution $U(0,1)$. Marsaglia and Marsaglia (2004) present several methods for deducing the probability $P(A^2 < k)$. The most practical one offers an approximation that has greater than 0.000002 accuracy for probabilities that are less than 0.9, and 0.0000008 accuracy for those that are above. The `ADGofTest` package (Gil Bellosta 2011) of the statistical software R includes the `ad.test` function, which gives p-values for the Anderson-Darling test that are calculated according to one of the methods proposed by Marsaglia and Marsaglia. The manual for this package does not clarify which of the paper's proposed methods is applied, but the accuracy of even the most quick-and-easy one is more than enough.

5.3.   *Normality Test using the mosaic of distributions*

As said before, given a sample on which we wish to test the normality, our proposal is testing whether it pertains not only to the Normal distribution but to all the distributions that appear in the mosaic. To show the results visually, each box in the mosaic is given a color that is a function of the p-value obtained by testing the distribution contained in it. These colors are:

p-value	Color
< 0.01	White
0.01 – 0.05	Light gray
0.05 – 0.10	Gray
> 0.10	Dark gray

In order to carry this out, we have created a program with the statistical software package R, which is available to interested readers as additional material to this article. The program allows you to choose:

- The size of the mosaic, up to a size of 49×49 (which would be easy to expand).

- Aspects related to the presentation, such as the cut-off points of the p-value in gray scale or whether or not the p-values are shown in each cell.

- The criterion for measuring the discrepancy. As we have mentioned, this may be that of Kolmogorov-Smirnov (KS) or of Anderson-Darling (AD).

- The sample to test. This may be a dataset whose normality we wish to test, or we can also use a randomly generated sample of a Normal, uniform or exponential distribution in order to observe the behavior of the program. In the first case, it is necessary to introduce the population parameters, which are presumably known.

If the criterion for measuring the discrepancy is that of KS, the p-value is calculated with the `ks.test` function of the `stats` package, which is already loaded by default. If the AD criterion is applied,

the `ADGofTest` package is used (it is necessary to install the library beforehand). These functions are based on Marsaglia *et al*. (2003) as well as Marsaglia and Marsaglia (2004). Furthermore, as we have mentioned previously, they provide a more than adequate approximation for our objectives.

Below, we will see some examples, all of which are reproducible using the seed indicated in each case. Let us take a sample of $n = 100$ observations of a distribution $N(0,1)$ generated with seed=1. Using the AD criteria, we obtain the mosaic in Figure 4 (right). Naturally, the null hypothesis of normality is not rejected, but neither will be the Laplace distribution (double exponential) located in the first row of the mosaic and in the same column as the normal distribution. However, the exponential distribution (both right- and left-skewed) would be rejected, as well as the uniform distribution. If the KS criterion is used, we have the mosaic in Figure 4 (left). It is clear that this test has less power than that of AD.

Unsurprisingly, the size of the shaded area on the mosaic tends to decrease as the sample size increases. This can be seen in Figure 5, where we represent the mosaics obtained using randomly generated samples (seed=1) of a distribution N (0,1) and with the indicated sample sizes.

This procedure can also be used to test whether the sample comes from any of the distributions that appear in the mosaic. It is enough to observe in which interval can be found the p-value of the box corresponding to the null hypothesis. For example, Figure 6 shows the shaded area obtained from a sample of 100 randomly generated observations (seed=1) from an exponential distribution (top) and a uniform distribution (bottom). Both cases include the histogram representing the distribution of the data.

5.4. Case Studies

5.4.1. Leghorn Chick Data

To illustrate normality tests based on EDF statistics, Stephens (1986, p. 98) uses data from Bliss (1946) on the weight in grams of twenty 21-day-old leghorn chicks (*Table 5.1*).

Table 5.1: Leghorn Chick Data (from Bliss, 1946)

156	162	168	182	186	190	190	196	202	210
214	220	226	230	230	236	236	242	246	270

Stephens states that – obviously – the data have been rounded, so in a strict sense the population is discrete. However, with values of this magnitude, the obtained approximation shows a negligible difference from the unrounded values. The values of μ and σ are assumed to be known, and we use those that are proposed by Stephens: $\mu = 200$ and $\sigma = 35$. Applying KS to test normality obtains a test statistic of $D = 0.171$, with a p-value = 0.60. An analogous result is obtained when applying the AD test: in this

case, the test statistic is $A^2 = 1.017$ and the p-value is 0.35. Therefore, we would assume that these data come from a normal distribution.
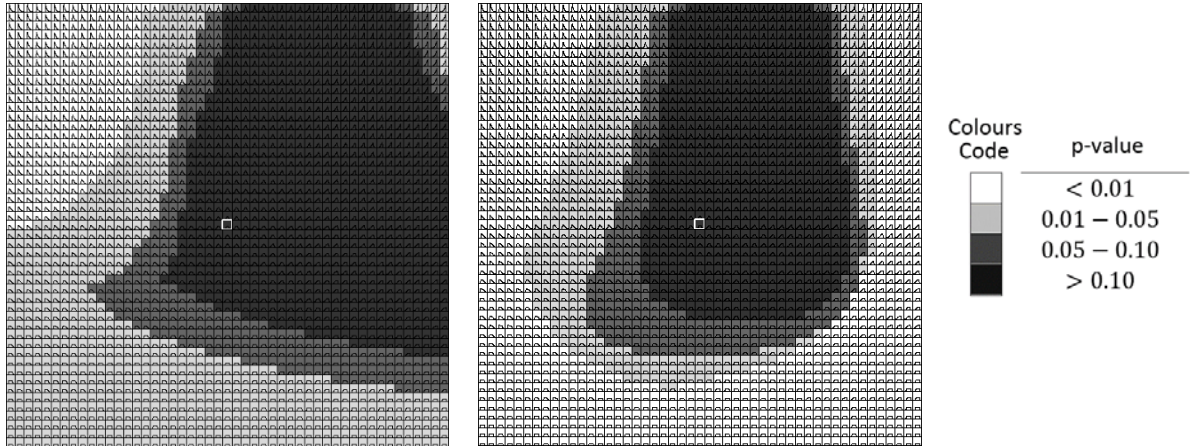
Figure 5.4: Mosaic test applied to a sample of 100 random numbers of N(0,1) generated with R (seed=1), using the criteria of Kolmogorov-Smirnov (left) and Anderson-Darling (right).
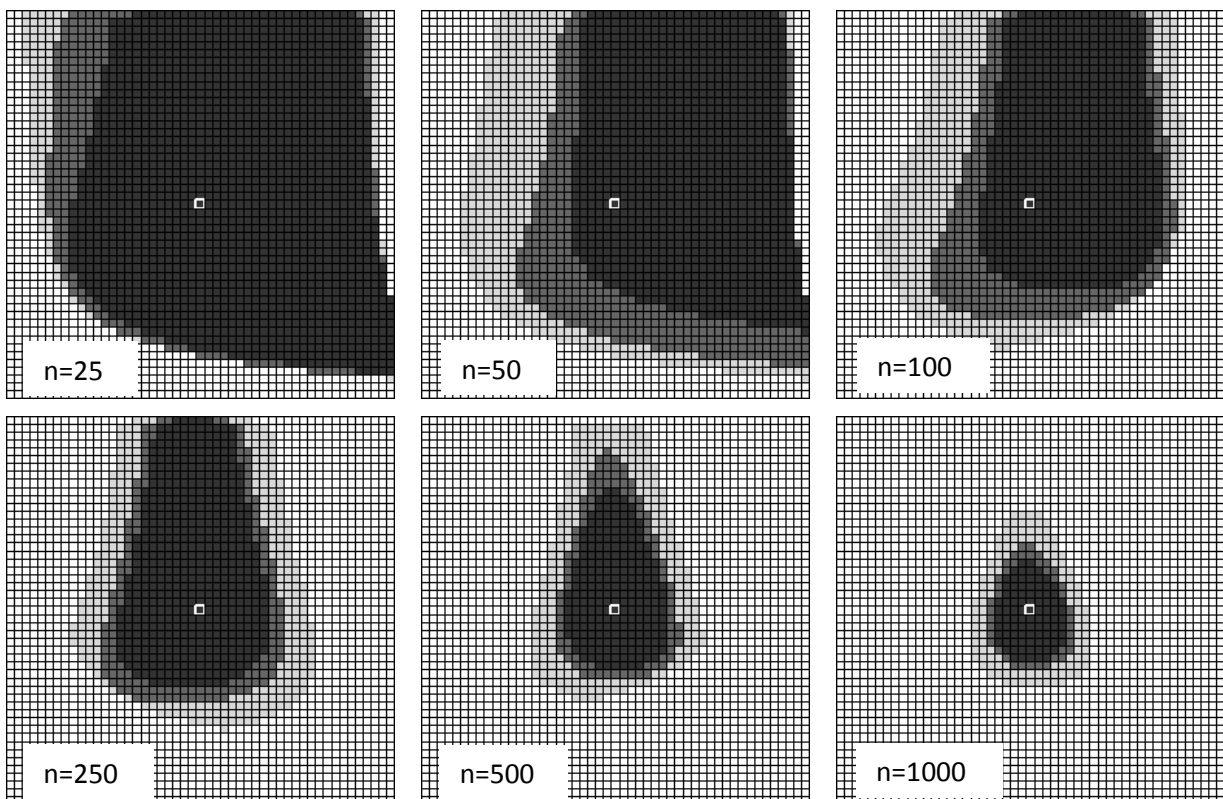
Figure 5.5: Results obtained using the AD criteria with samples of the indicated size, obtained randomly (with R, seed=1) from a population of N(0,1)
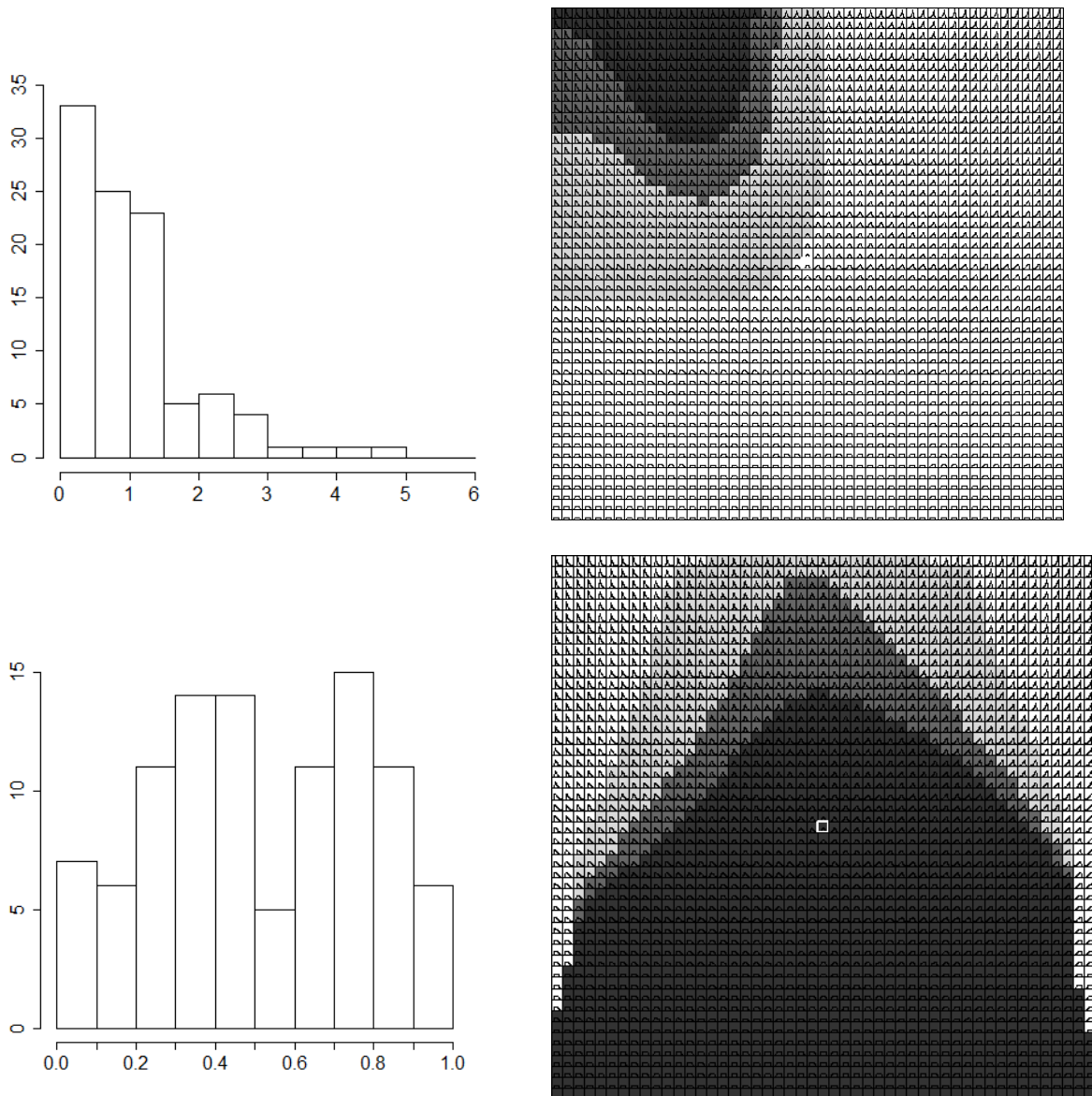
Figure 5.6: Results obtained using the AD criteria with a sample of n=100 from an exponential distribution (λ=1) (top) and uniform distribution (0,1) (bottom), using R (seed=1). The histogram corresponding to the data is attached.

However, when we test whether or not our sample comes from each of the distributions in the mosaic, all boxes take on a darker color when using the KS test, meaning that the obtained p-values are greater than 0.10 for all the distributions. Therefore, using the KS test at this level of significance precludes rejecting the hypothesis of normality, although neither can we reject that the data come from any of the other distributions in the mosaic (which, as we have seen, may be exponential, uniform or Laplace). This situation shows the low value of the KS test when using small sample sizes.

Using the AD test to these same data, we obtain a mosaic with distributions that would be rejected at the usual level of significance $\alpha = 0.05$, although all of them are very far from the normal distribution (Figure 7, left).

5.4.2. Chest circumferences of Scottish soldiers

Stigler (1986, p. 207) reproduces the famous dataset of chest circumferences of 5758 Scottish soldiers analyzed by Quetelet. Unfortunately, the original data were not preserved, and we have only the tabulated data (*Table 5.2*).

We have transformed each of the values in the table (X) into continuous values by replacing them with:

$$X - 0.5 + \frac{1}{f_X+1} \ , \ \ X - 0.5 + \frac{2}{f_X+1} \ , \ \cdots, \ X - 0.5 + \frac{f_X}{f_X+1}$$

where f_X is the observed frequency for the value X. Rounding to the nearest integer, the table's original values are once again obtained.

In this case, the data also pass the normality tests of both KS ($D = 0.009$, p-value= 0.74) and AD ($A^2 = 0.47$; p-value $= 0.48$).

Table 5.2 Quetelet's data on the chest circumferences of 5738 Scottish soldiers

Circumference of chest. English inches (X)	Frequency
33	3
34	18
35	81
36	185
37	420
38	749
39	1073
40	1079
41	934
42	658
43	370
44	92
45	50
46	21
47	4
48	1
Total	5738

For each distribution in the mosaic, we test the null hypothesis that our dataset pertains to that distribution, and it is rejected for all of them except for the normal distribution and those very close to it. Only a small spot surrounds the normal distribution in the center. Figure 7 (right) shows the spot obtained by applying the AD test. If the KS test is used, the spot is somewhat larger.
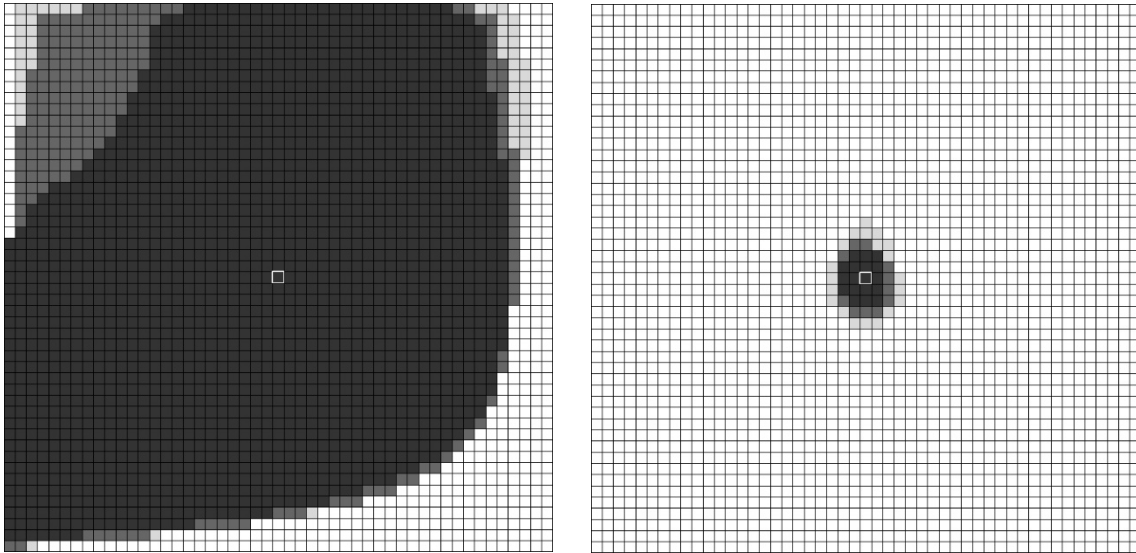
Figure 5.7: Mosaic obtained by applying the AD test to Bliss data (left) and Quetelet data (right)

Consequently, the result is unlike the previous example and we can state with certainty that these data come from a normal distribution, because we can definitively reject that the data come from distributions that are even very close to the normal. Applying a conventional normality test does not distinguish between the two analyzed cases. However, when using the mosaic of distributions, the two situations are shown to be clearly different.

5.5. Final remarks

The mosaic of distributions provides a clear idea of whether or not the sample is compatible not only with the distribution being tested as the null hypothesis, but also with all the unimodal distributions around it. A large shaded area on the mosaic (which occurs with small or moderately sized samples) shows that the sample could also come from other distributions that are far away from those being tested, and this is something that is generally not taken into account by analysts.

Note that, although the mosaics used to visualize the type II error (Sánchez-Espigares et al., 2018) and those we propose here are visually similar, their objectives and the information they represent are different. Table 3 summarizes the characteristics of each type of representation.

The proposed procedure requires knowing the parameters that define the population considered in the null hypothesis. For practical purposes, being able to use the estimates of those parameters

obtained from the available sample would be more useful. However, using these estimates changes the shape of the reference distributions, and we do not have their analytical expressions. Simulation techniques are necessary for deducing these expressions and/or calculating the critical values.

Deducing analytical expressions is not an easy task, and simulation requires great computational effort. The critical values depend on the values of α and p that are linked to the distribution, and also on the size of the sample n. Obtaining these critical values at a precision of 4 digits for each of the distributions in the mosaic is, for the moment, beyond our range of possibilities. Nevertheless, we believe this can be addressed in the near future.

Table 5.3 Comparison of mosaics for visualizing type II error and goodness of fit tests

Type II error display	Goodness of fit test
In each mosaic box, we generate k random samples of the distribution that appears in that box.	There is only one sample.
In each box, a normality test is carried out for each of the k generated samples.	The sample is tested to determine if it comes from each of the distributions in the mosaic.
Using a set level of significance for each test, the normality of the sample is or is not rejected. The proportion of times that the hypothesis of normality is not rejected is assigned to each box.	The p-value obtained by testing whether or not the sample comes from the distribution that appears in this box is assigned to that box.
Optionally, the proportion of times that the normality hypothesis has not been rejected is indicated over each box (distribution).	Optionally, the p-value obtained in the test is indicated over each box (distribution).
The boxes are colored in gray scale according to the proportion of times that the normality hypothesis is rejected.	The boxes are colored in gray scale according to the p-value obtained by testing whether the sample data come from the distribution of the box.

5.6. References

1. Anderson, T. W. and D. A. Darling. 1954. A Test of Goodness of Fit. *Journal of the American Statistical Association* 49(268):765-769.

2. Bliss, C. I. 1946. Collaborative comparison of three ratios for the chick assay of vitamin D. *Journal of the Association of Official Agricultural Chemists* 29:396-408.

3. Gil Bellosta, C. J. 2011. *ADGofTest: Anderson-Darling GoF test*. R package (version 0.3). https://CRAN.R-project.org/package=ADGofTest

4. Gibbons, J. D. and S. Chakraborti. 2003. *Nonparametric Statistical Inference*. 4 ed. New York: Marcel Dekker.

5. Marsaglia, G., W. W. Tsang and J. Wang. 2003. Evaluating Kolmogorov's distribution. *Journal of Statistical Software*, 8(18):1-4. doi: 10.18637/jss.v008.i18

6. Marsaglia, J. and G. Marsaglia. 2004. Evaluating the Anderson-Darling Distribution. *Journal of Statistical Software* 9(2):1-5. doi: 10.18637/jss.v009.i02.

7. R Core Team. 2017. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

8. Romão, X., R. Delgado and A. Costa. 2010. An empirical power comparison of univariate goodness-of-fit tests for normality. *Journal of Statistical Computation and Simulation* 80(5): 545-591. doi: 10.1080/00949650902740824

9. Sánchez-Espigares, J. A., P. Grima, L. Marco-Almagro. 2018. Visualizing Type II Error in Normality Tests. *The American Statistician.* 72(2):158-162. doi: 10.1080/00031305.2016.1278035

10. Stephens, M. A. 1986. Test Based on EDF Statistics. In *Goodness-of-Fit Techniques*, ed. R. B. D'Agostino and M. A. Stephens, 97-193. New York: Marcel Dekker.

11. Stigler, S. M. 1986. The History of Statistics. The Measurement of Uncertainty before 1900. Cambridge, MA: Harvard University Press.

12. Thode, H. C. 2002. *Testing for Normality*. New York: Marcel Dekker.

13. Yacini, B. and S. Yolacan. 2007. A comparison of various tests of normality. *Journal of Statistical Computation and Simulation.* 77(2): 175-183. doi: 10.1080/10629360600678310.

14. Yap, B. W. and C. H. Sim. 2011. Comparisons of various types of normality tests. *Journal of Statistical Computation and Simulation*, 81(12):2141-2155. doi: 10.1080/00949655.2010.520163

15. Zhu, D. and V. Zinde-Walsh. 2009. Properties and estimation of asymmetric exponential power distribution. *Journal of Econometrics.* 148(1): 89-99. doi: 10.1016/j.jeconom.2008.09.038

5.7.   *Supplementary Materials*

TestMosaic.Rmd

```
############################################################
# Supplementary material that accompanies the article:  #
#                Mosaic Normality Test                  #
############################################################
# The ADGofTest package must be installed
library(ADGofTest)
# You can change the values of the following parameters:
n=49          # n: mosaic size (nxn), with 1<n<50. If n is even, then n=n+1
color=c(0.01,0.05,0.10)  # P-value cut points for gray scale
label=F      # label: T=Shows p-value in each cell; F=don't show
bell=0       # 1: Draw the distribution in each box. 0: Do not draw the distribution


sample="NOR" # sample: SAM=from file, NOR=Random Normal(0,1), UNI=Random Uniform(0,1)
, EXP=Random Exponential(1)


# If sample="SAM" The sample has to be in an ASCCI file with all the values in a colu
mn without header and named sample.dat)
mu0=0        # Mean hipotized when data are from a file (sample="SAM"). For Quetelet
Data: mean = 40.332
sigma0=1     # Standard deviation hipotized when data are from a file (sample="SAM").
For Quetelet Data: sd = 2.070


# If sample !="SAM" The data are randomly generated
size=100     # Size for random samples
#set.seed(1)  # You can change the seed (or eliminate it) for random samples
test="AD"    # test: Type of test, AD=Anderson-Darling, KS=Kolmogorov-Smirnov
############################################################


# The following functions are needed
#### Density function for the skewed power exponential distribution
dsepd=function(x,mu=0,sigma=1,alpha=0.5,p=2){
  A=2*p^(1/p)*((1-alpha)^2-alpha^2)*gamma(2/p)/gamma(1/p)
  A2=(2*p^(1/p))^2*((1-alpha)^3+alpha^3)*gamma(3/p)/gamma(1/p)
  B=sqrt(A2-A^2)
  y=A+B*(x-mu)/sigma
  kp=1/(gamma(1+1/p)*(2*p^(1/p)))
```

```r
    kp*exp(-(abs(y/(2*ifelse(x<mu-sigma*A/B,alpha,1-alpha)))^p)/p)*B/sigma
}
#### Distribution function for the skewed power exponential distrib.
psepd=function(x,mu=0,sigma=1,alpha=0.5,p=2){
  A=2*p^(1/p)*((1-alpha)^2-alpha^2)*gamma(2/p)/gamma(1/p)
  A2=(2*p^(1/p))^2*((1-alpha)^3+alpha^3)*gamma(3/p)/gamma(1/p)
  B=sqrt(A2-A^2)
  x=A+B*(x-mu)/sigma
  alpaux=ifelse(x<0,alpha,1-alpha)
  alpha+sign(x)*alpaux*pgamma(((abs(x)/(2*alpaux))^p)/p,shape=1/p)
}
sam=switch(sample,
  SAM = list(values=unlist(read.table("sample.dat",header=F)),mu0=mu0,sigma0=sigma0),
  NOR = list(values=rnorm(size,0,1),mu0=0,sigma0=1),
  UNI = list(values=runif(size,0,1),mu0=0.5,sigma0=1/sqrt(12)),
  EXP = list(values=rexp(size,1),mu0=1,sigma0=1)
)
#### Mosaic function
mosaic=function(sam,n,mu,sigma,label=F){
  siz=length(sam)
  if (n<=1 | n>=50) {
    stop("n must be a value between 2 and 49\n")
  }
  pval=switch(test,
            AD = function(samp,mu=mu,sigma=sigma,alpha=alpha,p=p) ad.test(samp,dist
r.fun=psepd,mu=mu,sigma=sigma,alpha=alpha,p=p)$p.value,
            KS = function(samp,mu=mu,sigma=sigma,alpha=alpha,p=p) ks.test(samp,psep
d,mu=mu,sigma=sigma,alpha=alpha,p=p)$p.value
  )
  if (n%%2==0) n=n+1
  cex1=c(1.2,1,0.8,0.7,0.6)
  old.par=par(mfrow = c(n,n), xaxt="n", yaxt="n", mar=c(0,0,0,0), cex=cex1[findInterv
al(n,c(1,5,10,15,25,9999))], xaxs="i", yaxs="i")
  s=3.5
  ymax=c(0,1/sigma)
  j=(log10(50)/log10(2))^(1/(n/2-0.5))
  alfa=(0:(n-1))/(n-1)
  alfa[alfa==0]=1e-12
  alfa[alfa==1]=1-1e-12
  p=c(1,2,50)
  if (n>3) p=c(1,2^(1/(j^(((n-3)/2):1))),2,2^(j^((1:(n/2-0.5)))))
```

```r
  graf=function(param,label){
    x=seq(from=mu-s*sigma,to=mu+s*sigma,length.out=101)
    plot(0,0,xlim=c(mu-s*sigma,mu+s*sigma),ylim=ymax,type="n",axes=F)


    pvalue=round(pval(samp=sam,mu=mu,sigma=sigma,alpha=param[1],p=param[2]),2)
    id=findInterval(pvalue,c(0,color,1))
    colo=gray(c(0.9,0.85,0.4,0.2,0))[id]
    if(id>1) rect(mu-s*sigma,0,mu+s*sigma,ymax[2],col=colo)
    if (bell==1) curve(dsepd(x,mu=mu,sigma=sigma,alpha=param[1],p=param[2]),col="blac
k",add=T)
    if (param[1]==0.5 & param[2]==2) box(lwd=2,col="white") else box()
    if (label) text(mu,ymax[2]*0.5,round(pvalue,2),col=ifelse(id>2,"white","black"))
  }
  apply(expand.grid(alfa,p),1,graf,label=label)
  par(old.par)
}
mosaic(sam$values,n,sam$mu0,sam$sigma0,label)
```

Parte II

Aplicaciones de técnicas estadísticas en hidrología

TRANSFER FUNCTION AND TIME SERIES OUTLIER ANALYSIS: MODELLING SOIL SALINITY IN LOAMY SAND SOIL BY INCLUDING THE INFLUENCES OF IRRIGATION MANAGEMENT AND SOIL TEMPERATURE[†]

BASEM ALJOUMANI[1]*, JOSEP A. SÀNCHEZ-ESPIGARES[2], NURIA CAÑAMERAS[3], GERD WESSOLEK[1] AND RAMON JOSA[3]

[1]*Department of Ecology, Soil Conservation, Technische Universität Berlin, Berlin, Germany*
[2]*Department of Statistical and Operational Research, Universitat Politècnica de Catalunya (UPC), Barcelona, Catalunya, Spain*
[3]*Department of Agri-food Engineering and Biotechnology, Universitat Politècnica de Catalunya (UPC), Barcelona, Catalunya, Spain*

ABSTRACT

In variable interval irrigation, simply including soil salinity data in the soil salinity model is not valid for making predictions, because changes in irrigation frequency must also be taken into account. This study on variable interval irrigation used capacitance soil sensors simultaneously to obtain hourly measurements of bulk electrical conductivity (σ_b), soil temperature (t) and soil water content (θ). Observations of σ_b were converted so that the electrical conductivity of the pore water (σ_p) could be estimated as an indicator of soil salinity. Values of θ, t and σ_p were used to test a mathematical model for studying how σ_p cross-correlates with t and θ to predict soil salinity at a given depth. These predictions were based on measurements of σ_p, t, and θ at a shallow depth. As a result, prediction at shallow depth was successful after integrating intervention analysis and outlier detection into the seasonal autoregressive integrated moving average (ARIMA) model. We then used the (multiple-input/ one-output) transfer function models to logically predict soil salinity at the depths of interest. The model could also correctly determine the effect of the irrigation event on soil salinity. Copyright © 2017 John Wiley & Sons, Ltd.

KEY WORDS: capacitance device; pore water electrical conductivity; autoregressive integrated moving average (ARIMA) model; outlier detection; transfer function model

RÉSUMÉ

Dans le cas d'une irrigation à intervalle variable, il ne suffit pas d'inclure des données de salinité dans un modèle prédictif, car la fréquence des irrigations influe sur la salinité. Nous avons utilisé des capteurs capacitifs pour mesurer au pas de temps horaire la conductivité électrique apparente (σ_b), la température du sol (t) et la teneur en eau du sol (θ). Les observations σ de conductivité électrique ont été converties de façon à ce que la teneur en eau dans les pores (σ_p) puissent être estimées comme indicateur de la salinité. Les valeurs de θ, t et σ_p ont été utilisés pour tester un modèle mathématique pour étudier comment σ_p s'auto corrèle avec t et θ pour prédire la salinité du sol à une profondeur donnée. Ces prédictions sont basées sur des mesures de σ_p, t et θ à faible profondeur. En conséquence, la prédiction à faible profondeur a réussi après une analyse intégrante et la détection des valeurs aberrantes dans un modèle de moyenne mobile saisonnière autorégressive intégrée (ARIMA). Nous avons ensuite utilisé des modèles de fonction de transfert (plusieurs entrées, une seule sortie) pour prédire la salinité du sol aux profondeurs d'intérêt. Le modèle a pu également déterminer correctement l'effet de l'évènement d'irrigation sur la salinité du sol. Copyright © 2017 John Wiley & Sons, Ltd.

MOTS CLÉS: dispositif de mesure de capacitance; eau de conductivité électrique porale; modèle intégré autorégressif de moyenne mobile (ARIMA); détection des valeurs aberrantes; modèle de fonction de transfert

6. Transfer Function and time series outlier analysis: Modelling soil salinity in loamy sand soil by including the influences of irrigation management and soil temperature

Basem Aljoumani[1], José A. Sanchez-Espigares[2], Nuria Cañameras[3], Geerd Wessolek[1] and Ramon Josa[3]

[1]Department of Ecology, Soil Conservation, Technische Universität Berlin, Berlin, Germany
[2]Department of Statistical and Operational Research, UPC-BarcelonaTech, Barcelona, Spain
[3]Department of Agri-food Engineering and Biotechnology, UPC-BarcelonaTech, Barcelona, Spain

ABSTRACT

In variable interval irrigation, simply including soil salinity data in the soil salinity model is not valid for making predictions, because changes in irrigation frequency must also be taken into account. This study on variable interval irrigation used capacitance soil sensors simultaneously to obtain hourly measurements of bulk electrical conductivity (σ_b), soil temperature (t) and soil water content (ϑ). Observations of σ_b were converted so that the electrical conductivity of the pore water (σ_p) could be estimated as an indicator of soil salinity. Values of ϑ, t and σ_p were used to test a mathematical model for studying how σ_p cross-correlates with t and ϑ to predict soil salinity at a given depth. These predictions were based on measurements of σ_p, t, and ϑ at a shallow depth. As a result, prediction at shallow depth was successful after integrating intervention analysis and outlier detection into the seasonal autoregressive integrated moving average (ARIMA) model. We then used the (multiple–input/one-output) transfer function models to logically predict soil salinity at the depths of interest. The model could also correctly determine the effect of the irrigation event on soil salinity.

KEYWORDS: capacitance device; pore water electrical conductivity; autoregressive integrated moving average (ARIMA) model; outlier detection; transfer function model.

6.1.   Introduction

Salinity is shown to have a large impact on plants by reducing their ability to take up water. This creates an imbalance of plant nutrients, which ultimately leads to the degradation of land. Munns (2002) demonstrated that when salts accumulate excessively in older leaves, premature senescence occurs, causing a reduction in the leaf area available for photosynthesis and thus rendering the plant less able to sustain growth.

Determining the electrical conductivity of soil pore water (σ_p) conventionally requires extraction of the water from the soil by suction or measurement of saturated paste conductivity, both of which are labour-intensive methods. Also, there is always uncertainty as to whether all ions have been collected in the extract sample (Hilhorst, 2000).

A more recent method for temporally and spatially evaluating the σ_p is to convert the bulk electrical conductivity (σ_b) to σ_p by using methods, models and estimates like those described by Rhoades et al. (1990) or Mualem and Friedman (1991).

6.2.   Soil salinity measurement

New sensors have been developed to measure σ_b, such as time-domain reflectometry (TDR) and frequency-domain reflectometry (FDR).

Temperature and water content significantly affect the accuracy of determining σ_p. Therefore, a precise real-time measurement of σ_p depends on electrical conductivity sensors being able to simultaneously measure three variables: water content (θ), soil temperature (t) and σ_b.

The σ_b of the soil system is determined by estimating the conductance pathways in the system, namely: i) solid-liquid interphase; ii) solid phase; iii) liquid phase. In agricultural practices, it is beneficial to recognise the level of electrical conductivity of the liquid phase (σ_p) that is contained in the soil pores, as it provides a sound indicator of the solute concentration contained within the soil. A strong linear correlation between the values of the dielectric soil constant (ε_b) and σ_b in most soil types was discovered by Malicki et al. (1994). This discovery was further developed by Hilhorst (2000), whereby a successful conversion of σ_b to σ_p was made possible by applying a theoretical model that outlined the linear relationship between σ_b and ε_b.

σ_p was estimated by Hilhorst (2000) using the equation:

$$\sigma_p = \frac{\varepsilon_p \sigma_b}{\varepsilon_b - \varepsilon_{\sigma_b = 0}}$$

(1)

where σ_p is the pore water electrical conductivity (dS m^{-1}); ε_p is the real portion of the soil pore water's dielectric permittivity (unitless); σ_b represents the bulk electrical conductivity (dS m^{-1}); ε_b is the real portion of the bulk soil's dielectric permittivity (unitless); $\varepsilon_{\sigma b=0}$ is the real portion of the soil's dielectric permittivity when the bulk electrical conductivity has a value of 0 (unitless). It should be noted that $\varepsilon_{\sigma b=0}$ is an offset of the linear relationship between ε_b and σ_b. In the soils used for Hilhorst's study (2000), $\varepsilon_{\sigma b=0}$ depended on the soil type and it varied between 1.9 and 7.6. So, he recommended 4.1 as a generic offset.

Having simultaneous data on σ_p, t and θ enabled us to properly build models that could precisely predict soil salinity by taking into account the changes in θ and t.

6.2.1. Soil salinity movement models

In predicting solute transport between the land surface and groundwater table, deterministic convection-dispersion equations based on Fickian diffusion are convenient tools for describing solute movement, as they allow a limited number of field studies to be extracted to various soils, crops and climates, as well as to differing tillage and water management regimes (van Genuchten, 1991).

However, questions have been raised in the literature as to the utility of these equations when describing solute transport in structured soils where there exist large continuous voids, for example: natural inter-aggregate pores, inter-pedal voids, earthworm tunnels and gopher holes. The movements of solutes in these voids can be very different from those that occur through materials that are fairly homogeneous (Beven and Germann, 1982; White, 1985). Because most soils are heterogeneous, specific methods are required for simulating heterogeneous field-scale transport processes (van Genuchten, 1991).

As a way of managing the heterogeneous nature of soil and predicting the evolution of soil solutes, some experimenters prefer to adopt stochastic models instead of using constant values to describe the possible future evolution of soil solutes. These models assume that solute transport has random variables, and each variable is assigned a discrete value in accordance with a specified probability distribution.

The use of stochastic models has increased significantly over the last decade. For example, they have been employed in: artificial neural networks (ANN) (Huang et al., 2010); agronomic applications that model crop development (Zhang et al., 2009; Fortin et al., 2010); and predicting crop yields (Park et al., 2005; Green et al., 2007; Khazaei et al., 2008). Zou et al. (2010) collected silt loam soil profile data on a monthly basis from 2001 to 2006 and used it to compare the back propagation neural network (BPNN) model and the autoregressive integrated moving average (ARIMA) model. Their objective was to predict (1) the average moisture content in the top 1-meter profile by using the moisture content measured at 0.60 m depth, and (2) the average salt content measured at various depths of the soil profile (0.10, 0.20 and 0.45 m). Mishra and Desai (2005) used ARIMA and seasonal ARIMA models to forecast droughts.

Sarangi et al. (2006) used artificial neural networks (ANNs) to model root zone soil salinity and the salinity of subsurface drainage effluent in the coastal clay soils of rice fields in Andhra Pradesh, India. They found that drainage effluent salinity could be predicted better by feeding the input values to the ANNs after a time lag rather than by using the conceptual SALTMOD model.

In irrigated agriculture, the ARIMA model has the potential to predict soil salinity, in that it uses past observations as a means of predicting future patterns. Previous models assumed a fixed spacing between irrigation events and that therefore the ARIMA model could be applied suitably. One example could be that of a farmer who irrigates a field every ten days with the expectation that the field data set will provide an ARIMA model in which there is a decrease in soil salinity on the tenth day. This is known as a post-irrigation event. However, if the farmer decides to apply variable interval irrigation and thus uses a spacing of 9 days between irrigation events, the ARIMA models will not be capable of effective predictions in this case. In this instance, the previously identified ARIMA model would show a decrease in soil salinity after ten days instead of after 9 days. Wei (1989) states that –in order to successfully apply an ARIMA model to time series datasets with outliers– it is necessary to incorporate intervention analysis models and outlier detection.

Aljoumani et al. (2012) explained why the ARIMA models cannot predict soil water content in cases of variable interval irrigation. By including intervention analysis and outlier detection in the ARIMA model, they were able to predict the water content of the soil.

6.2.2. Outliers versus intervention variables

When considering the time series of soil salinity, soil moisture and soil temperature, a key distinction is made between outliers and intervention variables. If a situation arises where a priori information relating to a special event (in this case, an irrigation event) proves to cause possible abnormal observations, the effect of this event should be captured through intervention analysis. In this study, the abnormal observations of the soil salinity caused by irrigation events are captured through intervention analysis.

Conversely, if anomalies in the observations are represented without a priori information on their occurrence or on the dynamic patterns of their effects (i.e., a precipitation event), this represents an outlier. To predict soil salinity, we incorporated variable interval irrigation into the ARIMA model as a means of examining the effectiveness of the irrigation event by capturing, with intervention analysis, the abnormal observations of soil salinity time series caused by irrigation events. For this to be possible, it was necessary to perform two procedures. First, we detected and removed the outliers, which therefore caused an upward trend in soil salinity forecasts due to an absence of effect from the irrigation events (outliers). Second, we assessed the intervention effects of the irrigation event and included them in the

model, which thus resulted in a decrease of soil salinity at the time of irrigation. In this second procedure, the weight of the irrigation coefficient determined the extent of the decrease. This complementary analysis provides an advantage that can be attributed to the likelihood of having a well-realized irrigation schedule (which is based on a short duration of one day or even one hour) combined with the knowledge of its effect on the soil salinity. Therefore, the next irrigation event will be determined at the point where the prediction for soil salinity exceeds the plant's tolerance to salts. For example, to predict σ_p in a lettuce field, we determine the time of irrigation when the predicted value of σ_p exceeds 2 ds/m, even though the soil moisture value is at field capacity. This is because lettuce crops are sensitive to salinity.

Aljoumani et al. (2014) showed that it is important to capture the effect of irrigation events on bulk electrical conductivity and to then use the effect as an outlier for improving the fitted model. For the same reasons explained in those references, intervention analysis and outlier detection should complement the ARIMA model when modelling soil salinity in variable interval irrigation.

For the purpose of describing soil salinity fluctuations, incorporating the time series outlier and intervention analysis into the ARIMA model provides two advantages. First, if we employ intervention analysis results in the input series that is represented by a simple pulse or step indicator function, this will indicate whether or not an irrigation event is in fact present. This further improves the efficiency of irrigation scheduling and its effect on soil salinity by including the irrigation event's effectiveness in the ARIMA model. Outlier correction is employed primarily to modify the data in a way that accepts the normality hypothesis of the ARIMA model (Box et al., 1994).

The second advantage results from including the outlier analysis in the ARIMA model. This provides more precision from reducing the residual variance of the model.

Many studies have confirmed the effects of soil moisture and temperature when estimating soil salinity. However, McKenzie et al. (1989) and Slavich and Petterson (1990) found that soil texture, θ and t all affect the calibration measurements of electromagnetic (EM) induction when predicting σ_b. Sarangi et al. (2006) found that soil salinity correlates both with the content of soil water and with temperature. However, we found no studies that developed models for predicting soil salinity by taking into account future changes in soil water content and soil temperature.

This study seeks to fulfil the general objective of modelling soil salinity for predictive purposes while also considering future variations in soil water and soil temperature. The specific objectives of this study are as follows:

- to study the autocorrelation as well as the partial correlation functions of the estimated σ_p, ϑ and t, specifically when they are measured at shallow depths; the cross-correlation function between ϑ, σ_p,

and t at 0.10 m depth; the estimated σ_p at an interested depth (0.10 m); and the average soil salinity in the top 0.60 m of the soil profile;

- to develop models for predicting the soil salinity at various greater depths by measuring σ_p, ϑ and t at a single shallow depth;

- to employ outlier analysis and intervention analysis in examining how irrigation events affect soil salinity.

6.3. Materials and Methods

6.3.1. Hourly field observations of bulk electrical conductivity (σ_b), soil temperature (t), and soil water content (ϑ)

Commencing on 23 April 2010, the observations were carried out over a period of fifty-five days in the Agricultural Park of Baix Llobregat, which is situated five kilometers south of Barcelona, Spain. The set-up comprised an experimental field area on 275m2 (55m x 5m) of land with planted lettuce (Lactuca sativa) and which was irrigated by means of a furrow system. Four irrigation events were applied, with each irrigation dose set at 26 mm applied over a period of twenty to twenty-six minutes. The site's soil was fairly uniform, silty loam, and the bulk density ranged between 1.4 and 1.5 g cm-3as far down as 0.75 m of depth. The water table lay 4 m below the surface of the soil. We measured the distributions of σ_b, t and θ across the soil profile in the test furrow by using capacitance soil moisture sensors (5TE, Decagon Devices, Inc., Pullman, WA), which were installed at depths of 0.10, 0.20, 0.35, 0.50 and 0.60 m below the surface of the soil. Then, in order to convert σ_b to σ_p, we used the Hilhorst model (2000). To define the models for predicting σ_p, we used: a total of 1318 observations of estimated σ_p; measurements of t and θ at a depth of 0.10 m; and the σ_p averages for the upper 0.60 m. These were then validated by using an additional 659 observations. All the details about field observations and the data plot were described in the publications Aljoumani et al. (2012) and Aljoumani et al. (2014).

6.3.2. Model identification and forecast

The time series analysis of σ_p, t and θ was carried out in three steps. The first involved applying the Box-Jenkins method (Box et al., 1994) in order to identify an appropriate univariate model for the time series of σ_p, t and θ at 0.10 m depth. This was done using the seasonal ARIMA model:

$$(p, d, q) \times (P, D, Q)_S \qquad (2)$$

where p and q are the regular autoregressive and moving average orders; and P and Q are the seasonal autoregressive and moving average factors; d and D are the orders of differencing for the regular and seasonal part; and, finally, sub-index S represents the seasonal period, which in this study is 24 hours.

In the second step, we included irrigation duration in the model as an intervention analysis in order to evaluate its effects. Then, we conducted a search to establish whether or not the univariate series contains any outliers. In the third step, we modelled the linear system in order to identify the appropriate transfer function. This was done by using, as input, the time series of t and θ at a depth of 0.10 m, with the outputs being the time series of σ_p at a depth of 0.10 m and the average σ_p calculated for the top 0.60m of soil.

6.3.3. Univariate time series analysis

For elucidating the patterns of the σ_p, t and θ data at 0.10 m depth, and for finding the average σ_p in the top 0.60 m of soil, we implemented univariate seasonal (ARIMA) *(p, d, q) x (P, D, Q)*$_S$ modelling techniques. Moreover, for the purpose of identifying and fitting the ARIMA models, four phase approaches were adopted, namely: model identification, model parameter estimation, diagnostic checking, and forecasting. As part of the autoregressive (AR) process, each value of a time series was dependent on the preceding value, in addition to a random shock. The AR model for a centred time series with order p can be defined as:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + a_t$$

or

$$(1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p) X_t = a_t \quad (3)$$

where ϕ_j denotes the jth AR parameter, a_t the Gaussian white-noise error, and B the backshift operator in which $B^p X_t = X_{t-p}$. The errors for the moving average (MA) model are the average of the random errors for this period and previously. We define the MA time series of order q by means of:

$$X_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + .. + \theta_q a_{t-q} ,$$

or

$$X_t = (1 + \theta_1 B + \theta_2 B^2 + .. + \theta_q B^q) a_t \quad (4)$$

where θ_q is the qth MA parameter.

Identifying AR and MA models requires a stationary time series. Although this implies that the variance and mean values are constant, some transformation is needed in order to identify the model.

Successive differences of the data's regular and seasonal components were taken to ensure no trend in mean. The required numbers of differences for the stationary time series were denoted by d and D. Moreover, in applying the logarithmic transformation that is usually used, no trend in variance is obtained (Soebiyanto et al., 2010; Quinn, 1985; Vandaele, 1983).

In any time series, X_t, the ARIMA(p, d, q) × (P, D, Q)S of X_t is:

$$\phi_p(B)\,\Phi_P(B^s)(1-B)^d\,(1-B^s)^D\,X_t = \theta_q(B)\,\Theta_Q(B^s)a_t \qquad (5)$$

where $\phi_p(B)$ and $\theta_q(B)$ are the regular autoregressive and moving average factors, and $\Phi_P(B^s)$ and $\Theta_Q(B^s)$ are the seasonal autoregressive and moving average factors.

The autocorrelation function (ACF) and the partial autocorrelation function (PACF) were used to identify time series models (McCleary and Hay, 1980; Pankratz, 1983; Hoff, 1983). ACF measures the relation between X_t and X_{t+K}, where K is the time lag. We used PACF in order to take into account any dependence on intermediate elements (i.e., those inside the lag) (Box et al., 1994; McDowall et al., 1980; Wei, 1989).

The maximum likelihood method was chosen for this study in order to estimate the model parameters. In addition, by constructing the Wald test statistic, we were able to determine the significance of these parameters.

Diagnostic checking tests were used to ascertain whether the residuals showed any autocorrelation at any lags. The assumptions would be satisfied if the ACF and PACF of the residuals were non-significant at all lags.

6.3.4. Intervention analysis and outlier detection

Outliers in the σ_p, t and θ data refer to soil at 0.10 m depth and to the average σ_p in the top 0.60 m. They were removed using Grubbs' test for detecting outliers (Grubbs, 1969):

$$Z = \frac{/M - V/}{SD} \quad (6)$$

where Z is the test statistic, M is the mean of the values, V is the value being tested, and SD is the standard deviation of the values. A total of 1318 observations of σ_p, t and θ were available. Assuming an outlier

probability of 5%, we set the outlier test statistic at 4 (Grubbs, 1969). The values of σ_p, t and θ that yielded test statistics greater than or equal to 4 were eliminated from the data set. In order to properly assess how precipitation and other observed irregularities impacted the time series of σ_p, t and θ, two types of outliers were considered: additive outliers (AO) and temporary changes (TC). At the same time, level shift (LS) was used as an intervention analysis to assess the impact of the irrigation event on the time series of σ_p, t and θ. AO is a pulse that affects the time series at one period only. TC is an event that decays exponentially, according to a pre-specified dampening factor. LS is an event that permanently affects the subsequent level of a series (Chen and Liu, 1993).

Let Z_t denote the underlying time series process, which is free of the impact of outliers and occurs prior to the irrigation event; and let X_t denote the observed time series. We assume that Z_t follows the seasonal ARIMA*(p,d,q)(P,D,Q)*s model $\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D Z_t = \theta_q(B)\Theta_Q(B^s)a_t$. Based on these assumptions, the appropriate model for assessing the impact of the control is:

$$X_t = \sum_{r=1}^{n_r}\omega_r S_{T_r}^{(LS)} + \sum_{i=1}^{n_i}\omega_i P_{T_i}^{(TC)} + \sum_{j=1}^{n_j}\omega_j P_{T_j}^{(AO)} + Z_t$$

$$= \sum_{r=1}^{n_r}\omega_r S_{T_r}^{(LS)} + \sum_{i=1}^{n_i}\omega_i P_{T_i}^{(TC)} + \sum_{j=1}^{n_j}\omega_j P_{T_j}^{(AO)} + \frac{\theta_q(B)\,\Theta_Q(B^S)}{\phi_p(B)\Phi_p(B^S)(1-B)^d(1-B)^D}a_t \quad (7)$$

where ω_r is permanent change in the mean level following the intervention (i.e., the irrigation event), and $S_{T_r}^{(LS)}$ is a step indicator at the time of irrigation T_r, where:

$$S_{T_r}^{(LS)}\begin{cases} 0 & t < T_r \\ 1 & t \geq T_r \end{cases} \quad (8)$$

In this study, the level shift (LS) in the soil salinity time series is produced by the irrigation events; and, since the date of these events is known a priori, we were able to assess how irrigation events affect soil salinity by completing the ARIMA model with intervention analysis

ω_i is the transitory change in the mean level following any unusual observations (e.g., precipitation); $P_{T_i}^{(TC)}$ and $P_{T_j}^{(AO)}$ are pulse indicators taken at unusual observation times T_i and T_j, respectively, where:

$$P_{T_i}^{(TC)} \begin{cases} 0 & t < T_i \\ \dfrac{1}{(1 - \delta B)} & t = T_i \end{cases} \quad (9)$$

and

$$P_{T_j}^{(AO)} \begin{cases} 0 & t \neq T_j \\ 1 & t = T_j \end{cases} \quad (10)$$

δ is a dampening factor with the default value of 0.7 (Chen and Liu, 1993).

The exploratory method of visualizing the outliers in time series analyses has been well-established in other fields, and this seasonal-trend decomposition (commonly known as 'STL') uses locally-weighted regression (loess) (Cleveland et al., 1990; Hafen et al., 2009). The STL method is straightforward and flexible for specifying the amount of variation in the seasonal components and trends of time-series; furthermore, it provides robust estimates without any distortions resulting from transient outliers (Cleveland et al., 1990). STL was utilized to model the 24h-prevalant soil salinity, soil water and soil temperature time-series. STL is a filtering procedure for decomposing a time series into additive components of variation (trend, seasonality and the remainder); and it does so by applying loess smoothing models (Cleveland et al., 1990; Chaloupka, 2001).

6.3.5. Transfer function approach

We can use the observations and predictions from two-time series (input X_{1t} and X_{2t}) to estimate the outcome of another time series (output G_t). This is done by a relatively small number of parameters to model the linear system, which takes the form:

$$G_t = \frac{A_1(B)}{C_1(B)} X_{1,t-b_1} + \frac{A_2(B)}{C_2(B)} X_{2,t-b2} + a_t \quad (11)$$

where $A(B)$ and C(B) are, respectively, the polynomials of the s and r orders:

$$A(B) = (A_0 - A_1 B - A_2 B^2 - ... - A_s B^s) \quad (12)$$

$$C(B) = (1 - C_1 B - C_2 B^2 - ... - C_r B^r) \quad (13)$$

where A_0, A_1, A_2, …, A_s and C_1, C_2, …, C_r are the parameters of the model, b is the latent parameter, B is the backshift operator, and a_t is a disturbance (noise).

$A(B)/C(B)$ is the system's designated transfer function. In modelling a transfer function, the procedure involves three steps: a) identification, b) estimation and c) checking the model. The same filter can be applied to the output series G_t (pre-whitening) by employing a univariate model for the inputs of

X_{1_t} and X_{2_t} while using white noise residuals. Cross-correlating the two residuals identifies the transfer function form.

The transfer function in this study uses the soil water (θ) and soil temperature (t) that were observed at 0.10 m depth for the primary series (X_{1_t} and X_{2_t}).We chose the output series (G_t) from the time series of soil salinity (σp) that were observed at a depth of 0.10 m, as well as the average soil salinity (σp) from the upper 0.60 m of the soil profile. The formula used by Wu et al. (1997) was adopted for calculating the average soil salinity (σp) in the upper 0.60 m of the soil profile.

The software R, version 2.15.1 (R Development Core Team, 2012), executed all model identifications and subsequently predicted soil salinity at various depths (Cryer and Chan, 2008; Shumway and Stoffer, 2006).

6.4.   Results and discussion

Figure 1 shows the variation of θ, t, and σp at 0.10 m depth over time. Decreases in σp signalled the significant effects of irrigation events that occurred on days 4.29, 27.20, 32.04 and 46.33, as well as from precipitation events on days 9.33, 20.50 and 52.54. Figure 1 also shows that t increased after irrigation, due to the irrigation water being warmer than the soil before irrigation.
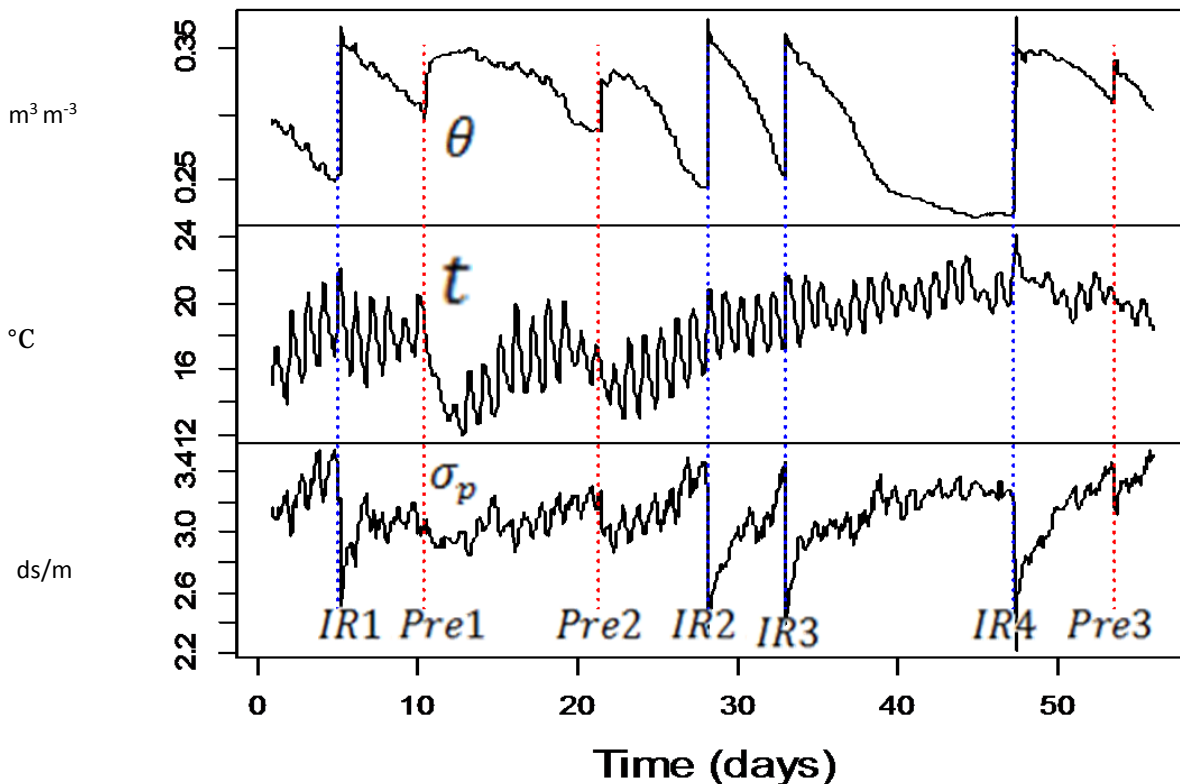
Figure 6.1 Variation of soil water content (ϑ m³m⁻³), soil temperature (t °C) and soil salinity(σP dS m⁻¹) at 0.10 m depth with time IR1, IR2, IR3 and IR4 are the irrigation events applied on days 4.29, 27.20, 32.04 and 46.33. Pre1, Pre2, and Pre3 are the precipitation event on days 9.33, 20.50 and 52.54.

The opposite occurred with precipitation: Figure 1 shows that t decreased after precipitation; acknowledgement of these fluctuations will help in modelling σ_p as a function of θ and t, as explained below. Later, we developed the ARIMA model for σ_p time series at 0.10 m depth and completed it by including the irrigation event as an intervention and the precipitation as outlier detection.

6.4.1. Univariate time series modelling of soil salinity at 0.10 m depth

We can see that the time series is non-stationary, as indicated by the significant slow convergence of ACF in the original σ_p time series at 0.10m depth (Figure 2.A). As a means for obtaining a stationary time series, we differentiated the original series (first-order difference and seasonal first-order difference). There is no requirement for applying a logarithmic transformation in this case, as there is no trend in variance observed in the series.
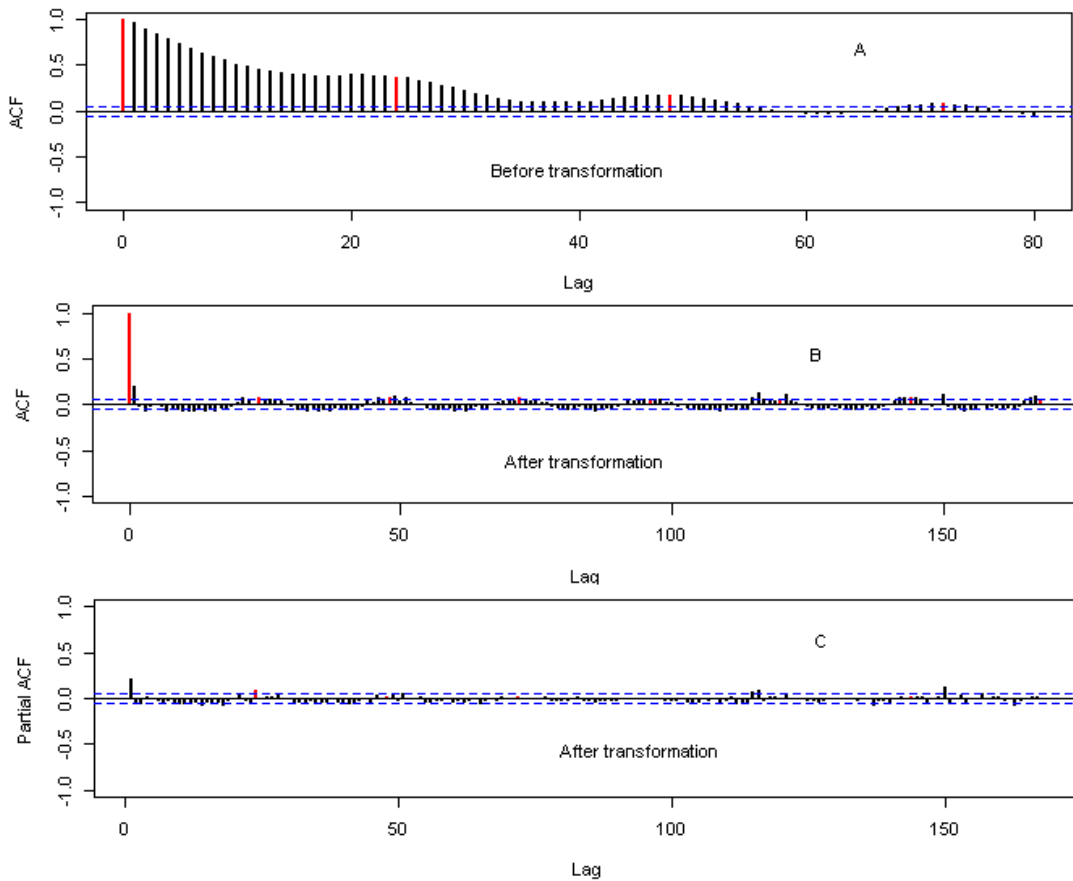
Figure 6.2 (A) Autocorrelation function (ACF) of the original data, (B) autocorrelation function, and (C) partial autocorrelation function (PACF) of the transformed time series of σ_p at 0.10 m depth. The ACF of the original data indicates that the series is not stationary. The dotted lines mark 2 x standard errors.

The ACF and PACF of the differentiated time series found that the regular component of the series was approximately AR*(3)* and that the seasonal component was MA*(1)*. This is because the ACF (Figure 2.B) found significant correlations only at the first three lags of PACF and the 24th lag of ACF.

The ARIMA *(p, d, q) (P, D, Q)*$_S$ model of the σ_p time series at a depth of 0.10 m resulted in ARIMA *(3, 1, 0) (0, 1, 1)*$_{24}$. The model is expressed with the usual notation by:

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)(1 - B)(1 - B^{24})X_t = (1 + \theta_{24} B^{24})a_t \quad (14)$$

where a_t is a white noise term that is independent and identically distributed with zero mean, and the variance = $2.8 \times 10\text{-}7$, ϕ_1 = 0.2088, ϕ_2 = -0.0468. ϕ_3 = -0.0883 are AR parameters. The θ_{24} = 0.99 parameter of the seasonal MA part indicates that the model is almost non-invertible. Therefore, it is inadequate and needs to be improved in structure.

By using a seasonal-trend decomposition (STL method), the large outliers of the remainder (random) return to the irrigation events. Because we already know the timing of the irrigation event, the model can be completed using intervention analysis (irrigation events) and outlier detection (model 10); in this way it becomes invertible and thus reduces its residual variance (Wei, 1989).

6.4.2. Outlier and intervention analysis in the σ_p time series of the ARIMA model at 0.10 m depth: Measuring the effect of the irrigation event on σ_p

We applied intervention analysis and automatic outlier detection to the previous ARIMA (3, 1, 0) (1, 0, 0)24 model in order to improve it and measure the effects of irrigation events on σ_p at 0.10 m depth. With Grubb's test (Eq. 6), 15 outliers were detected (Table I) for the time series of soil water content at 0.10 m depth.

Including the outlier detection and intervention analysis (effect of irrigation), the observed value of time series of σ_p at 0.10 m depth can be described according to Eq. 7 as:

$$X_t = \omega_r (S_{4.29}^{(t)} + S_{27.20}^{(t)} + S_{32.04}^{(t)} + S_{46.33}^{(t)}) + \sum_{i=1}^{23} \omega_i P_{T_i}^{(TC)} + \sum_{j=1}^{5} \omega_j P_{T_j}^{(AO)} + Z_t \qquad (15)$$

X_t is the observed time series; Z_t is the time series free of outliers; and ω_r =-0.759 represents the permanent change in the mean level after the irrigation event, which characterizes the effectiveness of the irrigation event on the soil salinity. In this study, the flow rate and cut-off time for the four applied irrigations were almost equal; therefore an average coefficient for ω_r was used for estimating the irrigation event's impact. In the $(S_{4.29}^{(t)} + S_{27.20}^{(t)} + S_{32.04}^{(t)} + S_{46.33}^{(t)})$ part of the equation, the step indicator is represented for four irrigation times T_r (days 4.29, 27.20, 32.04, and 46.33).

Table 6.1 Outlier detection and parameter estimation for time series of soil salinity at 0.10 m

Observation time (hour)	type	ω
103	AO 0.185	7.97
106	TC 0.125	4.55
153	AO 0.098	4.56
494	TC -0.170	6.04
653	AO 0.300	12.20
654	AO -0.111	5.146
770	TC -0.231	7.88
919	TC -0.111	4.18
962	TC -0.110	4.18
1001	TC -0.118	4.32
1029	TC -0.109	4.16
1089	TC 0.114	4.216
1112	TC -0.286	9.32
1113	TC -0.146	5.21
1262	TC -0.207	7.22

The $\omega_r (S_{4.29}^{(t)} + S_{27.20}^{(t)} + S_{32.04}^{(t)} + S_{46.33}^{(t)})$ part of the equation of a time series for soil salinity (X_t) shows that, at the time of irrigation, the evolution of soil salinity will decrease, and the coefficient ω_r will determine the degree to which it decreases. Moreover, the effects of the 15 detected outliers are represented by the $\sum_{i=1}^{23} \omega_i P_{T_i}^{(TC)} + \sum_{j=1}^{5} \omega_j P_{T_j}^{(AO)}$ part of the equation.

Applying a Box-Jenkins method to the time series of the soil salinity $(\sigma_p) Z_{t_t}$ from Eq. 15, we determine the ARIMA *(3, 1, 0) (0, 1, 1)*₂₄. In the usual notation, the model is written as:

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)(1 - B)(1 - B^{24})Z_t = (1 + \Theta_{24} B^{24})a_t \qquad (16)$$

This model (16) is free of outliers and is invertible, with non-significant ACF and PACF of residuals at all lags. Table 2 provides the comparison of statistical parameters of the two models (14) and (16).

After modelling σ_p at a depth of 0.10 m, our next step is to model the θ and t time series at a depth of 0.10 m as well as the average soil salinity in the top 0.60 m of soil. Table 3 shows the ARIMA models for θ and t and the average soil salinity in top upper 0.60 m of soil; these resulted from following the same steps applied to model σ_p at a depth of 0.10 m. The effects of the irrigation events on the time series of θ and t at 0.10 m were 0.0843 and 0.288, respectively.

After having identified all the models and estimating the parameters, diagnostic checks are then applied to the fitted model to verify whether the model is adequate.

6.4.1. Transfer function method

Looking at the cross-correlation between the pre-whitened, primary time series (soil water and temperature at a depth of 0.10 m) and target soil salinity (0.10 m depth and the average), we see an effect from the primary series on the target series, but not vice-versa. Figure 3 indicates that the current values of soil water, content and temperature at 0.10 m have a significant effect on the current soil salinity values at a depth of 0.10 m and on the average soil salinity in the top 0.60 m of soil.

Models were identified for predicting soil salinity based on soil water content and soil temperature at a depth of 0.10 m (Table 4). The coefficients of X_t in Table 4's equations indicate that the current values of soil water content and soil temperature at 0.10 m have effects, respectively, of -7.82 and -0.050 on the current values of soil salinity at a depth of 0.10 m. Furthermore, the current values of soil water content and soil temperature at 0.10 m depth have effects, respectively, of -1.68 and -0.004 on the current values of the average soil salinity in the soil's top 0.60 m.
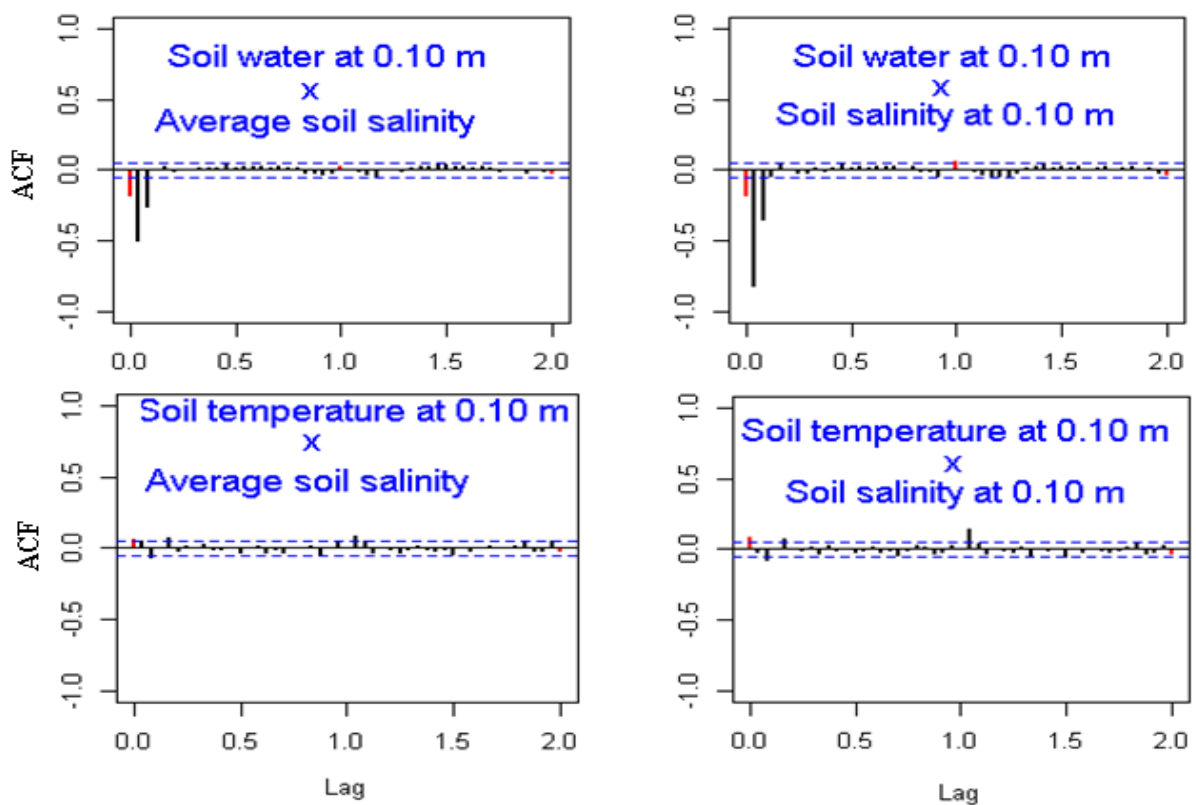
Figure 6.3 Cross-correlation function for soil water content and soil temperature hourly time series at 0.10 m and soil salinity at 0.10 m depth. Cross-correlation function for soil water content and soil temperature 0.10 m and soil salinity in the top 0.60 m of soil profile, respectively. Dashed lines indicate 95% confident limits.

Table 6.2 Comparison of the two models for soil salinity at 0.10 m depth in terms of statistical parameters (one based on observed data X_t and the second based on outlier-free data Z_t)

Model	ϕ_1	ϕ_2	ϕ_3	θ_{24}	σ^2
Model based on observed data X_t (10)	-0.0114	-0.0684			1.38×10^{-4}
Model based on Outlier free data Z_t (12)	-0.0467	-0.0108	0.0273	-0.923	7.43×10^{-5}

Table 6.3 Models of soil water content (θ), (t) at 0.10 m and soil salinity in the upper 0.60 m of soil

Model	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	Θ_1	θ_{24}	σ^2
Soil water	-0.03	-0.019						1.06×10^{-5}
Soil temperature	1.55	-0.641	0.027			-0.87	-0.88	1.83×10^{-5}
Soil salinity in the upper 0.60 m of soil	0.07	0.158	0.11	0.05	-0. 04	-0.54		1.56×10^{-5}

Table 6.4 Time series transfer function model for soil salinity at 0.10 m depth and in the top of 0.60 m of the soil profile

Soil water content $X_{1,t}$, soil temperature $X_{2,t}$ at 0.10 m and soil salinity Y_t at 0.10 m:

$$Y_t = \frac{(-7.8242 + 1.4053B - 0.2606B^2 + 0.7234B^3)X_{1,t} + (-0.0508 + 0.0153B)X_{2,t} + a_t}{(1-B)(1+0.680B^{24})(1+0.1597B+0.1266B^2+0.0502B^3)}$$

$$a_t \sim N(0, 6.489.8 \times 10^{-5})$$

Soil water content $X_{1,t}$, soil temperature $X_{2,t}$ at 0.10 m and average soil salinity Y_t

$$Y_t = \frac{(-1.6855 - 0.0548B + 0.4975B^2 + 0.0717B^3)X_{1,t} + (0.004 - 0.0111)X_{2,t} + a_t}{(1-B)(1-0.0358B^{24})(1-0.0627+0.0209B^2-0.1754B^3)}$$

$$a_t \sim N(0, 2.684 \times 10^{-5})$$

6.4.2. Forecasting

Figure 4 shows the calibration of the model as well as its predictions of average soil salinity and soil salinity at 0.10 m depth. The first 659 observations of each time series were used to define the model. The output of the calibrated model corresponded very well to the values before 659 for each depth. The predicted and observed values after the 659 observations agreed reasonably. There were sometimes large relative differences between the values that were predicted and those that were observed. However, the absolute difference between prediction and measurement never exceeded 0.27 dS m^{-1}.

Figure 5 provides an example of using the transfer function model presented in Table IV to predict soil salinity over two days at a depth of 0.10 m and in the top 0.60 m of soil. Soil salinity was observed at 55 days, and the prediction is for the 56th and 57th days (48h). It includes the effect of the next irrigation if the farmer chooses to irrigate after 36h. The figure shows how the irrigation event at day 57.5 affects soil salinity evolution, this effect is determined by the ω_r from the equation (15).

However, in this study, the 48-hour predicted value decreases, while the lead-time increases but still remains within the confidence interval (95%).
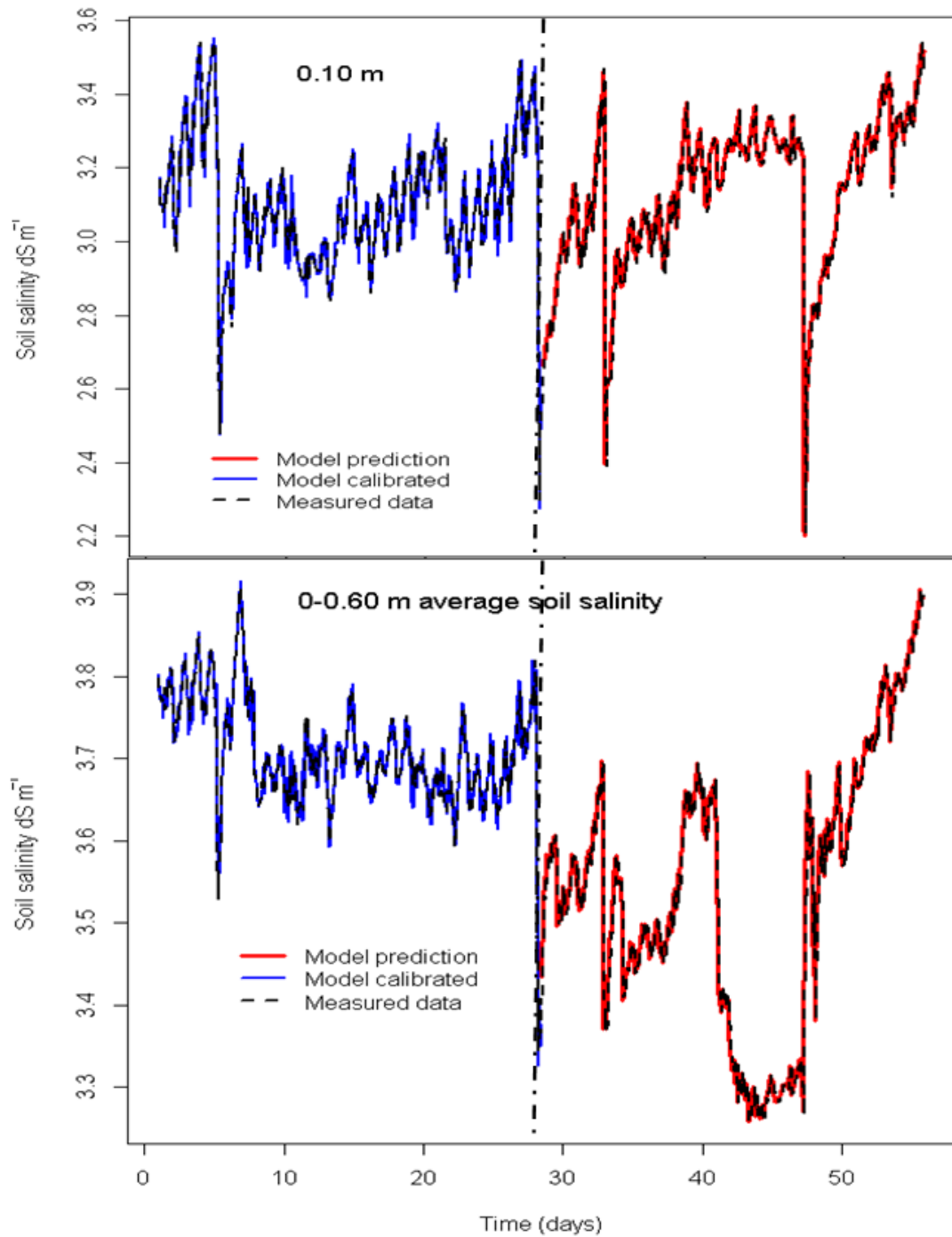
Figure 6.4 Measured and predicted soil salinity versus time at 0.10 m depth and in the top 0.60 m of soil profile. Prediction was based on the identified transfer function models for each one. The curve before the vertical dashed line refers to model calibration and after the vertical dashed line to model prediction.
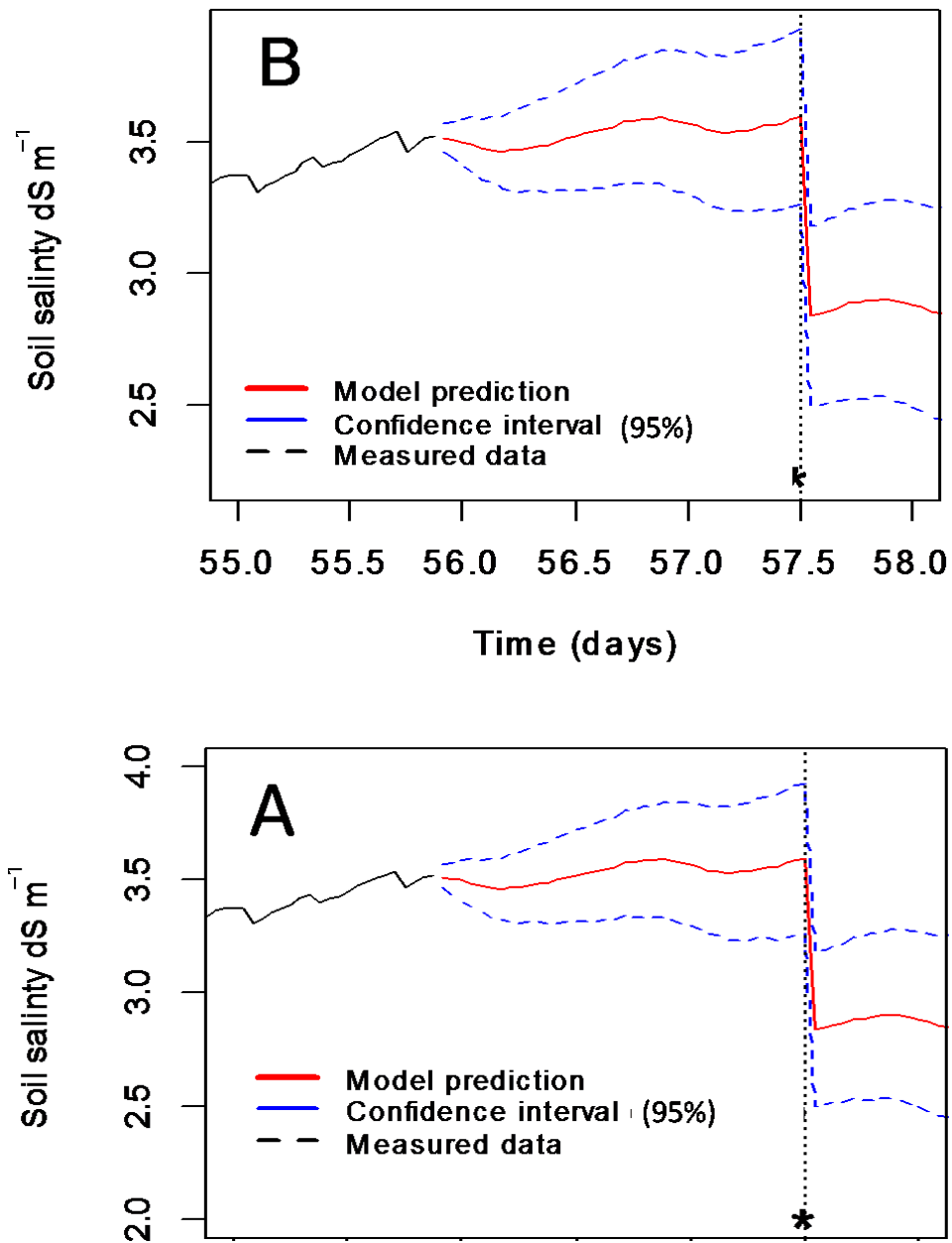
Figure 6.5 Prediction models for soil salinity at 0.10 m depth(A) and average soil salinity in the top 0.60 m of soil profile (B). Prediction was based on the indentified transfer function. We have observed data for 55 days, the model predicts the 56th and 57th days taking into account the effect of next irrigation if the farmer choose to irrigate on 57.5th day (* is the irrigation time at 57.5th day).

6.5. *Conclusion*

Modelling and predicting salinity in soils under variable interval irrigation are not valid if changes in irrigation frequency are not taken into account. We studied the time series under such a regime, specifically for soil salinity (σ_p), soil water content (θ) and soil temperature (t) –all of them at a depth of 0.10 m in a lettuce field of silty loam soil. Each time series was transformed into a stationary series; then, we constructed ARIMA models for studying each time series and for making predictions. The ARIMA model with an underlying normality assumption could not properly predict soil salinity. Therefore, the model was completed by means of intervention analysis (with the interventions being irrigation events) as well as outlier detection (in order to identify unusual observations). We then used transfer function models (multiple-input/one-output) to predict σ_p at our depths of interest (0.10 m and the average σ_p in the upper 0.60 m of soil); this was done using the measured water θ and t at 0.10 m depth and produced rational predictions. The subsequent irrigation and decrease in σ_p after that irrigation event were correctly estimated. Since the irrigation doses for the four irrigation events in this study were almost the same, we used an average mean level (ω_r = 0.087) to show an irrigation event's effectiveness on the soil salinity time series. In cases involving variable doses of irrigation, our suggestion is to study the effect of each irrigation event and to then include each effect separately in the model. Moreover, the time series analysis in our study applies mathematical models to one soil profile at different depths in order to find certain relationships among the observed variables (σ_p, θ and t). However, in order to validate the model we used for the whole area of study, we recommend using additional measurements of σ_p, θ and t from different soil profiles to take into account the spatial variability of the study area. Lastly, future studies should look toward further consideration of how these sensors and their data can be positioned in a way that makes them more accessible for practical use. At present, capacitance sensors of this type are available on the commercial market; however they are limited almost exclusively to use in scientific experiments. This can be attributed to the process involved, whereby the technician is required to travel to the field with a computer in order to transfer the sensor data from the datalogger to an Excel spreadsheet. From this, mathematics is then applied for studying the relationship between the variables. As an alternative, an electronic unit could be designed to be included in the sensor datalogger, which may result from telecommunication technicians incorporating our stochastic models. This would allow for the provision of a visual indication of the electrical conductivity of the soil solution. Our study finds that, with the proper programming, this low-cost sensor could be expanded in a way that provides further beneficial capabilities, such as allowing a normal farmer the opportunity to ascertain salt levels in the root zone.

Notation

σ_b	bulk electrical conductivity (dS m^{-1})	$\phi_p(B)$	regular autoregressive factor
T	soil temperature	$\theta_q(B)$	regular moving average factor
Θ	soil water content	$\Phi_P(B^s)$	seasonal autoregressive factor
σ_p	electrical conductivity of the pore water (dS m^{-1})	$\Theta_Q(B^s)$	seasonal moving average factor
ARIMA	seasonal autoregressive integrated moving average	ACF	autocorrelation function
ε_p	real portion of the soil pore water's dielectric permittivity (unitless)	PACF	partial autocorrelation function
ε_b	real portion of the bulk soil's dielectric permittivity (unitless)	Z	test statistic
$\varepsilon_{\sigma b=0}$	Offset of $\sigma_p \sim \varepsilon_b$ relationship	M	mean of the value
P	regular autoregressive factor	V	value being tested
Q	Regular moving average order	SD	standard deviation of the values
P	seasonal autoregressive order	AO	additive outliers
Q	seasonal moving average order	TC	temporary changes
D	orders of differencing for the regular part	LS	level shift
D	orders of differencing for the seasonal part	ω_r	permanent change in the mean level following the irrigation event
S	sub-index represents the seasonal period	$S_{T_r}^{(LS)}$	step indicator at the time of irrigation T_r
AR	part of the autoregressive process	ω_i	transitory change in the mean level following any unusual observations
ϕ_j	denotes the jth AR parameter	$P_{T_i}^{(TC)}$	pulse indicators taken at unusual observation times T_i
a_t	Gaussian white-noise error	$P_{T_j}^{(AO)}$	pulse indicators taken at unusual observation times T_j
B	backshift operator in which $B^p X_t = X_{t-p}$.	δ	dampening factor with the default value of 0.7
MA	moving average process	STL	seasonal-trend decomposition
θ_q	qth MA parameter	$A(B)/C(B)$	system's designated transfer function
θ_q	qth MA parameter	Z_t	time series free of outliers

6.6.   References

1. Aljoumani B, Sànchez-Espigares JA, Cañameras N, Josa R. 2014. An advanced process to evaluate the linear dielectric constant –bulk electrical conductivity model using capacitance sensor in field conditions. *Hydrological Science Journal* **60.** DOI:10.1080/02626667.2014.932053.

2. Aljoumani B, Sànchez-Espigares JA, Cañameras N, Josa R, Monserrat J. 2012. Time series outlier and intervention analysis: irrigation management influences on soil water content in silty loam soil. *Agricultural Water Management* **111**: 105-114. DOI: 10.1016/j.agwat.2012.05.008.

3. Beven K, Germann P. 1982. Macropores and water flow in soils. *Water Resources Research* **18** (5): 1311–1325. DOI: 10.1029/WR018i005p01311.

4. Box GEP, Jenkins GM, Reinsel GC. 1994. *Time series analysis: Forecasting and control*, 3rd ed. Englewood Clifs: Prentice Hall.

5. Chaloupka M. 2001. Historical trends, seasonality and spatial synchrony in green sea turtle egg production. *Biological Conservation*. **101** (3): 263–279. DOI: 10.1016/S0006-3207(00)00199-3.

6. Chen C, Liu L. 1993. Joint estimation of model parameters and outlier effects in time series. Journal of the American Statistical Association 88 (421): 284–297. DOI: 10.2307/2290724.Cleveland RB, Cleveland WS, McRae JE, Terpenning I. 1990. STL: a seasonal trend decomposition procedure based on loess. *Journal of Official Statistics* **6** (1): 3–73.

7. Cryer JD, Chan KS. 2008. *Time Series Analysis with Applications in R*, 2nd ed. New York: Springer. DOI: 10.1007/978-0-387-75959-3.

8. Fortin JG, Anctil F, Parent L, Bolinder MA. 2010. Aneural network experiment on the site-specific simulation of potato tuber growth in Eastern Canada. *Computers and Electronics in Agriculture* **73** (2): 126–132. DOI: 10.1016/j.compag.2010.05.011

9. Green TR, Salas JD, Martinez A, Erskine RH. 2007. Relating crop yield to topographic attributes using spatial analysis neural networks and regression. *Geoderma* **139** (1-2): 23–37. DOI: 10.1016/j.geoderma.2006.12.004.

10. Grubbs F. 1969. Procedures for detecting outlying observations in samples. *Technometrics* **11** (1): 1–21. DOI: 10.1080/00401706.1969.10490657.

11. Hafen R, Anderson D, Cleveland W, Maciejewski R, Ebert D, Abusalah A, Yakout M, Ouzzani M, Grannis S. 2009. Syndromic surveillance: STL for modeling, visualizing, and monitoring disease counts. *BMC Medical Information and Decision Mak*ing **9** (21). DOI: 10.1186/1472-6947-9-21.

12. Hilhorst MA. 2000. A pore water conductivity sensor. *Soil Science Society of American Journal* **64** (6): 1922-1925. DOI: 10.2136/sssaj2000.6461922x.

13. Hoff JC. 1983. *A Practical Guide to Box–Jenkins Forecasting*. London: Lifetime Learning Publications.

14. Huang Y, Lan Y, Thomson SJ., Fang A, Hoffmann WC, Lacey RE. 2010. Development of soft computing and applications in agricultural and biological engineering. *Computers and Electronics in Agriculture* **71** (2): 107-127. DOI: 10.1016/j.compag.2010.01.001.

15. Khazaei J, Naghavi MR, Jahansouz MR, Salimi-Khorshidi G. 2008. Yield estimation and clustering of chickpea genotypes using soft computing techniques. *Agronomy Journal* **100** (4): 1077–1087.DOI: 10.2134/agronj2006.0244.

16. Malicki MA, Walczak RT, Koch S, Flühler H. 1994. Determining soil salinity from simultaneous readings of its electrical conductivity and permittivity using TDR. In: *Proceedings of the Symposium on Time Domain Reflectometry in Environmental, Infrastructure, and Mining Applications*, Evanston, USA, 7-9 September, U.S. Bureau of Mines, Special Publication SP 19-94, NTIS PB95-105789; 328-336.

17. McCleary R, Hay RA. 1980. *Applied Time Series Analysis for the Social Sciences*. Thousand Oaks: Sage Publications Inc.

18. McDowall D, McCleary R, Meidinger EE, Hay RA. 1980. *Interrupted Time Series Analysis,* Beverly Hills: Sage Publications. McKenzie RC, Chomistek W, Clark NF. 1989. Conversion of electromagnetic inductance readings to saturated paste extract values in soils for different temperature, texture, and moisture conditions. *Canadian Journal of Soil Science.* **69** (1): 25-32. DOI: 10.4141/cjss89-003.

19. Mishra AK, Desai VR. 2005. Drought forecasting using stochastic models. *Stochastic Environmental Research and Risk Assessment* **19** (5): 326–339. DOI: 10.1007/s00477-005-0238-4.

20. Mualem Y, Friedman SP. 1991. Theoretical prediction of electrical conductivity in saturated and unsaturated soil. *Water Resources Res*earch **27** (10): 2771-2777. DOI: 10.1029/91WR01095.

21. Munns R. 2002. Comparative physiology of salt and water stress. *Plant Cell and Environ*ment **25** (2): 239-250. DOI: 10.1046/j.0016-8025.2001.00808.x.

22. Pankratz A. 1983. *Forecasting with Univariate Box–Jenkins Models: Concepts and Cases.* New York: John Wiley and Sons.

23. Park SJ, Hwang CS, Vlek PLG. 2005. Comparison of adaptive techniques to predict crop yield response under varying soil and land management conditions. *Agricultural Systems* **85** (1): 59–81.DOI:10.1016/j.agsy.2004.06.021.

24. Quinn II TJ. 1985. Catch-per-unit-effort: a statistical model for Pacific halibut (Hippoglossus stenolepis). *Canadian Journal of Fisheries and Aquatic Sciences* **42** (8): 1423–1429. DOI: 10.1139/f85-178.

25. R Development Core Team 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3- 900051-07-0, URL. http://www.R-project.org/.

26. Rhoades JD, Shouse PJ, Alves WJ, Manteghi NA, Lesch.SM. 1990. Determining soil salinity from electrical conductivity using different models and estimates. *Soil Science Society of American Journal* **54** (1): 45-54. DOI:10.2136/sssaj1990.03615995005400010007x.

27. Sarangi A, Man Singh AK, Bhattacharya AK, Singh AK. 2006. Subsurface drainage performance study using SALTMOD and ANN models. *Agricultural Water Managment* **84** (3): 240-148. DOI: 10.1016/j.agwat.2006.02.009.

28. Shumway RH, Stoffer DS. 2006. *Time Series Analysis and its Applications. With R Examples*. New York: Springer. DOI: 10.1007/0-387-36276-2.

29. Slavich PG, Peterson GH. 1990 Estimating the electrical conductivity of saturated paste extracts from 1:5 soil-water suspensions and texture. *Australian Journal of Soil Research* **31** (1):73-81. DOI: 10.1071/SR9930073.

30. Soebiyanto RP, Adimi F, Kiang RK. 2010. Modeling and predicting seasonal influenza transmission in warm regions using climatological parameters. *PLoS One*, **5** (3): e9450. DOI: 10.1371/journal.pone.0009450.

31. van Genuchten MT. 1991. Recent progress in modelling water flow and chemical transport in the unsaturated zone. In *Proceedings of the International Symposium. Hydrological Integrations Between Atmosphere, Soil and Vegetation*. Viena, Austria, 11-24 August, Kienitz G, Milly PCD, van Genuchten MTh, Rosbjerb D, Shuttleworth WJ (eds). IAHS Publication **204**:169-183.

32. Vandaele W. 1983. *Applied Time Series and Box–Jenkins Models*. New York: Academic Press.

33. Wei WWS. 1989. *Time Series Analysis: Univariate and Multivariate Methods*. New York: Person Addison-Wesley.

34. White RE. 1985. The influence of macrospores on the transport of dissolved and suspended matter through soil. In *Advanced in Soils Sciences*, Stewart BA (ed); 95-120. DOI: 10.1007/978-1-4612-5090-6.

35. Wu L, Jury WA, Chang AC, Allmaras RR, 1997. Time series analysis of field-measured water content of a sandy soil. Soil Science Society of American Journal**61** (3): 736–742. DOI: 10.2136/sssaj1997.03615995006100030005x.

36. Zhang JQ, Zhang LX, Zhang MH, Watson C. 2009. Prediction of soybean growth and development using artificial neural network and statistical models. *Acta Agronomica* Sinica **35** (2): 341–347. DOI: 10.1016/S1875-2780(08)60064-4.

37. Zou P, Yang J, Fu J, Liu, G, Li D. 2010. Artificial neural network and time series models for predicting soil salts and water content. *Agricultural Water Management* **97** (12): 2009-2019. DOI: 10.1016/j.agwat.2010.02.011.

An advanced process for evaluating a linear dielectric constant–bulk electrical conductivity model using a capacitance sensor in field conditions

Basem Aljoumani[1], Jose A. Sànchez-Espigares[2], Nuria Cañameras[1] and Ramon Josa[1]

[1]*Department of Agri-food Engineering and Biotechnology, Universitat Politècnica de Catalunya (UPC), Castelldefels, Spain*
baljoumani@gmail.com

[2]*Department of Statistical and Operational Research, Universitat Politècnica de Catalunya (UPC), Barcelona, Spain*

Editor D. Koutsoyiannis; **Associate editor** M.D. Fidelibus

Abstract The Hilhorst model was used to convert bulk electrical conductivity (σ_b) to pore water electrical conductivity (σ_p) under laboratory conditions by using the linear relationship between the soil dielectric constant (ε_b) and σ_b. In the present study, applying the linear relationship ε_b–σ_b to data obtained from field capacitance sensors resulted in strong positive autocorrelations between the residuals of that regression. We were able to derive an accurate offset of the relationship ε_b–σ_b and to estimate the evolution of σ_p over time by including a stochastic component to the linear model, rearranging it to a time-varying dynamic linear model (DLM), and using Kalman filtering and smoothing. The offset proved to vary for each depth in the same soil profile. A reason for this might be the changes in soil temperature along the soil profile.

Key words capacitance sensor; soil dielectric; time-varying dynamic linear modelling (DLM); Kalman filtering; offset; pore water electrical conductivity

Evaluation de la relation linéaire entre la constante diélectrique et la conductivité électrique apparente utilisant un capteur de capacitance dans des conditions de terrain
Résumé Le modèle de Hilhorst a été utilise afin de convertir la conductivité électrique apparente (σ_b) en conductivité électrique de l'eau porale (σ_p) au laboratoire en utilisant la relation linéaire entre la constante diélectrique du sol (ε_b) et σ_b. Dans cette étude, l'application de la relation linéaire ε_b–σ_b aux données fournies par les capteurs de capacitance de terrain aboutit à des autocorrélations fortement positives entre les résidus de cette régression. Nous avons pu trouver une compensation précise de la relation ε_b–σ_b et estimer l'évolution temporelle de σ_p en incluant une composante stochastique au modèle linéaire, le réarrangeant en un modèle dynamique linéaire transitoire (MDL), et en utilisant un filtre de Kalman et un lissage. La compensation varie selon la profondeur dans un même profil de sol, ce qui pourrait s'expliquer par les changements de température du sol dans le profil.

Mots clefs capteur de capacitance ; diélectrique du sol ; modèle dynamique linéaire transitoire (MDL) ; filtre de Kalman ; compensation ; conductivité électrique de l'eau porale

1 INTRODUCTION

Salinization is a significant cause of land degradation and nutrient deficiency. Munns (2002) showed that, if excessive amounts of salt enter a plant, salt will eventually rise to toxic levels in the older transpiring leaves, causing premature senescence, and reduce the photosynthetic leaf area of the plant to a level that cannot sustain growth. Thiruchelvam and Pathmarajah (1999), who studied the salinity problems in Sri Lanka's Mahaweli River

System "H" Irrigation Project, showed that, if left uncorrected, salinity can lead to the following agricultural problems: (a) reduced crop intensity; (b) decreased profitability; and (c) land scarcity. Salinity is most commonly measured with an electrical conductivity (EC) meter that estimates the concentration of soluble salts in soil slurry or a water solution by how well an electrical current passes through the medium. The ability of a solution to conduct electricity increases with a higher

7. An advanced process for evaluating a linear dielectric constant–bulk electrical conductivity model using a capacitance sensor in field conditions

Basem Aljoumani[1], José A. Sanchez-Espigares[2], Nuria Cañameras[1], Ramon Josa[1]

[1]Department of Agri-food Engineering and Biotechnology, UPC-BarcelonaTech, Barcelona, Spain
[2]Department of Statistical and Operational Research, UPC-BarcelonaTech, Barcelona, Spain

ABSTRACT

The Hilhorst (2000) model was able to convert σ_b to pore water electrical conductivity (σ_p) under laboratory conditions by using the linear relationship between the soil dielectric constant (ε_b) and bulk electrical conductivity (σ_b). In the present study, applying the linear relationship ε_b- σ_b to data obtained from field capacitance sensors resulted in strong positive autocorrelations between the residuals of that regression. We were able to derive an accurate offset of the relationship ε_b-σ_b and to estimate the evolution of σ_p over time by including a stochastic component to the linear model, rearranging it to a Time-varying Dynamic Linear Model (DLM), and using Kalman filtering and smoothing. The offset proved to vary for each depth in the same soil profile. A reason for this might be the changes in soil temperature along the soil profile.

KEYWORDS: capacitance sensor; soil dielectric; time-varying Linear Dynamic Model (LDM); Kalman filtering; offset. Pore water EC

7.1.   Introduction

Salinization is a significant cause of land degradation and nutrient deficiency. Munns (2002) showed that if excessive amounts of salt enter a plant, salt will eventually rise to toxic levels in the older transpiring leaves, causing premature senescence, and reduce the photosynthetic leaf area of the plant to a level that cannot sustain growth. Thiruchelvam and Pathmarajah (1999), who studied the salinity problems in Sri Lanka's Mahaweli River System "H" Irrigation Project, showed that salinity can lead to the following agricultural problems if left uncorrected: a) reduced crop intensity; b) decreased profitability and; c) land scarcity. Salinity is most commonly measured with an electrical conductivity (EC) meter that

estimates the concentration of soluble salts in soil slurry or a water solution by how well an electrical current passes through the medium. The ability of a solution to conduct electricity increases with a higher salt content; therefore, a high EC value corresponds to high amounts of soluble salts, and vice versa.

Determining the electrical conductivity of soil pore water (σ_p) requires extraction of the pore water from the soil by suction or measurement of the saturated paste conductivity, two conventional methods that are labour- intensive. Still, it is not certain that all ions are collected in the extracted sample (Hilhorst, 2000). Many other studies on soil salinity assessment concluded that it is important to assess soil salinity temporally and spatially in order to correctly evaluate its evolution and reasonably predict its values (Hajrasuliha et al. 1980, Mahmut et al. 2003, Rhoades et al. 1997, Shouse et al. 2010, Xiaoming et al. 2012). New devices have been developed, such as time-domain reflectometry (TDR) and frequency-domain reflectometry (FDR), which are able to measure bulk electrical conductivity (σ_b) temporally and spatially before converting σ_b to σ_p by using methods, models and estimates such as those described by Rhoades et al. (1990), Mualem and Friedman (1991) and Hilhorst (2000). The weakness of these methods is that they are applied in laboratory conditions with homogenous soil samples.

A soil water system's bulk electrical conductivity (σ_b) is ascertained by measuring three conductance pathways in the system. They are: (1) solid–liquid interphase; (2) solid phase; and (3) liquid phase. In agricultural application, it is advantageous to know the electrical conductivity of the liquid phase (σ_p) contained in the soil pores, which is a good indicator of the solute concentration in the soil. The σ_b of the soil depends on both the σ_p and water content (θ) (Persson, 2002). Thus, the σ_p can only be predicted if θ is constant, or if the relationship between σ_p, σ_b, and θ is determined. Several different models of the σ_p-σ_b-θ relationship have been developed (Rhoades et al. 1976, Mualem and Friedman 1991, Malicki and Walczak 1999). Malicki et al. (1994) discovered a high linear correlation between the dielectric constant (ε_b) and σ_b values by using time domain reflectometry for most soil types. Hilhorst (2000) took advantage of this relationship and was able to convert σ_b to σ_p by using a theoretical model describing a linear relationship between σ_b and ε_b.

Due to soil profile heterogeneity, some experimenters have found it more desirable to use stochastic models rather than constant values in describing the future evolution of soil water and soil solutes, where the parameters of stochastic transport models are treated as random variables with discrete values assigned according to a given probability distribution (Beven and Germann 1982, White 1985, Van Genuchten 1991). A stochastic process amounts to a sequence of random variables known as a time series. The time series method has been applied in several agricultural and hydrologic studies. During the last decade, there has been a significant increase in stochastic models for agronomic applications. Zou et al. (2010) worked on silt loam soil profile data, collected monthly from 2001 to 2006, to compare two mathematical models: the back propagation neural network (BPNN) model and the

autoregressive integrated moving average (ARIMA) model. The objective was to predict both the average water content in the top 1-meter profile by measuring water content at a 0.60 m depth, and also by measuring the average salt content at various depths of the soil profile (0.10, 0.20 and 0.45 m). Hoeben and Troch (2000) applied the Kalman filter to predict the soil moisture profile by using satellite data in a theoretical case. Their use of a Kalman filter improved the prediction of soil moisture profiles from surface measurements. Wendroth et al. (1999) described the temporal process of soil water content in different soil layers of a sandy loam soil using a simple transport equation in combination with the state-space theorem. The so-called Kalman filter (Kalman 1960, Katul et al. 1993, Parlange et al. 1993, Wendroth et al. 1993, Nielsen et al. 1994) is a stochastic term that weighs observations and model predictions via the Kalman gain (Gelb 1974) and it provides the variance-time behavior of predictions.

In this study, we will utilize capacitance soil sensors 5TE (Decagon Devices, Inc., Pullman, WA) to measure σ_b in field conditions. These types of sensors are commercially available and they use the Hilhorst (2000) model to convert σ_b to σ_p. Subsequently, we will show the weakness of applying the deterministic Hilhorst (2000) model to heterogeneous soil conditions, and then we will develop that model into a stochastic model capable of precisely estimating the σ_p in field conditions. In other words, the Hilhorst model used a deterministic linear relationship between σ_b and ε_b to estimate σ_p, while we will apply that relationship to field conditions by converting it to a stochastic model for precisely estimating σ_p and taking into account soil heterogeneity.

7.2. *Materials and methods*

7.2.1. The Linear σ_p-ϑ-σ_b Model:

According to Hilhorst (2000), σ_p can be determined from the equation

$$\sigma_p = \frac{\varepsilon_p \sigma_b}{\varepsilon_b - \varepsilon_{\sigma_b=0}} \qquad (1)$$

where σ_p is the pore water electrical conductivity (dS m^{-1}); ε_p is the real portion of the dielectric permittivity of the soil pore water (unitless); σ_b is the bulk electrical conductivity (dS m^{-1}); ε_b is the real portion of the dielectric permittivity of the bulk soil (unitless); $\varepsilon_{\sigma b=0}$ is the real portion of the dielectric permittivity of the soil when bulk electrical conductivity is 0 (unitless). However, $\varepsilon_{\sigma b=0}$ appears as an offset of the linear relationship between ε_b and σ_b. Hilhorst (2000) found that $\varepsilon_{\sigma b=0}$ depended on the soil type and varied between 1.9 and 7.6 in the soils used in his study. So he recommended 4.1 as a generic offset.

Many studies applied the Hilhorst (2000) model in their experiments to convert σ_b into σ_p. Persson (2002) applied it in time domain reflectometry (TDR) measurements for laboratory experiments using soil

columns with different θ and σ_p. By rearranging Eq. 1, the slope can be calculated theoretically

$$\varepsilon_b = (\varepsilon_p / \sigma_p)\sigma_b + \varepsilon_{\sigma_b=0}$$, i.e., slope = ε_p / σ_p . The value of the offset $\varepsilon_{\sigma b=0}$ was obtained as a fitting

parameter when the slope was fixed, assuming that ε_p equals the dielectric constant of free water at the specific temperature. He concluded his work by using different offsets (within the range of 3.67 to 6.38) according to the soil type. Moreover, the manufactured capacitance soil moisture sensors 5TE (Decagon Devices, Inc., Pullman, WA) also use the Hilhorst (2000) model to convert σ_b into σ_p. They recommend using offset $\varepsilon_{\sigma b=0}$ = 6 for all agricultural soils. Arquedas-Rodriguez (2009) used 5TE sensors in his study and found that offset $\varepsilon_{\sigma b=0}$ = 6 did not represent a very good the linear relationship between ε_b and σ_b. The WET sensor (Delta-T Device Ltd, Cambridge, UK) is a frequency domain dielectric sensor and designed for use with the standard offset $\varepsilon_{\sigma b=0}$ = 4.1 of the Hilhorst (2000) model. Bouksila et al. (2008) worked with a saline gypsiferous soil and found that the accuracy of the WET sensor in predicting σ_p was very poor when using the standard value of $\varepsilon_{\sigma b=0}$ =4.1. Despite the importance of computing σ_p from σ_b, it still has not been worked out very well (G. Campbell, Decagon Devices, personal communication, 2010). Moreover, we did not find studies used sensor techniques to estimate σ_p from measuring σ_b by using stochastic models in field conditions.

7.2.2. Field observations

Hourly field observations of bulk electrical conductivity (σ_b), soil dielectric (ε_b) and soil water dielectric (ε_p), as well as distributions of soil temperature (t) and soil water content (θ) were carried out for 55 days during the vegetative stage of lettuce (Lactuca sativa), starting on 23 April 2010 in the Agricultural Park of Baix Llobregat, 5 km south of Barcelona, Spain. A typical Mediterranean climate prevails in the region. Average annual rainfall is 629 mm. Evaporation exceeds rainfall throughout the year except during the rainy months. The dominant soil texture is silty loam. The field was irrigated by a furrow system in an experimental area of 275 m2 (55 m x 5 m). Four irrigation events were applied, with each irrigation dose at nearly 26 L m-2, and an application time ranging between 20-26 minutes. The site had a fairly uniform, bulk density ranging between 1.4 and 1.5 g cm-3 to a depth of 0.75 m, and the water table was 4 m below the soil surface. In the test furrow, we used capacitance soil moisture sensors (5TE, Decagon Devices, Inc., Pullman, WA) to measure the bulk electrical conductivity (σ_b), soil dielectric (ε_b) and soil water dielectric (ε_p) as well as the distributions of soil temperature (t) and soil water content (θ) in the soil profile. The installation depths were at 0.10, 0.20, 0.35, 0.50 and 0.60 m from the soil surface. The study focused on the root zone.

A total of 1318 observations were made of the bulk electrical conductivity (σ_b), soil water dielectric (ε_p) and soil dielectric (ε_b) for each depth, and these were used to estimate the offset $\varepsilon_{\sigma b=0}$ and evolution

of pore water EC (σ_p) at its corresponding depth, of which 659 observations were used to validate its forecast.

Bulk electrical conductivity (σ_b) responded at the interested depths –i.e., within the root zone of the test site, in the top of 0.60 m soil profile-- during the first hour of the irrigation event. Therefore, hourly capacitance sensor outputs (σ_b, soil water dielectric (ε_p) and soil dielectric (ε_b)) were measured in order to achieve the study objective. The observational experiment in this study is similar to the work done by Wu et al. (1997). They installed TDR sensors at different depths within the root zone and measured soil water content hourly for 55 days to predict soil water content at the interested depth by measuring the shallower depths. In our observational experiment, we use time series analysis of capacitance sensor data (σ_b, t, ε_p, ε_b, and θ) to apply mathematical models that find a certain relationship between σ_b and ε_b for estimating σ_p and offsetting $\varepsilon_{\sigma b=0}$ in field conditions. Because meteorological and precipitation data implicitly affect the values of sensor observations, we did not include them in our model.

7.2.3. Kalman Filter

The Kalman filter is a set of mathematical equations that provide an efficient computational (recursive) means for estimating the state of a process in a way that minimizes the mean of the squared error. The filter is very powerful in several aspects: it supports estimations of past, present, and even future states; and it can do so even when the precise nature of the modeled system is unknown. The purpose of the Kalman filter is to provide an estimate of the unobservable state vector based on model information and measurement information, balancing out the errors of both. It is a sequential algorithm for minimising the state error variance.

In this study, we will use capacitance soil sensor data (σ_b, ε_p, ε_b) and a Kalman filter statistical estimation technique that we apply on the Hilhorst (2000) model to estimate its unknown variables (σ_p, and $\varepsilon\sigma_{b=0}$). A Kalman filter soil state model is used to merge available soil physics data (σ_p, and $\varepsilon_{\sigma b=0}$) with data from capacitance sensors (ε_b). In order to obtain suitable σ_p and $\varepsilon_{\sigma b=0}$ for the study area, the model makes continuous estimates of σ_p, and $\varepsilon_{\sigma b=0}$, and it weights ε_b observations according to input and model-propagated error covariances.

The state-space model has three parts: unobservable data (σ_p and $\varepsilon_{\sigma b=0}$), observations (ε_b), and a Kalman filter that updates the unobservable data by assimilating observations into the dynamic soil state estimate. In this study, when observations of ε_b are available (observed by capacitance sensor), the Kalman filter uses the propagated state estimate (σ_p and $\varepsilon_{\sigma b=0}$) and a record of the propagation steps to adjust the state in a way that is proportional to the difference between the observed and the predicted value. The ratio of proportionality (the Kalman gain) is calculated from a propagated model state error

covariance matrix (for σ$_p$ and ε$_{σb=0}$ and an estimate of ε$_b$ measurement error). Together, these models produce continuous estimates of σ$_p$ and ε$_{σb=0}$ states, as well as their error covariances.

7.2.4. Time-varying dynamic linear model

Being linear and Gaussian, the Dynamic Linear Model (DLM) is presented as a special case of a general state space model. For dynamic linear models, estimation and forecasting can be recursively obtained by the well-known Kalman filter. Estimating unknown parameters in a DLM requires numerical techniques, but the Kalman filter can be used in this case as a building block for evaluating the likelihood function or for simulating the unobservable states.

The R (R Development Core Team 2010) package dlm (Petris 2010) provides an integrated environment for Bayesian inference using DLM, and the package includes functions for Kalman filtering and smoothing, as well as for maximum likelihood estimation.

7.2.5. Model identification

A time-varying DLM can be modelled as

Observation equation $\qquad y_t = \alpha_t + x_t\beta_t + v_t \qquad v_t \sim \mathcal{N}(0, V_t) \qquad$ (2)

Unobservable State equation $\qquad \begin{aligned} \alpha_t &= \alpha_{t-1} + w_{\alpha,t} \\ \beta_t &= \beta_{t-1} + w_{\beta,t} \end{aligned} \qquad \begin{aligned} w_{d,t} &\sim \mathcal{N}(0, w_{\alpha,t}) \quad (3) \\ w_{\beta,t} &\sim \mathcal{N}(0, w_{\beta,t}) \end{aligned}$

Here y_t is an m-dimensional vector, representing the observation at time t; in our study it represents ε$_b$ observations. x_t is an m.m-dimensional matrix of covariates. While α_t and β_t are unobservable m-dimensional vectors presenting the state of the system at time t, in our study they represent ε$_{σb=0}$ and σ$_p$, respectively. v_t , $w_{\alpha,t}$ and $w_{\beta,t}$ are the Gaussian white-noise errors. The only parameters of the model are the observations and evolution variances V_t, $w_{\alpha,t}$ and $w_{\beta,t}$. These are usually estimated from available data using maximum likelihood or Bayesian techniques.

7.2.6. Seasonality

When the model has a seasonal component, it is usual to include a DLM to describe this component. In the state-space expression, the seasonal component may have a stochastic error that allows changes for the seasonal pattern over time.

So Eq. 2 may have a seasonal component (s_t) and may be written as:

$$y_t = \alpha_t + x_t\beta_t + S_t + v_t \qquad (4)$$

Where yt represents the dielectric observation (ε_b); αt and βt represent the offset ($\varepsilon_{\sigma b=0}$) and the pore water (σ_p), respectively; and S_t represents the seasonal component.

7.3.   Results and discussion

7.3.1. Soil characterization

Table 1 shows the soil characterization of the study field beneath the furrow and ridge at various depths. It shows that the soil particles for clay, silt and sand have few variations in the root zone. The organic matter in the study field is representative of the area.

Table 7.1 Soil characterization.

Depth (m)	Clay (<0.002 mm) %	Silt (0.05–0.002 mm) %	Sand (2–0.05 mm) %
0.10	16.60	54.89	28.51
0.35	13.12	55.25	31.63
0.65	22.30	61.16	16.54

Figure 1 and figure 2 show the importance of hourly bulk electrical conductivity measurements. Figure 1 shows the soil water content at five depths over time (measured by capacitance sensor). Fluctuations in soil water content at deeper layers correspond to changing water content in the upper layers; this fluctuation diminishes as the layer becomes deeper. Irrigation events had significant effects on soil water content fluctuations when applied at 4.29, 27.20, 32.04 and 46.33 days, with precipitation occurring at 9.33, 20.50 and 52.54 days. Figure 2 shows the significant effect of soil water on bulk electrical conductivity at the time of irrigation, where the depths responded to the first hour of the irrigation event.

7.3.1. Deterministic model

We derive the offset using the method of Persson (2002), in which the Hilhorst (2000) model is rearranged as follows:

$$\varepsilon_b = (\varepsilon_p / \sigma_p)\sigma_b + \varepsilon_{\sigma_b=0}$$

(5)

After using hourly field measurements of ε_b and σ_b (1318 observation of each in Field 1), we can see in Table 2 the coefficients of the linear relationship between ε_b -σ_b. The offset of this relationship is 4.97 and the slope is $1/\sigma_p$= 0.33, so σ_p=5 dSm^{-1} is the average for all observations. After applying the Durbin–Watson test to see if there is an autocorrelation between the residuals of that regression, we can see in Table 3 that there is an extremely strong and positive autocorrelation, which indicates that the

result of that regression is not valid. Moreover, the linear model does not take into account the evolution of the unobservable variable over time. For this reason, it is reasonable to think that σ_p evolves with a stochastic component.
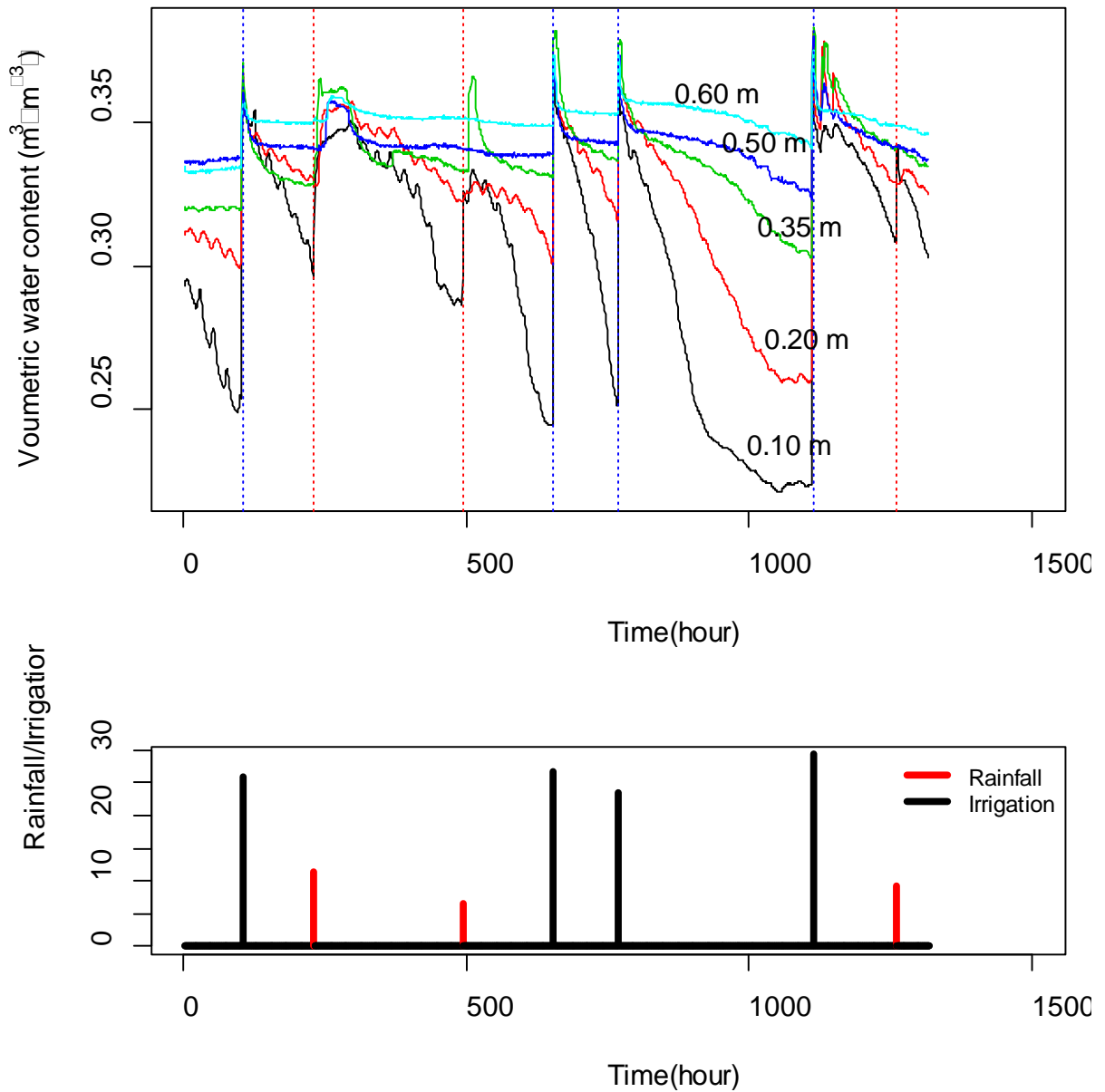
Figure 7.1 Measured soil volumetric water content, rainfall and irrigation at various depths
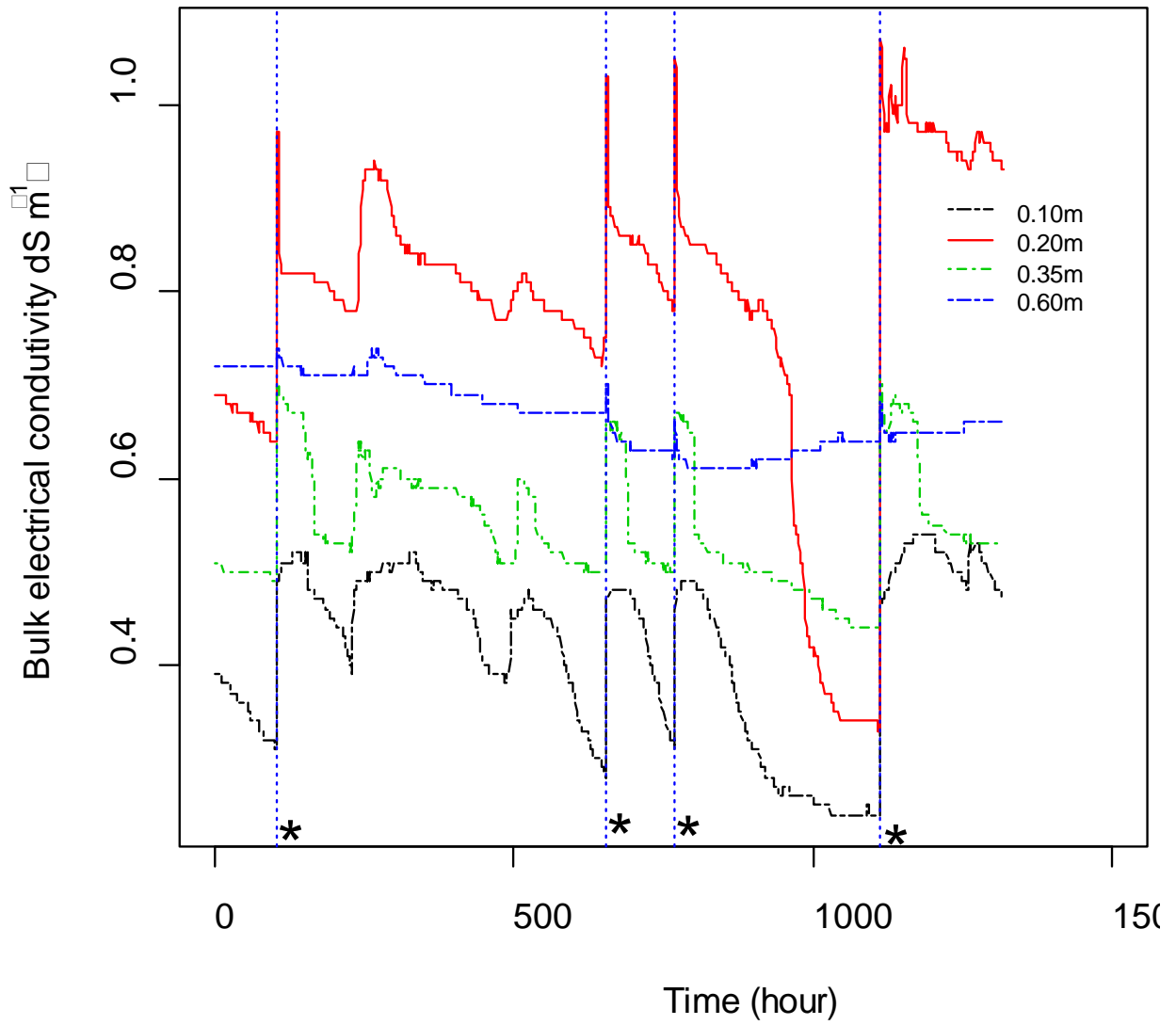
*Figure 7.2 Bulk electrical conductivity (σ_b) at four depths versus time. * indicates the irrigation time.*

Table 7.2 Estimated parameters from the linear regression ε_b - σ_b ,

	Estimate	Std. Error	t	Pr($>|t|$)
$\varepsilon_{\sigma b=0}$	4.978923	0.088208	56.45	<2 e-16 ***
1/ σ_p	0.354256	0.002546	139.15	<2 e-16 ***

Significance: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Table 7.3 Durbin–Watson test for linear regression ε_b - σ_b

lag	Autocorrelation	D-W Statistic	p-value
1	0.9524539	0.09079999	0

The known parameters for the Hilhorst (2000) model are: ε_b, σ_b and ε_p. They are simultaneously and hourly measured by capacitance sensors, while σ_b is obtained directly from the data logger. For ε_b and ε_p, we applied the equations provided by the manufacturer (Decagon Devices 2008) as follows:

$$\varepsilon_b = \frac{\varepsilon_{raw}}{50} \tag{6}$$

$$\varepsilon_p = 80.3 - 0.37(T_{soil} - 20) \tag{7}$$

where ε_{raw} represents the raw data output values from the datalogger (raw soil water content counts), and T_{soil} is the soil temperature measured by the sensor directly.

Figure 3 shows the evolution of the soil dielectric constant (ε_b), the water dielectric constant ε_p and the soil bulk electrical conductivity (σ_b) at 0.10 m depth. It also shows that the irrigation events have a significant effect on σ_b and ε_b.
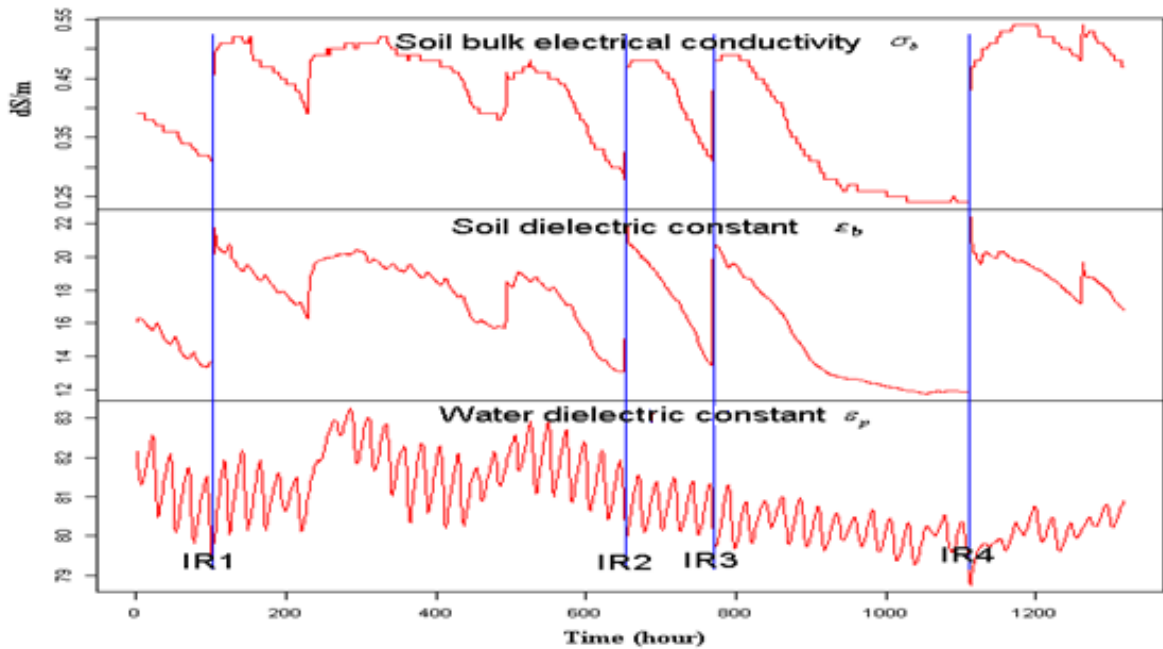
Figure 7.3 Known variables for the Hilhorst model (σ_b, ε_b and ε_p); IR1, IR2, IR3 and IR4 are the times of the irrigation events (0.10m depth).

7.3.2. Time-varying Linear Dynamic Model (LDM)

The deterministic equation 5 can be modified into the time-varying DLM for observation and unobservable (state) models. In this case, the observation data is the soil dielectric constant ε_b observed by the capacitance sensor, while the unobservable data are offset ($\varepsilon_{\sigma b=0}$) and pore water electrical conductivity (σ_p). Equation 5 can be modified to the time-varying DLM as follows:

- The observation equation can be obtained by modifying the Hilhorst (2000) model (written in equation 5) into a stochastic equation, in accordance with equation 4. This is achieved by adding the seasonal component for each 24 h (s_t) and the Gaussian white-noise errors (v_t):

$$(\varepsilon_b)_t = \left(\varepsilon_{\sigma_b=0}\right)_t + \left(\varepsilon_p *\sigma_b\right)_t \left(\frac{1}{\sigma_p}\right)_t + s_t + v_t \qquad v_t \sim \mathcal{N}(0, \sigma_v{}^2) \qquad (8)$$

where ε_b, ε_p and σ_b are observed by capacitance sensors, as mentioned above.

- Unobservable data in Equation 8 are $\varepsilon_{\sigma b=0}$, and the slope $1/\sigma_p$ can be converted to the unobservable state equation of the time-varying DLM according to equation 3. The unobservable state equation can be arranged as follows:

$$\begin{cases} (\varepsilon_{\sigma_b=0})_t = (\varepsilon_{\sigma_b=0})_{t-1} \\ (\dfrac{1}{\sigma_p})_t = (\dfrac{1}{\sigma_p})_{t-1} + w_t \qquad w_t \sim \mathcal{N}(0, (\sigma_w)_t^2) \qquad\qquad (9) \\ s_t + s_{t-1} + s_{t-2} \dots\dots + s_{t-23} = 0 \end{cases}$$

Here we consider $\varepsilon_{\sigma b=0}$ as a constant and its present value is related only to its past value. The slope $1/\sigma_p$ changes over time, and its present value is related to its past value plus the Gaussian white-noise errors (w_t). st is the seasonal component (every 24 h), which changes over time.

In a study by Aljoumani et al. (2012), we used the same experimental data to show that irrigation events have a significant effect on the behaviour of soil water content and should be captured as outliers to improve the fitted model. Figure 1 shows that irrigation events also have a significant effect on the behaviour of ε_b. To capture the time of irrigation event as an outlier, we increase the state variance $(\sigma_w)_t^2$ by a constant factor (k>1) at hours 103, 654, 770, and 1112 h, which are the times of irrigation. This change in the model gives better estimates of how irrigation time affects the state values. Once we estimate the parameters, we apply the Kalman filter to get the offset $\varepsilon_{\sigma b=0}$ and the slope $\dfrac{1}{\sigma_p}$.

For the Kalman filter discussed in this paper, the state equation (equation 3 and its application in this study are included in equation 9) consists of two terms: (1) the constant offset ($\varepsilon_{\sigma b=0}$) of the Hilhorst (2000) model, with its present value related to its past value; and (2) the pore water EC (σ_p), which is a stochastic difference equation that incorporates a random component for noise in the system. The measurement equations (equations 2, 4 and their application in this study are included in equation 8) are defined so that they can handle indirect measurements, gaps in the sequence of measurements, and measurement errors. In this study, the measurements correspond to soil dielectric (ε_b), bulk electrical conductivity (σ_b), water dielectric (ε_p) and a random component for noise in the system. The Kalman filter operates recursively to predict forward the state of the system one step at a time, based on the previously

predicted state and the next measurement. Its predictions are optimal in the sense that they have minimum variance among all unbiased predictors. In this respect, the filter is similar to kriging. Moreover, in this study, we add a seasonal component (every 24 h) to the stochastic equations for state and measurement.

The equations can also be applied in reverse order to estimate the state variable at all time points from a complete series of measurements, including past, present and future measurements. We applied this in our study to estimate the state variables ($\varepsilon_{\sigma b=0}$ and σ_p) at all time points from a complete series of the soil dielectric constant (ε_b). This process is known as smoothing.

7.3.3. DLM validation

After using equation 8, we can see in Figure 4 the observed and predicted time series of the soil dielectric constant ε_b at 0.10 m depth. The predicted and observed values agreed reasonably after 1318 observations.
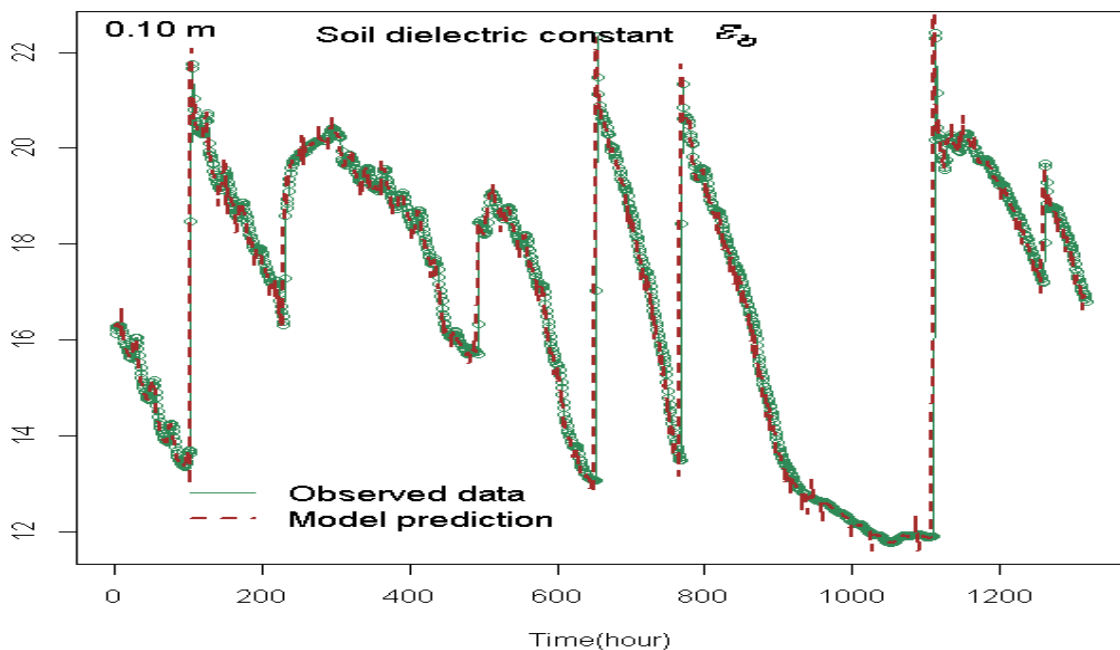
Figure 7.4 Observed and predicted data of soil dielectric constant at 0.10 m depth (furrow, lettuce).

The performance of all models for each depth was tested by two tools, by using the mean absolute prediction error (MAPE) and by estimating the coefficient of determination R2 between the predicted versus the observed values. The mean absolute prediction error (MAPE) of the model for the time series never exceeded 0.02. The predicted values, for both the fit and forecast periods, were compared to the observed ones and the estimated adjusted R2 varied between 0.96 and 0.99 (Fig. 5).
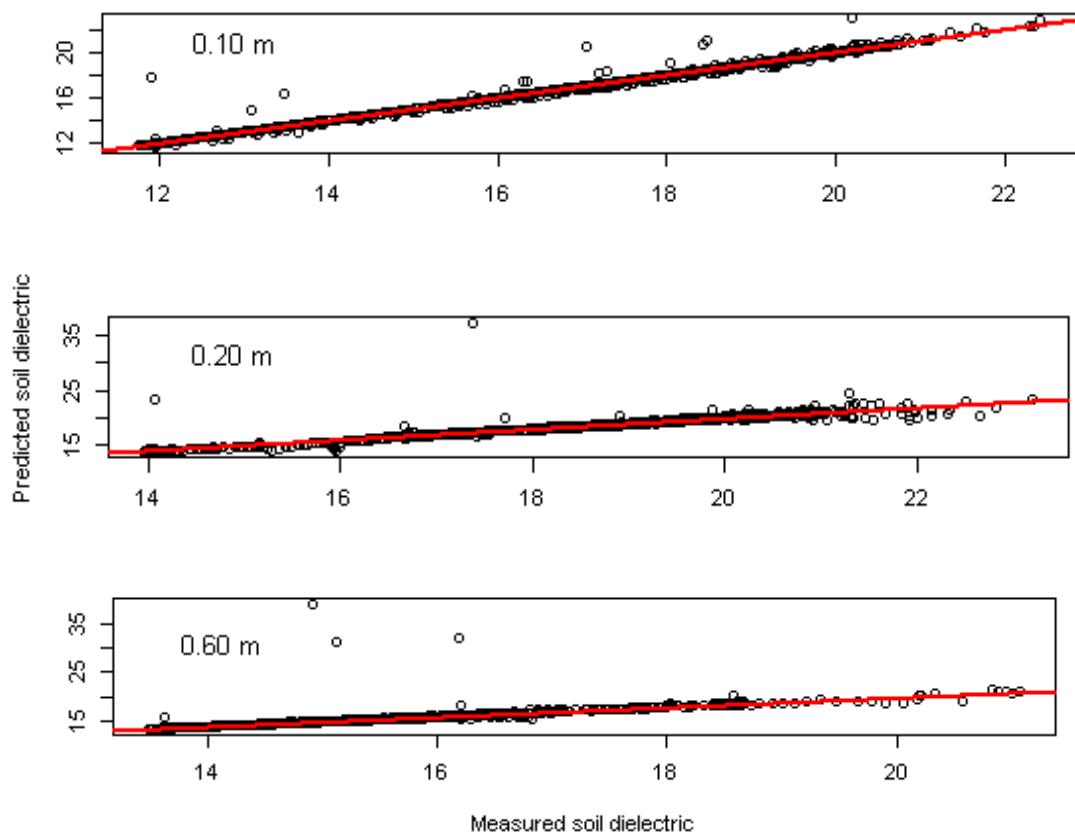
Figure 7.5 Measured versus Predicted soil dielectric constant

Since prediction of the soil dielectric constant ε_b is valid, then the estimation of electrical conductivity is also valid for the soil pore water σ_p; and offset $\varepsilon_{\sigma b=0}$ (equation 9) is also valid because it is used in predicting the dielectric constant ε_b. Figure 6 shows the electrical conductivity values of the soil pore water σ_p, and offset $\varepsilon_{\sigma b=0}$ after applying the time-varying DLM to the data at 0.10 m depth. At this depth, the offset is 3.8 and σ_p varies over time. The figure shows a clear decrease in σ_p at the time of irrigation, which is expected since irrigation leaches the salts downward.

7.3.1. Field estimation of $\varepsilon_{\sigma b=0}$ and σ_p

By applying the time-varying DLM to the observed data at the various depths (field conditions), the values of $\varepsilon_{\sigma b=0}$ were within the range of 3.8 to 8.5 needed to estimate the evolution of σ_p over time at each depth. Figure 7 shows the values of σ_p and $\varepsilon_{\sigma b=0}$ for 0.20 m and 0.60 m depths, respectively. The following questions arise. Why do these differences exist between the offset values at different depths? Are these differences statistically significant? While investigating these questions, many studies found that calibration measurements of electromagnetic induction for predicting σ_b are affected by soil texture, water content, and soil temperature (McKenzie et al. 1989, Slavich and Petterson 1993). Yuanshi et al. (2003) showed that ε_b changes when soil compaction and temperature vary.
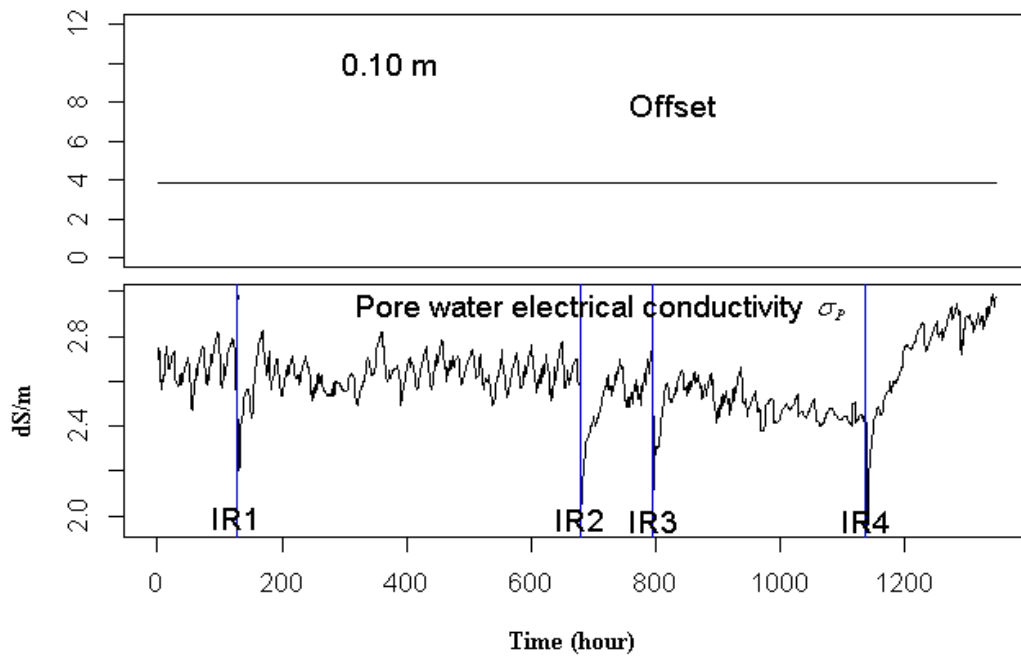
Figure 7.6 Estimation of the unobservable data ($\varepsilon_{\sigma b=0}$ and σ_P) by applying the Time-varying DLM to data (lettuce, furrow, 0.10 m depth).

In this study, the value of the $\varepsilon_{\sigma b=0}$ was derived from the ε_b observations. Since temperature affects ε_b, we can consider the null hypothesis, which states: the soil temperature has no effect on the $\varepsilon_{\sigma b=0}$ value. The alternative hypothesis states: the soil temperature has an effect on the $\varepsilon_{\sigma b=0}$ values.

For this analysis, we took 30 measurements of soil temperature three days after one irrigation event. All data were subjected to analysis of variance (ANOVA) procedures using R (R Development Core Team 2010). Table 4 shows that the univariate ANOVA produced statistically significant results. Since we estimate a constant offset at shallow depths, it is difficult to determine the effect of temperature on the offset value, due to the high variability of the temperature at those depths. But in the deeper depths, the soil temperature affected the offset values, and we can see that the offset increases when the temperature increases.

Table 7.4 Effect of the mean soil temperature (ºC) on the offset at various depths.

Mean soil temperature (ºC)	Depth				
	0.10 m	0.20 m	0.35 m	0.50 m	0.60 m
18.14	3.8				
16.25					5.8
16.94				7.1	
18.04		7.8			
17.36			8.2		
significance	*	*	*	*	*

Significance: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.
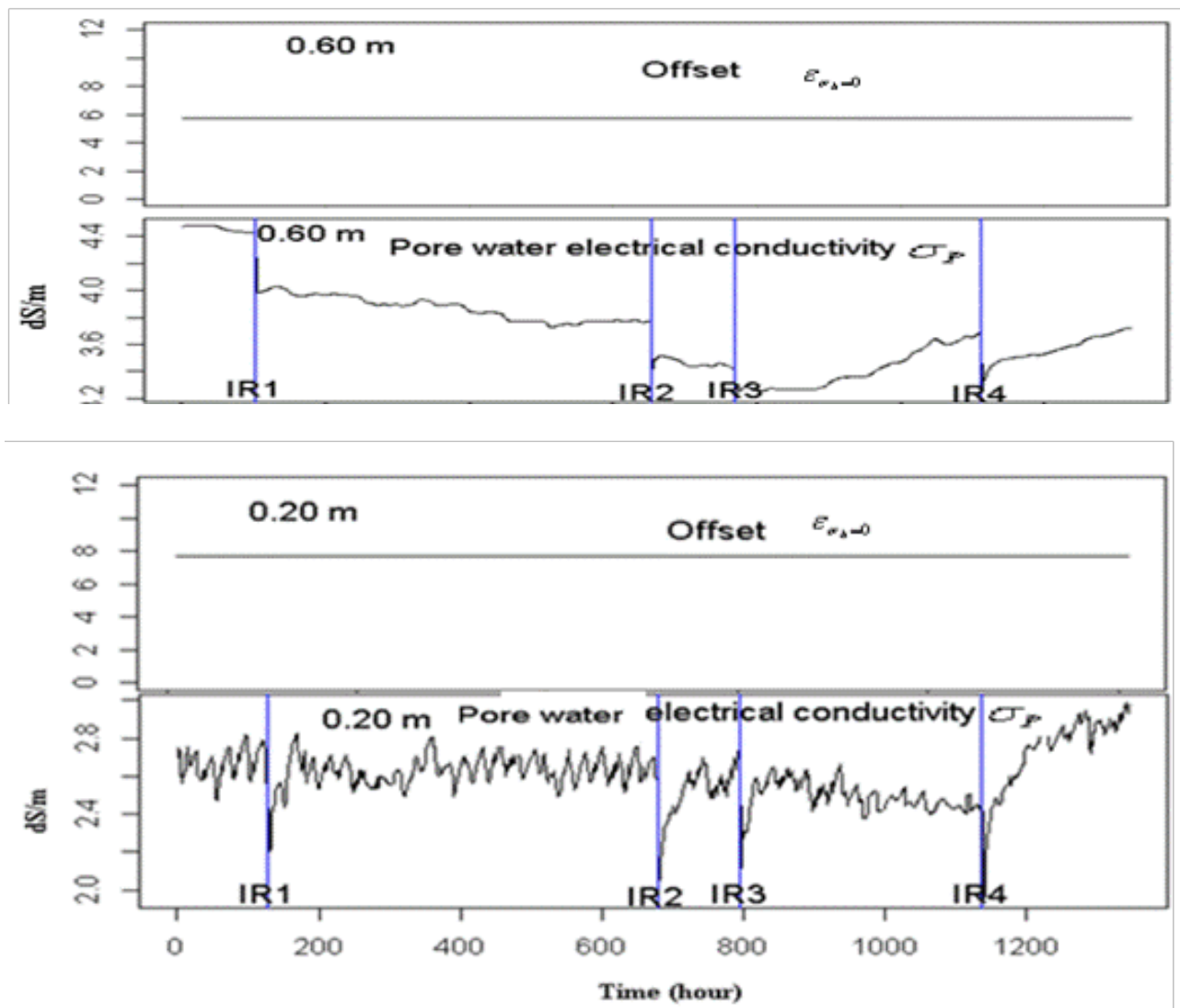
Figure 7.7 Estimation of the unobservable data ($\varepsilon_{ob}=0$ and σ_p) by applying the Time-varying DLM to data (lettuce, furrow, 0.20 m and 0.60 m depth).

In their empirical σ_b-σ_p- $\varepsilon_{\sigma_b=0}$ model, Malicki et al. (1994) and Malicki and Walczak (1999) included sand content in % by weight. Table 1 shows that there is different sand content (%) at each depth in Field 1, but this study could not conclude that the sand content has an effect on the value of $\varepsilon_{ob=0}$ because more data would be required to statistically confirm this effect. Moreover, due to the fact that most soils are heterogeneous, this could indicate a need to adapt an offset for each depth.

7.4.   Conclusions

Several models have been studied to assess σ_p from the ε_b -σ_b relationship (Rhoades et al. 1976, Mualem and Friedman1991, Malicki and Walczak 1999). Recently, Hilhorst (2000) presented a theoretical

model describing a linear relationship between σ_b and ε_b in moist soil. By using this linear relationship, Hilhorst (2000) found that σ_p can be measured in a wide range of soil types without soil-specific calibrations.

In this present study, we applied the ε_b-σ_b linear relationship to the field condition data obtained from capacitance sensors and found an extremely strong positive autocorrelation between the residuals of that regression. By including a stochastic component to the linear model and rearranging it to a time-varying Dynamic Linear Model (DLM), combined with Kalman filtering and smoothing, we were able to derive an accurate offset of the relationship ε_b-σ_b and to estimate the evolution of σ_p over time. The offset was shown to vary for each depth in the same soil profile. A reason for this might be changes in soil temperature throughout the soil profile.

In our work, we have shown the limitations of using one generic offset in the Hilhorst (2000) model when applying it to field conditions. We confirmed our hypothesis by taking just one soil profile in the same soil type. However, we recommend additional measurements in different soil types, taking into account the spatial variability to validate the model which we have developed. In our work, we studied in more detail the relationship of ε_b-σ_b to estimate σ_p. Moreover, to enhance our findings, it could be useful to compare the estimated σ_p (obtained by our stochastic form of the Hilhorst (2000) model) with other independent methods, such as measurements of soil solution EC (EC-sol) σ_p collected in ceramic suction cups (SCs) installed at each depth near the sensor. Finally, future studies should consider how to make these sensors and their data more accessible for practical use. Currently, these types of capacitance sensors are commercially available, but they are limited almost exclusively to scientific experiments because of the process involved: the technician must go to the field with a computer to transfer the sensor data from the datalogger to an Excel sheet, and then apply mathematics to study the relationship between the variables. Alternatively, our stochastic models could be programmed in collaboration with telecommunications technicians, and an electronic unit could be designed for inclusion in the sensor datalogger in order to provide a visual indication of the electrical conductivity of the soil solution. With a device such as this, these types of low-cost sensors could be expanded so that a normal farmer can ascertain salt levels in the root zone.

7.5.   *References*

1. Aljoumani, B., Sànchez-Espigares, J.A., Cañameras, N., Josa, R. and Monserrat, J., 2012. Time series outlier and intervention analysis: irrigation management influences on soil water content in silty loam soil. Agricultural Water Management, 111, 105-114,doi: dx.doi.org/10.1016/j.agwat.2012.05.008.

2. Arquedas-Rodriguez, F. R., 2009. Calibrating capacitance sensors to estimate water content, matric potential, and electrical conductivity in soilless substrates. Thesis (PhD) University of Maryland. Available from: http://www.psla.umd.edu/faculty/lea-cox/Dissertations,%20Theses/F.%20R.%20Arguedas- Rodriguez%20MS%20Thesis%202009.pdf [10 February 2012].

3. Beven, K. andGermann, P., 1982. Macropores and water flow in soils. Water Resource Reseearch,18, 1311–1325. Available from: http://onlinelibrary.wiley.com/doi/10.1029/WR018i005p01311/pdf.

4. Bouksila, F., Persson, M., Berndtsson, R. and Bahri, A., 2008. Soil water content and salinity determination using different dielectric methods in saline gypsiferous soil. Hydrological Science Journal, 53, 253-265. doi:10.1623/hysj.53.1.253.

5. Decagon Devices, 2008. 5TE Water content, EC, and temperature sensors operator'smanual. Version No. 3., Pullman, WA.

6. Gelb, A., 1974. Applied Optimal Estimation. Cambridge (MA). Massachusetts Institute of Technology Press.

7. Hajrasuliha, S., Baniabbassi, N., Metthey, J. and Nilsen, D. R., 1980. Spatial variablility of soil sampling for salinity studies in southwest Iran. Irrigation Science 1, 197-208.

8. Hilhorst, M. A., 2000. A pore water conductivity sensor. Soil Science Society of America Journal 64, 1922-1925, doi:10.2136/sssaj2000.6461922x.

9. Hoeben, R., and Troch, P.A., 2000. Assimilation of active microwave observation data for soil moisture profile estimation. Water Resources Research, 36, 2805–2819, doi:10.1029/2000WR900100.

10. Kalman, R.E., 1960. A new approach to linear filtering andprediction problems.Transaction of the ASME-Journal of Basic Engineering, 8, 35-45, doi:10.1115/1.3662552.

11. Katul, G.G., Wendroth, O., Parlange, M.B., Puente, C.E. and Nielsen,D.R., 1993. Estimation of in situ hydraulic conductivity function from nonlinear filtering theory. Water Resources Research, 29, 1063-1070, doi:10.1029/92WR02593.

12. Mahmut, C. and Cevat, K., 2003. Spatial and temporal changes of soil salinity in cotton field irrigated with low-quality water. Journal of Hydrology, 272, 238-249, doi:10.1016/S0022-1694(02)00268-8.

13. Malicki, M. A. and Walczak, R. T., 1999. Fvalualimj soil salinity status from bulk electrical conductivity and permittivity. Science Society of America Journal,   50, 505-514.

14. Malicki, M. A., Walczak, R. T., Koch, S. and Flühler, H., 1994. Determining soil salinity from simultaneous readings of its electrical conductivity and permittivity using TDR. In; Proceedings of the Symposium on Time Domain Reflectometry in Environmental, Infrastructure, and Mining Applications, Evanston, Illinois, Sept 7-9, U.S. Bureau of Mines, Special Publication SP 19-94, NTIS PB95-105789, 328-336.

15. McKenzie, R. C., Chomistek, W. and Clark, N. F., 1989. Conversion of electromagnetic inductance readings to saturated paste extract values in soils for different temperature, texture, and moisture conditions. Canadian Journal of Soil Science, 69, 25– 32. Available from: http://pubs.aic.ca/doi/pdf/10.4141/cjss89-003.

16. Mualem. Y. and Friedman, S. P.,1991. Theoretical prediction of electrical conductivity in saturated and unsaturated soil. Water Resource Research, 27, 2771-2777, doi:10.1029/91WR01095.

17. Munns, R., 2002. Comparative physiology of salt and water stress. Plant, Cell and Environment, 25, 239-250, doi:10.1046/j.0016-8025.2001.00808.x.

18. Nielsen, D.R., Katul, G.G., Wendroth, O., Folegatti, M.V. and Parlange, M.B., 1994. State-space approaches to estimate soil physical properties from field measurements, p. 61-85. Proc. 15th conf. ISSS: Vol. 2a. July 10-16, 1994. Acapulco, Mexico.

19. Parlange, M.B., Katul, G.G., Folegatti, M.V. and Nielsen, D.R., 1993. Evaporation and the field scale soil water diffusivity function. Water Resources Resarch. 29, 1279-1286, doi:10.1029/93WR00094.

20. Persson, M., 2002. Evaluating the linear dielectric constant –electrical conductivity model using time domain reflectometry. Hydrological Science Journal, 47, 269-278, doi:10.1080/02626660209492929.

21. Petris, G., 2010. dlm: Bayesian and Likelihood Analysis of Dynamic Linear Models. R package version 1.1-1, Available from http://CRAN.R-project.org/package=dlm [Accessed 4 April 2012].

22. R Development Core Team, 2010. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.ISBN 3- 900051-07-0. Available from http://www.R-project.org/. [Accessed 15 February 2012].

23. Rhoades, J. D., Shouse, P. J., Alves, W. J., Manteghi, N. A. and Lesch, S. M., 1990. Determining soil salinity from soil electrical conductivity using different models and estimates. Soil Science Society of America Journal, 54, 64-54.

24. Rhoades, J.D., Lesch, S.M., LeMert, R.D. and Alves, W.J., 1997. Assessing irrigation/drainage/salinity management using spatially referenced salinity measurements. Agricultural Water Management, 35, 147–165.

25. Rhoades, J.D., Raats, P.A.C. and Prather, R.J., 1976. Effects of liquid-phase electrical conductivity, water-content, and surface conductivity on bulk soil electrical conductivity. Soil Science Society of America Journal, 40, 651-655.

26. Shouse, P.J., Goldberg, S., Skaggs, T. H., Soppe, R. W. O. and Ayars, J. E., 2010.Changes in spatial and temporal variability of SAR affected by shallow groundwater management of an irrigated field, California. Agricultural Water Management, 97, 673-680, doi:10.1016/j.agwat.2009.12.008.

27. Slavich, P. G. and Peterson, G. H., 1993. Estimating the electrical conductivity of saturated paste extracts from 1:5 soil-water suspensions and texture. Australian Journal of Soil Research, 31, 73-81.

28. Thiruchelvam, S. and Pathmarajah. S., 1999. Economic feasibility of controlling salinity problems in the Mahaweli H area. A research paper presented at the monthly seminar, IWMI (formally IIMI), March, 1999, Colombo, Sri Lank.

29. Van Genuchten, M. TH., 1991. Recent progress in modelling water flow and chemical transport in the unsaturated zone. IAHS Publication, 169-183 Available from: http://ks360352.kimsufi.com/redbooks/a204/iahs_204_0169.pdf

30. Wendroth, O., Katul, G.G., Parlange, M.B., Puente, C.E. and Nielsen, D.R., 1993. A nonlinear filtering approach for determininghydraulic conductivity functions in field soils. Soil Science, 156, 293-301.

31. Wendroth, O., Rogasik, H., Kozinski, S., Ritsema, C. J., Dekker, L. W. and Nielsen, D. R., 1999. State-Space prediction of field –scale soil water content time series in a sandy loam. Soil Tillage Research, 50, 85-93.

32. White, R. E., 1985. The influence of macropores on the transport of dissolved matter through soil. Advances in Soil Sciences, 1, 95-120.

33. Wu, L., Jury, W.A., Chang, A.C. and Allmaras, R.R., 1997. Time series analysis of field-measured water content of a sandy soil. Soil Scicience Society of American Journal, 61, 736–742, doi:10.2136/sssaj1997.03615995006100030005x

34. Xiaoming, L. I., Jingsong, Y.,Meixian, L. I. U., Guangming, L. I. U. and Mei, Y. U., 2012. Spatio-temporal changes of soil salinity in arid areas of South Xinjiang using electromagnetic induction. Journal of Integrative Agriculture, 11, 1365-1376, doi:10.1016/S2095-3119(12)60135-X.

35. Yuanshi, G., Qiaohong, C. and Zongjia, S., 2003. The effects of soil bulk density, clay content and temperature on soil water content measurement using time-domain reflectometry. Hydrological Process, 17, 3601-3614.

36. Zou, P., Yang, J., Fu, J., Liu, G. and Li, D., 2010. Artificial neural network and time series models for predicting soil salt and water content. Agricultural Water Management, 97 (12), 2009-2019, doi:10.1016/j.agwat.2010.02.011.

sensors

MDPI

Article

Estimating Pore Water Electrical Conductivity of Sandy Soil from Time Domain Reflectometry Records Using a Time-Varying Dynamic Linear Model

Basem Aljoumani [1,*] **, Jose A. Sanchez-Espigares** [2] **and Gerd Wessolek** [3]

[1]   Department of Ecology, Ecohydrology and Landscape Evaluation, Technische Universität Berlin Ernst-Reuter Platz 1, 10587 Berlin, Germany

[2]   Department of Statistical and Operational Research, Universitat Politècnica de Catalunya (UPC), Jordi Girona, 31, 08034 Barcelona, Spain; josep.a.sanchez@upc.edu

[3]   Department of Ecology, Technische Universität Berlin Ernst-Reuter Platz 1, 10587 Berlin, Germany; gerd.wessolek@tu-berlin.de

*   Correspondence: basem.aljoumani@tu-berlin.de

check for **updates**

Abstract: Despite the importance of computing soil pore water electrical conductivity (σ_p) from soil bulk electrical conductivity (σ_b) in ecological and hydrological applications, a good method of doing so remains elusive. The Hilhorst concept offers a theoretical model describing a linear relationship between σ_b, and relative dielectric permittivity (ε_b) in moist soil. The reciprocal of pore water electrical conductivity ($1/\sigma_p$) appears as a slope of the Hilhorst model and the ordinary least squares (OLS) of this linear relationship yields a single estimate $\widehat{(1/\sigma_p)}$ of the regression parameter vector (σ_p) for the entire data. This study was carried out on a sandy soil under laboratory conditions. We used a time-varying dynamic linear model (DLM) and the Kalman filter (Kf) to estimate the evolution of σ_p over time. A time series of the relative dielectric permittivity (ε_b) and σ_b of the soil were measured using time domain reflectometry (TDR) at different depths in a soil column to transform the deterministic Hilhorst model into a stochastic model and evaluate the linear relationship between ε_b and σ_b in order to capture deterministic changes to ($1/\sigma_p$). Applying the Hilhorst model, strong positive autocorrelations between the residuals could be found. By using and modifying them to DLM, the observed and modeled data of ε_b obtain a much better match and the estimated evolution of σ_p converged to its true value. Moreover, the offset of this linear relation varies for each soil depth.

Keywords: electrical conductivity; relative dielectric permittivity; time domain reflectometry; kalman filter; dynamic linear model

1. Introduction

Salinization reduces crop productivity, decreases profitability, and causes land scarcity [1]. Thus, it decreases the world's agricultural productivity and causes a global income loss of US$ 12 billion per year [2]. Extracting soil solution by suction or using saturated paste conductivity measurements are the common methods to determine the electrical conductivity of soil pore water (σ_p) as an indicator of the soil salinity; however, they are labour- and cost intensive. There is no evidence that all ions are collected in the sample extract [3]. For soil salinity assessment, it is important to look for practical methods that evaluate the soil salinity state temporally and spatially. These methods help to correctly evaluate soil salinity evolution and reasonably predict its values [4–9]. In recent times, soil electromagnetic sensors have been used to estimate bulk electrical conductivity (σ_b). Then, methods are required to transform σ_b to σ_p [3,6,10].

8. Estimating pore water electrical conductivity of sandy soil from time domain reflectometry records using time-varying dynamic linear model

Basem Aljoumani[1], Jose A. Sanchez-Espigares[2], Gerd Wessolek[3]

[1]*Department of Ecology, Ecohydrology and Landscape Evaluation, Technische Universität Berlin Ernst-Reuter Platz 1, 10587 Berlin, Germany*

[2]*Department of Statistical and Operational Research, Universitat Politècnica de Catalunya (UPC) Spain.*

[3]*Department of Ecology, Technische Universität Berlin Ernst-Reuter Platz 1, 10587 Berlin, Germany*

ABSTRACT

Despite the importance of computing soil pore water electrical conductivity (σ_p) from soil bulk electrical conductivity (σ_b) in ecological and hydrological applications, it still has not been worked out very well. The Hilhorst concept offers a theoretical model describing a linear relationship between σ_b, and relative dielectric permittivity (ε_b) in moist soil. The reciprocal of pore water electrical conductivity ($1/\sigma_p$) appears as a slope of the Hilhorst model and the ordinary least squares (OLS) of this linear relationship yields in a single estimate ($\widehat{1/\sigma_p}$) of the regression parameter vector (σ_p) for the entire data.

This study was carried out on a sandy soil under laboratory conditions. We used a time-varying dynamic linear model (DLM) and the Kalman filter (KF) to estimate the evolution of σ_p over time. Time series of soil relative dielectric permittivity (ε_b) and σ_b were measured using Time domain reflectometry (TDR) at different depths in a soil column to transform the deterministic Hilhorst model into a stochastic model and evaluate the linear relationship between ε_b and σ_b in order to capture deterministic changes to ($1/\sigma_p$).

Applying the Hilhorst model, strong positive autocorrelations between the residuals could be found. By using and modifying them to DLM, the observed and modeled data of ε_b match much better and the estimated evolution of σ_p converged to its true value. Moreover, the offset of this linear relation varies for each soil depth.

KEYWORDS: Electrical conductivity; Relative dielectric permittivity; Time Domain Reflectometry; Kalman Filter; Dynamic linear model

8.1.   Introduction

Salinization reduces crop productivity, decreases profitability and causes land scarcity [1]. Thus, it decreases the world's agricultural productivity and cause a global income loss by US$12 billion per year [2]. Extracting soil solution by suction or using saturated paste conductivity measurement are the common methods to determine the electrical conductivity of soil pore water (σ_p) as an indicator of the soil salinity, but they are labour- and cost intensive. There is no evidence that all ions are collected in the extract sample [3]. For soil salinity assessment it is important to look for practical methods allowing to evaluate the soil salinity state temporally and spatially. These methods help for correctly evaluating its evolution and for reasonably predicting its values [4-9]. In recent times, soil electromagnetic sensors are used to estimate bulk electrical conductivity (σ_b). Then, methods are required to transform σ_b to σ_p [3,6,10].
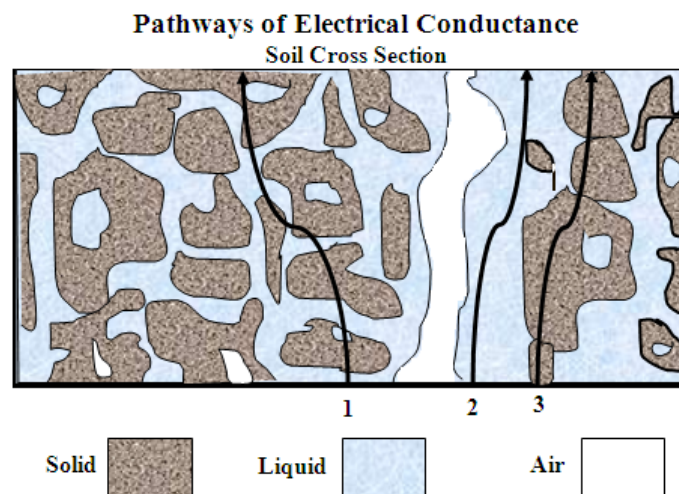
Figure 8.1 Three conductance pathways for the σ_b measurements, inspired byWyllie and Southwick [11]

According to Wyllie and Southwick [11], three conductance pathways (Fig. 1) contribute to the σ_b of a soil: (i) solid phase pathway through soil particles that are continuously contacted, (ii) liquid phase pathway through dissolved ions in the soil water inhabiting the large pores, and (iii) liquid-solid interphase pathway through exchangeable cations like surfaces of clay minerals. Electrical conductivity (EC) in the liquid phase (σ_p) is used to estimate the soil salinity, high EC refers to high concentration of soluble salts, and vice versa. The σ_p could be estimated if the relationship between σ_p, σ_b and water content (θ) is fixed [12-14]. The discovered linear correlation between the soil relative dielectric permittivity (ε_b) and σ_b values [15] enabled Hilhorst [3] to convert σ_b to σ_p by using a theoretical model. According to Hilhorst, σ_p can be determined from the equation:

$$\sigma_p = \frac{\varepsilon_p \sigma_b}{\varepsilon_b - \varepsilon_{\sigma_b=0}} \qquad (1)$$

Where σ_p is the pore water electrical conductivity (dS/m); ε_p is the relative dielectric permittivity of the soil pore water (dimensionless), ε_b is the relative dielectric permittivity of the bulk soil (dimensionless, relative dielectric permittivity is dimensionless since it is a ratio of permittivity of medium to the permittivity of free space), σ_b is the bulk electrical conductivity (dS/m), $\varepsilon_{\sigma b=0}$ is the relative dielectric permittivity of the soil when bulk electrical conductivity is 0 (dimensionless). However, $\varepsilon_{\sigma b=0}$ appears as an offset of the linear relationship between ε_b and σ_b. Hilhorst model [3] concluded that his method could be validated for water contents between 0.10 and saturation and for the conductivity of the pore water up to 0.3 S m-1. He found that $\varepsilon\sigma_b=0$ depends on soil type and varies between 1.9 and 7.6. He recommended using 4.1 as a generic offset. Many studies applied the deterministic Hilhorst model [3] in their experiments to convert σ_b into σ_p, but they did not use the same offset to achieve their study objective. For example, some studies concluded their work by using different offsets (within the range of 3.67 to 6.38) according to the soil type [10]. The producer of capacitance soil moisture sensors 5TE [16] recommend to use an offset $\varepsilon_{\sigma b=0}$ of 6, another study found that an offset $\varepsilon_{\sigma b=0}$ = 6 does not present a good linear relationship between ε_b and σ_b [17]. The WET sensor (Delta-T Device Ltd, Cambridge, UK) is a frequency domain dielectric sensor and has been designed to estimate the σ_p based on Hilhorst model [3] and incorporate the standard offset $\varepsilon_{\sigma b=0}$ = 4.1 of the model in the software of the device. By applying the Hilhorst model [3] in a saline gypsum influenced soil, the accuracy of the WET sensor in predicting σ_p was very poor when using the offset model = 4.1 [18]. Another study used WET sensor for experimental measurements in the laboratory using four different soils (sand, sandy loam, loam, and clay) [9], they found that the offset depends on both soil type and σ_p where it becomes larger for larger σ_p. Moreover, oscillator frequency and sensor circuitry could affect the estimation of ε_b and water content (θ) [19].

There are three elementary causes why deterministic system and control theories do not produce a totally sufficient means of performing this analysis and design:

(i) many effects are left unknown since the objective of the model is to represent the main modes of system response,

(ii) deterministic models are driven not by only our own control inputs, but also disturbances which we can neither control nor model deterministically, and

(iii) sensors do not offer exact readings of chosen quantities, but present their own system dynamics and distortions as well and these devices are noise corrupted [20]. Despite the importance of computing σ_p from σ_b, it still has not been worked out very well (G. Campbell [16], personal communication)

Solute transport and water flow in the unsaturated zone are normally derived from the classical Richard and the convection-dispersion equations. Deterministic. deterministic explanations of these equations are important implements in research, due to soil heterogeneity at a variety of spatial scales, these equations for predicting actual field- scale processes are increasingly being questioned [26]. Some researchers working on soil heterogeneity therefore concluded that for the evolution of soil water and solutes, it is more desired to use stochastic models rather than constant values, where the parameters of stochastic transport models are treated as random variables with discrete values assigned according to a given probability distribution [21-26]. Among stochastic models, many studies used Kalman filtering in hydrological applications. Kalman filter is an optimal recursive data processing algorithm, that recursively couples the most recent measurements into the linear model to update the model state output [27]. Under the assumption that the linear system is a stochastic process with Gaussian noises, it produces the best estimation with minimum mean square error and it has been widely used in hydrological models to optimally merge information from the model simulations and the independent observations with appropriate modeling [28-32].

In previous work, we installed frequency domain reflectometry (FDR) sensors (5 TE), which are commercially available from METER Group, Inc. USA, in field conditions at different depths where the soil is heterogeneous to estimate σ_p [22]. We used ε_b and σ_b observations to modify the Hilhorst deterministic model [3] to stochastic model using time varying dynamic linear model and Klaman filter and study the linear relationship between them.

In this study, we used Time Domain Reflectometry (TDR) sensors (FP/mts), which are commercially available from Easy Test, Poland, to measure ε_b and σ_b in laboratory conditions where the soil is homogeneous. Then, we tried to use the Hilhorst model [3] to convert σ_b to σ_p. Later, we could show the weakness of applying the deterministic Hilhorst model [3] even in homogenous soils. Thus, we are aiming to adapt this approach into a stochastic model under laboratory conditions. Thus we used one homogeneous soil type to estimate precisely the changes in σ_p over time, and to conclude whether the model offset is constant or it changes in one soil profile.

8.2. Material and methods

8.2.1. The column experiment

To achieve the objective of this study, we used two soil columns with a height of 55cm provided by sprinkler (Fig. 2). The lower boundary was controlled using vacuum pump at a constant pressure head of -30hPa. The columns were packed with a density of 1.4 g/cm3. The substrate is a sand with 80% of fine sand. The water content during packing was approximately 4 m3/m3 . The TDR and soil temperatures

sensors were installed in four depths: 7, 21, 35 and 48 cm. Since the soil is a sand, soil relative dielectric permittivity (ε_b), bulk electrical conductivity (σ_b) and temperature were measured each 5 minutes to get enough observations needed for modeling
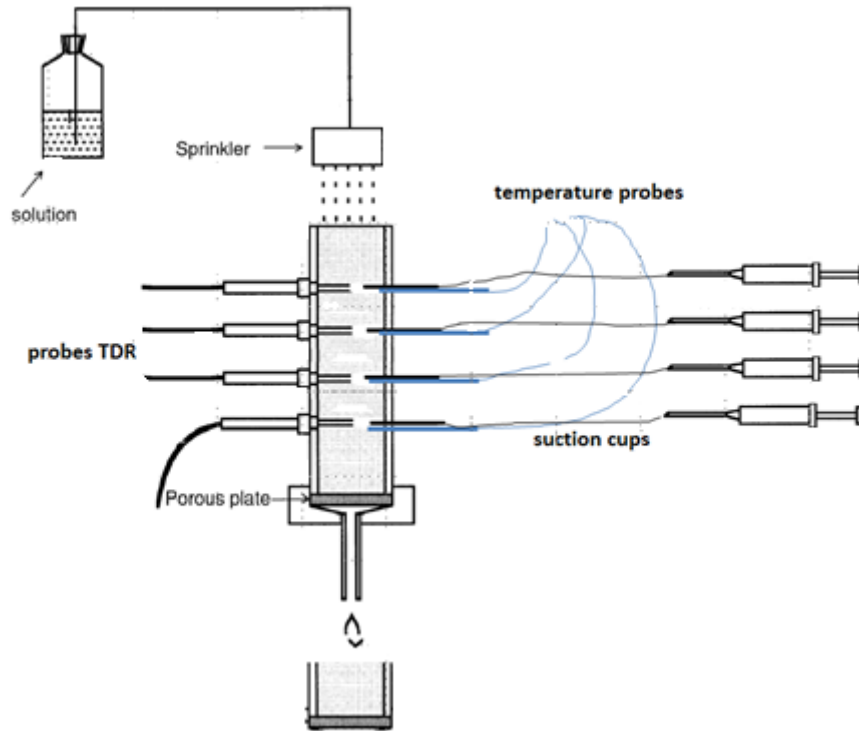
Figure 8.2 Measurement set.

Additionally, porous suction cups for taking soil solution samples were installed at each depth to validate the results of our model. The lower boundary of the column uses a membrane to let the water drain. Drainage water was collected in a bottle under -30hPa vacuum, which is supplied in the range from -20 to -30 hPa. The sprinkler is 5cm above the soil surface lets water dropping through small nozzles. Five irrigation events with KCl solution with different electrical conductivity were applied. The first three events irrigated with 20 dS/m of KCl, then the fourth, and fifth events with 30 dS/m of KCl. The flux was approximately 1l/h. The columns were free of salt at the beginning and before the irrigations events started. The TDR probes are FP/mts commercially available from Easy Test, Poland, have been calibrated in air and deionized water. The temperature probes are Thermistors of the type 2k252 (type fenewal UUA 32J3) with a range of -20 up to 60 °C. Soil temperature data (Tsoil) used to estimate the relative dielectric permittivity of the soil pore water directly (ε_p):

$$\varepsilon_p = 80.3 - 0.37(T_{soil} - 20) \qquad (1)$$

To apply dynamic linear model and the Kalman filter, time series of interested variable is needed [27]. In our study, time series of ε_b, σ_b and ε_p are required to estimate σ_p. Therefore, we used five irrigation events with two level of KCL solution to get variation of these interested variables over time for each

depth. In total, 289 observations were made of σ_b, ε_p and ε_b for each soil depth, and these were used to estimate both, the offset $\varepsilon_{\sigma b=0}$ of the our modified Hilhorst model [3], and the evolution of σ_p at its corresponding depth, of which 144 observations were used to validate their forecasts.

8.2.2. Time-varying dynamic linear model

In general, the state space model is identified by two assumptions, (i) there is a hidden or latent process xt called the state process. The state process is assumed to be a Markov process, where past and future values of xt are independent conditional on the present xt, ({xs, S > t}, and {xs, S< t} are independent on the xt), (ii) the observations, yt are independent given the states xt. This means that the dependence among the observations is generated by states. The dynamic linear model (DLM) or linear Gaussian state space model, in its simple form, employs an order one, p-dimensional vector autoregression as the state equation:

$$x_t = x_{t-1} + w_t \qquad w_t \sim \mathcal{N}(0, W_t) \tag{2}$$

We do not observe the state vector xt directly, but only a linear transformed version of them with noise added, say:

$$y_t = A_t x_t + v_t \qquad v_t \sim \mathcal{N}(0, V_t) \tag{3}$$

yt is an m-dimensional vector, representing the observation at time t, A_t is a q×p measurement or observation matrix. Equation (4) is called as observation equation, in which v_t ,w_t and are the Gaussian white-noise errors. The evolution variances V_t,W_t and can be estimated from available data using maximum likelihood or Bayesian techniques.

In this study, we modified the deterministic Hihlorst model (1) to a stochastic one. The model (1) has the variables σ_b, ε_p, σ_p, ε_b and $\varepsilon_{\sigma b=0}$. The σ_p and $\varepsilon_{\sigma b=0}$ are unobserved and they need to be estimated by the state equation (3) as x_t , while σ_b, ε_p and ε_b are observed by the sensors (ε_p calculated from equation (2) using soil temperature sensor data) and represent by observation equation (4) as y_t.

The R [33] and package dlm [34] provide an integrated environment for Bayesian inference using DLM, and the package includes functions for Kalman filtering and smoothing, as well as for maximum likelihood estimation.

8.3.   Results and discussion

8.3.1. Deterministic model:

The offset of Hilhorst model [3] can be calculated from the equation (1):

$$\varepsilon_b = (1/\sigma_p)\varepsilon_p * \sigma_b + \varepsilon_{\sigma_b=0} \qquad\qquad (4)$$

We derived the offset ($\varepsilon_{\sigma b=0}$) from this linear model after using measurements of ε_b and σ_b . For example, applying the ordinary least squares (OLS) on measurements of ε_b and σ_b obtained from soil column 2 data during the third irrigation at depth 21 cm, Table 1 shows that the offset of the linear relationship between ε_b-σ_b is 9.41. Further, the single estimate of the slope ($\widehat{1/\sigma_p}$) of the regression parameter vector ($1/\sigma_p$) for the entire data set is very small. Thus, the estimated soil pore water electrical conductivity (σ_p) is to high compared with the EC meter value (Table 2). Afterwards we applied the Durbin–Watson in order to test if there is any autocorrelation between the residuals of the regression. Table 3 shows that there is an extremely strong and positive autocorrelation, meaning, that the result of that regression is not valid.

Table 8.1 Estimated parameters gained from the linear regression analysis.

	Estimate	Std. Error	t value	Pr(>|t|)
$\varepsilon_{\sigma b=0}$	9.411	8.591e-03	1095.4	<2e-16 ***
$1/\sigma_p$	6.963e-04	4.461e-06	156.1	<2e-16 ***

Significance: * P < 0.05, ** P < 0.01, *** P < 0.001.

Table 8.2 Electrical conductivity of the soil solution (dS/m) according to soil column number, irrigation event and depth (cm), it is collected by porous suction cups and measured by EC meter device.

Soil column 1				Soil column 2			
Irrigation event 3		Irrigation event 4		Irrigation event 3		Irrigation event 4	
Dep.21cm	Dep.35cm	Dep.21cm	Dep.35cm	Dep.21cm	Dep.35cm	Dep.21cm	Dep.35cm
15.96	18	21.89	22.35	18.61	14.97	25.67	22

For each irrigation event we got one solution sample at each depth by using porous suction cups. Unfortunately, some samples did not have enough solution to measure its electrical conductivity by EC meter device. Table 2 shows the values of EC measured by EC meter device. The Table shows 8 EC values from EC meter according to the depth and irrigations event number for each soil column. Due to the variability in water flow in unsaturated soil, we observed in our experiment a variation in the time needed to collect the solution sample, more time required to collect enough solution for EC meter device more amount of ions gathered in the sample and as a result high EC value of the sample. Therefore, there is a difference in the EC values between the soil columns at the same depth (table 2). We applied modified Hilhorst model on the 8 times series data corresponding to table 2 (depth, irrigation event and soil

column). The reason for that is to compare our finding of σ_p obtained from our modified model to the values of σ_p obtained by EC meter device (Table 2).

Table 8.3 Durbin-Watson test for linear regression ε_b–σ_b.

lag	Autocorrelation	D-W Statistic	p-value
1	0.852	0.278	0

The reason for choosing 1l/h for irrigation rate and 5 minutes for irrigation interval is visualized in figure 3. At each depth we could see how the bulk electrical conductivity responds to the irrigation event.
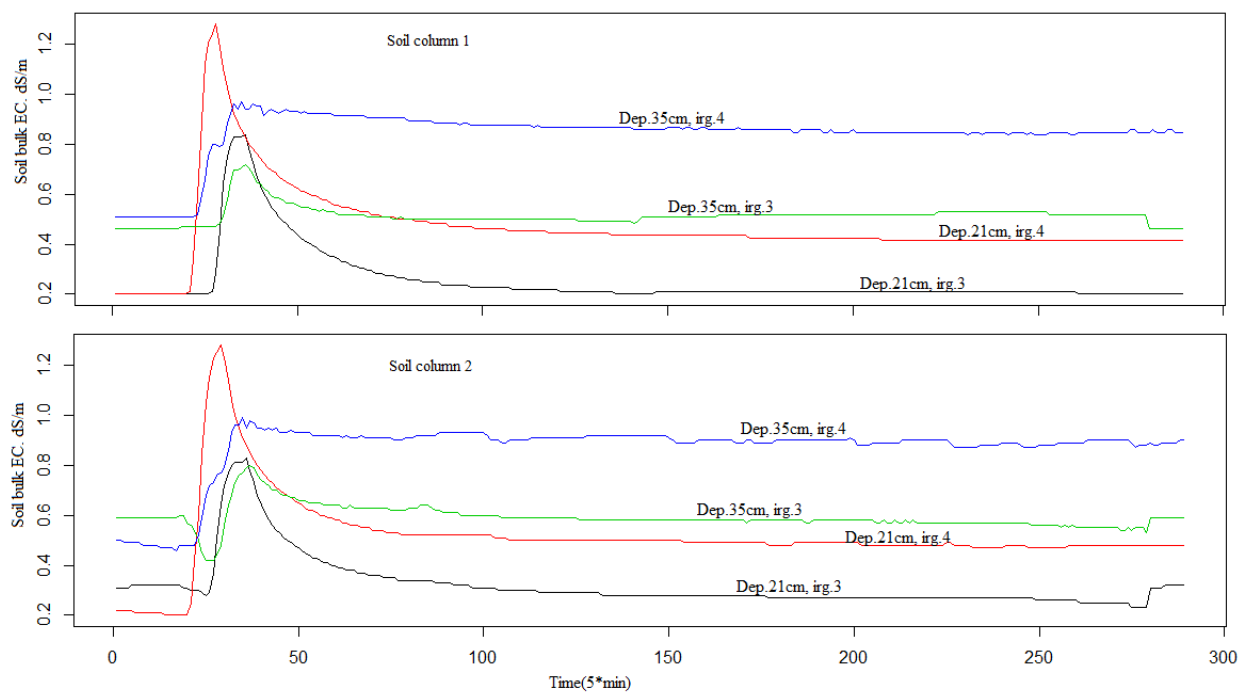
Figure 8.3 Bulk electrical conductivity (σ_b) in the two soil columns, for two irrigation events (N°.3 and N°.4) at two depths (21cm and 35cm). Series peaks related to time irrigation.

8.3.2. Time-varying Linear Dynamic Model (LDM)

The deterministic (Eq.5) can be modified into the time-varying DLM for observation and unobservable (state) models. In this case, the observation data are the soil relative dielectric permittivity (ε_b), bulk electrical conductivity (σ_b) and the relative dielectric permittivity (ε_p), while the unobservable data are the offset ($\varepsilon_{\sigma b=0}$) and pore water electrical conductivity (σ_p). Equation 4 can be modified to the time-varying DLM as follows:

- The observation equation can be obtained by modifying the Hilhorst model [3] (written in equation 5) into a stochastic equation, in accordance with equation 4 as following:

$$(\varepsilon_b)_t = \left(\varepsilon_{\sigma_b=0}\right)_t + \left(\varepsilon_p * \sigma_b\right)_t \left(\frac{1}{\sigma_p}\right)_t + v_t \text{ (6)}$$

$$v_t \sim \mathcal{N}(0, \sigma_v{}^2)$$

- The state equation (unobservable data) in Eq.3 are $\varepsilon_{\sigma b=0}$, and the slope $1/\sigma_p$. They can be converted to the unobservable state equation of the time-varying DLM according to equation 3. The unobservable state equation can be arranged as follows:

$$\begin{cases} (\varepsilon_{\sigma_b=0})_t = (\varepsilon_{\sigma_b=0})_{t-1} \\ \left(\frac{1}{\sigma_p}\right)_t = \left(\frac{1}{\sigma_p}\right)_{t-1} + w_t \quad w_t \sim \mathcal{N}(0, (\sigma_w)_t^2) \end{cases} \text{ (7)}$$

Here, we consider $\varepsilon_{\sigma b=0}$ as a constant. The actual value is related only to its past value. The slope $1/\sigma_p$ changes over time, and its actual value is related to its past value plus the Gaussian white-noise errors (w_t). We applied the equation in reverse order to estimate the state variables ($\varepsilon_{\sigma b=0}$ and σ_p) at all-time points from a complete series of the soil relative dielectric permittivity (ε_b). This process is known as smoothing.

An example of the evolution of ε_b, ε_p, and σ_b data needed for Hilhorst model [3] is shown in figure 4. Applying the equation (6) and (7) using DLM and the Kalman filter on the 8-time series data, we see in Fig. 5 the observed and predicted time series of the soil relative dielectric permittivity (ε_b). The predicted and observed values of ε_b agree reasonably well. The mean absolute prediction error (MAPE) for the time series never exceeded 0.02.
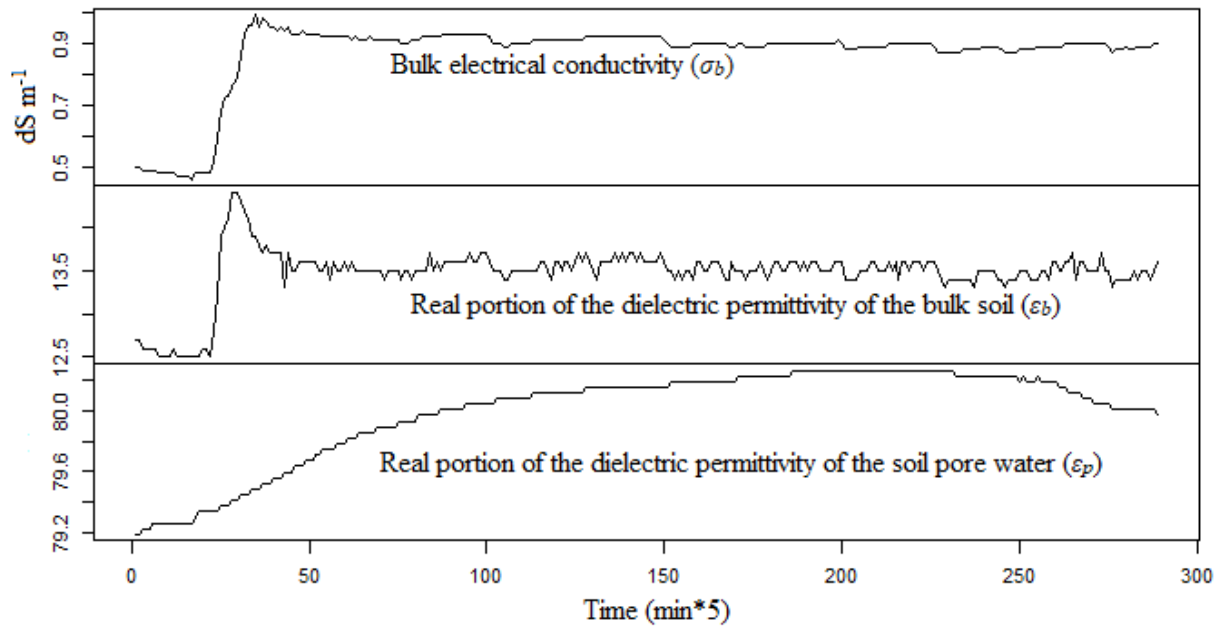
Figure 8.4 Known variables for the Hilhorst model (σ_b, ε_b and ε_p); data form soil column 2, depth 35cm and irrigation event N°4.

Since prediction of the soil relative dielectric permittivity (ε_b) is valid, the estimation of electrical conductivity of pore water (σ_p) and the offset $\varepsilon_{\sigma b=0}$ (equation 7) are also valid because they are used in the prediction the soil relative dielectric permittivity (ε_b) and they are converged to their true values. The evolution of σ_p over time obtained by DLM presented in figure 6, it shows the importance of using DLM since it obtained all the changes of σ_p over time and not a single value for entire data set. Another interesting aspect is that the figure 6 shows the changes in model offset for each irrigation event at each depth. This finding is very important since it shows that the offset does not depend on soil type [10,14,15, 16,] neither on soil type and salinity [17] where in this study we used two columns with the same type soil. Moreover, in Figure 6 we put the corresponding value of σ_p measured by EC meter device for each depth according the irrigation event and soil column number a.

Comparing the mean evolution of σ_p values obtained from our modified Hilhorst model (Fig. 6) with the single corresponding EC value obtained from porous suction cups and measured by EC meter device (Table 2), we found that they agree very well (R^2= 72%).

From these results, three advantages of using DLM and Kalman filter to estimate σ_p in two homogenous soil columns, first, we observed that the offset value of Hilhorst model does not depend on the soil type and σ_p and it changes in the same soil profile. Secondly, we obtained the changes in the estimated σ_p over time and not just a single value as a coefficient for the entire data set. Third, the estimated changes in σ_p occur instantly and save time and labor costs.

col.1, dep.21cm, 3rd irg. event

col.1, dep.21cm, 4th irg. event

col.2, dep.21cm, 3rd irg. event

col.2, dep.21cm, 4th irg. event

col.1, dep.35cm, 3rd irg. event

col.1, dep.35cm, 4th irg. event

col.2, dep.35cm, 3rd irg. event

col.2, dep.35cm, 4th irg. event
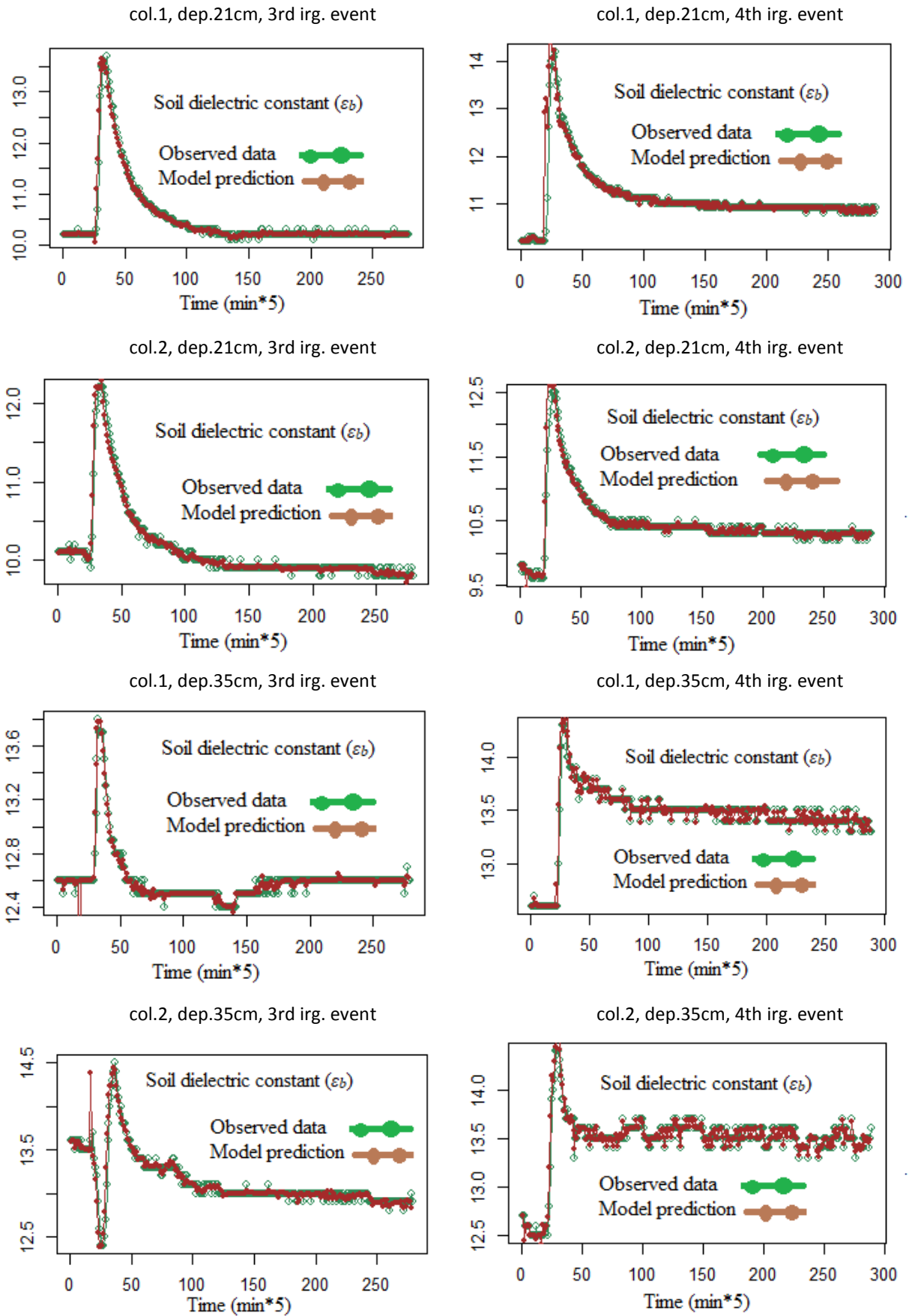
Figure 8.5 Observed and predicted soil relative dielectric permittivity according to the soil column number, depth and irrigation event.

col.1, dep.21cm, 3rd irg. Event   col.1, dep.21cm, 4th irg. Event   col.2, dep.35cm, 3rd irg. Event

σ_p **by EC meter =15.96 dS/m**

σ_p **by EC meter =21.89 dS/m**

σ_p **by EC meter =14.97 dS/m**

col.2, dep.21cm, 3rd irg. Event
σ_p **by EC meter =18.61 dS/m**

col.2, dep.21cm, 4th irg. Event
σ_p **by EC meter =25.67 dS/m**

col.2, dep.35cm, 4th irg. Event
σ_p **by EC meter =22 dS/m**

col.1, dep.35cm, 3rd irg. Event
σ_p **by EC meter =18 dS/m**

col.1, dep.35cm, 4th irg. Event
σ_p **by EC meter =22.35 dS/m**

Figure 8.6 Estimation of the unobservable data ($\varepsilon_{\sigma b=0}$ and σ_p) by applying the time-varying DLM and the Kalman filter on the data according to the soil column number (col), depth (dep) and irrigation event (irg. Event, the corresponding of σ_p by EC meter device is given for each estimated σ_p.

8.4.   *Conclusion*

In this study we applied the ε_b-σ_b linear relationship to homogeneous soil columns data obtained from TDR sensors. We found an extremely strong positive autocorrelation between the residuals of the regression analysis. When residuals are correlated, the least squares method is not the most efficient model coefficient estimator. By modifying the regression by a time varying-dynamic linear model (DLM), the observed and modeled data of ε_b match much better, and the estimated evolution of σ_p converges to its true value. Moreover, in this study we used two homogeneous soil columns with the same condition to show that the offset of the Hilhorst model [3] is not constant as suggested for all moist soil or as others suggested that it is soil type depended [10, 14, 15, 16] or soil type and salinity depended [17], we repeat the experiment to show that the offset changes even in the same soil type and the same conditions. Dynamic linear model enables to capture the offset changes and it shows the importance to calculate it simultaneously when estimating σ_p by Hilhorst model. The next promising step would be programing and inserting these models into the TDR software in order to estimate directly the soil pore water electrical conductivity (σ_p) from senor records.

Acknowledgement

8.5.   *References*

1. Thiruchelvam, S.; Pathmarajah, S. An Economic Analysis of Salinity Problems in the Mahaweli River System H Irrigation Scheme in Sri Lanka. 47.
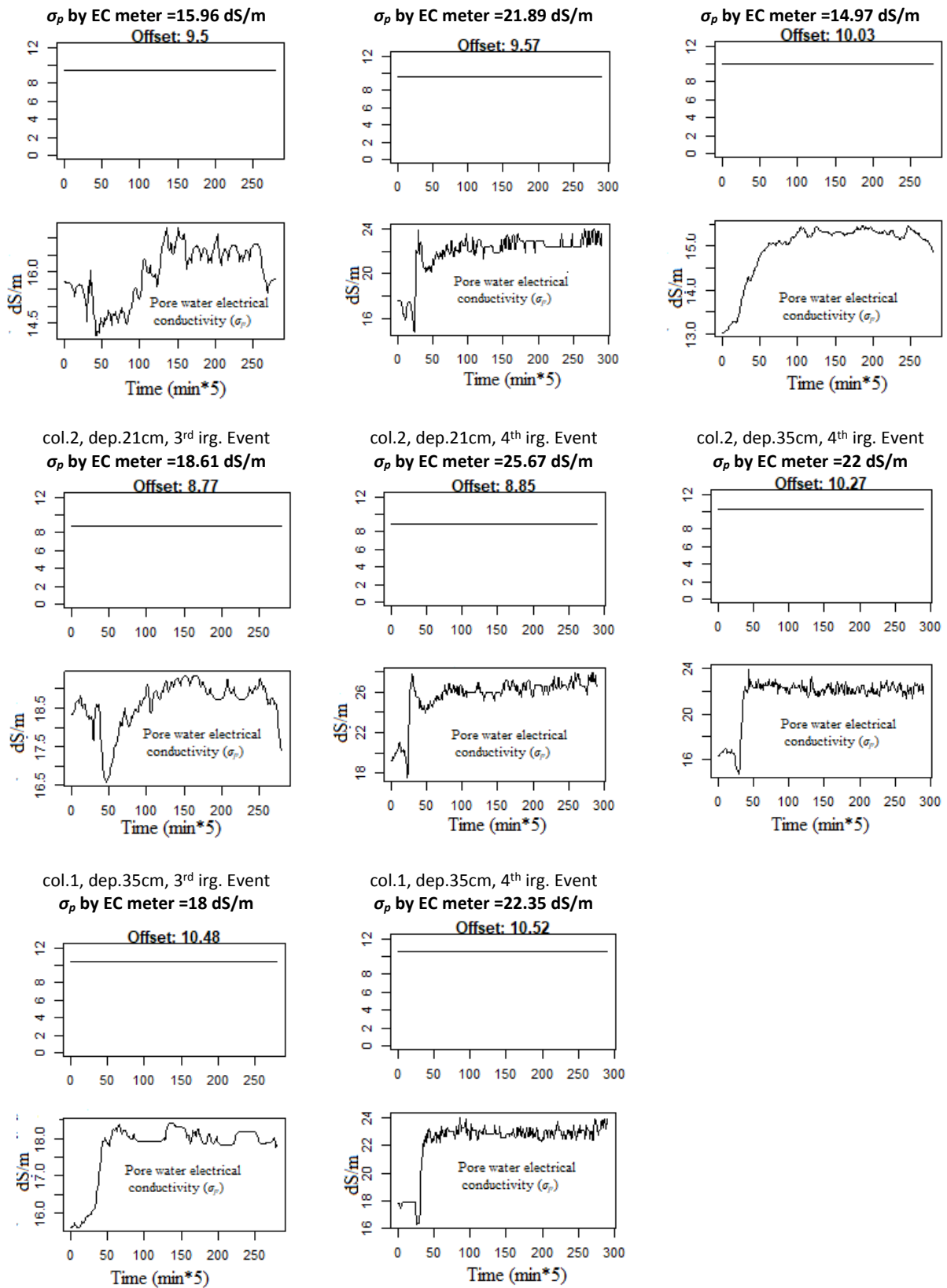
2. Ghassemi, F.; Jakeman, A. J.; Nix, H. A. Salinisation of Land and Water Resources: Human Causes, Extent, Management and Case Studies.; CAB International: Wallingford, 1995.

3. Hilhorst, M. A. A Pore Water Conductivity Sensor. Soil Science Society of America Journal 2000, 64 (6), 1922–1925.

4. Hajrasuliha, S.; Baniabbassi, N.; Metthey, J.; Nielsen, D. R. Spatial Variability of Soil Sampling for Salinity Studies in Southwest Iran. Irrigation Science 1980, 1 (4), 197–208.

5. Cetin, M.; Kirda, C. Spatial and Temporal Changes of Soil Salinity in a Cotton Field Irrigated with Low-Quality Water. Journal of Hydrology 2003, 272 (1–4), 238–249.

6. Rhoades, J. D.; Shouse, P. J.; Alves, W. J.; Manteghi, N. A.; Lesch, S. M. Determining Soil Salinity from Soil Electrical Conductivity Using Different Models and Estimates. Soil Science Society of America Journal 1990, 54 (1), 46–54.

7. Shouse, P. J.; Goldberg, S.; Skaggs, T. H.; Soppe, R. W. O.; Ayars, J. E. Changes in Spatial and Temporal Variability of SAR Affected by Shallow Groundwater Management of an Irrigated Field, California. Agricultural Water Management 2010, 97 (5), 673–680.

8. Li, X.; Yang, J.; Liu, M.; Liu, G.; Yu, M. Spatio-Temporal Changes of Soil Salinity in Arid Areas of South Xinjiang Using Electromagnetic Induction. Journal of Integrative Agriculture 2012, 11 (8), 1365–1376.

9. Kargas George; Persson Magnus; Kanelis George; Markopoulou Ioanna; Kerkides Petros. Prediction of Soil Solution Electrical Conductivity by the Permittivity Corrected Linear Model Using a Dielectric Sensor. Journal of Irrigation and Drainage Engineering 2017, 143 (8), 04017030. https://doi.org/10.1061/(ASCE)IR.1943-4774.0001210.

10. Persson, M. Evaluating the Linear Dielectric Constant-Electrical Conductivity Model Using Time-Domain Reflectometry. Hydrological Sciences Journal 2002, 47 (2), 269–277.

11. Wyllie, M. R. J.; Southwick, P. F. An Experimental Investigation of the S.P. and Resistivity Phenomena in Dirty Sands. Journal of Petroleum Technology 1954, 6 (02), 44–57. https://doi.org/10.2118/302-G.

12. Rhoades, J. D.; Raats, P. A. C.; Prather, R. J. Effects of Liquid-Phase Electrical Conductivity, Water Content, and Surface Conductivity on Bulk Soil Electrical Conductivity1. Soil Science Society of America Journal 1976, 40 (5), 651–655.

13. Mualem, Y.; Friedman, S. P. Theoretical Prediction of Electrical Conductivity in Saturated and Unsaturated Soil. Water Resources Research 1991, 27 (10), 2771–2777.

14. Malicki, M. A.; Walczak, R. T. Evaluating Soil Salinity Status from Bulk Electrical Conductivity and Permittivity. European Journal of Soil Science 1999, 50 (3), 505–514.

15. Malicki MA, W. R., Koch S, Fluhler H. Determining Soil Salinity from Simultaneous Readings of Its Electrical Conductivity and Permittivity Using TDR. InProceedings of the Symposium and Workshop on Time Domain Reflectometry in Environmental, Infrastructure, and Mining Applications, Evanston 1994, 101, 7–9.

16. Devices, Decagon. 5TE Water Content, EC, and Temperature Sensors Operator's Manual. 2008.

17. Rodriguez, A.; Ruben, F. CALIBRATING CAPACITANCE SENSORS TO ESTIMATE WATER CONTENT, MATRIC POTENTIAL, AND ELECTRICAL CONDUCTIVITY IN SOILLESS SUBSTRATES. Thesis, 2009.

18. Bouksila, F.; Persson, M.; Berndtsson, R.; Bahri, A. Soil Water Content and Salinity Determination Using Different Dielectric Methods in Saline Gypsiferous Soil/Détermination de La Teneur En Eau et de La Salinité de Sols Salins Gypseux à l'aide de Différentes Méthodes Diélectriques. Hydrological Sciences Journal 2008, 53 (1), 253–265.

19. Visconti, F.; MartÃ-nez, D.; Molina, M. J.; Ingelmo, F.; Miguel de Paz, J. A Combined Equation to Estimate the Soil Pore-Water Electrical Conductivity: Calibration with the WET and 5TE Sensors. Soil Res. 2014, 52 (5), 419–430.

20. Maybeck, P. S. Stochastic Models, Estimation, and Control; Academic Press, 1982.

21. Aljoumani, B.; Sànchez-Espigares, J. A.; Cañameras, N.; Wessolek, G.; Josa, R. Transfer Function and Time Series Outlier Analysis: Modelling Soil Salinity in Loamy Sand Soil by Including the Influences of Irrigation Management and Soil Temperature. Irrigation and Drainage 2018, 67 (2), 282–294.

22. Aljoumani, B.; Sànchez-Espigares, J. A.; Cañameras, N.; Josa, R. An Advanced Process for Evaluating a Linear Dielectric Constant–Bulk Electrical Conductivity Model Using a Capacitance Sensor in Field Conditions. Hydrological Sciences Journal 2015, 60 (10), 1828–1839.

23. Aljoumani, B.; Sànchez-Espigares, J. A.; Cañameras, N.; Josa, R.; Monserrat, J. Time Series Outlier and Intervention Analysis: Irrigation Management Influences on Soil Water Content in Silty Loam Soil. Agricultural Water Management 2012, 111, 105–114.

24. Beven, K.; Germann, P. Macropores and Water Flow in Soils. Water Resources Research 1982, 18 (5), 1311–1325.

25. White, R. E. The Influence of Macropores on the Transport of Dissolved and Suspended Matter Through Soil. In Advances in Soil Science; Stewart, B. A., Ed.; Springer New York, 1985; pp 95–120.

26. Genuchten M, V. Recent Progress in Modelling Water Flow and Chemical Transport in the Unsaturated Zone. , 08/11-24/91. :. 20 th General Assembly of the International Union of Geodesy and Geophysics, Vienna, Austria 1991, 08/11, 169–183.

27. Kalman, R. E. A New Approach to Linear Filtering and Prediction Problems. Journal of Basic Engineering 1960, 82 (1), 35–45.

28. Wendroth, O.; Rogasik, H.; Koszinski, S.; Ritsema, C. J.; Dekker, L. W.; Nielsen, D. R. State±space Prediction of ®eld±scale Soil Water Content Time Series in a Sandy Loam. 1999, 9.

29. Hoeben, R.; Troch, P. A. Assimilation of Active Microwave Observation Data for Soil Moisture Profile Estimation. Water Resources Research 2000, 36 (10), 2805–2819.

30. Moradkhani, H.; Sorooshian, S.; Gupta, H. V.; Houser, P. R. Dual State–Parameter Estimation of Hydrological Models Using Ensemble Kalman Filter. Advances in Water Resources 2005, 28 (2), 135–147.

31. Anderson, J. L. An Ensemble Adjustment Kalman Filter for Data Assimilation. Monthly Weather Review 2001, 129 (12), 2884–2903.

32. Liu, Y.; Weerts, A. H.; Clark, M.; Hendricks Franssen, H.-J.; Kumar, S.; Moradkhani, H.; Seo, D.-J.; Schwanenberg, D.; Smith, P.; van Dijk, A. I. J. M.; et al. Advancing Data Assimilation in Operational

Hydrologic Forecasting: Progresses, Challenges, and Emerging Opportunities. Hydrology and Earth System Sciences 2012, 16 (10), 3863–3887.

33. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria 2013.

34. Petris, G. An R Package for Dynamic Linear Models. Journal of Statistical Software 2010, 36 (12).

Evaluating the Variation of Dissolved Metals on a Highway Roadside Using a Generalized Additive Mixed Model (GAMM)

Basem Aljoumani · Björn Kluge · Josep Sanchez · Gerd Wessolek

Abstract Assessing metal concentrations in roadside soils requires a better understanding of the extent to which they are affected by different environmental factors such as soil texture, depth, pH, runoff concentration, and precipitation. Monthly data of dissolved Cd, Ni, Cr, Pb, Cu, and Zn concentrations in three different roadside soils (sandy loam, gravel (0–32 mm) and a mixture of sandy loam and gravel) were measured during a 2-year lysimeter field study at different depths. The data was used to assess the variation of trace elements and how they were affected by environmental factors. For data interpretation, generalized additive mixed models (GAMMs) were used to explore the complex behavior of metals in heterogeneous soils by detecting linear and nonlinear trends of metal concentrations in the soil solution. As a result, the modeling approach showed that Cd, Ni, Cr, Pb, Cu, and Zn concentrations are functions of different environmental variables, which have either linear or nonlinear behavior. All investigated metals showed that pH could explain their variation. With exception of precipitation, Ni and Cr variations can nearly be explained by the same environmental factors used in this study (time, pH, infiltration volume, roadside soil type, runoff concentrations, and depth). During the study period, we found that Zn variation can be explained by its nonlinear relationship with all the significant studied environmental factors. As the depth increases from the surface to 30 cm of depth, the metal concentration of Cd, Ni, Cr, Pb, and Zn increases. Surprisingly, the roadside soil consisting of gravel has the lowest organic carbon and showed the lowest median concentration of Cd, Ni, Pb, Cu, and Zn at 30 cm. Moreover, the model showed that the surface runoff volume has no effect on the metal variation in the soil solution.

Basem Aljoumani and Björn Kluge contributed equally to this work.

Capsule: We were able to explore the complex behavior of metals in heterogeneous soil systems by detecting the linear and nonlinear dynamics of metals in soil solutions.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11270-019-4137-6) contains supplementary material, which is available to authorized users.

B. Aljoumani (✉) · B. Kluge
Department of Ecology, Ecohydrology and Landscape Evaluation, Technische Universität Berlin, Ernst-Reuter Platz 1, 10587 Berlin, Germany
e-mail: basem.aljoumani@tu-berlin.de

B. Kluge (✉) · G. Wessolek
Department of Ecology, Soil Conservation, Technische Universität Berlin, Berlin Ernst-Reuter Platz 1, 10587 Berlin, Germany
e-mail: bjoern.kluge@tu-berlin.de

J. Sanchez
Department of Statistical and Operational Research, Universitat Politècnica de Catalunya (UPC), Barcelona, Spain

1 Introduction

Road traffic has long been known to be a major source of pollutants in the environment. The worldwide growth

9. Evaluating the Variation of Dissolved Metals on a Highway Roadside Using a Generalized Additive Mixed Model (GAMM)

Basem Aljoumani[1], Björn Kluge[1,2], Josep Sanchez[3], Gerd Wessolek[2]

[1]Department of Ecology, Ecohydrology and Landscape Evaluation, Technische Universität Berlin Ernst-Reuter Platz 1, 10587 Berlin, Germany
[2]Department of Ecology, Soil Conservation, Technische Universität Berlin Ernst-Reuter Platz 1, 10587 Berlin, Germany
[3]Department of Statistical and Operational Research, UPC-BarcelonaTECH Spain

ABSTRACT

Assessing metal-concentrations in roadside soils requires a better understanding of the extent to which they are affected by different environmental factors such as soil-texture, depth, pH, runoff concentration and precipitation Monthly data of dissolved Cd, Ni, Cr, Pb, Cu and Zn concentrations in three different roadside soils (sandy loam, gravel (0-32 mm) and a mixture of sandy loam and gravel) were measured during a two-year lysimeter field study at different depths. The data was used to assess the variation of trace elements and how they were affected by environmental factors. For data interpretation, generalized additive mixed models (GAMMs) were used to explore the complex behavior of metals in heterogeneous soils by detecting linear and nonlinear trends of metal concentrations in the soil solution. As a result, the modeling approach showed that Cd, Ni, Cr, Pb, Cu and Zn concentrations are functions of different environmental variables, which have either linear or nonlinear behavior. All investigated metals showed that pH could explain their variation. With exception of precipitation, Ni and Cr variations can nearly be explained by the same environmental factors used in this study (time, pH, infiltration volume, roadside soil type, runoff concentrations and depth). During the study period, we found that Zn variation can be explained by its nonlinear relationship with all the significant studied environmental factors. As the depth increases from the surface to 30 cm of depth, the metal concentration of Cd, Ni, Cr, Pb, and Zn increases. Surprisingly, the roadside soil consisting of gravel has the lowest organic carbon and showed

the lowest median concentration of Cd, Ni, Pb, Cu and Zn at 30 cm. Moreover, the model showed that the surface runoff volume has no effect on the metal variation in the soil solution.

KEYWORDS: metals, roadside soil, generalized additive mixed models (GAMMs), environmental factors Capsule: We were able to explore the complex behavior of metals in heterogeneous soil systems by detecting the linear and nonlinear dynamics of metals in soil solutions.

9.1.   Introduction

Road traffic has long been known to be a major source of pollutants in the environment. The worldwide growth in traffic volume (OECD, 2013) has led to increasing emission rates, which in turn result in increasing accumulation rates, especially in road dust, runoff water and in roadside soils (Monks et al., 2009; Werkenthin et al. 2014, Padoan et al. 2017). The distribution and amount of pollutants introduced in the environment is determined by different factors, such as traffic intensity (Ward, 1990; Arslan and Gizir, 2006), climatic conditions and events, vegetation cover and highway design (Barbosa and Hvitved-Jacobsen, 1999; Pagotto et al., 2001; Huber et al., 2016, Liu et al., 2016). Moreover, galvanized crash barriers, traffic control devices and road signs release a considerable amount of pollutants due to splashing and rainwater (Van Bohemen and Van de Laak, 2003; Kluge et al. 2014). The composition of traffic-derived pollutants is various and multifaceted. The most well-known pollutants emitted are carbon monoxide, nitrogen oxide, hydrocarbon, sulfur dioxide, polycyclic aromatic hydrocarbons and metals (Kayhanian et al., 2012).

Of the traffic-derived pollutants, metals are those that have the utmost importance because they are not degradable and have particularly high toxicity for plants, microorganisms, animals and humans (Wuana and Okieimen, 2011). Among other metals, Zn, Cd, Pb, Cu, Cr and Ni are the most well-known and studied metals in roadside soils (Münch, 1993; Folkeson et al., 2009; Kayhanian et al., 2012; Men et al., 2018). The metal concentrations in roadside soils decrease mostly with distance from the road edge and at increasingly greater soil depth (e.g., Harrison et al., 1985; Pagotto et al., 2001; Modrzewska and Wyszkowski, 2014; Werkenthin et al., 2014; Huber et al., 2016). The mobility of metals depends greatly on soil pH and organic matter content, as has been noted in many studies (e.g., Turer and Maynard, 2003; Walraven et al., 2014). Turer and Maynard (2003) showed a strong positive correlation between soil organic matter and certain metal concentrations. Hjortenkrans et al. (2008) estimated a mobilization of the easily exchangeable metal fraction after extreme climatic conditions, such as dry periods in connection with oxidation or intensive rain events with reducing processes. Furthermore, the application of de-icing salts in the winter often leads to higher dissolution, dispersion and leaching of organic matter in roadside soils, which in turn causes an increase in metal mobilization rates in roadside soils (Bäckström et al., 2004; Ramakrishna and Viraraghavan, 2005, Fay and Shi, 2012). High metal accumulation and leaching by the

previously described processes may lead to high concentrations in roadside soil solutions (Pagotto et al. 2010; Werkenthin et al. 2014), which in turn could result in contamination of groundwater and/or receiving waters, with all the corresponding harmful effects on the environment.

From an ecological and hydrological point of view, it is therefore highly important to evaluate the road runoff metal concentrations and soil solution concentrations of roadside environments, but this is a complex task because the transport mechanisms are potentially nonlinear. To the best of our knowledge, only a few studies have dealt with the on-site soil solution concentrations beside the affected roads and highways (Reinirkens, 1996; Dierkes and Geiger, 1999; Bäckström et al., 2004; Kocher et al., 2005; Kluge and Wessolek, 2012; Kluge et al., 2014; Werkenthin et al., 2016). All these studies used classical summary statistics to present the variability of dissolved metal concentrations in soil solutions, and they did not at all address the temporal dependencies and multivariate environmental effects.

Varied methods have been used to predict the metal concentrations in road runoff, sediments and soils. These have included classical summary statistics (Wang et al., 2013; Pal et al., 2011), regression models (Yanet al., 2013), geospatial analysis (Guo et al., 2012; Raulinaitis et al., 2012; Ignatavičius et al., 2017). Sauvé et al. (2000) collected different soil-liquid partitioning coefficients (Kd) in soils for the metals cadmium, copper, lead, nickel, and zinc, which they derived from over 70 studies of various origins. They used multiple linear regression models and showed that the Kd Coefficients are a function of soil solution pH, -organic matter, and -total metal concentrations. However, the R2 values of their model were low. One reason for this might be the nonlinear behavior of some metals.

Casey et al. (2006) compared mean invertebrate metal concentration between two periods to capture the temporal trend of trace metals, using mixed models to compare sediment metal levels across years and land-use types. However, their models did not mention the nature of nonlinear patterns of trace metals. Wen et al. (2012) studied the temporal variation of dissolved heavy metal concentrations by using only the effects of rainfall events. Moreover, most previous studies did not include the influence of other environmental variables like soil properties, which could also contribute to the variability of dissolved metals in soil. Recently, Lequy et al. (2017) assessed the temporal trends of trace metal concentrations in mosses by using a generalized additive mixed model (GAMMs) approach that included the sampling protocol (sampling period, collector and moss species) to describe the variability of the concentrations in mosses. As they mentioned, it was the first attempt to analyze the temporal trend and variability of metal concentrations by using GAMMs. The authors highlighted the importance of accounting for nonlinear temporal variations in metal concentrations, but only the parameters of time, collector, and moss species were used as explanatory variables in their models.

The objective of this study is to identify the environmental data used in this study that could explain the variation of dissolved heavy metal concentrations in roadside soil solutions. To achieve this

objective, we use as response variables mixed modeling on repeated measurements of concentrations over two years while basing the explanatory variables on different environmental variables such as soil properties and meteorological factors.

9.2. Material and Methods

9.2.1. Study area

In this study, data from a section of the Highway BAB A115 near the city of Berlin, Germany (52° 23' 27.5" N, 13° 09' 42.8" E) is selected as an example. The highway has an established speed limit of 120km/h and is surrounded by a mixed pine-oak forest stand on sandy soil beginning at a distance of about 15 m from the road. The traffic volume ranges from 63,000 to 80,000 vehicles per day, with 6-7% of them constituting heavy traffic (Fitschen and Nordmann, 2012)

Along this section, the highway is 34 m wide (including the central reservation) and has three concrete lanes in each direction. The sealed road surface has a width of 15 m in each direction and is drained across the adjacent embankment. The relative elevation of the highway is about 2 m; the longitudinal inclination is 1% and no crash barriers are installed. Annual average temperature is 8°C and precipitation is 580 mm (see also the Supplementary Data in Fig. 1.).

9.2.2. Highway Lysimeter

9.2.2.1. Soil solution, road runoff and climate data

At this section of the highway were installed three lysimeters of polyethylene (PE), each with a length of 150 cm, a width of 100 cm and a height of 60 cm, together with three runoff lysimeters.

The lysimeters were placed directly beside the road edge and filled with different soil materials that were recently used for embankment construction in Germany:

Sm1: Reference embankment material (surrounding topsoil of arable land) – Sandy loam

Sm2: Mixture of Sm2 with 15% of Sm1 – Gravel (0-32 mm) mixed with sandy loam

Sm3: Base course construction material from natural broken rock and sand – Gravel (0-32 mm)

The investigated embankment soils were obtained from a nearby highway construction site. The thickness of the embankment layer was 20 cm, equivalent to a soil volume of 0.3 m3. The lysimeters had three discharges: one for collecting the surface runoff and two for collecting the soil solution at depths of 15 cm and 30 cm (Fig.1). The water volume was measured by tipping buckets (type: V2A, UP- GmbH). Aliquots were collected as composite samples in PE vessels. The runoff lysimeter had the same surface

area as the embankment lysimeters (1.5 m²); but they were only 15 cm high (length: 150 cm, width: 100 cm, height: 15 cm) and were filled with lime-free, washed pebble gravel (Ø 5.6-8 mm) to meet the road safety regulations. They had only one discharge at the bottom and were placed between the embankment lysimeters SM1-SM3 to quantify the variability of runoff quantity. Water volumes of road runoff were also measured by pre-described tipping buckets for automatic recording. In addition, a weather station was installed to measure precipitation, temperature, humidity and potential evaporation at 15 min. intervals. For more details of the lysimeter study site and construction of the highway lysimeters, please see Werkenthin et al., 2016.

Sampling of soil solution at three depths for each embankment lysimeter and runoff lysimeter was conducted once a month from November 2012 to December 2014. Samples were analyzed in the laboratory for Cadmium (Cd), Copper (Cu), Zinc (Zn), lead (Pb), Nickel (Ni), Chromium (Cr), pH, electrical conductivity (EC) and dissolved organic carbon (DOC). Part of the data (July 2013 to July 2014) is already published (Werkenthin et al., 2016).
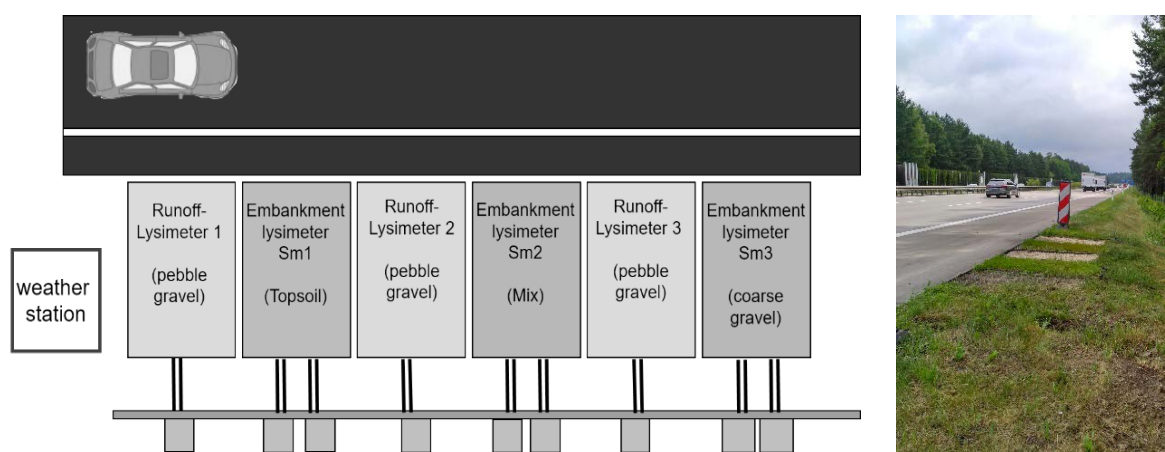
Figure 9.1 Schematic sketch of the embankment lysimeter and view of the monitoring site at the Highway A115, Germany

9.2.3. Laboratory analysis

Dissolved concentrations of Cd, Cu, Cr, Ni, Pb and Zn were determined in the soil solution and road runoff. The samples were filtered through a 0.45 µm filter (Satorius; cellulose-acetate) and acidified with 5M HNO_3 to pH < 2. All samples were measured with ICP-OES (Thermo Fisher ICAP 6000) in accordance with DIN EN ISO 11885. Limits of quantification (LOQ) were 0.09 (Cd), 0.3 (Cr), 0.8 (Cu), 0.8 (Ni), 2 (Pb) and 0.1 (Zn) µg L-1. In order to ensure good measurement quality, the maximum permissible deviation for individual metal recovery from the reference material (BRM 06C, Germany) was set to ± 8% (Kluge et al.2014).

9.2.4. Methodology

9.2.4.1. Nonparametric models

Monthly repeated measurements for two years on sampling lysimeters induces a structure in the data that violates the assumption of independence between samples. Therefore, we used generalized additive mixed models (GAMMs), which allow for dependence between samples and for the nonlinearity of covariants (Zuur et al., 2009). GAMMs are an extension of generalized additive models (GLMs) that allow the predictor function to also include random effects.

The GAMM could be described as:

$$g(\mu_{it}) = X_{it}^T \beta + \sum_{j=1}^{m} \alpha_{(j)}(u_{itj}) + Z_{it}^T b_i$$

where g is a link function, $X_{it}^T\beta$ is a linear parametric term with the parameter vector $\beta^T = (\beta_0, \beta_1, ..., \beta_p)$ that includes the intercept; $\sum_{j=1}^{m} \alpha_{(j)}(u_{itj})$ represents the smooth and nonlinear function, it is an additive term with unspecified influence functions $a_{(1)}, ..., a_{(m)}$; and $Z_{it}^T b_i$ contains the cluster-specific random effects.

If no linear component is included, the model is considered to be nonparametric. A model whose predictions consist of both linear and unspecified nonlinear functions of predictor variables is considered to be a semi parametric model.

For GAMM, we used the metal concentrations (log scale) as a response variable and the following predictors as fixed effects: time (represented by a numeric vector from 0 to 743 days after 2012-11-21); EC (microSiemens/cm); pH; rainfall (mm); soil temperature (C°); ; infiltration (l); soil moisture (mm-3/mm-3); surface runoff volume (l); and runoff concentration (microgram/l). The factors of soil type (Sm1, Sm2, and Sm3) and soil depth are also included as fixed effects. While the random variables refer to the points at which the repeated measurements were taken. These points are identified by the combination of soil depth and soil type. Visual inspection of residual plots did not reveal any major deviations from homoscedasticity or normality. Specifically, the metal data was log-transformed to meet the demand for residual normality and homoscedasticity. In all the models, the time covariant was kept to adjust time trends, even when this predictor was not significant.

In the first run of the model, we hypothesized that the response of metal concentrations would have nonlinear relationships with the predictors, as we mentioned above. Using the GAMMs of the MGCV package (Wood, 2006) implemented in R version 3.3.0 (R Core Team, 2017), the GAMM model for the first run could be expressed as:

$$form = as.formula(\log(metal\ concentraion)) \sim (time\ )$$
$$+materila * depth + s(EC) + s(pH) + s(precipitation)$$
$$+s(soil\ temperature) + s(infiltration_{15cm})$$
$$+s(infiltration_{30cm}) + s(soil\ moisture) \qquad (1)$$
$$+s(surface\ runoff\ volume) + s(runoff\ concentratoin)$$

$$mod < -gamm4(form, data, random = \sim(1/ID))$$

where ID refers to the point where we obtain the sample from either the runoff or the soil solution according to the soil type and depth. The s() represents the spline functions that allow the curve to bend in order to describe the observed metal data. We also hypothesize that the metal variation is explained by the depth for each soil type. Therefore, we show in the model the interaction between soil type (material) and depth.

The p-value is used to determine which nonlinear relationship in model (1) can be considered linear and which explanatory variables should be excluded from model (1). When these processes are accompanied by a decrease in the Akaike information criterion (AIC), then the model is preferred. We choose the model when the lower AIC is obtained with significant explanatory variables (Burnham and Anderson, 2004).

9.3.   Results

9.3.1. Descriptive analyses of the dissolved concentrations at the different highway embankments

The raw data of the dissolved soil solution concentrations of the three embankment soils and the runoff concentrations during the measurement period is presented in fig. 2. The concentrations at 30 cm depths of the three different embankment soils often show higher ranges than at shallow depths (Fig. 2).

The soil solution concentrations of SM1 show the highest range at 15 cm and decrease at 30 cm, while the concentration range at SM2 increases with depth (Tab.1). The median dissolved metal concentrations in soil solution are highest in the SM1 embankment at soil depths of 15 and 30 cm when compared with SM2 and SM3, with the exception of Cd (Tab. 1). SM3 has the lowest median metal concentration at 30 cm, except for Cr, where it is slightly higher than SM2. The runoff concentrations of Pb, Cd and Ni are mostly lower than the concentrations in the soil solution at 15 and 30 cm of the three different embankment soils (Fig. 2), but they are nearly in the same range for SM2 and SM3.

The metal concentrations of all lysimeters at 30 cm show an increase during the wintertime, whereas the concentrations decrease in spring before increasing again in summertime (Fig.2). Comparing

the three different embankment soil types, the SM1 embankment soil type seems to react as a source for these metals.

Figure 9.2 Raw data of the dissolved soil solution concentrations of the three different embankment soils and the runoff concentrations during the measurement period

The descriptive analysis mentioned above visually confirms the existence of variation in metal concentration according to soil type and depth over time. To understand these metal variations and the influence of other environmental factors, an advanced statistical method will be used: the generalized additive mixed model (GAMM), which we are going to explain in the following sections.

Table 9.1 Median, min and max dissolved metal concentrations for three embankment soils at three different depths during the measurement period

	Depth (cm)	Cd	Cr	Cu	Ni	Pb	Zn
SM1	0	0.13 (0.01-0.53)	3.360 (1.33-15.03)	15.88 (7.55-45.70)	1.40 (0.71-4.97)	0.80 (0.01-3.80)	30.35 (6.55-233.90)
	15	0.15 (0.01-0.66)	4.55 (1.31-12.75)	35.85 (7.10-163.40)	2.60 (0.01-24.97)	3.80 (0.30-52.80)	24.45 (4.45-387.55)
	30	0.30 (0.01-4.06)	7.01 (0.1-19.05)	73.60 (9.60-393.60)	3.94 (1.38-35.20)	11.20 (0.01-59.84)	42.40 (9.50-96.65)
SM2	0	0.10 (0.01-1.75)	4.77 (1.90-9.20)	16.90 (5.55-123.20)	1.80 (0.01-8.10)	1.02 (0.01-3.80)	41.55 (17.40-151.35)
	15	0.14 (0.01-0.55)	4.23 (1.20-12.06)	13.43 (6.70-50.30)	2.53 (0.01-22.42)	0.74 (0.01-5.10)	25.18 (5.95-72.30)
	30	0.22 (0.01-1.34)	4.12 (1.30-18.15)	6.55 (0.10-26.36)	3.20 (0.01-16.32)	1.89 (0.01-8.20)	6.60 (0.20-49.30)
SM3	0	0.08 (0.01-0.36)	3.60 (0.71-13.31)	12.50 (7.10-44.90)	1.80 (0.82-10.73)	0.57 (0.01-4.50)	27 (6.40-412.10)
	15	0.30 (0.01-2.15)	3.65 (1.59-13.55)	13.90 (4.50-83.80)	3.88 (0.01-50.60)	2.25 (0.01-9.50)	12.30 (3.75-55.95)
	30	0.30 (0.01-1.75)	3.61 (1.25-24.60)	11.80 (3.35-90.70)	5.82 (1.09-29.10)	2.27 (0.01-9.90)	10.70 (0.45-95.55)

9.3.2. *Effects of environmental factors on metal variations according to GAMM*

An assessment of the effects from environmental factors on the variation of dissolved metals at the roadside was conducted for each metal by selecting the best GAMM for each one. The significance analysis of the model shows the explanatory variables for each metal to be the following: time for Cr, Cu, Pb, Ni and Zn; pH for Cd, Cr, Cu, Pb, Ni and Zn; depth for Cd, Cr, Pb, Ni and Zn, EC for Cd and Cu; precipitation for Cu, Ni and Zn; soil type for Cr, Cu and Ni; soil temperature for Pb; soil moisture for Zn; infiltration for Pb, Ni, Cu, Cr and Cd; and runoff concentration for Ni and Cr (Table 3). For all metals, the model showed no significance in the interaction between soil type and depth. Moreover, the surface runoff volume could not explain the metal variation in the soil solution.

The following provides more details about the how the explanatory variables could explain the metal variation investigated in this study over the experiment period.
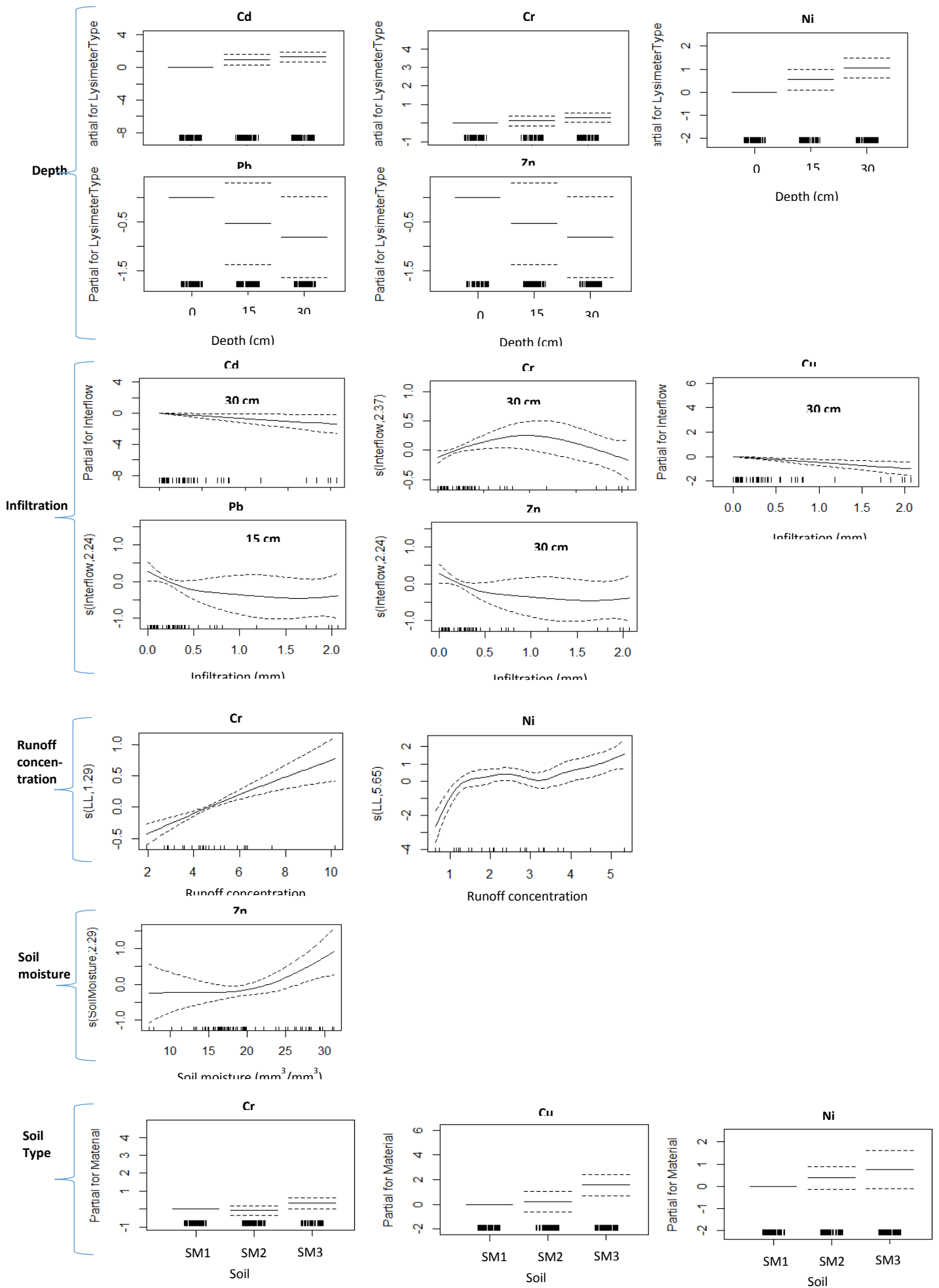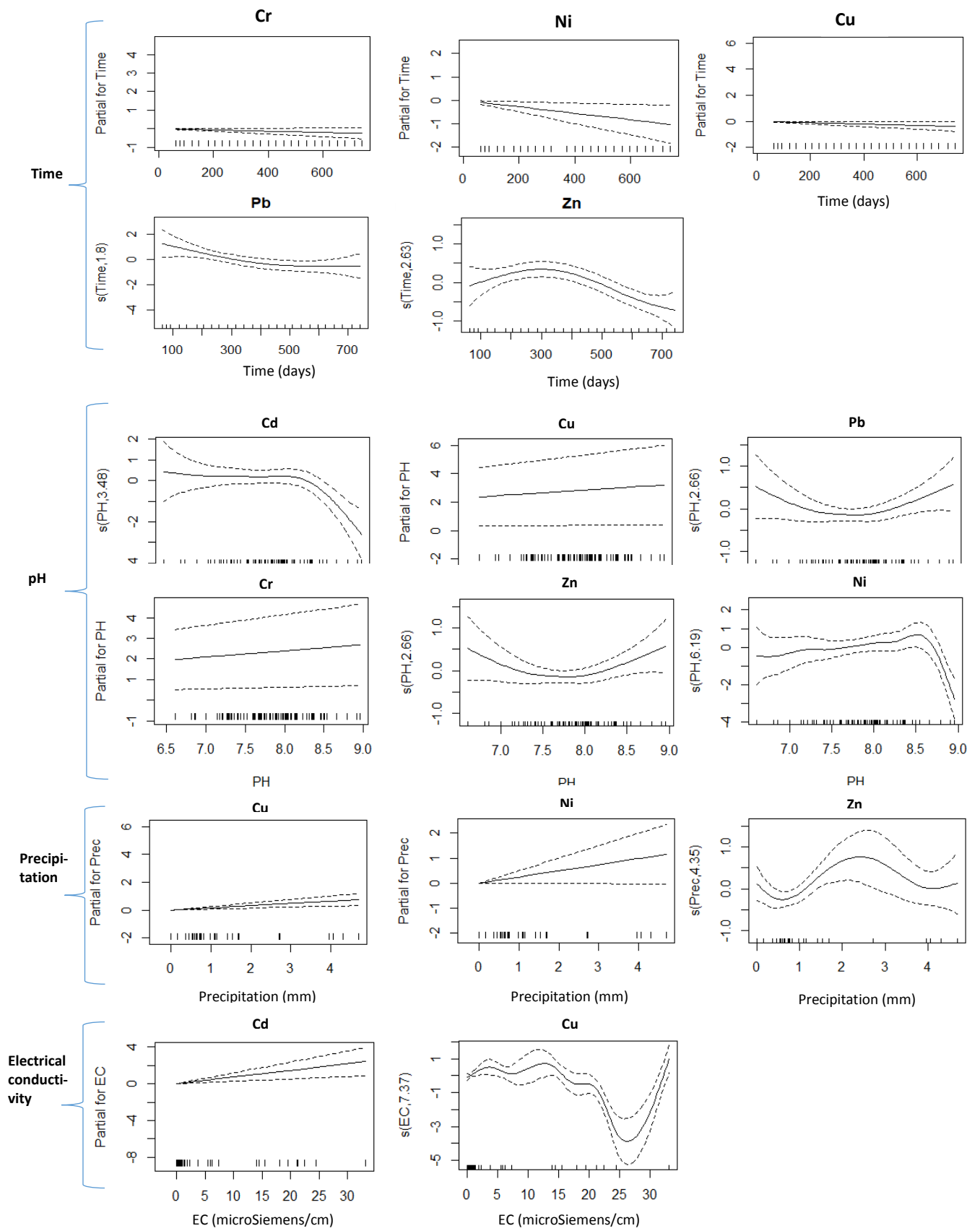
Figure 9.3 Explanatory variables that explain the variation of dissolved metal concentration used in GAMM

9.3.2.1. Time

Taking into account that the constructed embankment needs time to settle, we observed that the concentrations of Cu and Cr somehow decreased linearly after two years of the field experiment when compared with the initial concentrations (Tab. 2). This might be an effect of leaching from the soil matrix (see Tab.3 and Fig.3). On the other hand, the time variable showed a nonlinear effect on the dissolved Pb and Zn concentrations in the soil solutions of the roadside embankments. Pb and Zn have higher solution concentrations at the beginning and lower ends of the sampling period (Fig. 3). In this study, only the Cd variation could not be explained by the time variable.

Table 9.2 Metal concentrations of the embankment soils at the beginning of the field study

	Cd	Cr	Cu	Ni	Pb	Zn	pH	C org
mg*kg^{-1}							_	M.%
Sm1	< 0.2	5	20	<5	25	33	7.67	0.88
Sm2	< 0.2	< 5	5	5	5	19	7.57	0.40
Sm3	< 0.2	< 5	11	6	13	30	7.63	0.13

9.3.2.2. pH

The pH shows linear effects on Cu and Cr concentrations. It is generally agreed that pH is the key factor that affects concentrations of soluble and plant-available metals (Brallier et al., 1996). The metal solubility in soils is inclined to increase at lower pH and decrease at higher pH values, as has been demonstrated in numerous laboratory studies (Tills and Alloway, 1983, Sanchez-Camazano et al., 1994 and Rieuwerts et al., 1998), but this is only partially true for our data, which was generated from a lysimeter field study. We can see in Table 3 that the coefficients of Cr and Cu with pH are positive. The Cu concentrations increase with increasing pH within the range of 6.5 to 9.0. This finding could be explained by soluble metal-organic complexes forming above pH 6 (Brümmer et al. 1986). The organic matter preferably immobilizes Cu in soils. The solid and dissolved fraction of organic matter can substantially affect mobility (Temminghoff, 1997). A calculation by Tipping and Woof (1990) shows that increasing soil pH by 0.5 units could lead to an increase in approximately 50% in the mobilized organic matter. We therefore assume that increase in the Cu increasing with the pH is related to enhanced leaching of DOC. Sauvé et al. (2000) showed that total Cu, pH, and DOC are the most

significant variables controlling Cu solubility in soils. Adsorption of Cr characteristically decreases with increasing pH due to the decrease in the positive surface charge of the soils (Rai et al., 1986; Zachara, 1989; Khaodhiar et al., 2000). Moreover, the influence of pH on the concentrations shows a nonlinear trend for Cd, Zn, Ni and Pb (Fig. 3). In the case of Cd, no changes in the concentrations were observed when pH ranges between 7.0 and 8.2. The concentrations of Pb decrease with increasing pH. Brümmer et al. (1986) showed the strong relationship between Pb concentrations and soil pH. They described a strong decrease in soil solution concentrations with increasing pH. The correlation was even higher when the concentrations were higher. The Ni concentrations decrease sharply when the pH value is higher than 8.5 (Fig. 3). This may be associated with the start of the hydroxide precipitation reactions with the Ni^{2+} and $NiOH^+$ species (Sen Gupta and Bhattacharyya, 2006). Zn decreased drastically under alkaline conditions and then increased after pH > 8, because soluble metal-organic complexes are formed in this pH range (Brümmer et al. 1986).

It is interesting to see that metals which have a nonlinear relationship with pH also have a nonlinear relationship with the time covariant. pH varies over the year due to rainfall events. Häring et al. (2017) reported that water quantity significantly affected pH over time.

9.3.2.3. Precipitation

The increase in Cu and Ni concentrations with precipitation (Tab. 3 and Fig. 3) might be explained by temporal and spatial variations of dissolved organic carbon (DOC) in runoff water and higher rates of microbial decomposition of organic compounds. High pH (in our study 6-9) and DOC increases the mobility of Cd, Ni, and Zn (Impellitteri et al., 2002). Orlović-Leko et al. (2009) and Pan et al. (2010) reported strong seasonal variations of organic carbon in rainwater samples and explained these findings through a combination of different climatic conditions, i.e.: the amount of precipitation, temperature, humidity, site specific parameters and proximity to anthropogenic and/or biogenic emission sources. Only Zn shows nonlinear behavior with the amount of rainfall. Runoff concentrations of Zn increase during intensive rainfall events (Fig. 2). Explanation: Metal road installations such as crash barriers are generally galvanized with Zinc. Due to corrosion by rain and spray water, Zn in particular as well as Cd and Cu are released into roadside soils (Dierkes and Geiger, 1999; Barbosa and Hvitved-Jacobsen, 1999; Kluge and Wessolek, 2012. They all reported higher Zn concentrations in runoff and soils as a result of the corrosion of crash barriers.

Table 9.3 Best selection GAMMs for each element according to AIC including its coefficient estimates for linear and non-linear (nl), with: Time(T), depth (D), EC (E), pH (P), humidity (H), soil material (S), evapotranspiration (Ev), runoff concentration (R), precipitation (Pr), soil temperature (St), soil moisture (sm). s() refers to spline function (non-linear), i15 and i30 are infiltration rates at 15 cm and 30 cm respectively. The term n.l. refers to non-linear relationship.

	Cd	Cr	Cu	Ni	Pb	Zn
	Fixed effect: s(T)+s(P)+D+H+E+i30	Fixed effect: T+P+D+S+Ev+s(R)+s(i30)	Fixed effect: T+P+S+s(E)+Pr+s(H)+i30	Fixed effect: s(T)+s(P)+D+S+Pr+s(R)+i30	Fixed effect: s(T)+s(P)+D+s(St)+i15	Fixed effect: s(T)+s(P)+D+s(Ev)+s(Pr)+s(sm)+s(i30)
	Random effect: 1/S_D	Random effect: 1/S_D	Random effect: 1/S_D	Random effect: 1/S_D	Random effect: 1/S_D	Random effect: 1/S_D
(Intercept)	0.6416	-0.8517	1.137	1.040	-0.9389	3.511
Time		-0.0003	-0.0005	-0.0014	n.l.	n.l.
EC	0.0721		n.l.			
pH	n.l.	0.2983	0.3587	n.l.	n.l.	n.l.
Precipitation			0.1595	0.2457		n.l.
Soil.temperature					n.l.	
Infiltration_15cm					-0.0638	
Infiltration_30cm	-0.6792	n.l.	-0.4817	-0.4432		n.l.
Soil.moisture						n.l.
SurfaceRunoff						
Runoff.concentration		n.l.		n.l.		
Soil.materialSm2		-0.397	-1.332	-0.3761		
Soil.materialSm3		-0.3162	-1.566	-0.7541		
Depth15	0.9511	0.1228		0.5519	1.411	-0.5298
Depth30	1.273	0.2924		1.046	1.851	-0.8094

9.3.2.4. EC

The variable EC shows linear effects on the soil solution concentrations of Cd. This is easily explained by the strong increase in EC concentrations in runoff and the extended time during the winter period the when the soil solution was exposed to de-icing salts (Fig. 2). The presence of high Cl concentrations leads to a strong decrease in the adsorption of Cd ions in soils (Doner, 1978; Boekhold et al., 1993; Lumsdon et al., 1995). The strong influence of salt concentration on the mobilization of metals has already been observed by many other authors dealing with the topic of road runoff and roadside soils (e.g., Bauske & Goetz, 1993; Amrhein and Strong, 1990, Amrhein et al., 1992; Bäckström et al., 2004, Ramakrishna and Viraraghavan, 2005). Li et al. (2015) determined an extensive mobilization of Cd soil leachate by salts, accounting for about 21% of the total Cd in the soil. Bauske & Goetz (1993) examined concentrations of Cd and Zn that had increased by a factor of 20 in highway soil solutions during the utilization of de-icing salts in winter. Salimi et al. (2015) found that the Cl^- ion formed complex ions with Cd to $CdCl^+$, $CdCl_3^-$, and $CaCl_4^-$. The application of Cl^- might dissolve the adsorbed soil Cd to form complex ions. The regular relationship between EC and metals concentration is positively linear. In this study, the GAMM model shows a nonlinear relationship between EC and Cu (Fig 3). The use of road salt may result in an increased mobilization of Cu due to complexion with chloride ions (Doner, 1978; Lumsdon et al., 1995). An increase in soluble Cu by extraction tests with calcium magnesium acetate solutions has been reported by Elliott and Linn (1987).

The nonlinear findings here between Cu and EC also agree well with Makarychev et al. (2013). They used a conductometric and potentiometric titration of water extracts (Cu, Al and Pb). A nonlinear dependence of electrical conductivity on water extracts was observed with the added concentration of metals. They explained their findings by metal organic complexes forming with the participation of metal ions. Our results show that the highway runoff in wintertime combined with high amounts of road de-icing agents (high EC) might lead to a higher release of sorbed metals from the organic matter of the SM1 and SM2 (Fig.2;Tab.2).

9.3.2.5. Depth

The positive coefficients of Cd, Cr, Ni and Pb at greater depth indicate an increase in the concentration with soil depth, and the negative coefficient of Zn indicates a decrease in Zn concentration with soil depth (Table 3 and Fig. 3). Cd is the most mobile metal in soils because a large proportion is associated with easily exchangeable and carbonated chemical fractions, as several studies have pointed out (Harrison et al., 1981; Gibson and Farmer, 1984; Chlopecka et al., 1996).

9.3.2.6. Soil Type

The coefficients in Table 3 indicate how much the concentrations of Cr, Cu and Ni decrease from Sm1 to Sm2 and from Sm1 to Sm3. This is because the Sm1 has higher organic matter than Sm2 and Sm3 (Table 2), and thus the dissolved Cu concentrations are higher as we explained above. Cu in soils is known to be dominant in the organic and residual phases (Harrison et al., 1981, Gibson and Farmer, 1984; Hamilton et al., 1984; Ramos et al., 1994; Ma and Rao, 1997).

The results in Table 3 regarding soil material agree well with the results found by Turer et al. (2003). They concluded that the mobility of Ni and Cr appears to be controlled by the original soil material rather than by anthropogenic sources.

9.3.2.7. Infiltration volume

There is a negative relationship between Cd, Cu, Ni and Pb and the infiltration volume (Tab. 3 and Fig. 3). Pb was largely associated with particles (90%), as reported in the previous study on the same data (Werkenthin et al., 2016). The dilution could be one reason for a decrease in dissolved Pb concentrations with infiltration volumes at 15 cm.

9.3.2.8. Soil temperature

The soil temperature has a significant nonlinear effect on the Pb, the concentration of Pb increases from 0 to 15 with an increase in temperature and then decreases (Fig. 3). A decrease in dissolved Pb at higher temperatures could be explained by Cornu et al. (2016), who found that the affinity that dissolved organic matter (DOM) has for metals was indeed suspected to increase with soil temperature.

9.3.2.9. Runoff concentration

Runoff concentration has a nonlinear effect on the dissolved Cr and Ni concentrations (Fig.3). The concentrations increase sharply at the beginning, and then continue to increase but slowly. The pH and ionic strength (the total ion concentration in the solution) affect the Ni adsorption. Scheidegger et al. (1996) observed an increasing Ni adsorption at pH values lower than 7, with a decrease in the ionic strength. At pH > 7.0, Ni adsorption also seems to be affected by the ionic strength. This finding could explain the nonlinear relationship that Ni and Cr have with runoff concentration. Moreover, as Turer et al. (2003) reported that soil material has a significant effect on dissolved Ni and Cr, the runoff concentrations of dissolved Ni and Cr affect the mobility of dissolved Ni and Cr in roadside soils.

In general, the results show that time and pH explanatory variables have the most significant effects on the variation of dissolved metal concentrations in the soil solutions studied here. Regarding the issue of choosing an adequate embankment soil that provides enough safety and metal retardation to

prevent groundwater contamination, GAMM allows distinguishing between three roadside soil types in order to establish which has the best possible retardation. The model demonstrated that the embankment soils with higher OM show higher dissolved soil solution concentrations of Ni and Cr than the embankment soils with lower OM. Therefore, the OM content of the embankment soils affect the solubility of Ni, Cr and Cu, because they are mostly bounded at the OM fraction and thus might be more easily remobilized. The higher Cu concentrations in our study might be due to the higher initial concentrations of the soil at the time when we began setting up the field lysimeter experiment (Table 2). Moreover, the runoff concentrations do not significantly affect Cd, Pb, Cu and Zn concentrations in the soil solution; only the Ni and Cr concentrations in runoff correlated with the dissolved Ni and Cr concentrations in roadside soil solutions. In addition, the modeling approach shows that the metal variation can be explained by the depth: as the depth increases from 0 to 30 cm, the concentration of dissolved metals increases (Fig. 3). Table 1 shows that, with the exception of Cr, the SM3 at 30 cm depth has the lowest concentration of dissolved metals.

Our model results mostly agree with other studies that used controlled laboratory experiments to investigate individual metal behaviors. Furthermore, partially higher Ni, Pb and Cr concentrations in soil solutions at increasing soil depths suggest that a metal translocation occurs from the upper to the lower horizons within the soil profiles, thus leading also to the assumption that transport hot spots exist.

9.4.   Conclusions

This study assesses the effects of different environmental factors on the variation of dissolved metals in roadside soils. Monthly data of dissolved Cd, Ni, Cr, Pb, Cu and Zn in the soil solution concentrations of three different roadside soils were measured at different depths over a two-year period in a lysimeter field study.

Generalized Additive Mixed Models (GAMMs) have been used to interpret the multiple data derived from a lysimeter field study at a roadside. The chosen approach enabled us to explore the complex behavior of metals in heterogeneous soil systems by detecting the linear and nonlinear dynamics of metals in soil solutions.

Regarding the EC, the GAMM model allows us to show only a nonlinear relationship between EC and Cu. Generally, metal solubility tends to increase at lower pH and decrease at higher pH values. However, this is only partly true for our data, which are gained from a lysimeter field study. The GAMM model results show only that Cu and Cr soil solution concentrations have a positive linear relationship with pH. In our case, the modeling approach showed that pH and soil material composition play the most important roles in terms of metal release in roadside soil. The study shows that variations of metal concentrations can be explained by the depth: as the depth increases from the surface to 30 cm, the metal

concentrations increase, with SM3 having the lowest median of dissolved concentrations of Cd, Ni, Pb, Cu and Zn at 30 cm. In our case, SM3 has the lowest content of organic carbon and is the most effective embankment material for reducing soil water pollution from dissolved metal concentrations. This is a surprising finding, because existing recommendations suggest using twenty or more centimeters of thick topsoil with a high ratio of organic carbon to ensure high sorption capacity for infiltrated runoff. The highway runoff in wintertime and high amounts of road de-icing agents might lead to a higher release of sorbed metals from the organic carbon of the SM1 and SM2 when combined with higher amounts of organic matter.

One interesting matter shown by the model is that the surface runoff volume does not affect the metal variation in soil solutions.

Using the GAMM for interpreting heterogeneous roadside soil solution data has two main strengths. First, the mixed part of the model takes into account the dependencies in the data. This is because our measurements were repeated and thus induced a structure in the data that would violate the assumption of independence between samples. Second, due to the nature of the irregular fluctuations of metal concentrations over time, polynomials fail to describe the observed curvature in the data. By including smoothing functions for the explanatory variables, e.g., spline functions, the additive part of the model allows the curve to bend in order to describe the observed data.

Overall, when using GAMM to analyze environmental data such as soil solution concentrations at the field scale, it gives a strong advantage to assessing the variation of metal translocation that is affected by different environmental factors.

Acknowledgements

9.5.   References

1. Amrhein, C., Strong, J. E. (1990). The effect of deicing salts on trace metal mobility in roadside soils. Journal of Environmental Quality, 19(4), 765-772.

2. Amrhein, C., Strong, J. E., & Mosher, P. A. (1992). Effect of deicing salts on metal and organic matter mobilization in roadside soils. Environmental Science & Technology, 26(4), 703-709.

3. Arslan, H., & Gizir, A. M. (2006). Heavy-metal content of roadside soil in Mersin, Turkey. Fresenius Environmental Bulletin, 15(1), 15-20.

4. Bäckström, M., Karlsson, S., Bäckman, L., Folkeson, L., & Lind, B. (2004). Mobilisation of heavy metals by deicing salts in a roadside environment. Water research, 38(3), 720-732.

5. Barbosa, A. E., Hvitved-Jacobsen, T. (1999). Highway runoff and potential for removal of heavy metals in an infiltration pond in Portugal. Science of the Total Environment, 235, 151–159.

6. Bauske, B., & Goetz, D. (1993). Effects of Deicing-Salts on Heavy Metal Mobility Zum Einfluß von Streusalzen auf die Beweglichkeit von Schwermetallen. CLEAN–Soil, Air, Water, 21(1), 38-42.

7. Bizikova, L., Roy, D., Swanson, D., Venema, H. D., & McCandless, M. (2013). The water-energy-food security nexus: Towards a practical planning and decision-support framework for landscape investment and risk management. Winnipeg, Manitoba: International Institute for Sustainable Development.

8. Boekhold, A. E., Temminghoff, E. J. M., & Zee, S. E. A. T. M. (1993). Influence of electrolyte composition and pH on cadmium sorption by an acid sandy soil. European Journal of Soil Science, 44(1), 85-96.

9. Brallier, S., Harrison, R. B., Henry, C. L., & Dongsen, X. (1996). Liming effects on availability of Cd, Cu, Ni and Zn in a soil amended with sewage sludge 16 years previously. Water, Air, and Soil Pollution, 86(1-4), 195-206.

10. Brümmer, G. W. (1986). Heavy metal species, mobility and availability in soils. In The importance of chemical "speciation" in environmental processes (pp. 169-192). Springer, Berlin, Heidelberg.

11. Casey, R. E., Simon, J. A., Atueyi, S., Snodgrass, J. W., Karouna-Reiner, N., & Sparling, D. W. (2006). Temporal trends of trace metals in sediment and invertebrates from stormwater management ponds. Water, Air, and Soil Pollution, 178, 69–77.

12. Chlopecka, A., Bacon, J. R., Wilson, M. J., & Kay, J. (1996). Forms of cadmium, lead, and zinc in contaminated soils from southwest Poland. Journal of Environmental Quality, 25(1), 69-79.

13. Cong Men, Ruimin Liu, Fei Xu, Qingrui Wang, Lijia Guo, Zhenyao Shen, Cornu, J. Y., Denaix, L., Lacoste, J., Sappin-Didier, V., Nguyen, C., & Schneider, A. (2016). Impact of temperature on the dynamics of organic matter and on the soil-to-plant transfer of Cd, Zn and Pb in a contaminated agricultural soil. Environmental Science and Pollution Research, 23(4), 2997-3007.

14. Dierkes, C., Geiger, W. F. (1999). Pollution retention capabilities of roadside soils. Water Science and Technology, 39(2), 201–208.

15. DIN EN ISO 11885. (1998). Determination of 33 Elements by Inductively Coupled Plasma Atomic Emission Spectrosopy.

16. Doner, H. E. (1978). Chloride as a Factor in Mobilities of Ni (II), Cu (II), and Cd (II) in Soil 1. Soil Science Society of America Journal, 42(6), 882-885.

17. Elliott, H. A., & Linn, J. H. (1987). Effect of Calcium Magnesium Acetate on Heavy Metal Mobility in Soils 1. Journal of environmental quality, 16(3), 222-226.

18. Fay, L., and X. Shi. 2012. Environmental impacts of chemicals for snow and ice control: State of the knowledge. Water, Air, & Soil Pollution 223: 2751–2770.

19. Fitschen A, Nordmann H (2012) Traffic development on federal highways 2010. Report No. V 223. Federal Highway Research Institute, p. 260 (in German).

20. Gibson, M. J., & Farmer, J. G. (1986). Multi-step sequential chemical extraction of heavy metals from urban soils. Environmental Pollution Series B, Chemical and Physical, 11(2), 117-135.

21. Green, S. M., Machin, R., & Cresser, M. S. (2008). Effect of long-term changes in soil chemistry induced by road salt applications on N-transformations in roadside soils. Environmental pollution, 152(1), 20-31.

22. Grigoratos, T., & Martini, G. (2015). Brake wear particle emissions: a review. Environmental Science and Pollution Research, 22(4), 2491-2504.

23. Guo.G.H., Wu.F.C., F.Z. Xie.F.Z., Zhang.R.Q. (2012). Spatial distribution and pollution assessment of heavy metals in urban soils from southwest ChinaJ. Environ. Sci., 24 (3) (2012), pp. 410–418.

24. Gupta, S.S., K.G. Bhattacharyya,K. G., 2006. Adsorption of Ni(II) on clays. J. Colloid Int. Sci., 295 (2006), pp. 21-32.

25. Hamilton, R. S., Revitt, D. M., & Warren, R. S. (1984). Levels and physico-chemical associations of Cd, Cu, Pb and Zn in road sediments. Science of the Total Environment, 33(1-4), 59-74.

26. Häring, V., Manka'abusi, D., Akoto-Danso, E. K., Werner, S., Atiah, K., Steiner, C., ... & Marschner, B. (2017). Effects of biochar, wastewater irrigation and fertilization on soil properties in West African urban agriculture. Scientific Reports, 7(1), 10738.

27. Harrison, R. M., Laxen, D. P., & Wilson, S. J. (1981). Chemical associations of lead, cadmium, copper, and zinc in street dusts and roadside soils. Environmental Science & Technology, 15(11), 1378-1383.

28. Hjortenkrans, D. S. T., Bergbäck, B. G., & Häggerud, A. V. (2007). Metal emissions from brake linings and tires: case studies of Stockholm, Sweden 1995/1998 and 2005. Environmental Science & Technology, 41, 5224–5230.

29. Huber, M., Welker, A. Brigitte, Helmreich, B. (2016). Critical review of heavy metal pollution of traffic area runoff: Occurrence, influencing factors, and partitioning, Science of The Total Environment 541,895-919.

30. Ignatavicius, G., Valskys, V., Bulskaya, I., Paliulis, D., Zigmontiene, A., & Satkunas, J. (2017). Heavy metal contamination in surface runoff sediments of the urban area of Vilnius, Lithuania. Estonian journal of earth sciences, 66(1).

31. Impellitteri, C.A., Lu, Y., Saxe, J.K., Allen, H.E., Peijnenburg, W.J.G.M. (2002). Correlation of the partitioning of dissolved organic matter fractions with the desorption of Cd, Cu, Ni, Pb and Zn from 18 Dutch soils. Environment International 28, 401–410.

32. Kalbitz, K. & Wennrich, R. (1998). Mobilization of heavy metals and arsenic in polluted wetland soils and its dependence on dissolved organic matter. Science of the Total Environment, 209, 27-39.

33. Kayhanian, M., Fruchtman, B. D., Gulliver, J. S., Montanaro, C., Ranieri, E., & Wuertz, S. (2012). Review of highway runoff characteristics: Comparative analysis and universal implications. Water research, 46(20), 6609-6624.

34. Khaodhiar, S., Azizian, M. F., Osathaphan, K., & Nelson, P. O. (2000). Copper, chromium, and arsenic adsorption and equilibrium modeling in an iron-oxide-coated sand, background electrolyte system. Water, air, and soil pollution, 119(1-4), 105-120.

35. Khoshgoftar, A. H., Shariatmadari, H., Karimian, N., Kalbasi, M., Van der Zee, S. E. A. T. M., & Parker, D. R. (2004). Salinity and zinc application effects on phytoavailability of cadmium and zinc. Soil Science Society of America Journal, 68(6), 1885-1889.

36. Kluge, B., Wessolek, G. (2012). Heavy metal pattern and solute concentration in soils along the oldest highway of the world–the AVUS Autobahn. Environmental Monitoring and Assessment, 184(11), 6469-6481.

37. Kocher, B., Wessolek, G., & Stoffregen, H. (2005). Water and heavy metal transport in roadside soils. Pedosphere, 15(6), 746–753.

38. Legret, M., & Pagotto, C. (1999). Evaluation of pollutant loadings in the runoff waters from a major rural highway. Science of the Total Environment, 235(1-3), 143-150.

39. Lequy, E., Dubos, N., Witté, I., Pascaud, A., Sauvage, S., & Leblond, S. (2017). Assessing temporal trends of trace metal concentrations in mosses over France between 1996 and 2011: A flexible and robust method to account for heterogeneous sampling strategies. Environmental pollution, 220, 828-836.

40. Li, F., Zhang, Y., Fan, Z., & Oh, K. (2015). Accumulation of de-icing salts and its short-term effect on metal mobility in urban roadside soils. Bulletin of environmental contamination and toxicology, 94(4), 525-531.

41. Liu, A., Gunawardana, C., Gunawardena, J., Egodawatta, P., Ayoko, G. A., & Goonetilleke, A. (2016). Taxonomy of factors which influence heavy metal build-up on urban road surfaces. Journal of hazardous materials, 310, 20-29.

42. Lumsdon, D. G., Evans, L. J., & Bolton, K. A. (1995). The influence of pH and chloride on the retention of cadmium, lead, mercury, and zinc by soils. Soil and Sediment Contamination, 4(2), 137-150.

43. Ma, L. Q., & Rao, G. N. (1997). Chemical fractionation of cadmium, copper, nickel, and zinc in contaminated soils. Journal of Environmental Quality, 26(1), 259-264.

44. Makarychev, I. P.AU., Motuzova, G. V. (2013).Complexation between metal ions and water soluble organic soil substances as analyzed by electrochemical techniques. Moscow University Soil Science Bulletin. Allerton Press, Vol. 68, No. 1, pp. 41–47.

45. Men, C., Liu, R., Xu, F., Wang, Q., Guo, L., & Shen, Z. (2018). Pollution characteristics, risk assessment, and source apportionment of heavy metals in road dust in Beijing, China. Science of The Total Environment, 612, 138-147.

46. Modrzewska, B., & Wyszkowski, M. (2014). Trace metals content in soils along the state road 51 (northeastern Poland). Environmental monitoring and assessment, 186(4), 2589-2597.

47. Münch, D. (1993). Concentration profiles of arsenic, cadmium, chromium, copper, lead, mercury, nickel, zinc, vanadium and polynuclear aromatic hydrocarbons (PAH) in forest soil beside an urban road. Science of the Total Environment, 138(1-3), 47-55.

48. OECD (2013), Environment at a Glance 2013: OECD Indicators, OECD Publishing, Paris, https://doi.org/10.1787/9789264185715-en.

49. Orlović-Leko, P., Plavšić, M., Bura-Nakić, E., Kozarac, Z., & Ćosović, B. (2009). Organic matter in the bulk precipitations in Zagreb and Šibenik, Croatia. Atmospheric Environment, 43(4), 805-811.

50. Ozkutlu, F., Ozturk, L., Erdem, H., McLaughlin, M., & Cakmak, I. (2007). Leaf-applied sodium chloride promotes cadmium accumulation in durum wheat grain. Plant and soil, 290(1-2), 323-331.

51. Padoan, E., Romè, C., & Ajmone-Marsan, F. (2017). Bioaccessibility and size distribution of metals in road dust and roadside soils along a peri-urban transect. Science of The Total Environment, 601, 89-98.

52. Pagotto, C., Remy, N., Legret, M., & Le Cloirec, P. (2001). Heavy metal pollution of road dust and roadside soil near a major rural highway. Environmental Technology, 22(3), 307-319.

53. Pal, S., Wallis, S., & Arthur, S. (2011). Assessment of heavy metals emission from traffic on road surfaces. Open Chemistry, 9(2), 314-319.

54. Pan, Y., Wang, Y., Xin, J., Tang, G., Song, T., Wang, Y., ... & Wu, F. (2010). Study on dissolved organic carbon in precipitation in Northern China. Atmospheric Environment, 44(19), 2350-2357.

55. Rai, D., J.M. Zachara, L.E. Eary, D.C. Girvin, D.A. Moore, C.T. Resch, B.M. Sass and R.L. Schmidt (1986). Geochemical Behavior of Chromium Species. Electric Power ResearchInstitute, Palo Alto, California, EA-4544.

56. Ramakrishna, D. M., & Viraraghavan, T. (2005). Environmental impact of chemical deicers–a review. Water, Air, and Soil Pollution, 166(1-4), 49-63.

57. Ramos, L., Hernandez, L. M., & Gonzalez, M. J. (1994). Sequential fractionation of copper, lead, cadmium and zinc in soils from or near Donana National Park. Journal of environmental Quality, 23(1), 50-57.

58. Raulinaitis, M., Ignatavičius, G., Sinkevičius, S., & Oškinis, V. (2012). Assessment of heavy metal contamination and spatial distribution in surface and subsurface sediment layers in the northern part of Lake Babrukas. Ekologija, 58(1).

59. Reinirkens, P. (1996). Analysis of emissions through traffic volume in roadside soils and their effects on seepage water. Science of the total environment, 189, 361-369.

60. Rieuwerts, J. S., Thornton, I., Farago, M. E., & Ashmore, M. R. (1998). Factors influencing metal bioavailability in soils: preliminary investigations for the development of a critical loads approach for metals. Chemical Speciation & Bioavailability, 10(2), 61-75.

61. Salimi, M., Amin, M. M., Ebrahimi, A., Ghazifard, A., & Najafi, P. (2012). Influence of electrical conductivity on the phytoremediation of contaminated soils to Cd 2+ and Zn 2+. International Journal of Environmental Health Engineering, 1(1), 11.

62. Sanchez-Camazano, M., Sanchez-Martin, M. J., & Lorenzo, L. F. (1994). Lead and cadmium in soils and vegetables from urban gardens of Salamanca (Spain). Science of the Total Environment, 146, 163-168.

63. Sauve, S., Hendershot, W., & Allen, H. E. (2000). Solid-solution partitioning of metals in contaminated soils: dependence on pH, total metal burden, and organic matter. Environmental science & technology, 34(7), 1125-1131.

64. Scheidegger, A. M., Sparks, D. L., & Fendorf, M. (1996). Mechanisms of nickel sorption on pyrophyllite: Macroscopic and microscopic approaches. Soil Science Society of America Journal, 60(6), 1763-1772.

65. Strobel, B.W., Hansen, H.C.B., Borggaard, O.K., Andersen, M.K. & Raulund-Rasmussen, K. (2001). Cadmium and copper release kinetics in relation to afforestation of cultivated soil. Geochimica et Cosmochimica Acta, 65, 1233-1242.

66. Temminghoff, E.J.M., Van Der Zee, S.E.A.T.M. & De Haan, F.A.M. (1998). Effects of dissolved organic matter on the mobility of copper in a contaminated sandy soil. European Journal of Soil Science, 49, 617-628.

67. Thorpe, A., & Harrison, R. M. (2008). Sources and properties of non-exhaust particulate matter from road traffic: a review. Science of the total environment, 400(1-3), 270-282.

68. Tills, A.R. and Alloway, B.J. (1983). The speciation of lead in soil solution from very polluted soils. EnvironmentalTechnology Letters,4, 529–534.

69. Tipping, E., & Woof, C. (1990). Humic substances in acid organic soils: modelling their release to the soil solution in terms of humic charge. European Journal of Soil Science, 41(4), 573-586.

70. Turer, D. G., & Maynard, B. J. (2003). Heavy metal contamination in highway soils. Comparison of Corpus Christi, Texas and Cincinnati, Ohio shows organic matter is key to mobility. Clean Technologies and Environmental Policy, 4(4), 235-245.

71. Van Bohemen, H. D., & Van de Laak, W. J. (2003). The influence of road infrastructure and traffic on soil, water, and air quality. Environmental Management, 31(1), 0050-0068.

72. Walraven, N., Van Os, B. J. H., Klaver, G. T., Middelburg, J. J., & Davies, G. R. (2014). The lead (Pb) isotope signature, behaviour and fate of traffic-related lead pollution in roadside soils in The Netherlands. Science of the Total Environment, 472, 888-900.

73. Wang, Y. J., Chen, C. F., & Lin, J. Y. (2013). The Measurement of dry deposition and surface runoff to quantify urban road pollution in Taipei, Taiwan. International journal of environmental research and public health, 10(10), 5130-5145.

74. Ward, N. I. (1990). Multielement contamination of British motorway environments. Science of the Total Environment, 93, 393-401.

75. Werkenthin, M., Kluge, B., & Wessolek, G. (2014). Metals in European roadside soils and soil solution–A review. Environmental Pollution, 189, 98-110.

76. Wuana, R. A., & Okieimen, F. E. (2011). Heavy metals in contaminated soils: a review of sources, chemistry, risks and best available strategies for remediation. Isrn Ecology, 2011.

77. Yan, X., Gao, D., Zhang, F., Zeng, C., Xiang, W., & Zhang, M. (2013). Relationships between heavy metal concentrations in roadside topsoil and distance to road edge based on field observations in the Qinghai-Tibet Plateau, China. International journal of environmental research and public health, 10(3), 762-775.

78. Zachara, J. M., Ainsworth, C. C., Cowan, C. E., & Resch, C. T. (1989). Adsorption of chromate by subsurface soil horizons. Soil Science Society of America Journal, 53(2), 418-428.

10

Conclusiones

Las principales contribuciones de la presente tesis se pueden agrupar en dos bloques, correspondientes a las dos partes en que está estructurada:

Aportaciones metodológicas a la estadística

- En el contexto de los modelos mixtos, se explora la eficiencia de métodos de remuestreo para mejorar la inferencia sobre los parámetros, principalmente ligados a la estructura de varianzas.

- En concreto, se consideran métodos de remuestreo semiparamétricos (*residual resampling*), ajustado a la estimación y la variante Wild Bootstrap. Se realiza un escenario de simulación en que se aplican especificaciones no gaussianas tanto para los efectos aleatorios como para los residuos. Se consideran distribuciones alternativas en asimetría y Kurtosis. Los principales resultados indican que no parece existir un método de remuestreo mejor en todos los escenarios.

- Como consecuencia del análisis de los resultados de la anterior simulación, se plantea el desarrollo de una herramienta visual que permita explorar el efecto del tamaño de muestra en la potencia de los tests de bondad de ajuste. Se define e implementa un mosaico gráfico en que las celdas se asimilan a distribuciones de una familia de distribuciones parametrizadas según los parámetros de asimetría y Kurtosis.

- En base a esta herramienta, se aplica inicialmente en la representación del error de tipo 2 asociado a tests de bondad de ajuste a la distribución Normal.

- Como segunda contribución de esta herramienta, se compara la potencia de diferentes alternativas clásicas para la realización de los tests de normalidad. Respecto a la familia de distribuciones considerada, el test que presenta una mayor potencia es el de Shapiro-Wilks.

- El siguiente desarrollo se centra en el objetivo inicial para el que fue diseñado, que consiste en visualizar el conjunto de distribuciones del mosaico compatibles con una muestra dada. Se observa que el tamaño de la muestra es claramente determinante de la extensión en el mosaico de las distribuciones compatibles con la muestra.

Aplicaciones de técnicas estadísticas en hidrología

Como resultado de una colaboración con un grupo de expertos en el ámbito de la hidrología, se plantea la aplicación de técnicas estadísticas avanzadas para resolver problemáticas concretas de los datos que analizan.

Inicialmente, ante la necesidad de obtener predicciones del contenido de agua y salinidad de terrenos de regadío, se plantea la utilización de modelos de series temporales ARIMA pero incluyendo extensiones por tratamiento de atípicos, análisis de intervención y función de transferencia. El objetivo es conseguir modelos válidos que permitan obtener predicciones de contenido de agua en el terreno para automatizar el riego.

- Los modelos ARIMA se basan en reproducir con modelos matemáticos la estructura de autocorrelación.

- El tratamiento de atípicos permite mejorar los modelos incorporando componentes asociados a eventos no esperados (episodios de lluvia).

- El análisis de intervención permite evaluar e incorporar en el modelo componentes correspondientes a diseños experimentales (eventos de riego).

- Finalmente, la relación espacial de las series medidas a diferentes profundidades permite mejorar la precisión de las predicciones mediante la incorporación de funciones de transferencia en el modelo.

- El desarrollo de estos métodos fue objeto de un curso impartido en la conferencia EGU 2016 (European Geosciences Union 2016, Viena) "*Modelling soil water time series data using R program*"

Otro problema tratado se basa en la obtención de la serie temporal de salinidad (conductividad aparente) a partir de la medición de la conductividad de la componente acuosa del terreno. Habitualmente los sensores utilizan una relación física entre ambas variables (modelo de Hilhorst) que precisa de la imputación de un parámetro denominado offset. Este

parámetro depende de la composición y estado del terreno, por lo que se suele utilizar un valor estándar resultante de experimentación en laboratorio.

- Se propone la definición de un espacio de estado para el sistema, que incluye el offset y la salinidad como variables de estado no observadas, siendo la conductividad del agua la serie observada. El modelo resultante es un modelo lineal dinámico.

- El filtro de Kalman permite evaluar la verosimilitud del modelo a partir de la serie observada y por tanto, en combinación con un método numérico de optimización es posible obtener el valor de los parámetros del modelo de máxima verosimilitud.

- Una vez estimado el modelo, la aplicación del alisado procedente del filtro de Kalman permite obtener el valor del offset y la serie de salinidad de mayor verosimilitud.

- Este procedimiento está actualmente en proceso de patente

Por último, se dispone de datos provenientes de un diseño experimental que recoge datos de contenido de metales pesados en la proximidad de una autopista en Alemania. Así mismo, se recogen variables climatológicas y ambientales de la zona. La recogida de datos no se realiza de forma equiespaciada y el objetivo del estudio se centra en la influencia de las covariables en la abundancia de dichos metales.

- En este caso, la técnica propuesta para el análisis de las medidas repetidas no se basa en modelos de serie temporal sino que se proponen el ajuste mediante modelos GAMM.

- Los modelos GAM permiten relaciones flexibles entre covariables y la respuesta.

- Es habitual incluir una componente temporal para ajustar la autocorrelación de los datos.

- Puesto que se recogen datos en diferentes puntos, es conveniente añadir una componente aleatoria para trabajar con los datos agrupados por la localización (modelo mixto).

Futuras extensiones

- Se pretende avanzar en la implementación y evaluación de métodos bootstrap tanto en modelos lineales mixtos como en modelos lineales generalizados mixtos.

- En estos momentos está en fase de revisión un cuarto artículo sobre el mosaico de distribuciones unimodales, en el que los parámetros de localización y escala se estiman a partir de la muestra (en el artículo incluido en la presente tesis, estos parámetros se consideraban conocidos).

- La colaboración con el equipo de la Universidad Técnica de Berlín continua con el análisis de nuevos conjuntos de datos. En concreto, el siguiente problema se basa en la modelización del proceso de evaporación de agua en función del tipo de pavimento.

Bibliografía

Agresti, Alan, Brian Caffo, and Pamela Ohman-Strickland. 2004. Examples in Which Misspecification of a Random Effects Distribution Reduces Efficiency, and Possible Remedies. Computational Statistics & Data Analysis 47: 639–53.

Aldor-Noiman, S., Brown, L.D., Buja, A., Rolke, W. and Stine, R.A. 2013, The Power to See: A New Graphical Test of Normality, The American Statistician, 67, 249-260.

Aljoumani B, Sànchez-Espigares JA, Cañameras N, Josa R, Monserrat J. 2012. Time series outlier and intervention analysis: irrigation management influences on soil water content in silty loam soil. Agricultural Water Management 111: 105-114. DOI: 10.1016/j.agwat.2012.05.008.

Aljoumani, B.; Sànchez-Espigares, J. A.; Cañameras, N.; Josa, R. 2015. An Advanced Process for Evaluating a Linear Dielectric Constant–Bulk Electrical Conductivity Model Using a Capacitance Sensor in Field Conditions. Hydrological Sciences Journal 2015, 60 (10), 1828–1839.

Aljoumani, B.; Sànchez-Espigares, J. A.; Cañameras, N.; Wessolek, G.; Josa, R. 2018. Transfer Function and Time Series Outlier Analysis: Modelling Soil Salinity in Loamy Sand Soil by Including the Influences of Irrigation Management and Soil Temperature. Irrigation and Drainage 2018, 67 (2), 282–294.

Amrhein, C., Strong, J. E. 1990. The effect of deicing salts on trace metal mobility in roadside soils. Journal of Environmental Quality, 19(4), 765-772.

Amrhein, C., Strong, J. E., & Mosher, P. A. 1992. Effect of deicing salts on metal and organic matter mobilization in roadside soils. Environmental Science & Technology, 26(4), 703-709.

Anderson, J. L. 2001. An Ensemble Adjustment Kalman Filter for Data Assimilation. Monthly Weather Review 2001, 129 (12), 2884–2903.

Anderson, T. W. and D. A. Darling. 1954. A Test of Goodness of Fit. Journal of the American Statistical Association 49(268):765-769.

Arellano-Valle, Reinaldo Boris, Heleno Bolfarine, and Victor H. Lachos. 2005. Skew-Normal Linear Mixed Models. J. Data Sci. 3: 415–38.

Arquedas-Rodriguez, F. R., 2009. Calibrating capacitance sensors to estimate water content, matric potential, and electrical conductivity in soilless substrates. Thesis (PhD)

180

Arslan, H., & Gizir, A. M. 2006. Heavy-metal content of roadside soil in Mersin, Turkey. Fresenius Environmental Bulletin, 15(1), 15-20.

Bäckström, M., Karlsson, S., Bäckman, L., Folkeson, L., & Lind, B. 2004. Mobilisation of heavy metals by deicing salts in a roadside environment. Water research, 38(3), 720-732.

Barbosa, A. E., Hvitved-Jacobsen, T. 1999. Highway runoff and potential for removal of heavy metals in an infiltration pond in Portugal. Science of the Total Environment, 235, 151–159.

Bartko, John. 1966. The Intraclass Correlation Coefficient as a Measure of Reliability. Psychological Reports 19 (September): 3–11. doi:10.2466/pr0.1966.19.1.3.

Bauske, B., & Goetz, D. 1993. Effects of Deicing-Salts on Heavy Metal Mobility Zum Einfluß von Streusalzen auf die Beweglichkeit von Schwermetallen. CLEAN–Soil, Air, Water, 21(1), 38-42.

Beven K, Germann P. 1982. Macropores and water flow in soils. Water Resources Research 18 (5): 1311–1325. DOI: 10.1029/WR018i005p01311.

Bizikova, L., Roy, D., Swanson, D., Venema, H. D., & McCandless, M. 2013. The water-energy-food security nexus: Towards a practical planning and decision-support framework for landscape investment and risk management. Winnipeg, Manitoba: International Institute for Sustainable Development.

Bliss, C. I. 1946. Collaborative comparison of three ratios for the chick assay of vitamin D. Journal of the Association of Official Agricultural Chemists 29:396-408.

Boekhold, A. E., Temminghoff, E. J. M., & Zee, S. E. A. T. M. 1993. Influence of electrolyte composition and pH on cadmium sorption by an acid sandy soil. European Journal of Soil Science, 44(1), 85-96.

Bouksila, F., Persson, M., Berndtsson, R. and Bahri, A., 2008. Soil water content and salinity determination using different dielectric methods in saline gypsiferous soil. Hydrological Science Journal, 53, 253-265. doi:10.1623/hysj.53.1.253.

Box GEP, Jenkins GM, Reinsel GC. 1994. Time series analysis: Forecasting and control, 3rd ed. Englewood Clifs: Prentice Hall.

Brallier, S., Harrison, R. B., Henry, C. L., & Dongsen, X. 1996. Liming effects on availability of Cd, Cu, Ni and Zn in a soil amended with sewage sludge 16 years previously. Water, Air, and Soil Pollution, 86(1-4), 195-206.

Brümmer, G. W. 1986. Heavy metal species, mobility and availability in soils. In The importance of chemical speciation in environmental processes (pp. 169-192). Springer, Berlin, Heidelberg.

Butler, S. M., and T. A. Louis. 1992. Random Effects Models with Non-Parametric Priors. Stat. Med. 11: 1981–2000.

Carrasco, Josep, and Lluis Jover. 2004. Estimating the Generalized Concordance Correlation Coefficient Through Variance Components. Biometrics 59 (January): 849–58. doi:10.1111/j.0006-341X.2003.00099.x.

Carrasco, Josep, Tonya King, and Vernon Chinchilli. 2009. The Concordance Correlation Coefficient for Repeated Measures Estimated by Variance Components. Journal of Biopharmaceutical Statistics 19 (February): 90–105. doi:10.1080/10543400802527890.

Casey, R. E., Simon, J. A., Atueyi, S., Snodgrass, J. W., Karouna-Reiner, N., & Sparling, D. W. 2006. Temporal trends of trace metals in sediment and invertebrates from stormwater management ponds. Water, Air, and Soil Pollution, 178, 69–77.

Cetin, M.; Kirda, C. 2003. Spatial and Temporal Changes of Soil Salinity in a Cotton Field Irrigated with Low-Quality Water. Journal of Hydrology 2003, 272 (1–4), 238–249.

Chaloupka M. 2001. Historical trends, seasonality and spatial synchrony in green sea turtle egg production. Biological Conservation. 101 (3): 263–279. DOI: 10.1016/S0006-3207(00)00199-3.

Chatterjee, Snigdhansu, and Arup Bose. 2005. Generalized Bootstrap for Estimating Equations. Ann. Statist. 33: 414–36.

Chee, Chew-Seng, and Yong Wang. 2013. Estimation of Finite Mixtures with Symmetric Components. Statistics and Computing 23 (2): 233–49. doi:10.1007/s11222-011-9305-5.

Chen C, Liu L. 1993. Joint estimation of model parameters and outlier effects in time series. Journal of the American Statistical Association 88 (421): 284–297. DOI: 10.2307/2290724.Cleveland RB, Cleveland WS, McRae JE, Terpenning I. 1990. STL: a seasonal trend decomposition procedure based on loess. Journal of Official Statistics 6 (1): 3–73.

Chen, Chia-Cheng, and Huiman X. Barnhart. 2013. Assessing Agreement with Intraclass Correlation Coefficient and Concordance Correlation Coefficient for Data with Repeated Measures. Comput. Stat. Data Anal. 60 (April). Amsterdam, The Netherlands, The Netherlands: Elsevier Science Publishers B. V.: 132–45. doi:10.1016/j.csda.2012.11.004.

Chlopecka, A., Bacon, J. R., Wilson, M. J., & Kay, J. 1996. Forms of cadmium, lead, and zinc in contaminated soils from southwest Poland. Journal of Environmental Quality, 25(1), 69-79.

Cong Men, Ruimin Liu, Fei Xu, Qingrui Wang, Lijia Guo, Zhenyao Shen, Cornu, J. Y., Denaix, L., Lacoste, J., Sappin-Didier, V., Nguyen, C., & Schneider, A. 2016. Impact of temperature on the dynamics of organic matter and on the soil-to-plant transfer of Cd, Zn and Pb in a contaminated agricultural soil. Environmental Science and Pollution Research, 23(4), 2997-3007.

Cryer JD, Chan KS. 2008. Time Series Analysis with Applications in R, 2nd ed. New York: Springer. DOI: 10.1007/978-0-387-75959-3.

D'Agostino, Ralph B.; Albert Belanger; Ralph B. D'Agostino, Jr 1990. "A suggestion for using powerful and informative tests of normality" (PDF). The American Statistician. 44 (4): 316–321. doi:10.2307/2684359.

Dallal, G.E. and Wilkinson, L. 1986: An analytic approximation to the distribution of Lilliefors' test for normality. The American Statistician, 40, 294–296.

Davidson, Russell, and Emmanuel Flachaire. 2008. The Wild Bootstrap, Tamed at Last. Journal of Econometrics 146 (1): 162–69.

Davis, C. S. 2002. Statistical Methods for the Analysis of Repeated Measurements. Springer Science & Business Media.

Davison, A. C., and D. V. Hinkley. 1997. Bootstrap Methods and Their Application. Cambridge University Press.

Decagon Devices, 2008. 5TE Water content, EC, and temperature sensors operator'smanual. Version No. 3., Pullman, WA.

Desgagné, A. and de Micheaux, P.L. 2017 A powerful and interpretable alternative to the Jarque–Bera test of normality based on 2nd-power skewness and kurtosis, using the Rao's score test on the APD family, Journal of Applied Statistics, DOI: 10.1080/02664763.2017.1415311.

Dierkes, C., Geiger, W. F. 1999. Pollution retention capabilities of roadside soils. Water Science and Technology, 39(2), 201–208.

Diethelm Wuertz, Tobias Setz and Yohan Chalabi 2017. fBasics: Rmetrics - Markets and Basic Statistics. R package version 3042.89. https://CRAN.R-project.org/package=fBasics

DIN EN ISO 11885. 1998. Determination of 33 Elements by Inductively Coupled Plasma Atomic Emission Spectrosopy.

Doner, H. E. 1978. Chloride as a Factor in Mobilities of Ni (II), Cu (II), and Cd (II) in Soil 1. Soil Science Society of America Journal, 42(6), 882-885.

Elliott, H. A., & Linn, J. H. 1987. Effect of Calcium Magnesium Acetate on Heavy Metal Mobility in Soils 1. Journal of environmental quality, 16(3), 222-226.

Farrell, P.J., Rogers-Stewart, K. 2006, Comprehensive study of tests for normality and symmetry: extending the Spiegelhalter test, Journal of Statistical Computation and Simulation, 76, 803-816.

Fay, L., and X. Shi. 2012. Environmental impacts of chemicals for snow and ice control: State of the knowledge. Water, Air, & Soil Pollution 223: 2751–2770.

Field, Chris, and Alan Welsh. 2007. Bootstrapping Clustered Data. J. R. Stat. Soc. Ser. B Stat. Methodol. 69: 369–90.

Field, Chris, Zhen Pang, and Alan Welsh. 2008. Bootstrapping Data with Multiple Levels of Variation. Canad. J. Statist. 36: 521–39.

Field, Chris, Zhen Pang, and Alan Welsh. 2010. Bootstrapping Robust Estimates for Clustered Data. J. Amer. Statist. Assoc. 105: 1606–16.

Filliben, J.J. 1975: The Probability Plot Correlation Coefficient Test for Normality. Technometrics, Vol. 17, No. 1, pp.: 111-116.

Fitschen A, Nordmann H 2012 Traffic development on federal highways 2010. Report No. V 223. Federal Highway Research Institute, p. 260 (in German).

Fortin JG, Anctil F, Parent L, Bolinder MA. 2010. Aneural network experiment on the site-specific simulation of potato tuber growth in Eastern Canada. Computers and Electronics in Agriculture 73 (2): 126–132. DOI: 10.1016/j.compag.2010.05.011

Friedrich, Sarah, Frank Konietschke, and Markus Pauly. 2016. A Wild Bootstrap Approach for Nonparametric Repeated Measurements. Computational Statistics & Data Analysis 113 (July). doi:10.1016/j.csda.2016.06.016.

Gelb, A., 1974. Applied Optimal Estimation. Cambridge (MA). Massachusetts Institute of Technology Press.

184

Genuchten M, V. 1991. Recent Progress in Modelling Water Flow and Chemical Transport in the Unsaturated Zone. , 08/11-24/91. :. 20 th General Assembly of the International Union of Geodesy and Geophysics, Vienna, Austria 1991, 08/11, 169–183.

Ghassemi, F.; Jakeman, A. J.; Nix, H. A. 1995. Salinisation of Land and Water Resources: Human Causes, Extent, Management and Case Studies.; CAB International: Wallingford, 1995.

Ghidey, Wendimagegn, Emmanuel Lesaffre, and Geert Verbeke. 2010. A Comparison of Methods for Estimating the Random Effects Distribution of a Linear Mixed Model. Statistical Methods in Medical Research 19: 575–600.

Gibbons, J. D. and S. Chakraborti. 2003. Nonparametric Statistical Inference. 4 ed. New York: Marcel Dekker.

Gibson, M. J., & Farmer, J. G. 1986. Multi-step sequential chemical extraction of heavy metals from urban soils. Environmental Pollution Series B, Chemical and Physical, 11(2), 117-135.

Gil Bellosta, C. J. 2011. ADGofTest: Anderson-Darling GoF test. R package (version 0.3). https://CRAN.R-project.org/package=ADGofTest

Goldstein, Harvey. 2011. Multilevel Statistical Models. 4th ed. Wiley.

González-Manteiga, Wenceslao, María José Lombardía, Isabel Molina, Domingo Morales, and Laureano Santamaría. 2008. Analytic and Bootstrap Approximations of Prediction Errors Under a Multivariate Fay-Herriot Model. Computational Statistics & Data Analysis 52: 5242–52.

Green TR, Salas JD, Martinez A, Erskine RH. 2007. Relating crop yield to topographic attributes using spatial analysis neural networks and regression. Geoderma 139 (1-2): 23–37. DOI: 10.1016/j.geoderma.2006.12.004.

Green, S. M., Machin, R., & Cresser, M. S. 2008. Effect of long-term changes in soil chemistry induced by road salt applications on N-transformations in roadside soils. Environmental pollution, 152(1), 20-31.

Grigoratos, T., & Martini, G. 2015. Brake wear particle emissions: a review. Environmental Science and Pollution Research, 22(4), 2491-2504.

Gross, J. and bug fixes by Ligges, U. 2012. nortest: Tests for Normality. R package version 1.0-2. http://CRAN.R-project.org/package=nortest

Grubbs F. 1969. Procedures for detecting outlying observations in samples. Technometrics 11 (1): 1–21. DOI: 10.1080/00401706.1969.10490657.

Guo.G.H., Wu.F.C., F.Z. Xie.F.Z., Zhang.R.Q. 2012. Spatial distribution and pollution assessment of heavy metals in urban soils from southwest ChinaJ. Environ. Sci., 24 (3) 2012, pp. 410–418.

Gupta, S.S., K.G. Bhattacharyya,K. G., 2006. Adsorption of Ni(II) on clays. J. Colloid Int. Sci., 295 2006, pp. 21-32.

Hafen R, Anderson D, Cleveland W, Maciejewski R, Ebert D, Abusalah A, Yakout M, Ouzzani M, Grannis S. 2009. Syndromic surveillance: STL for modeling, visualizing, and monitoring disease counts. BMC Medical Information and Decision Making 9 (21). DOI: 10.1186/1472-6947-9-21.

Hajrasuliha, S.; Baniabbassi, N.; Metthey, J.; Nielsen, D. R. 1980. Spatial Variability of Soil Sampling for Salinity Studies in Southwest Iran. Irrigation Science 1980, 1 (4), 197–208.

Hamilton, R. S., Revitt, D. M., & Warren, R. S. 1984. Levels and physico-chemical associations of Cd, Cu, Pb and Zn in road sediments. Science of the Total Environment, 33(1-4), 59-74.

Häring, V., Manka'abusi, D., Akoto-Danso, E. K., Werner, S., Atiah, K., Steiner, C., ... & Marschner, B. 2017. Effects of biochar, wastewater irrigation and fertilization on soil properties in West African urban agriculture. Scientific Reports, 7(1), 10738.

Harrison, G.A., and G. Brush. 1990. On Correlations Between Adjacent Velocities and Accelerations in Longitudinal Growth Data. Annals of Human Biology 17 (1).

Harrison, R. M., Laxen, D. P., & Wilson, S. J. 1981. Chemical associations of lead, cadmium, copper, and zinc in street dusts and roadside soils. Environmental Science & Technology, 15(11), 1378-1383.

Hedeker, D., and R.D. Gibbons. 2006. Longitudinal Data Analysis. Wiley Series in Probability and Statistics 451. John Wiley & Sons.

Hilhorst, M. A., 2000. A pore water conductivity sensor. Soil Science Society of America Journal 64, 1922-1925, doi:10.2136/sssaj2000.6461922x.

Hjortenkrans, D. S. T., Bergbäck, B. G., & Häggerud, A. V. 2007. Metal emissions from brake linings and tires: case studies of Stockholm, Sweden 1995/1998 and 2005. Environmental Science & Technology, 41, 5224–5230.

Ho, Hsiu J., and Tsung-I Lin. 2010. Robust Linear Mixed Models Using the Skew T Distribution with Application to Schizophrenia Data. Biometrical Journal. Biometrische Zeitschrift 52 (4): 449–69.

Hoeben, R.; Troch, P. A. 2000. Assimilation of Active Microwave Observation Data for Soil Moisture Profile Estimation. Water Resources Research 2000, 36 (10), 2805–2819.

Hoff JC. 1983. A Practical Guide to Box–Jenkins Forecasting. London: Lifetime Learning Publications.

Howell, D. 2013. Fundamental Statistics for the Behavioral Sciences. Cengage Learning.

Huang Y, Lan Y, Thomson SJ., Fang A, Hoffmann WC, Lacey RE. 2010. Development of soft computing and applications in agricultural and biological engineering. Computers and Electronics in Agriculture 71 (2): 107-127. DOI: 10.1016/j.compag.2010.01.001.

Huber, M., Welker, A. Brigitte, Helmreich, B. 2016. Critical review of heavy metal pollution of traffic area runoff: Occurrence, influencing factors, and partitioning, Science of The Total Environment 541,895-919.

Ignatavicius, G., Valskys, V., Bulskaya, I., Paliulis, D., Zigmontiene, A., & Satkunas, J. 2017. Heavy metal contamination in surface runoff sediments of the urban area of Vilnius, Lithuania. Estonian journal of earth sciences, 66(1).

Ilya Gavrilov and Ruslan Pusev 2014. normtest: Tests for Normality. R package version 1.1. https://CRAN.R-project.org/package=normtest

Impellitteri, C.A., Lu, Y., Saxe, J.K., Allen, H.E., Peijnenburg, W.J.G.M. 2002. Correlation of the partitioning of dissolved organic matter fractions with the desorption of Cd, Cu, Ni, Pb and Zn from 18 Dutch soils. Environment International 28, 401–410.

Jacqmin-Gadda, Hélène, Solenne Sibillot, Cécile Proust-Lima, Jean-Michel Molina, and Rodolphe Thiébaut. 2007. Robustness of the Linear Mixed Model to Misspecified Error Distribution. Computational Statistics & Data Analysis 51: 5142–54.

Jarque, C.M. and Bera, A.K. 1987: A Test for Normality of Observations and Regression Residuals. International Statistical Review, 55, 2, pp. 163-172.

Johnson, R. A., and D. W. Wichern. 2007. Applied Multivariate Statistical Analysis. Pearson.

Juergen Gross and Uwe Ligges 2015. nortest: Tests for Normality. R package version 1.0-4. https://CRAN.R-project.org/package=nortest

Kalbitz, K. & Wennrich, R. 1998. Mobilization of heavy metals and arsenic in polluted wetland soils and its dependence on dissolved organic matter. Science of the Total Environment, 209, 27-39.

Kalman, R.E., 1960. A new approach to linear filtering andprediction problems.Transaction of the ASME-Journal of Basic Engineering, 8, 35-45, doi:10.1115/1.3662552.

Kargas George; Persson Magnus; Kanelis George; Markopoulou Ioanna; Kerkides Petros. 2017. Prediction of Soil Solution Electrical Conductivity by the Permittivity Corrected Linear Model Using a Dielectric Sensor. Journal of Irrigation and Drainage Engineering 2017, 143 (8), 04017030. https://doi.org/10.1061/(ASCE)IR.1943-4774.0001210.

Katul, G.G., Wendroth, O., Parlange, M.B., Puente, C.E. and Nielsen,D.R., 1993. Estimation of in situ hydraulic conductivity function from nonlinear filtering theory. Water Resources Research, 29, 1063-1070, doi:10.1029/92WR02593.

Kayhanian, M., Fruchtman, B. D., Gulliver, J. S., Montanaro, C., Ranieri, E., & Wuertz, S. 2012. Review of highway runoff characteristics: Comparative analysis and universal implications. Water research, 46(20), 6609-6624.

Khaodhiar, S., Azizian, M. F., Osathaphan, K., & Nelson, P. O. 2000. Copper, chromium, and arsenic adsorption and equilibrium modeling in an iron-oxide-coated sand, background electrolyte system. Water, air, and soil pollution, 119(1-4), 105-120.

Khazaei J, Naghavi MR, Jahansouz MR, Salimi-Khorshidi G. 2008. Yield estimation and clustering of chickpea genotypes using soft computing techniques. Agronomy Journal 100 (4): 1077–1087.DOI: 10.2134/agronj2006.0244.

Kherad-Pajouh, Sara, and Olivier Renaud. 2014. A General Permutation Approach for Analyzing Repeated Measures Anova and Mixed-Model Designs. Statistical Papers 56 (August): 1–21. doi:10.1007/s00362-014-0617-3.

Khoshgoftar, A. H., Shariatmadari, H., Karimian, N., Kalbasi, M., Van der Zee, S. E. A. T. M., & Parker, D. R. 2004. Salinity and zinc application effects on phytoavailability of cadmium and zinc. Soil Science Society of America Journal, 68(6), 1885-1889.

Kluge, B., Wessolek, G. 2012. Heavy metal pattern and solute concentration in soils along the oldest highway of the world–the AVUS Autobahn. Environmental Monitoring and Assessment, 184(11), 6469-6481.

Kocher, B., Wessolek, G., & Stoffregen, H. 2005. Water and heavy metal transport in roadside soils. Pedosphere, 15(6), 746–753.

Konietschke, Frank, Arne Bathke, Solomon Harrar, and Markus Pauly. 2015. Parametric and Nonparametric Bootstrap Methods for General Manova. Journal of Multivariate Analysis 140 (May). doi:10.1016/j.jmva.2015.05.001.

Lachos, Victor H., Pulak Ghosh, and Reinaldo Boris Arellano-Valle. 2010. Likelihood Based Inference for Skew-Normal Independent Linear Mixed Models. Statist. Sinica 20: 303–22.

Laird, Nan M. 1978. Nonparametric Maximum Likelihood Estimation of a Mixing Distribution. Journal of the American Statistical Association 73: 805–11.

Laird, Nan M., and James H. Ware. 1982. Random-Effects Models for Longitudinal Data. Biometrics 38 (4): 963–74.

Lange, Nicholas, and Louise Ryan. 1989. Assessing Normality in Random Effects Models. Ann. Statist. 17 (2). The Institute of Mathematical Statistics: 624–42. doi:10.1214/aos/1176347130.

Legret, M., & Pagotto, C. 1999. Evaluation of pollutant loadings in the runoff waters from a major rural highway. Science of the Total Environment, 235(1-3), 143-150.

Lequy, E., Dubos, N., Witté, I., Pascaud, A., Sauvage, S., & Leblond, S. 2017. Assessing temporal trends of trace metal concentrations in mosses over France between 1996 and 2011: A flexible and robust method to account for heterogeneous sampling strategies. Environmental pollution, 220, 828-836.

Li, F., Zhang, Y., Fan, Z., & Oh, K. 2015. Accumulation of de-icing salts and its short-term effect on metal mobility in urban roadside soils. Bulletin of environmental contamination and toxicology, 94(4), 525-531.

Li, X.; Yang, J.; Liu, M.; Liu, G.; Yu, M. 2012. Spatio-Temporal Changes of Soil Salinity in Arid Areas of South Xinjiang Using Electromagnetic Induction. Journal of Integrative Agriculture 2012, 11 (8), 1365–1376.

Liang, Kung Yee, and Scott Zeger. 1986. Longitudinal Data Analysis Using Generalized Linear Models. Biometrika 73: 13–22.

Lin, Lawrence. 1989. A Concordance Correlation-Coefficient to Evaluate Reproducibility. Biometrics 45 (April): 255–68. doi:10.2307/2532051.

Liu, A., Gunawardana, C., Gunawardena, J., Egodawatta, P., Ayoko, G. A., & Goonetilleke, A. 2016. Taxonomy of factors which influence heavy metal build-up on urban road surfaces. Journal of hazardous materials, 310, 20-29.

Liu, Y.; Weerts, A. H.; Clark, M.; Hendricks Franssen, H.-J.; Kumar, S.; Moradkhani, H.; Seo, D.-J.; Schwanenberg, D.; Smith, P.; van Dijk, A. I. J. M.; et al. 2012. Advancing Data Assimilation in Operational Hydrologic Forecasting: Progresses, Challenges, and Emerging Opportunities. Hydrology and Earth System Sciences 2012, 16 (10), 3863–3887.

Lumsdon, D. G., Evans, L. J., & Bolton, K. A. 1995. The influence of pH and chloride on the retention of cadmium, lead, mercury, and zinc by soils. Soil and Sediment Contamination, 4(2), 137-150.

Ma, L. Q., & Rao, G. N. 1997. Chemical fractionation of cadmium, copper, nickel, and zinc in contaminated soils. Journal of Environmental Quality, 26(1), 259-264.

Mahmut, C. and Cevat, K., 2003. Spatial and temporal changes of soil salinity in cotton field irrigated with low-quality water. Journal of Hydrology, 272, 238-249, doi:10.1016/S0022-1694(02)00268-8.

Makarychev, I. P.AU., Motuzova, G. V. 2013.Complexation between metal ions and water soluble organic soil substances as analyzed by electrochemical techniques. Moscow University Soil Science Bulletin. Allerton Press, Vol. 68, No. 1, pp. 41–47.

Malicki MA, Walczak RT, Koch S, Flühler H. 1994. Determining soil salinity from simultaneous readings of its electrical conductivity and permittivity using TDR. In: Proceedings of the Symposium on Time Domain Reflectometry in Environmental, Infrastructure, and Mining Applications, Evanston, USA, 7-9 September, U.S. Bureau of Mines, Special Publication SP 19-94, NTIS PB95-105789; 328-336.

Malicki, M. A.; Walczak, R. T. 1999. Evaluating Soil Salinity Status from Bulk Electrical Conductivity and Permittivity. European Journal of Soil Science 1999, 50 (3), 505–514.

Mammen, Enno. 1993. Bootstrap and Wild Bootstrap for High Dimensional Linear Models. Ann. Statist. 21 (1). The Institute of Mathematical Statistics: 255–85. doi:10.1214/aos/1176349025.

Marsaglia, G., W. W. Tsang and J. Wang. 2003. Evaluating Kolmogorov's distribution. Journal of Statistical Software, 8(18):1-4. doi: 10.18637/jss.v008.i18

Marsaglia, J. and G. Marsaglia. 2004. Evaluating the Anderson-Darling Distribution. Journal of Statistical Software 9(2):1-5. doi: 10.18637/jss.v009.i02.

Maybeck, P. S. 1982. Stochastic Models, Estimation, and Control; Academic Press, 1982.

McCleary R, Hay RA. 1980. Applied Time Series Analysis for the Social Sciences. Thousand Oaks: Sage Publications Inc.

McCulloch, C. E., and S. R. Searle. 2001. Generalized, Linear, and Mixed Models. Wiley.

McCulloch, Charles E, and John M Neuhaus. 2011. Misspecifying the Shape of a Random Effects Distribution: Why Getting It Wrong May Not Matter. Stat. Sci. 26: 388–402.

McDowall D, McCleary R, Meidinger EE, Hay RA. 1980. Interrupted Time Series Analysis, Beverly Hills: Sage Publications. McKenzie RC, Chomistek W, Clark NF. 1989. Conversion of electromagnetic inductance readings to saturated paste extract values in soils for different temperature, texture, and moisture conditions. Canadian Journal of Soil Science. 69 (1): 25-32. DOI: 10.4141/cjss89-003.

McKenzie, R. C., Chomistek, W. and Clark, N. F., 1989. Conversion of electromagnetic inductance readings to saturated paste extract values in soils for different temperature, texture, and moisture conditions. Canadian Journal of Soil Science, 69, 25– 32. Available from: http://pubs.aic.ca/doi/pdf/10.4141/cjss89-003.

Men, C., Liu, R., Xu, F., Wang, Q., Guo, L., & Shen, Z. 2018. Pollution characteristics, risk assessment, and source apportionment of heavy metals in road dust in Beijing, China. Science of The Total Environment, 612, 138-147.

Mishra AK, Desai VR. 2005. Drought forecasting using stochastic models. Stochastic Environmental Research and Risk Assessment 19 (5): 326–339. DOI: 10.1007/s00477-005-0238-4.

Modrzewska, B., & Wyszkowski, M. 2014. Trace metals content in soils along the state road 51 (northeastern Poland). Environmental monitoring and assessment, 186(4), 2589-2597.

Moradkhani, H.; Sorooshian, S.; Gupta, H. V.; Houser, P. R. 2005. Dual State–Parameter Estimation of Hydrological Models Using Ensemble Kalman Filter. Advances in Water Resources 2005, 28 (2), 135–147.

Mualem Y, Friedman SP. 1991. Theoretical prediction of electrical conductivity in saturated and unsaturated soil. Water Resources Research 27 (10): 2771-2777. DOI: 10.1029/91WR01095.

Münch, D. 1993. Concentration profiles of arsenic, cadmium, chromium, copper, lead, mercury, nickel, zinc, vanadium and polynuclear aromatic hydrocarbons (PAH) in forest soil beside an urban road. Science of the Total Environment, 138(1-3), 47-55.

Munns R. 2002. Comparative physiology of salt and water stress. Plant Cell and Environment 25 (2): 239-250. DOI: 10.1046/j.0016-8025.2001.00808.x.

Nguyen, Hien Duy, and Geoffrey J. McLachlan. 2016. Linear Mixed Models with Marginally Symmetric Nonparametric Random Effects. Computational Statistics & Data Analysis 103: 151–69.

Nielsen, D.R., Katul, G.G., Wendroth, O., Folegatti, M.V. and Parlange, M.B., 1994. State-space approaches to estimate soil physical properties from field measurements, p. 61-85. Proc. 15th conf. ISSS: Vol. 2a. July 10-16, 1994. Acapulco, Mexico.

O'Shaughnessy, P. Y., and Alan H. Welsh. 2018. Bootstrapping Longitudinal Data with Multiple Levels of Variation. Computational Statistics & Data Analysis 124: 117–31.

OECD 2013, Environment at a Glance 2013: OECD Indicators, OECD Publishing, Paris, https://doi.org/10.1787/9789264185715-en.

Orlović-Leko, P., Plavšić, M., Bura-Nakić, E., Kozarac, Z., & Ćosović, B. 2009. Organic matter in the bulk precipitations in Zagreb and Šibenik, Croatia. Atmospheric Environment, 43(4), 805-811.

Ozkutlu, F., Ozturk, L., Erdem, H., McLaughlin, M., & Cakmak, I. 2007. Leaf-applied sodium chloride promotes cadmium accumulation in durum wheat grain. Plant and soil, 290(1-2), 323-331.

Padoan, E., Romè, C., & Ajmone-Marsan, F. 2017. Bioaccessibility and size distribution of metals in road dust and roadside soils along a peri-urban transect. Science of The Total Environment, 601, 89-98.

Pagotto, C., Remy, N., Legret, M., & Le Cloirec, P. 2001. Heavy metal pollution of road dust and roadside soil near a major rural highway. Environmental Technology, 22(3), 307-319.

Pal, S., Wallis, S., & Arthur, S. 2011. Assessment of heavy metals emission from traffic on road surfaces. Open Chemistry, 9(2), 314-319.

Pan, Y., Wang, Y., Xin, J., Tang, G., Song, T., Wang, Y., … & Wu, F. 2010. Study on dissolved organic carbon in precipitation in Northern China. Atmospheric Environment, 44(19), 2350-2357.

Pankratz A. 1983. Forecasting with Univariate Box–Jenkins Models: Concepts and Cases. New York: John Wiley and Sons.

Park SJ, Hwang CS, Vlek PLG. 2005. Comparison of adaptive techniques to predict crop yield response under varying soil and land management conditions. Agricultural Systems 85 (1): 59–81.DOI:10.1016/j.agsy.2004.06.021.

Parlange, M.B., Katul, G.G., Folegatti, M.V. and Nielsen, D.R., 1993. Evaporation and the field scale soil water diffusivity function. Water Resources Resarch. 29, 1279-1286, doi:10.1029/93WR00094.

Persson, M., 2002. Evaluating the linear dielectric constant –electrical conductivity model using time domain reflectometry. Hydrological Science Journal, 47, 269-278, doi:10.1080/02626660209492929.

Petris, G. 2010. An R Package for Dynamic Linear Models. Journal of Statistical Software 2010, 36 (12).

Petris, G., 2010. dlm: Bayesian and Likelihood Analysis of Dynamic Linear Models. R package version 1.1-1, Available from http://CRAN.R-project.org/package=dlm [Accessed 4 April 2012].

Pinheiro, J. C., and D. M. Bates. 2000. Mixed-Effects Models in S and S-Plus. Springer.

Pinheiro, José C., Chuanhai Liu, and Ying Nian Wu. 2001. Efficient Algorithms for Robust Estimation in Linear Mixed-Effects Models Using the Multivariate T Distribution. J. Comput. Graph. Statist. 10: 249–76.

Pinheiro, Jose, Douglas Bates, Saikat DebRoy, Deepayan Sarkar, and R Core Team. 2019. nlme: Linear and Nonlinear Mixed Effects Models. https://CRAN.R-project.org/package=nlme.

Quinn II TJ. 1985. Catch-per-unit-effort: a statistical model for Pacific halibut (Hippoglossus stenolepis). Canadian Journal of Fisheries and Aquatic Sciences 42 (8): 1423–1429. DOI: 10.1139/f85-178.

R Core Team 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rai, D., J.M. Zachara, L.E. Eary, D.C. Girvin, D.A. Moore, C.T. Resch, B.M. Sass and R.L. Schmidt 1986. Geochemical Behavior of Chromium Species. Electric Power ResearchInstitute, Palo Alto, California, EA-4544.

Ramakrishna, D. M., & Viraraghavan, T. 2005. Environmental impact of chemical deicers–a review. Water, Air, and Soil Pollution, 166(1-4), 49-63.

Ramos, L., Hernandez, L. M., & Gonzalez, M. J. 1994. Sequential fractionation of copper, lead, cadmium and zinc in soils from or near Donana National Park. Journal of environmental Quality, 23(1), 50-57.

Raulinaitis, M., Ignatavičius, G., Sinkevičius, S., & Oškinis, V. 2012. Assessment of heavy metal contamination and spatial distribution in surface and subsurface sediment layers in the northern part of Lake Babrukas. Ekologija, 58(1).

Reinirkens, P. 1996. Analysis of emissions through traffic volume in roadside soils and their effects on seepage water. Science of the total environment, 189, 361-369.

Rhoades JD, Shouse PJ, Alves WJ, Manteghi NA, Lesch.SM. 1990. Determining soil salinity from electrical conductivity using different models and estimates. Soil Science Society of American Journal 54 (1): 45-54. DOI:10.2136/sssaj1990.03615995005400010007x.

Rhoades, J.D., Lesch, S.M., LeMert, R.D. and Alves, W.J., 1997. Assessing irrigation/drainage/salinity management using spatially referenced salinity measurements. Agricultural Water Management, 35, 147–165.

Rhoades, J.D., Raats, P.A.C. and Prather, R.J., 1976. Effects of liquid-phase electrical conductivity, water-content, and surface conductivity on bulk soil electrical conductivity. Soil Science Society of America Journal, 40, 651-655.

Rieuwerts, J. S., Thornton, I., Farago, M. E., & Ashmore, M. R. 1998. Factors influencing metal bioavailability in soils: preliminary investigations for the development of a critical loads approach for metals. Chemical Speciation & Bioavailability, 10(2), 61-75.

Rodriguez, A.; Ruben, F. 2009. CALIBRATING CAPACITANCE SENSORS TO ESTIMATE WATER CONTENT, MATRIC POTENTIAL, AND ELECTRICAL CONDUCTIVITY IN SOILLESS SUBSTRATES. Thesis, 2009.

Romão, X., R. Delgado and A. Costa. 2010. An empirical power comparison of univariate goodness-of-fit tests for normality. Journal of Statistical Computation and Simulation 80(5): 545-591. doi: 10.1080/00949650902740824

Rosenkrantz, W.A. 2000, Confidence Bands for Quantile Functions: A Parametric and Graphic Alternative for Testing Goodness of Fit, The American Statistician, 54, 185-190.

Royall, Richard M. 1986. Model Robust Confidence Intervals Using Maximum Likelihood Estimators. Internat. Statist. Rev. 2: 221–26.

Royston, P. 1991: Approximating the Shapiro-Wilk W-test for non-normality. Statistics and Computing 1992 2, 117-119.

Salimi, M., Amin, M. M., Ebrahimi, A., Ghazifard, A., & Najafi, P. 2012. Influence of electrical conductivity on the phytoremediation of contaminated soils to Cd 2+ and Zn 2+. International Journal of Environmental Health Engineering, 1(1), 11.

Samanta, Mayukh, and Alan H. Welsh. 2013. Bootstrapping for Highly Unbalanced Clustered Data. Computational Statistics & Data Analysis 59: 70–81.

Sanchez-Camazano, M., Sanchez-Martin, M. J., & Lorenzo, L. F. 1994. Lead and cadmium in soils and vegetables from urban gardens of Salamanca (Spain). Science of the Total Environment, 146, 163-168.

Sánchez-Espigares, J. A., P. Grima, L. Marco-Almagro. 2018. Visualizing Type II Error in Normality Tests. The American Statistician. 72(2):158-162. doi: 10.1080/00031305.2016.1278035

Sarangi A, Man Singh AK, Bhattacharya AK, Singh AK. 2006. Subsurface drainage performance study using SALTMOD and ANN models. Agricultural Water Managment 84 (3): 240-148. DOI: 10.1016/j.agwat.2006.02.009.

Sauve, S., Hendershot, W., & Allen, H. E. 2000. Solid-solution partitioning of metals in contaminated soils: dependence on pH, total metal burden, and organic matter. Environmental science & technology, 34(7), 1125-1131.

Scheidegger, A. M., Sparks, D. L., & Fendorf, M. 1996. Mechanisms of nickel sorption on pyrophyllite: Macroscopic and microscopic approaches. Soil Science Society of America Journal, 60(6), 1763-1772.

Schützenmeister, A., and H.-P. Piepho. 2011. Residual Analysis of Linear Mixed Models Using a Simulation Approach. Computational Statistics and Data Analysis 56: 1405–16. doi:10.1016/j.csda.2011.11.006.

Shapiro, S.S and Wilk, M.B. 1965: An Analysis of Variance Test for Normality (Complete Samples). Biometrika. Vol. 52, No. 3/4 (Dec., 1965), pp. 591-611

Shapiro, S.S. and Francia, R.S. "An approximate analysis of variance test for normality", Journal of the American Statistical Association 67 1972 215–216.

Shouse, P.J., Goldberg, S., Skaggs, T. H., Soppe, R. W. O. and Ayars, J. E., 2010.Changes in spatial and temporal variability of SAR affected by shallow groundwater management of an irrigated field, California. Agricultural Water Management, 97, 673-680, doi:10.1016/j.agwat.2009.12.008.

Shrout, P. E., and J. L. Fleiss. 1979. Intraclass Correlations: Uses in Assessing Rater Reliability. Psychol Bull 86 (2): 420–38.

Shumway RH, Stoffer DS. 2006. Time Series Analysis and its Applications. With R Examples. New York: Springer. DOI: 10.1007/0-387-36276-2.

Slavich PG, Peterson GH. 1990 Estimating the electrical conductivity of saturated paste extracts from 1:5 soil-water suspensions and texture. Australian Journal of Soil Research 31 (1):73-81. DOI: 10.1071/SR9930073.

Soebiyanto RP, Adimi F, Kiang RK. 2010. Modeling and predicting seasonal influenza transmission in warm regions using climatological parameters. PLoS One, 5 (3): e9450. DOI: 10.1371/journal.pone.0009450.

Song, P. X.-K., P. Zhang, and A. Qu. 2007. Maximum Likelihood Inference in Robust Linear Mixed-Effects Models Using Multivariate T Distributions. Statist. Sinica 17: 929–43.

Stephens, M. A. 1986. Test Based on EDF Statistics. In Goodness-of-Fit Techniques, ed. R. B. D'Agostino and M. A. Stephens, 97-193. New York: Marcel Dekker.

Stevens, J. P. 2012. Applied Multivariate Statistics for the Social Sciences. Routledge.

Stigler, S. M. 1986. The History of Statistics. The Measurement of Uncertainty before 1900. Cambridge, MA: Harvard University Press.

Strobel, B.W., Hansen, H.C.B., Borggaard, O.K., Andersen, M.K. & Raulund-Rasmussen, K. 2001. Cadmium and copper release kinetics in relation to afforestation of cultivated soil. Geochimica et Cosmochimica Acta, 65, 1233-1242.

Suo, Chen, Timothea Toulopoulou, Elvira Bramon, Muriel Walshe, Marco Picchioni, Robin Murray, and J Ott. 2013. Analysis of Multiple Phenotypes in Genome-Wide Genetic Mapping Studies. BMC Bioinformatics 14 (May).

Temminghoff, E.J.M., Van Der Zee, S.E.A.T.M. & De Haan, F.A.M. 1998. Effects of dissolved organic matter on the mobility of copper in a contaminated sandy soil. European Journal of Soil Science, 49, 617-628.

Thiruchelvam, S. and Pathmarajah. S., 1999. Economic feasibility of controlling salinity problems in the Mahaweli H area. A research paper presented at the monthly seminar, IWMI (formally IIMI), March, 1999, Colombo, Sri Lank.

Thiruchelvam, S.; Pathmarajah, S. 1999. An Economic Analysis of Salinity Problems in the Mahaweli River System H Irrigation Scheme in Sri Lanka. 47.

Thode, H. C. 2002. Testing for Normality. New York: Marcel Dekker.

Thorpe, A., & Harrison, R. M. 2008. Sources and properties of non-exhaust particulate matter from road traffic: a review. Science of the total environment, 400(1-3), 270-282.

Thorsten Pohlert 2017. ppcc: Probability Plot Correlation Coefficient Test. R package version 1.0. https://CRAN.R-project.org/package=ppcc

Tills, A.R. and Alloway, B.J. 1983. The speciation of lead in soil solution from very polluted soils. EnvironmentalTechnology Letters,4, 529–534.

Tipping, E., & Woof, C. 1990. Humic substances in acid organic soils: modelling their release to the soil solution in terms of humic charge. European Journal of Soil Science, 41(4), 573-586.

Tsai, Miao-Yu, and Chaochun Lin. 2017. Concordance Correlation Coefficients Estimated by Variance Components for Longitudinal Normal and Poisson Data. Computational Statistics & Data Analysis 121 (December). doi:10.1016/j.csda.2017.12.003.

Tsai, Miao-Yu. 2015. Comparison of Concordance Correlation Coefficient via Variance Components, Generalized Estimating Equations and Weighted Approaches with Model Selection. Computational Statistics & Data Analysis 82 (February): 47–58. doi:10.1016/j.csda.2014.08.005.

Turer, D. G., & Maynard, B. J. 2003. Heavy metal contamination in highway soils. Comparison of Corpus Christi, Texas and Cincinnati, Ohio shows organic matter is key to mobility. Clean Technologies and Environmental Policy, 4(4), 235-245.

Urzua, C. M. 1996: On the correct use of omnibus tests for normality. — Economics Letters, vol. 53, pp. 247–251.

Van Bohemen, H. D., & Van de Laak, W. J. 2003. The influence of road infrastructure and traffic on soil, water, and air quality. Environmental Management, 31(1), 0050-0068.

Van Genuchten, M. TH., 1991. Recent progress in modelling water flow and chemical transport in the unsaturated zone. IAHS Publication, 169-183 Available from: http://ks360352.kimsufi.com/redbooks/a204/iahs_204_0169.pdf

Vandaele W. 1983. Applied Time Series and Box–Jenkins Models. New York: Academic Press.

Vangeneugden, Tony, Annouschka Laenen, Helena Geys, Didier Renard, and Geert Molenberghs. 2005. Applying Concepts of Generalizability Theory on Clinical Trial Data to Investigate Sources of Variation and Their Impact on Reliability. Biometrics 61 (1): 295–304. https://ideas.repec.org/a/bla/biomet/v61y2005i1p295-304.html.

Verbeke, G., and G. Molenberghs. 2000. Linear Mixed Models for Longitudinal Data. Springer.

Verbeke, Geert, and Emmanuel Lesaffre. 1997. The Effect of Misspecifying the Random-Effects Distribution in Linear Mixed Models for Longitudinal Data. Computational Statistics & Data Analysis 23: 541–56.

Visconti, F.; MartÃ-nez, D.; Molina, M. J.; Ingelmo, F.; Miguel de Paz, J. A 2014. Combined Equation to Estimate the Soil Pore-Water Electrical Conductivity: Calibration with the WET and 5TE Sensors. Soil Res. 2014, 52 (5), 419–430.

Walraven, N., Van Os, B. J. H., Klaver, G. T., Middelburg, J. J., & Davies, G. R. 2014. The lead (Pb) isotope signature, behaviour and fate of traffic-related lead pollution in roadside soils in The Netherlands. Science of the Total Environment, 472, 888-900.

Wang, Y. J., Chen, C. F., & Lin, J. Y. 2013. The Measurement of dry deposition and surface runoff to quantify urban road pollution in Taipei, Taiwan. International journal of environmental research and public health, 10(10), 5130-5145.

Ward, N. I. 1990. Multielement contamination of British motorway environments. Science of the Total Environment, 93, 393-401.

Wei WWS. 1989. Time Series Analysis: Univariate and Multivariate Methods. New York: Person Addison-Wesley.

Wendroth, O., Katul, G.G., Parlange, M.B., Puente, C.E. and Nielsen, D.R., 1993. A nonlinear filtering approach for determininghydraulic conductivity functions in field soils. Soil Science, 156, 293-301.

Wendroth, O., Rogasik, H., Kozinski, S., Ritsema, C. J., Dekker, L. W. and Nielsen, D. R., 1999. State-Space prediction of field –scale soil water content time series in a sandy loam. Soil Tillage Research, 50, 85-93.

Werkenthin, M., Kluge, B., & Wessolek, G. 2014. Metals in European roadside soils and soil solution–A review. Environmental Pollution, 189, 98-110.

White RE. 1985. The influence of macrospores on the transport of dissolved and suspended matter through soil. In Advanced in Soils Sciences, Stewart BA (ed); 95-120. DOI: 10.1007/978-1-4612-5090-6.

Wu L, Jury WA, Chang AC, Allmaras RR, 1997. Time series analysis of field-measured water content of a sandy soil. Soil Science Society of American Journal61 (3): 736–742. DOI: 10.2136/sssaj1997.03615995006100030005x.

Wuana, R. A., & Okieimen, F. E. 2011. Heavy metals in contaminated soils: a review of sources, chemistry, risks and best available strategies for remediation. Isrn Ecology, 2011.

Wyllie, M. R. J.; Southwick, P. F. 1954. An Experimental Investigation of the S.P. and Resistivity Phenomena in Dirty Sands. Journal of Petroleum Technology 1954, 6 (02), 44–57. https://doi.org/10.2118/302-G.

Xiaoming, L. I., Jingsong, Y.,Meixian, L. I. U., Guangming, L. I. U. and Mei, Y. U., 2012. Spatio-temporal changes of soil salinity in arid areas of South Xinjiang using electromagnetic induction. Journal of Integrative Agriculture, 11, 1365-1376, doi:10.1016/S2095-3119(12)60135-X.

Xu, Jin, and Xinping Cui. 2008. Robustified Manova with Applications in Detecting Differentially Expressed Genes from Oligonucleotide Arrays. Bioinformatics (Oxford, England) 24 (May): 1056–62. doi:10.1093/bioinformatics/btn053.

Yacini, B. and S. Yolacan. 2007. A comparison of various tests of normality. Journal of Statistical Computation and Simulation. 77(2): 175-183. doi: 10.1080/10629360600678310.

Yan, X., Gao, D., Zhang, F., Zeng, C., Xiang, W., & Zhang, M. 2013. Relationships between heavy metal concentrations in roadside topsoil and distance to road edge based on field observations in the Qinghai-Tibet Plateau, China. International journal of environmental research and public health, 10(3), 762-775.

Yap, B. W. and C. H. Sim. 2011. Comparisons of various types of normality tests. Journal of Statistical Computation and Simulation, 81(12):2141-2155. doi: 10.1080/00949655.2010.520163

Yazici, B and Yolacan, S. 2007. A comparison of various tests of normality. Journal of Statistical Computation and Simulation. Vol. 77, No. 2, February 2007, 175–183

Yuanshi, G., Qiaohong, C. and Zongjia, S., 2003. The effects of soil bulk density, clay content and temperature on soil water content measurement using time-domain reflectometry. Hydrological Process, 17, 3601-3614.

Zachara, J. M., Ainsworth, C. C., Cowan, C. E., & Resch, C. T. 1989. Adsorption of chromate by subsurface soil horizons. Soil Science Society of America Journal, 53(2), 418-428.

Zhang JQ, Zhang LX, Zhang MH, Watson C. 2009. Prediction of soybean growth and development using artificial neural network and statistical models. Acta Agronomica Sinica 35 (2): 341–347. DOI: 10.1016/S1875-2780(08)60064-4.

Zhang, Daowen, and Marie Davidian. 2001. Linear Mixed Models with Flexible Distributions of Random Effects for Longitudinal Data. Biometrics 57 (3): 795–802.

Zhu, D. and V. Zinde-Walsh. 2009. Properties and estimation of asymmetric exponential power distribution. Journal of Econometrics. 148(1): 89-99. doi: 10.1016/j.jeconom.2008.09.038

Zou P, Yang J, Fu J, Liu, G, Li D. 2010. Artificial neural network and time series models for predicting soil salts and water content. Agricultural Water Management 97 (12): 2009-2019. DOI: 10.1016/j.agwat.2010.02.011.