

[10.1071/CP22280](https://doi.org/10.1071/CP22280)

Crop & Pasture Science

Supplementary Material

Modelling spatial and temporal correlation in multi-assessment perennial crop variety selection trials using a multivariate autoregressive model

J. De Faveri^{A,}, A. P. Verbyla^A, and R. A. Culvenor^B*

^AThe University of Queensland, Queensland Alliance for Agriculture & Food Innovation (QAAFI), Brisbane, Qld 4001, Australia.

^BCSIRO Agriculture & Food, Black Mountain, ACT 2601, Australia.

*Correspondence to: J. De Faveri The University of Queensland, Queensland Alliance for Agriculture & Food Innovation (QAAFI), Brisbane, Qld 4001, Australia Email: j.defaveri@uq.edu.au

Supplementary Information

ASReml-R code for fitting the models:

EXAMPLE CODE FOR 2DIMVAR1 MODELS

```
# Example is a field trial with 30 varieties with data measured at 5
times
```

```
# Data (freq) recorded on each plot with 120 Plots in 15 Columns by 8
Rows
```

```
# Need to model the spatial (row x col) by temporal
covariance/correlation structure
```

```
# First fit a separate spatial ar1(Col)xar1(Row) model for each time
(Year).
```

```
# This gives an idea of the spatial correlation for each Year and
whether the spatial parameters differ
```

```
# This model does not account for the correlation between Years but does
allow for a different residual variance for each Year. It also allows
for different spatial correlation parameters for each Year.
```

```
diag.asr <- asreml(freq ~ Year ,
                  random = ~ Genetic model + at(Year):Rep
                  + at(Year):Row+ at(Year):Col,
                  residual = ~ dsum(~ar1(Col):ar1(Row)|Year),
                  data=pasture_multitime, maxiter=20)
```

```
# A comparison with a separable diag(Year):ar1(Col):ar1(Row) model shows
whether assuming common row and col parameters may be ok.
```

```
# This model assumes common spatial parameters for each Year and allows
for a different residual variance for each Year.
```

```
# But this still does not account for the correlation between Years or
any temporal by spatial interaction
```

```
diag1.asr <- asreml(freq ~ Year ,
                   random = ~ Genetic model + at(Year):Rep
```

```

        + at(Year):Row+ at(Year):Col,
        residual = ~ diag(Year):ar1(Col):ar1(Row),
        data=pasture_multitime, maxiter=20)

# Now fit 2DIMVAR1 residual model based on FA models to allow for
different spatial correlation parameters for each Year. This model
allows for different residual variance for each Year and correlations
between Years and models the spatio-temporal interaction.

# make 5 (number of times) copies of Year - Year1, Year2, Year3, Year4,
Year5

# use lmmtools package and mvarlinit function for starting values

MV1.sv <- asreml(freq~ Year,

        random =~Genetic model + at(Year):Rep
+at(Year):Row+at(Year):Col#+Col:Row

+ fa(Year1,1):ar1(Col):ar1(Row)
+ fa(Year2,1):ar1(Col):ar1(Row)
+ fa(Year3,1):ar1(Col):ar1(Row)
+ fa(Year4,1):ar1(Col):ar1(Row)
+ fa(Year5,1):ar1(Col):ar1(Row),

        family=asr_gaussian(link="identity",dispersion=1e-04),
        data=pasture_multitime,maxit=50,start.values=T,
        R.param=diag.asr$R.param, G.param=diag.asr$G.param)

# take any variance parameters that have gone to the boundary and put as
small starting value

gam<-MV1.sv$vparameters.table
gam$Value[gam$Constraint=="B"]<-0.1
gam$Constraint[gam$Constraint=="B"]<-"P"

# get initial starting values using mvarlinit
init.var1<-diag.asr$vparameters[igrep(list("Col:Row","!Year"),
names(diag.asr$vparameters))]

```

```

names (init.var1) <- paste0 ("Year", 1:5)

new.gam <- mvarlinit (gam, init.var = init.var1, spatial = list ("Col", "Row"),
Traits = paste0 ("Year", 1:5), n.trait = 5)

MV1.asr <- asreml (freq ~ Year,

    random = ~ Genetic model + at (Year) : Rep
    + at (Year) : Row + at (Year) : Col # + Col : Row

    + fa (Year1, 1) : ar1 (Col) : ar1 (Row)

    + fa (Year2, 1) : ar1 (Col) : ar1 (Row)

    + fa (Year3, 1) : ar1 (Col) : ar1 (Row)

    + fa (Year4, 1) : ar1 (Col) : ar1 (Row)

    + fa (Year5, 1) : ar1 (Col) : ar1 (Row),

    family = asr_gaussian (link = "identity", dispersion = 1e-04),

    data = pasture_multitime, maxit = 20, G.param = new.gam)

# calculate Sigma and Omega_r Omega_c

Pinv <- matrix (nrow = 5, ncol = 5,
data = c (t11fa1, t21fa1, t31fa1, t41fa1, t51fa1, t12fa1, t22fa1, t32fa1, t42fa1, t5
2fa1, t13fa1, t23fa1, t33fa1, t43fa1, t53fa1, t14fa1, t24fa1, t34fa1, t44fa1, t54f
a1, t15fa1, t25fa1, t35fa1, t45fa1, t55fa1))

Dr <-
matrix (nrow = 5, ncol = 5, data = c (phir1, 0, 0, 0, 0, 0, phir2, 0, 0, 0, 0, 0, phir3, 0, 0, 0,
0, 0, phir4, 0, 0, 0, 0, 0, phir5))

Dc <-
matrix (nrow = 5, ncol = 5, data = c (phic1, 0, 0, 0, 0, 0, phic2, 0, 0, 0, 0, 0, phic3, 0, 0, 0,
0, 0, phic4, 0, 0, 0, 0, 0, phic5))

Sigma <- Pinv %*% t (Pinv)

Omegar <- Pinv %*% Dr %*% solve (Pinv)

Omegac <- Pinv %*% Dc %*% solve (Pinv)

```

Estimated spatial dependency and temporal matrices from 2DIMVAR1 residual model (M12).

Σ

```
      [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 36.727    7.612    3.693    4.075    3.765
[2,]  7.612  112.629   83.786  106.119   47.109
[3,]  3.693   83.786   79.969   93.677   54.716
[4,]  4.075  106.119   93.677  131.132   60.150
[5,]  3.765   47.109   54.716   60.150  311.056
```

Correlations (**Σ**)

```
> round(cov2cor(Sigma), 3)
      [,1] [,2] [,3] [,4] [,5]
[1,] 1.000 0.118 0.068 0.059 0.035
[2,] 0.118 1.000 0.883 0.873 0.252
[3,] 0.068 0.883 1.000 0.915 0.347
[4,] 0.059 0.873 0.915 1.000 0.298
[5,] 0.035 0.252 0.347 0.298 1.000
```

Ω_r

```
      [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 0.352    0.170    0.078   -0.187    0.005
[2,] 0.238   -0.026   -0.035   -0.024    0.106
[3,] 0.119    0.014   -0.123    0.025    0.098
[4,] 0.062    0.017    0.007   -0.102    0.127
[5,] 0.061    0.049   -0.030    0.183    0.133
```

Ω_c

```
      [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 0.212    0.107   -0.062   -0.070    0.001
[2,] -0.042    0.068    0.293    0.298   -0.051
[3,] -0.126    0.235   -0.005    0.271    0.005
[4,] -0.167    0.191    0.037    0.450   -0.013
[5,] -0.080   -0.126   -0.142    0.179    0.614
```

spatio-temporal covariance matrix between neighbouring plots in row direction

$\Omega_r \Sigma$

```
> ST1<- round(Omegar*%Sigma, 3)
> ST1
      [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 13.778    8.715    4.495    2.510    3.751
[2,]  8.715   -1.651   -0.587   -1.890   29.206
[3,]  4.495   -0.587   -0.546   -0.414   26.394
[4,]  2.510   -1.890   -0.414   -3.041   34.684
[5,]  3.751   29.206   26.394   34.684   53.273
```

> # spatio-temporal covariance matrix between neighbouring plots in col direction

$\Omega_c \Sigma$

```

> ST2<- round(Omegac%*%Sigma,3)
> ST2
      [,1]  [,2]  [,3]  [,4]  [,5]
[1,]  8.103  1.067 -1.733 -2.741 -1.377
[2,]  1.067 61.090 54.092 70.490 21.203
[3,] -1.733 54.092 44.485 59.796 28.144
[4,] -2.741 70.490 59.796 81.292 33.391
[5,] -1.377 21.203 28.144 33.391 187.703

```

Calculate spatial correlation between neighbouring plots at each time from M12(to compare with original spatial correlation parameters estimated from each timepoint - in Table 2). We can see these spatial correlations from M12 are quite similar to the correlations from the separate analyses.

In a similar way we can also estimate the spatial correlation between neighbouring plots at different times (not presented here).

Row dir

```

ST1[1,1]
# time 1:5 Row dir          ## from table 2 (separate time)
ST1[1,1]/Sigma[1,1] #  0.375  ## 0.385
ST1[2,2]/Sigma[2,2] # -0.015  ##-0.048
ST1[3,3]/Sigma[3,3] # -0.007  ##-0.118
ST1[4,4]/Sigma[4,4] # -0.023  ## 0.046
ST1[5,5]/Sigma[5,5] #  0.171  ## 0.049

```

time 1:5 Col dir

```

ST2[1,1]/Sigma[1,1] #  0.220  ## 0.261
ST2[2,2]/Sigma[2,2] #  0.542  ## 0.528
ST2[3,3]/Sigma[3,3] #  0.556  ## 0.562
ST2[4,4]/Sigma[4,4] #  0.620  ## 0.582
ST2[5,5]/Sigma[5,5] #  0.603  ## 0.622

```

Model terms used in Table 1 (from De Faveri et al 2015)

Name	Model Function	Correlation or variance	Example	correlation or covariance matrix
identity	id	correlation	id(Harvest)	I_h
identity variance	idv	variance	idv(Harvest)	$\sigma_h^2 I_h^1$
diagonal	diag	variance	diag(Harvest)	$\text{diag}(\sigma_{hj}^2)^{1,4}$
heterogeneous correlation	corh	variance	corh(Harvest)	$\text{diag}(\sigma_{hj}) \{I_h + \rho_t (J_h - I_h)\} \text{diag}(\sigma_{hj})^{2,3}$
unstructured	us	variance	us(Harvest)	G_h^5
factor analytic, order k	fa	variance	fa(Harvest,k)	$\Lambda_h \Lambda_h^T + \Psi_h^6$
autoregressive, order 1	ar1	correlation	ar1(Row)	$\Sigma_r^{\text{ar1}} = I_r + \sum_{j=1}^{r-1} \rho_r^j F_j^7$
autoregressive variance, order 1	ar1v	variance	ar1v(Row)	$\sigma_r^2 \Sigma_r^{\text{ar1},3,7}$
autoregressive heterogeneous variance, order 1	ar1h	variance	ar1h(Row)	$D \Sigma_r^{\text{ar1}} D^{1,3,7,8}$
exponential heterogeneous variance	exph	variance	exph(HarvestTime)	$D \Sigma_h^{\text{exp}} D^8$ where $\Sigma_h^{\text{exp}}_{ij} = \phi^{ t_r - t_j }$
antependence, order s	ante	variance	ante(HarvestTime,s)	Σ_h^{ante} where $\Sigma_h^{\text{ante}-1} = U^T D^* U^9$
at	at	compound	at(Trait)	separate structure for each level of Trait

¹ σ_h^2 , σ_{hj}^2 and σ_r^2 are variances.

² σ_{hj} are standard deviations.

³ ρ_t , ρ_r and ϕ are correlations.

⁴ $\text{diag}()$ is a diagonal matrix with elements specified.

⁵ G_t is a fully parameterized covariance matrix of order t .

⁶ Λ_t is a matrix of factor loadings, Ψ_t is a diagonal matrix of order t .

⁷ F_j is a matrix which has 'ones' on the j th sub and super-diagonals and zeros elsewhere.

⁸ D is a diagonal matrix of standard deviations.

⁹ U is a lower triangular matrix with 'ones' on the diagonal and $-u_{ij}$ on $(i-j)$ th sub-diagonal (where $(i-j) \leq s$) and zeroes elsewhere and where $Y_j = \sum_{k=1}^{s_j} u_{j(j-k)} Y_{j-k} + e_j$, with $s_j = \min(j-1, s)$. D^* is a diagonal matrix.