**Comments on: A Review on the Optimal Fingerprinting Approach in Climate Change Studies by Chen, Chen and Mu (2021)**

It is good to see authors with statistical expertise weighing in on the issues pertinent to Allen and Tett (1999, "AT99" herein) and Optimal Fingerprinting as raised in McKitrick (2021, "M21" herein). The authors have clearly grasped many of the core issues but have not provided enough of an advance on the M21 paper to merit publication. The material up to the end of Section 2.2 is really just a restatement of M21 supplemented with a few minor comments and conjectures. In other places they tend to assume away the issues raised in M21, and at certain points their arguments are incorrect.

Detailed comments are as follows.

**Section 2**

Lines 071-073: "optimality" requires not only minimum variance but also unbiasedness.

Line 107 Equation (1) is not an identification condition. It is usually called the "error independence" condition, and it is also known as a moment condition.

Lines 110-112: Referring to the error independence condition, the authors say "Our reading is that AT99 implicitly assumed such a condition." But this is mere speculation. AT99 did not state the condition, nor is it implied by the assumptions they did state, nor did they ever propose a test for whether it fails to hold. It is more likely that they overlooked it, as did every other author in the field. The authors should therefore acknowledge that AT99 failed to list the assumption and the implications of not testing for violations of it have never been considered.

The present authors go on simply to assume (line 229) that the condition holds, which makes much of their subsequent analysis irrelevant since the potential for the error independence condition to fail is one of the key problems for the AT99 method.

Lines 113-116 This is a repetition of lines 070-078.

Lines 126-131 The authors observe that AT99 ignored the distinction between the dimensions of **P** and **I**, as was discussed in M21 (Section (ii)). They also note that AT99 did not state the additional requirement that $\kappa' > m$. This is a trivial observation: it merely states that in any regression model there can't be more variables than observations. That's not a GM condition specifically, that's just common sense. There is no need to fault AT99 for not stating that assumption.

Line 132 The authors say they will ignore the randomness of **P** "for the moment" but never come back to it even when they get to topics where the randomness of **P** interferes with the analysis. This is another shortcoming of the paper. The author's analysis would greatly benefit from a detailed exploration of the difficulties in establishing unbiasedness and consistency if **P** is a random matrix.

Line 133 Reference to equation (4) should be to (AT1)

Lines 148-151 The authors need to clarify here what they mean by the "null setting" on the climate model. The models in use are constructed with specific tuning to replicate the 20th century surface record in response to observed forcing (including GHG's) subject to parameterized representation of known physical processes. The "null" run involves using such a model without external forcing. But the model still embeds the assumptions necessary to achieve the target fit, including constraints on internal variability, so it is not truly "null" for the purpose of Optimal Fingerprinting. A truly null model would be a climate model tuned as best as possible to reproduce the 20th century surface record without greenhouse forcing, which might imply different parameterizations and mechanisms to represent internal variability.

Lines 159-163 This section is a bit unclear but it seems the authors are claiming, without proof, that the only requirement on the model-generated $\widehat{\boldsymbol{C}}_N$ matrix for $\tilde{\beta}$ to be BLUE is independence of the ensemble runs with each other and with $\boldsymbol{X}$. I would first note that by referring to an ensemble they appear to have in mind an ensemble mean, but if so they should make that clear in the notation. Currently it is ambiguous.

More generally, there are a few missing elements in this discussion. First, the authors need to explain what they mean by "independence" of the ensemble runs: it can refer to linear independence, orthogonality, zero covariance, covariance less than some cut-off value, exogeneity, predeterminedness or asymptotic independence. Depending on the use to be made of the independence condition some of these may be stronger than necessary. In the OLS case we require $plim\left(l^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\right) = S_X$ where $l$ is the sample size and $S_X$ is a deterministic, finite matrix. This in turn requires limits on the degree of dependence among the columns of $\boldsymbol{X}$ but not zero covariance or orthogonality, for instance.

Second, independence—however defined— is not enough. For the purpose of taking probability limits of least squares solutions some kind of stationarity condition is required, an assumption not mentioned by the authors, and one which may be a problem in climate model runs especially when generating the simulated response patterns since it is well-known that climate models run under GHG forcing are non-stationary (they exhibit "stochastic trends" or unit roots – see Kaufmann et al. 2013).

We also generally require that $\boldsymbol{X}$ be non-stochastic, which is a problem since $\boldsymbol{X}$ is derived from climate model runs and much of the literature subsequent to AT99 (such as Allen and Stott 2003) was concerned with dealing with the fact that $\boldsymbol{X}$ is a random matrix.

Third, in the AT99 case the corresponding condition would be $plim\left(l^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{P}^{\boldsymbol{T}}\boldsymbol{P}\boldsymbol{X}\right) = S_X$. Even if $\boldsymbol{X}$ satisfies the conditions to yield a deterministic and finite plim, we need additionally to know what the required conditions are on $\boldsymbol{P}$, especially when it is assumed to be stochastic.

Consequently the authors have not established that independence of the $\widehat{\boldsymbol{C}}_N$ matrix is sufficient to remedy the problems in the AT99 framework.

Lines 163-165 While it is true that using a climate model to generate $\widehat{\boldsymbol{C}}_N$ would mean errors arising from a misspecified regression model are not introduced into $\widehat{\boldsymbol{C}}_N$, such errors are obviously still present in the Optimal Fingerprinting regression itself, so this is not a helpful point. Moreover additional errors may be introduced by misspecification of the climate model, as the authors note in lines 167-169.

Lines 165-167 The sentence "Let $\widetilde{\boldsymbol{C}}_N$ be the underlying covariance matrix of the ensembled null simulations to which $\widehat{\boldsymbol{C}}_N$ converges to in probability" should be stated as a requirement, along with all necessary conditions for such a matrix to exist. There is no guarantee a single climate model converges in this way, let alone an ensemble of many climate models. What is needed here is one or more testable restrictions which allow us to assume that a matrix $\widetilde{\boldsymbol{C}}_N$ exists such that, for instance,

$$\lim_{l/p\to\infty} \widehat{\boldsymbol{C}}_N = \widetilde{\boldsymbol{C}}_N$$

where $l$ is the sample length and $p$ is the dimension of $\widetilde{\boldsymbol{C}}_N$. Of course one of the challenges, as noted in M21, is that the rank of the estimate $\widehat{\boldsymbol{C}}_N$ is constrained by $\kappa$ regardless of $l$ so ordinary asymptotic theory is unavailable. The present authors do not address how to get around this.

More generally, the condition required is not merely that $\widehat{\boldsymbol{C}}_N$ converges to *something*, but that it converges to $E(\boldsymbol{u}\boldsymbol{u}^T)$ because if it doesn't then there is no justification for using a climate model in the first place. This is precisely the point raised in M21 in the discussion of necessary condition [N3]. The present authors note this issue in passing but otherwise ignore it (and in lines 377 to 379 explicitly wish it away), which amounts to assuming [N3] holds. But then it is trivial to say that the failure to prove [N3] doesn't matter because we assume it holds.

Lines 191-193 The statement "AT99 seems to suggest that $\widehat{\boldsymbol{C}}_N$ was designed to consistently estimate $\Sigma_{i-1}^{k}\lambda_i v_i v_i^T$, the first κ-terms in the spectral decomposition of $\boldsymbol{C}_N$" should be the other way around: $\Sigma_{i-1}^{k}\lambda_i v_i v_i^T$ was designed to estimate $\widehat{\boldsymbol{C}}_N$ in the hopes it would correspond to $\boldsymbol{C}_N$. The comment in line 194 still applies.

Line 196 The authors claim that "Our understanding is that $\widehat{\boldsymbol{C}}_N$ is a consistent estimator of $\widetilde{\boldsymbol{C}}_N$". They haven't even shown that $\widetilde{\boldsymbol{C}}_N$ exists, nor have they offered any proof that $\widehat{\boldsymbol{C}}_N$ estimates it consistently. They can't simply make this assumption.

Lines 198-200 M21 already pointed this out (Section 2: "If GM condition (2) fails OLS coefficient estimates will still be unbiased but potentially inefficient")

Lines 204-215 Here the authors ignore the problem that $\boldsymbol{P}$ is a random matrix so the expectation as stated does not hold. More generally, equation (4) simply restates necessary condition [N4] in M21 and assumes it holds, including the conditional independence between $\boldsymbol{X}$ and $\boldsymbol{u}$. Consequently the "proof" of unbiasedness in lines 216-221 adds nothing to the discussion. It is trivial to state that if we assume [N4] holds, the failure to prove whether [N4] holds no longer matters.

Lines 228-230 In this section the authors proceed with an analysis that assumes $\widehat{\boldsymbol{C}}_N$ converges to $\widetilde{\boldsymbol{C}}_N$, assumes away any distinction between $\widetilde{\boldsymbol{C}}_N$ and $\boldsymbol{C}_N$, and assumes the missing GM condition holds. Under these assumptions $\tilde{\beta}$, the AT99 estimator, is unbiased. Then they show that if another unbiased estimator $\hat{\beta}$ exists, it must have the same expectation as $\tilde{\beta}$ and at least as large a variance. This is an obvious, but pointless, observation.

Lines 251-254 The authors say that if all their assumptions hold but $\kappa < l$ then $\tilde{\beta}$ is a "restricted BLUE" estimator. There is no such thing in the literature as a "restricted BLUE" estimator—the estimator still fails to be BLUE even if they assume away all the other problems, because of the rank restriction on $\boldsymbol{P}$. But this is already clear from M21. Here the authors should develop their idea more formally. They could try to make an argument that if $\boldsymbol{P}$ is a $\kappa \times l$ matrix defined such that $E\left(\boldsymbol{P}\boldsymbol{u}\boldsymbol{u}^{\boldsymbol{T}}\boldsymbol{P}^{\boldsymbol{T}}\right) = I_\kappa$ where the $\kappa$ subscript denotes the dimension of the identity matrix and $\kappa < l$, $\tilde{\beta}$ may have some attractive properties. However they would need to note that efficiency is lost due to the restriction on $\kappa$, and consistency is not guaranteed, which raises the question of why bother with an AT99 method when White's (1980) method is easier to use and is consistent. A formal comparison of the properties of the AT99 covariance matrix estimator to that of White (1980) would be illuminating.

Lines 255-259 This statement is wrong as presented. Failure to ensure conditional independence (equation (3) in M21) means $\tilde{\beta}$ will be biased and inconsistent even if $\widehat{\boldsymbol{C}}_N$ is reasonable. The appeal to feasible GLS does not help. The authors appear to be saying that $\tilde{\beta}$ will be unbiased and consistent if an estimator can be found that remedies all the regression misspecifications, which again is a trivial statement (even beyond the fact that feasible GLS doesn't necessarily fix the specification errors). The point of M21 was that no one has paid any attention to these issues so they are very much a matter of concern for interpreting past results.

**Section 2.3**

Lines 280-2294 The authors don't explain how they concluded the null hypothesis in AT99 is $H_0\colon \widetilde{\boldsymbol{C}}_N = \boldsymbol{C}_N = Var(\boldsymbol{u})$. The language in AT99 is so vague there are several different ways $H_0$ could be written. And if that truly is the null AT99 had in mind, the test described in AT99 and written out in lines 290 to 295 does not test it. The quadratic form in line 290 can be used to test that the mean of $\widetilde{\boldsymbol{u}}$ is 0, but not that $\widetilde{\boldsymbol{C}}_N = \boldsymbol{C}_N$. Tests of the equality of empirical covariance matrices are discussed in Gupta and Tang (1984). If $S_u$ denotes the empirical covariance matrix of $\widetilde{\boldsymbol{u}}$ and $S_C$ denotes the empirical covariance of the climate model control run (e.g. $\widetilde{\boldsymbol{C}}_N$) then Gupta and Tang (1984) show that if the underlying data generating process is Gaussian, a test of $H_0\colon \boldsymbol{C}_N = Var(\boldsymbol{u})$ would be given by

$$\Lambda = \frac{|S_u|^{\frac{l-1}{2}}|S_C|^{\frac{\kappa-1}{2}}}{|S|^{\frac{l+\kappa-1}{2}}}$$

where $S = \left(\frac{1}{l+\kappa-2}\right)\left((l-1)S_u + (\kappa-1)S_C\right)$. However, this statistic does not have a standard $\chi^2$ distribution and its critical values need to be computed using the method given in Gupta and Tang (1984). Gupta and Xu (2006) discuss modifications when the underlying distributions are not assumed to be Gaussian. The test statistics take various other forms unrelated to the AT99 RC test. And the derivation relies on asymptotic theory which does not apply when $\kappa$ is fixed and strictly less than $l$, as is the case in AT99. Consequently the RC test has no known connection to a test of equality of covariance matrices.

Line 294 The same-order notation term $o_p(1)$ is inadmissible here because the degrees of freedom $\kappa - m$ does not change as $l \to \infty$.

Lines 305-325 Notwithstanding the lack of clarity about $H_0$ the authors make a valid observation that under the alternative, the RCT is no longer distributed $\chi^2$. This section raises important questions about the power of the RC Test. Power analyses of tests of this kind require assumptions about the relative magnitudes of, in this case, $\kappa$ and $p$ (the number of signal vectors in $\boldsymbol{X}$). Both tend to be small in optimal fingerprinting applications. I encourage the authors to develop this line of research into a proper power analysis.

**Section 3**

Lines 335-345 There seems to be some incorrect notation here. In equation (10) the $\hat{u}_i's$ are scalars so $l^{-1}\Sigma(\hat{u}_i\hat{u}_i^T) = l^{-1}\Sigma(\hat{u}_i^2)$ which is also a scalar and therefore cannot be a consistent estimator of the $l \times l$ matrix $\boldsymbol{C_N}$. If they meant to write the outer product of $\tilde{u}$ then line 347 is incorrect since in this case since $\boldsymbol{C}_{N1}^{-1}$ does not exist therefore neither does $\tilde{\beta}_{FGLS}$. A truly feasible FGLS estimator requires using an estimated heteroskedastic function which is presented in most introductory econometrics textbooks (e.g. Wooldridge 2019). M21 discusses as an alternative to FGLS White's (1980) heteroskedasticity-consistent matrix which is simple to compute and generally preferred to FGLS because it is always unbiased and consistent whereas FGLS may be biased if the wrong heteroskedastic function is used. Hence this section is undermined by errors in presentation and the fact that a simpler and superior option is already shown in M21.

**Conclusion**

Overall the authors have not advanced the examination of the AT99 method beyond what is in M21. They have confirmed the basic points in M21 and they propose that if the main problems can be assumed away, the method survives the M21 criticism. This is a tautology.

Despite these criticisms I hope that the authors will pursue the topic in more depth, for instance by reviewing the literature on testing equality of covariance matrices and undertaking a power analysis so as to work towards constructing a valid test of the null that $\widehat{\boldsymbol{C}}_{\boldsymbol{N}}$ as computed by a climate model yields a valid covariance matrix for $\tilde{\beta}$ under the assumption of a fixed $\kappa$. However, I would also note that the issues raised thereby will tend to involve rather advanced theoretical concepts in econometrics and statistics and may need to be explored in the statistics literature rather than in climatology journals.

Signed review: Ross McKitrick

**References:**


Allen, M.R. and S.F..B. Tett (1999) Checking for model consistency in optimal fingerprinting. *Climate Dynamics* 15: 419–434 https://doi.org/10.1007/s003820050291.

Allen, M.R. and P.A. Stott (2003) Estimating Signal Amplitudes in Optimal Finger-Printing, Part I: Theory. Climate Dynamics 21:477—491. DOI 10.1007/s00382-003-0313-9

Gupta, A.K. and J. Tang (1984) Distribution of likelihood ratio statistic for testing equality of covariance matrices of multivariate Gaussian models. *Biometrika* 71(3): 555-9.

Gupta, A.K. and J. Xu (2006) On some tests of the covariance matrix under general conditions. *AISM* 58:101-114.

Kaufmann, R.K., Kauppi, H., Mann, M.L. et al. (2013) Does temperature contain a stochastic trend: linking statistical results to physical mechanisms. *Climatic Change* 118, 729–743. https://doi.org/10.1007/s10584-012-0683-2

McKitrick, Ross R (2021) Checking for Model Consistency in Optimal Fingerprinting: A Comment. Climate Dynamics August 2021 https://doi.org/10.1007/s00382-021-05913-7

White, Halbert (1980) A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica* 48(4) pp. 817—838.

Wooldridge, Jeffrey (2019) *Introductory Econometrics: A Modern Approach 7th ed.* Boston: Cengage.