Piqué Muntané, P. & Fernández-Getino García, M. J. (25-28 June 2023). *Parametric model and estimator classifier for optimal averaging in mobile OFDM systems with superimposed training* [proceedings]. In 2023 International Technical Conference on Circuits/ Systems, Computers, and Communications (ITC-CSCC), Jeju, Republic of Korea.

DOI: 10.1109/ITC-CSCC58803.2023.10212534

Parametric Model and Estimator Classifier for Optimal Averaging in Mobile OFDM Systems with Superimposed Training

Ignasi Piqué Muntané
Department of Signal Theory and Communications
Carlos III University of Madrid
Madrid, Spain
ignasip@tsc.uc3m.es

M. Julia Fernández-Getino García
Department of Signal Theory and Communications
Carlos III University of Madrid
Madrid, Spain
mjulia@tsc.uc3m.es

***Abstract*—Superimposed training (ST) is an attractive technique for channel estimation in orthogonal frequency division multiplexing (OFDM) modulation. However, its main challenge is the intrinsic interference due to the joint transmission of pilot and data symbols, which can be mitigated by averaging the received signal. Previous works analyzed the mean square error (MSE) of the channel estimation, for both least squares (LS) and minimum MSE (MMSE) estimators, and showed that, under realistic channel models, the optimum number of averaged symbols could be computed by solving a transcendental equation. In this paper, as a practical implementation proposal, these optimum averaging values are parametrically approximated with a multilinear regression model. Also, it is proposed an accurate classifier that, under delay and performance tolerances, is able to select the most suitable estimator between LS and MMSE.**

***Index Terms*—OFDM, Superimposed Training, Time-Variant Channel, Channel Estimation, Minimum Mean Squared Error, Least Squares, Averaging, Regression, Classifier.**

I. Introduction

The fifth generation (5G) of wireless communications has been designed to enhance, by several orders of magnitude, the capacity of data transmission and the amount of devices supported by the network [1]. For this purpose, orthogonal frequency division multiplexing (OFDM) has been chosen as the main modulation scheme, however, to properly operate, accurate estimations of the channel model are required [2].

A common approach to perform this estimation is based on the pilot symbol assisted modulation (PSAM) technique, which reserves resources from the OFDM grid to be used as pilot symbols [3]. Even though this method is very reliable, if the channel coefficients change very frequently or many devices are in the system, the spectral efficiency can be very hampered [1]. Under these conditions, superimposed training (ST), which consists in a scheme that embeds the pilot symbols with the data stream, has proven to provide better results [3]–[5]. In particular, ST showed a great potential to mitigate the pilot contamination issue due to overcrowded cells [6].

This work has been supported by a fellowship from the Spanish Ministry of Science and Innovation, under grant PRE2018-084315, and, by the Spanish National Project IRENE-EARTH (PID2020-115323RB-C33 / AEI / 10.13039/501100011033).

Then, since the potential of ST is very attractive for future schemes, in [7] and [8], its performance was analyzed when realistic time-variant channel coefficients from 5G New Radio (NR) standard models were used. In these works, the mean square error (MSE) of the ST-based channel estimation was derived for the least squares (LS) and minimum MSE (MMSE) estimators. Among the results, it was shown that the number of received symbols that needs to be averaged, which is a common procedure that mitigates an intrinsic interference of ST [3], was no longer the trivial state-of-the-art (SOA) coherence time of the channel model. In fact, the averaging parameter could be obtained for every scenario by solving a transcendental equation numerically.

To avoid this costly computation, especially for real time applications, in [9], a parametrical implementation of LS estimator was proposed following a multilinear regression model. To extend this proposal for the MMSE case, a more general approach is presented in this paper.

The contributions of this paper are:

- A multilinear regression model is proposed to parametrically compute the optimum averaging length of ST for MMSE, and it is compared with the reported regression model for LS.
- The robustness of the regression model is shown in terms of final MSE discrepancy and the best polynomial configuration is determined.
- A novel classification method is proposed, which selects the most suitable estimator between LS and MMSE depending on delay and performance design tolerances.

Notation: x, $\mathbf{x}$ and $\mathbf{X}$ refer to a scalar, a vector and a matrix, respectively, with $\left[\mathbf{X}\right]_{(m,:)}$ being the elements from the m-th row; $(\cdot)^T$ is the transpose and $(\cdot)^{-1}$ is the inverse of a matrix; $\otimes$ is the matrix Kronecker product and $||$ is the concatenation of two vectors; $\mathbb{E}\{\cdot\}$ is the mean, $\lceil\cdot\rceil$ is the ceil and $\lfloor\cdot\rceil$ is the nearest integer operation; $\mathbb{N}_0$ and $\mathbb{N}_1$ are the sets of non-negative and positive integers, respectively, and $\mathbb{R}_{>0}$ is the set of positive real numbers; $\cap$ corresponds to the intersection operation between sets.

II. System model

Following the line of work of [7]–[9], the considered system model is an uplink (UL) single-antenna transmitter and receiver. Obviously, under proper notation, the proposed analysis is valid for more complex scenarios.

A. Transmitter

In the transmitter, all the resources from the OFDM grid have implemented a ST scheme. The complete symbol (x) for the k-th subcarrier at m-th time index is [3],

$$x_m^k = s_m^k + c_m^k, \quad \begin{cases} \forall\, k \ \in [0, \cdots, K-1] \\ \forall\, m \ \in \mathbb{N}_0 \end{cases}, \tag{1}$$

where c and s are pilot and data symbols, whose powers are P_c and P_s, respectively, and K is the number of subcarriers. The intrinsic interference between c and s is limited by β, the power allocation factor. This way, the power of x is,

$$P = P_s + P_c \begin{cases} P_s = (1-\beta)P \\ P_c = \beta P \end{cases}, \quad \text{with } \beta = [0,1]\ . \tag{2}$$

After that, x is processed by the OFDM transmitter scheme which modulates the signal with the proper guard bands and cyclic-prefix (CP).

B. Receiver

In the receiver, a perfect synchronization of the received signal is assumed. The signal is processed with the OFDM receiver scheme and the symbol at the k-th subcarrier and m-th time index is recovered as the following flat fading model,

$$y_m^k = H_m^k x_m^k + w_m^k = H_m^k s_m^k + H_m^k c_m^k + w_m^k\,, \tag{3}$$

in which $H_m^k \sim \mathcal{CN}\left(0, \frac{\sigma_h^2}{K}\right)$ and $w_m^k \sim \mathcal{CN}\left(0, \sigma_w^2\right)$, are the channel coefficient and the additive white Gaussian noise (AWGN) component, both in the frequency domain. Without loss of generality, only small-scale fading effects of the channel model have been taken into account.

Also, the power relation between components of the system can be expressed with the signal-to-noise ratio (SNR) as,

$$SNR = \frac{P\sigma_h^2}{\sigma_w^2}\ , \quad SNR^{\,dB} = 10\,\log\left(SNR\right)\,, \tag{4}$$

where σ_h^2 and σ_w^2 are the powers of the channel and the AWGN, respectively, and $SNR^{\,dB}$ is the logarithmic SNR[(1)].

C. Channel Model in Time Domain

Since the aim of this paper is to optimize the ST scheme under realistic time variant channels, only the temporal evolution of the channel coefficients will be considered[(2)]. Therefore, as it was explained in [7]–[9], the correlation between the m-th and the m'-th time index of the channel coefficients can be expressed as,

$$\mathbb{E}\left\{(H_m)^*\, H_{m'}\right\} = \frac{\sigma_h^2}{K}\rho_t(\gamma\Delta m)\ ,\ \text{with } \Delta m = |m-m'|\ , \tag{5}$$

in which $\rho_t(\cdot)$ is the correlation profile and γ is a constant that embeds the configuration of the system as follows [10], [11],

$$\gamma = 2\pi f_d T_{sym}\,, \text{ with } T_{sym} = \frac{\left(1+\frac{L_{cp}}{K}\right)}{\Delta f} \text{ and } f_d = \frac{v}{3.6\,c} f_c\,, \tag{6}$$

where T_{sym} is the duration of the OFDM symbol, computed with the subacarrier spacing (Δf) and the extra samples appended as CP (L_{cp}); and f_d is the Doppler frequency computed with the relative speed between the transmitter and the receiver (v in km/h for convenience), the speed of light constant (c in m/s) and the carrier frequency (f_c).

From the definition of T_{sym} and f_d in (6), the coherence time of the channel model can be approximated with $T_{coh} \sim \frac{0.423}{f_d}$ [12]. This value is commonly used as the duration that channel coefficients remain almost constant (at least with a correlation higher than 70%) [7]. Analogously, T_{coh} can also be expressed in terms of number of symbols as,

$$N_c \sim \left\lfloor \frac{T_{coh}}{T_{sym}} \right\rceil = \left\lfloor \frac{\kappa}{v} \right\rceil \sim N_c(v)\,, \text{ with } \kappa = \frac{0.423 \cdot 3.6\,c}{T_{sym}\, f_c}\,. \tag{7}$$

III. Optimal ST-based Channel Estimation: LS & MMSE

In this section, the optimal averaging of ST is presented when realistic time-variant channels are considered. Detailed derivations of following expressions can be found in previous works such as [7], [8].

Before analyzing any channel estimation technique, it should be appreciated that both c and s from (3) are affected by the same channel coefficient. In [3], [5], [7], to enhance the MSE of the estimation, the authors proved that this self-interference could be mitigated if a group of received symbols is averaged as follows,

$$\bar{y}_\xi^\chi = \sum_{k'}\sum_{m'} \frac{1}{N_t N_f} y_{m'}^{k'}\,, \text{ with } \begin{cases} (\chi-1)\,N_f \le k' \le \chi N_f - 1 \\ (\xi-1)\,N_t \le m' \le \xi N_t - 1 \end{cases}, \tag{8}$$

where N_t and N_f are the size of the averaged group in time and frequency domain, and $\xi \in \mathbb{N}_1$ and $\chi \in \left[1, \left\lceil \frac{K}{N_f} \right\rceil\right]$ are time and frequency index positions for each group, respectively. After that, $\bar{y}_\xi^\chi$ is used as the input for both LS or MMSE estimators to estimate the corresponding channel coefficients.

A. Derivation of the Estimator and the MSE

The LS estimator can be computed by minimizing the following cost function [13],

$$J_{cost} = \left| \frac{c^*}{\beta P}\,\bar{y}_\xi^\chi - \left(\widehat{H}_{\text{LS}}\right)_\xi^\chi \right|^2 \ \rightarrow \left(\widehat{H}_{\text{LS}}\right)_\xi^\chi = \frac{c^*}{\beta P}\,\bar{y}_\xi^\chi\,. \tag{9}$$

Since there is no distinction between different (ξ, χ) groups, to simplify notation, the analysis will be focused on $\xi = \chi = 1$ and both indices will be omitted.

[(1)] All logarithmic values from this paper are defined with "log" in base 10.
[(2)] From now on, the k-th index from channel coefficients will be omitted.

$$\Psi = \frac{\sigma_h^2}{K}\left(1 - \frac{1}{N_t}\sum_{m'=0}^{N_t-1} A_{m'}^2 \left(\overbrace{\frac{1}{N_t^2}\sum_{m_1=0}^{N_t-1}\sum_{m_2=0}^{N_t-1}\rho_t(\gamma\,|m_1-m_2|)}^{\text{Contribution of the channel coefficients' variability}} + \frac{B}{N_t}\Upsilon_0\right)\right), \tag{11}$$

$$\text{where } \{A_{m'}, B\} = \begin{cases} \{1\,, -1\} & \text{for the LS estimator} \\ \{\zeta_{m'} \text{ from (10)}\,,\ 1\} & \text{for the MMSE estimator} \end{cases}$$

Using (9), the MMSE can be derived as follows,

$$\widehat{H}_{m,\,\text{MMSE}} = \frac{\text{Cov}\,\{H_m, \bar{y}\}}{\text{Var}\,\{\bar{y}\}}\bar{y} = \zeta_m \widehat{H}_{\text{LS}}$$
$$\text{where } \zeta_m = \frac{\sum_{m_1=0}^{N_t-1}\rho_t(\gamma\,|m-m_1|)}{\frac{1}{N_t}\sum_{m_1=0}^{N_t-1}\sum_{m_2=0}^{N_t-1}\rho_t(\gamma\,|m_1-m_2|) + \Upsilon_0}\,,$$
$$\text{and } \Upsilon_0 = \frac{1}{N_f}\left(\left(\frac{1}{\beta}-1\right) + \frac{\sigma_w^2}{\beta P}\frac{K}{\sigma_h^2}\right). \tag{10}$$

Unlike LS, MMSE depends on ζ_m, which shows the effect of the correlation between different channel coefficients. Finally, the respective MSE can be derived as (11).

B. Optimization of the MSE

From (11), it should be noted that the MSE (Ψ) can be minimized if the proper number of averaged symbols (N_t) is selected. Obviously, this computation is affected by the correlation profile $\big(\rho_t(\cdot)\,\big)$. For example, if a simplified block-fading model is assumed, in which the channel coefficients remain constant for a specific T_{coh}, the optimum averaging is actually the corresponding N_c [3], [4]. However, in a more realistic channel model this trivial averaging is no longer valid.

In [7] and [8], to derive the optimum averaging under realistic channel models, an approximation of a correlation profile from 5G-NR standard was proposed. The results provided the optimum MSE performance and showed to be very accurate under different scenario conditions. To obtain these solutions, for both LS and MMSE, the non-negative real-valued roots ($n_{t,\text{LS}}^{opt}$, $n_{t,\text{MMSE}}^{opt} \in \mathbb{R}_{>0}$) of the following equations must be computed,

$$\frac{\gamma}{n_{t,\text{LS}}^{opt}}\phi\Big(n_{t,\text{LS}}^{opt}\Big) - \text{Si}\left(\gamma n_{t,\text{LS}}^{opt}\right) - \frac{\gamma}{2}\Upsilon_0 = 0\,, \tag{12}$$

$$\left(\phi\Big(n_{t,\text{MMSE}}^{opt}\Big) - \frac{n_{t,\text{MMSE}}^{opt}}{\gamma}\,\text{Si}\left(\gamma n_{t,\text{MMSE}}^{opt}\right)\right)\cdot \cdot\left(2 + \frac{1}{n_t\Upsilon_0}\phi\Big(n_{t,\text{MMSE}}^{opt}\Big)\right) - \frac{1}{2}\phi\Big(n_{t,\text{MMSE}}^{opt}\Big) = 0\,, \tag{13}$$

where,

$$\phi(n_t) = \frac{2}{\gamma^2}\left(\gamma n_t\,\text{Si}\,(\gamma n_t) + \cos\left(\gamma n_t\right) - 1\right). \tag{14}$$

Finally, the optimum averaging values are their closest integers, $N_{t,\text{LS}}^{opt} = \Big\lfloor n_{t,\text{LS}}^{opt} \Big\rceil$ and $N_{t,\text{MMSE}}^{opt} = \Big\lfloor n_{t,\text{MMSE}}^{opt} \Big\rceil$.

IV. Proposed Regression Model

As it was shown in previous Subsection III-B, the optimum averaging can be numerically computed by finding the solutions to transcendental equations (12) and (13). Since this computation relies in numerical methods, which may become time consuming in real-time applications, a parametric approach, based on a regression model, can be an attractive implementation.

In this section, the multiple linear regression model from [9] is revisited and it is extended to MMSE. Since the same approach will be applied independently to each estimator, the "LS" and "MMSE" labels from the variables will be omitted.

In this manner, the fitting polynomial that tries to match the i-th data point $\big(y_i = \log(N_{t,\text{LS}}^{opt})_i$, or $y_i = \log(N_{t,\text{MMSE}}^{opt})_i\big)$, can be expressed as [14],

$$y_i = \alpha_0 + \sum_{j=1}^{p}\alpha_j\, x_{i,j} + \varepsilon_i\ ,\ \forall\, i = 1, 2, ..., n\ ,\ \text{with } p < n \tag{15}$$

where, α_j are weight values, $x_{i,j}$ are the input variables of the model, ε_i is the approximation error, and n and p are the size of the data points set and the order of the polynomial, respectively[3].

From the scalar fitting of previous equation (15), a general expression can be written in a matrix form [14]. This way, the whole data points set can be approximated at once, which ultimately can be used to estimate the weights of the regression model. A common way to proceed with this computation is by applying the pseudoinverse of the regression matrix as follows,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}\ \rightarrow\ \widehat{\boldsymbol{\alpha}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}\,, \tag{16}$$

in which the previous scalar values are gathered in a matrix form as,

$$\mathbf{y} = \begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix}^T \qquad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 & \cdots & \varepsilon_n \end{bmatrix}^T$$
$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_0 & \cdots & \alpha_p \end{bmatrix}^T \qquad \widehat{\boldsymbol{\alpha}} = \begin{bmatrix} \widehat{\alpha}_0 & \cdots & \widehat{\alpha}_p \end{bmatrix}^T$$
$$\mathbf{X} = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,p} \end{bmatrix}. \tag{17}$$

and $\widehat{\alpha}_j$ are the estimated weights of the polynomial.

As it was shown in [9] for the LS case, the values of N_t^{opt} showed an almost linear relationship with $N_c(v)$ in log-log

[3]Despite the reuse of variables (x, y) from previous Section II, from now on, they will refer only to the regression model variables from (15).

scale, and a less trivial relationship with SNR^{dB}. Now, it has been found that, for the MMSE case, the same almost linear relationship is followed. Thus, y_i can be approximated with a polynomial whose inputs ($\mathbf{X}$) are constructed with SNR_i^{dB} and $\log\big(N_c(v_i)\big)$ values as follows,

$$\mathbf{x}_i = [\mathbf{X}]_{(i,:)} = \big(\mathbf{x2}_i \otimes \mathbf{x1}_i\big)^T \tag{18}$$

$$\text{where } \begin{cases} \mathbf{x1}_i = \Big[1 \;\; \log\big(N_c(v_i)\big) \;\; \cdots \;\; \log\big(N_c(v_i)\big)^{p'}\Big]^T \\ \mathbf{x2}_i = \Big[1 \;\; SNR_i^{dB} \;\; \cdots \;\; \big(SNR_i^{dB}\big)^{p'}\Big]^T \end{cases},$$

and p' is the maximum order of each variable.

In this case, the order of the polynomial is $p = (p'+1)^2$, however, simpler models, which do not have cross-terms between SNR_i^{dB} and $\log\big(N_c(v_i)\big)$, are constructed with,

$$\mathbf{x}_i = \big(\mathbf{x1}_i \,||\, \mathbf{x2'}_i\big)^T \tag{19}$$

$$\text{where } \mathbf{x2'}_i = \Big[SNR_i^{dB} \quad \cdots \quad \big(SNR_i^{dB}\big)^{p'}\Big]^T.$$

Here, the order of the polynomial is reduced to $p = 2\,p' + 1$.

As a last step, the optimum averaging can be computed parametrically using the estimated weights of each estimator. This way, the resultant polynomial model is,

$$\widehat{y}_i = \log(\widehat{n}_{t,i}^{opt}) = \widehat{\alpha}_0 + \sum_{j=1}^{p} \widehat{\alpha}_j \, x_{i,j} \;, \quad \widehat{N}_{t,i}^{opt} = \Big\lfloor \widehat{n}_{t,i}^{opt} \Big\rceil . \tag{20}$$

V. Numerical Results

In this paper, simulations have been computed using the same 5G-NR channel model and the same system deployment as in [7]. It basically consists in a mmWave transmission model modulated at 28 GHz (f_c), with an OFDM scheme of 512 subcarriers (K), whose subcarrier spacing (Δf) is 120 kHz and the overhead of the CP $\left(\frac{L_{cp}}{K}\right)$ is 7%. As a convention, the coloring of the results will be orange for LS and blue for MMSE.

A. Performance Analysis

To start with, in Fig. 1 it is plotted the MSE performance of an ST when the averaging is implemented at the trivial SOA coherence time or at the optimum averaging from [7] and [8].

In general, the SOA averaging (empty markers with dotted lines) provides worse results, particularly in high SNR which shows an error floor. On the other hand, the optimum averaging (full markers with solid lines) consistently improves the MSE for all the speed scenarios. As expected, the MMSE shows the best performance, which is more apparent in low SNR environments with medium-fast speeds.

B. Delay Analysis and Accuracy of the Regression Model

In Fig. 2, N_t^{opt} values for different SNR conditions are plotted against the coherence time of their respective scenario (expressed as N_c in the bottom axis or v in the top axis).

From this figure, it can be appreciated the apparent linear relationship between the data points. For this reason, it has been illustrated the multilinear regression model from Section IV (colored lines from red to green) that matches the original data

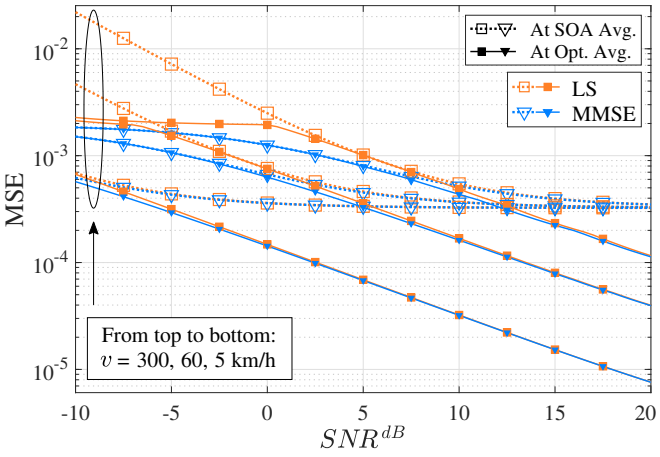

Fig. 1. MSE performance of optimum and SOA averaging for both LS and MMSE estimators.

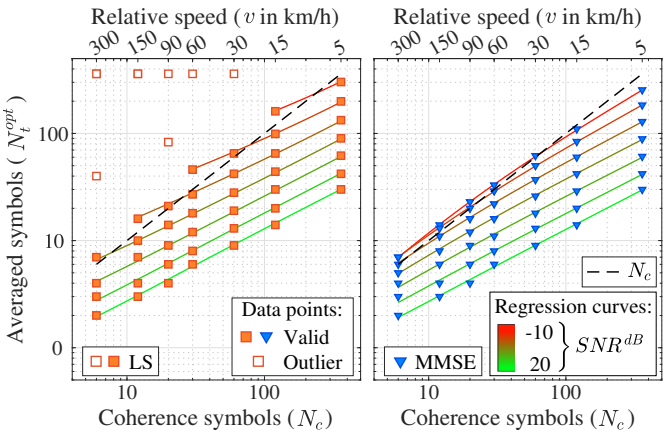

Fig. 2. Relationship between $N_{t,\text{LS}}^{opt}$ (left) and $N_{t,\text{MMSE}}^{opt}$ (right) with respect to N_c under different scenarios. Also, parametric approximation (colored solid lines) of a polynomial with $p' = 5$ and cross-terms input variables.

points (colored markers) with high accuracy. Also, to train the parametric model, the outlier points from the LS case (empty squares in Fig. 2-left) were omitted [9].

It must be noted that the polynomial regression from Fig. 2 (colored solid lines) is computed with the input variables of (18), which has cross-terms between $\log\big(N_c(v_i)\big)$ and SNR_i^{dB}, and their maximum order is $p' = 5$. The reason to use this configuration is depicted in the following Fig. 3, in which it is plotted the discrepancy between the MSE at the approximated $\widehat{N}_{t,i}^{opt}$ and the MSE at the original $N_{t,i}^{opt}$. In the top figure it is represented the maximum error, while in the bottom it is depicted the mean error, which are computed as follows,

$$\epsilon^{Max} = \frac{\max(\epsilon)}{\Psi\big(N_t^{opt}\big)} \cdot 100 \;, \quad \epsilon^{Avg} = \frac{\mathbb{E}\{\epsilon\}}{\Psi\big(N_t^{opt}\big)} \cdot 100$$
$$\text{with } \epsilon = \Big|\Psi\big(N_t^{opt}\big) - \Psi\big(\widehat{N}_t^{opt}\big)\Big| \;. \tag{21}$$

From these results, when $p' > 2$, ϵ^{Max} is very similar for all regression models, with the best results for $p' = 4$ (Fig. 3-top).

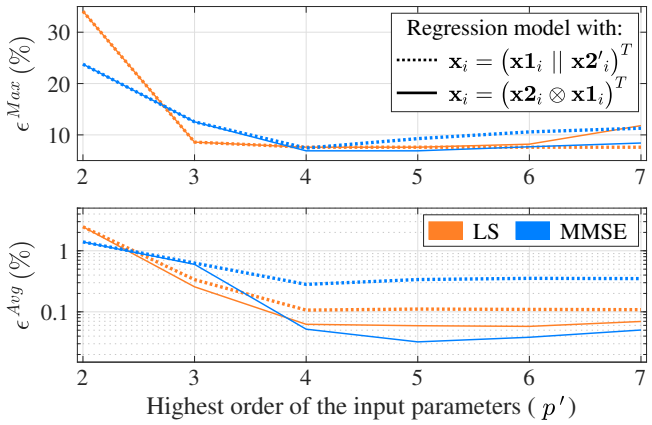

Fig. 3. Maximum (top) and averaged (bottom) error in terms of the MSE due to different regression models to approximate the optimum averaging.

On the other hand, when $p' > 3$, ϵ^{Avg} is minimal for both estimators (Fig. 3-bottom). Specifically, results are the most accurate if the polynomial model is constructed considering cross-terms (solid lines) as in (18).

C. Proposed Estimator Classifier

Once the parametric model is shown to provide accurate results, in this section it is proposed a classification method to select the most suitable estimator for each scenario.

To start with, it is true that a trivial solution consists in using always the MMSE estimator since it provides the best performance in terms of MSE (Fig. 1) and delay (Fig. 2)[4]. However, there are occasions in which the LS estimator can be a more attractive solution, particularly because its simpler expression does not rely on ζ_m from (10), and it does not need to be updated at each m-th time index.

Thus, to select an estimator, the following inequalities have be proposed as classification rules related to,

$$\text{Delay:} \quad N^{opt}_{t,\text{LS}} - N^{opt}_{t,\text{MMSE}} \underset{\text{MMSE}}{\overset{\text{LS}}{\lessgtr}} t_d \cdot N^{opt}_{t,\text{MMSE}} \quad (t_d \in [0,1]), \tag{22}$$

$$\text{Performance:} \quad \Psi^{opt}_{\text{LS}} - \Psi^{opt}_{\text{MMSE}} \underset{\text{MMSE}}{\overset{\text{LS}}{\lessgtr}} t_p \cdot \Psi^{opt}_{\text{MMSE}} \quad (t_p \in [0,1]), \tag{23}$$

where $\Psi^{opt} = \Psi\left(N^{opt}_t\right)$, and t_d, t_p are design thresholds.

In Fig. 4, these classifiers are depicted for different speeds and SNR values when $t_d = t_p = 10\%$. As it can be seen, in general, when there are high speeds and low SNR conditions (left diagonal side), the selected estimator is MMSE, whereas under low speeds and high SNR values (right diagonal side) the selected estimator is LS. In essence, this means that if the channel conditions are favorable, the LS estimator is good enough to be selected instead of the MMSE.

It is worth noticing that altough the *performance classifier* (Fig. 4-middle) shows a clear boundary between estimator

[4]Note that Fig. 2-right does not have outlier points.

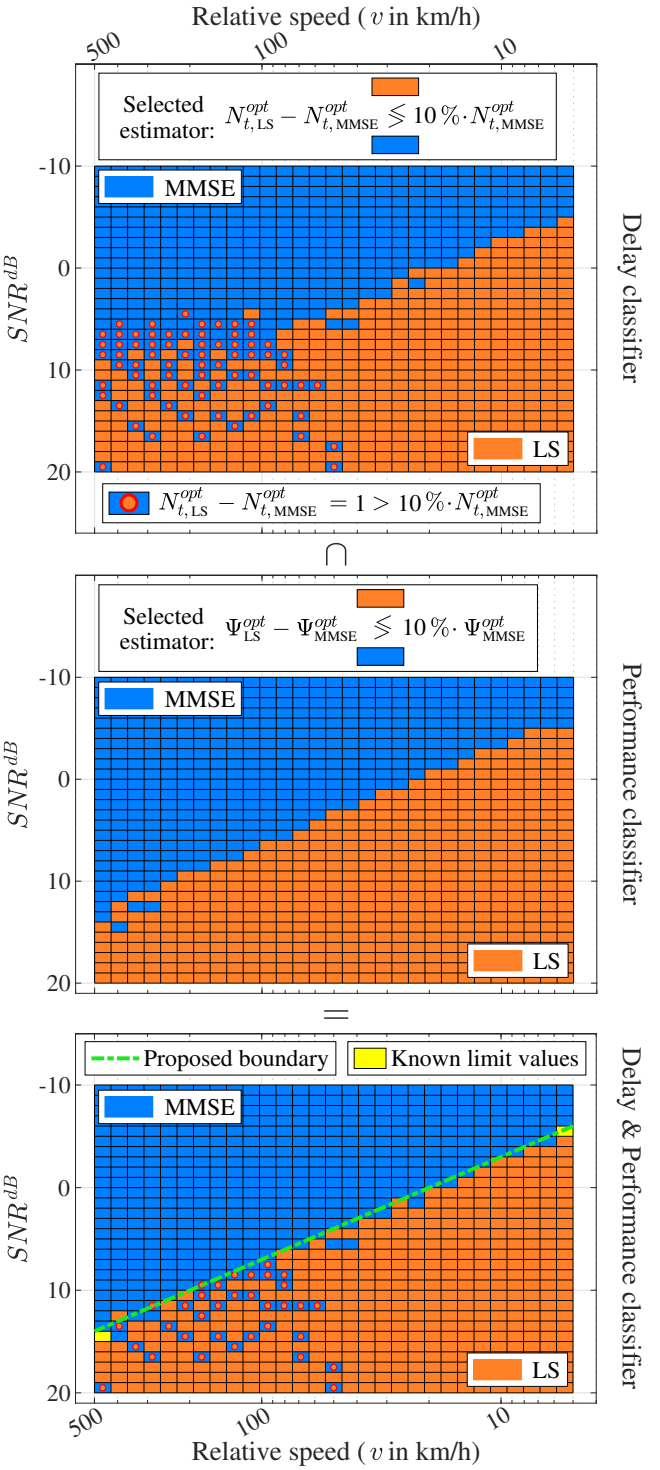

Fig. 4. Classification criterions for different speeds and SNR values: *delay* (top), *performance* (middle) and the intersection of both (bottom). Also, linear approximation of LS and MMSE regions (green dashed line).

regions, the *delay classifier* (Fig. 4-top) is more irregular with sparsed blue regions in the bottom-left part of the grid. This behavior happens because under high speed scenarios, $N^{opt}_{t,\text{MMSE}} < 10$. Then, since $N^{opt}_{t,\text{LS}} - N^{opt}_{t,\text{MMSE}}$ is usually equal to 1, which is greater than $10\% \cdot N^{opt}_{t,\text{MMSE}}$, the *delay classifier*

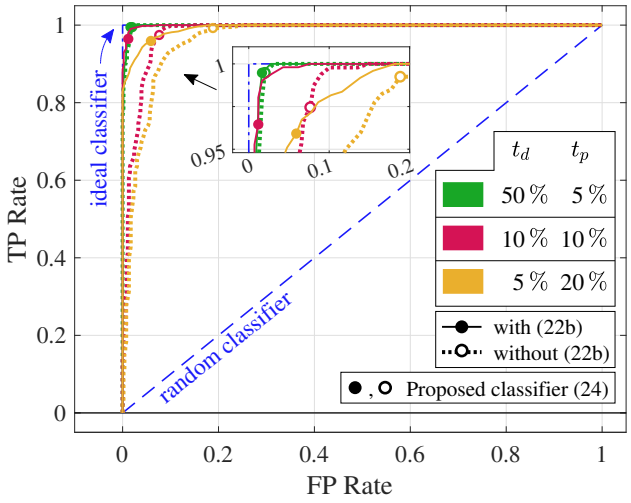

Fig. 5. ROC of the proposed linear boundary between LS and MMSE regions for different t_p and t_d values.

is not fulfilled by LS. To avoid this "blue" sparsity, the *delay classifier* from (22) can be adjusted by treating the following *special case* apart (blue squares with orange circles),

$$\text{If} \;\; N_{t,\text{LS}}^{opt} - N_{t,\text{MMSE}}^{opt} = 1 > t_d \cdot N_{t,\text{MMSE}}^{opt} \;\rightarrow\; \text{LS}\,. \qquad (22b)$$

Then, in Fig. 4-bottom the intersection ($\cap$) of both sets of selected estimators is computed and a complete classifier based on *delay* & *performance* criterions is depicted.

Now, the main issue about these results is that for each scenario, conditions (22), (22b) and (23), which depend on N_t^{opt} and Ψ^{opt}, need to be computed. As an alternative, since the boundary of LS and MMSE regions in Fig. 4-bottom is very distinct, the following simplified decision method (dashed green line) is proposed,

$$a\log(v_i)+b-SNR_i^{dB} \underset{\text{MMSE}}{\overset{\text{LS}}{\lessgtr}} 0\,, \text{with} \begin{cases} a = \frac{SNR_2^{dB}-SNR_1^{dB}}{\log(v_2)-\log(v_1)} \\ b = SNR_2^{dB} - a\log(v_2) \end{cases}. \qquad (24)$$

where a and b are the slope and the offset of the line, and (v_1, SNR_1^{dB}) and (v_2, SNR_2^{dB}) are two known configuration parameters (yellow squares), whose only condition is to belong to MMSE region and to be contiguous to LS region. This way, the *delay* & *performance classifier* can be simplified as a decision method constructed with a linear boundary.

Finally, in Fig. 5, the accuracy of (24) is shown under different t_d and t_p tolerances. In addition, the robustness of the proposed boundary is tested when (22b) is considered (solid lines) or when it is omitted (dotted lines). This figure compares the rate of true positives (TP-vertical axis) in front of the rate of false positives (FP-horizontal axis), which is known as the receiver operating characteristic curve (ROC) [15]. An ideal classifier (dash-dotted blue line) reaches the upper-left corner, whereas a random guess classifier (dashed blue line) provides a 45º diagonal.

From these results, it can be seen that, in general, the linear boundary from (24) is reliable since its TP values are close to the upper-left corner (full and empty circles). Additionally, results are optimum when $t_p < t_d$ (green curves), particularly, when the *special case* (22b) is considered (solid lines). Overall, it can be stated that the implementation of the linear approximation (24) is promising to select the most suitable estimator for each scenario.

VI. Conclusion

This paper proposes a practical implementation of ST technique in two application areas. On one hand, a multilinear regression model is presented to parametrically approximate the optimum averaging for the MMSE estimator. This way, previous solutions that required numerical methods can be avoided. On the other hand, an estimator classifier for LS and MMSE, which selects the most suitable technique depending on delay and performance tolerances, is proposed. Finally, a simplified version of this classifier is designed and its accuracy is verified under different configurations.

References

[1] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, 2014.

[2] A. A. Zaidi, R. Baldemair, H. Tullberg, H. Bjorkegren, L. Sundstrom, J. Medbo, C. Kilinc, and I. Da Silva, "Waveform and numerology to support 5G services and requirements," *IEEE Communications Magazine*, vol. 54, no. 11, pp. 90–98, 2016.

[3] W. Huang, C. Li, and H. Li, "On the power allocation and system capacity of OFDM systems using superimposed training schemes," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 4, pp. 1731–1740, 2009.

[4] J. C. Estrada-Jiménez and M. J. Fernández-Getino García, "Partial-data superimposed training with data precoding for OFDM systems," *IEEE Transactions on Broadcasting*, vol. 65, no. 2, pp. 234–244, 2019.

[5] K. Chen-Hu, M. J. Fernández-Getino García, A. M. Tonello, and A. García Armada, "Low-complexity power allocation in pilot-pouring superimposed-training over CB-FMT," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 12, pp. 13 010–13 021, 2021.

[6] D. Verenzuela, E. Björnson, and L. Sanguinetti, "Spectral and energy efficiency of superimposed pilots in uplink massive MIMO," *IEEE Transactions on Wireless Communications*, vol. 17, no. 11, pp. 7099–7115, 2018.

[7] I. Piqué Muntané and M. J. Fernández-Getino García, "Optimum averaging of superimposed training schemes in OFDM under realistic time-variant channels," *IEEE Access*, vol. 9, pp. 115 620–115 631, 2021.

[8] ——, "Superimposed training aware of realistic time-variant channel models in OFDM: Optimization of the averaging in MMSE estimation," *Submitted to IEEE Global Communications Conference (GLOBECOM)*, 2023.

[9] ——, "Parametric approximation to optimal averaging in superimposed training schemes under realistic time-variant channels," in *13th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, 2022, pp. 738–743.

[10] R. H. Clarke, "A statistical theory of mobile-radio reception," *The Bell System Technical Journal*, vol. 47, no. 6, pp. 957–1000, 1968.

[11] M. C. Jeruchim, P. Balaban, and K. S. Shanmugan, *Simulation of Communication Systems: Modeling, Methodology and Techniques*, 2nd ed. USA: Kluwer Academic Publishers, 2000.

[12] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*, 1st ed. USA: Cambridge University Press, 2008.

[13] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. USA: Prentice-Hall, Inc., 1993.

[14] N. Draper and H. Smith, *Applied Regression Analysis*, 3rd ed. New York: Wiley, 1998. [Online]. Available: https://doi.org/10.1002/9781118625590

[15] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012.