Your quick start guide to p-rep designs and how to generate them in CycDesigN

Dr. Vanessa Cave and Dr. Emlyn Williams

What is p-rep design?

A ***single-location*** partially replicated, or p-rep, design is a special type of unequally replicated block or row-column design in which some treatments are replicated twice, whilst the other treatments are unreplicated.

Below is an example of a single-location p-rep ***block*** design in which six of the 12 treatments occur in two blocks (i.e., are replicated twice), and the other six treatments occur in only one block (i.e., are unreplicated).

Block 1	Block 2	Block 3
5	6	1
12	1	12
11	11	2
4	8	7
7	4	10
3	9	6

But wait … doesn't partial replication run counter to the fundamental experimental design principle of replication?

Replication is a crucial feature of sound experimental design. It is essential for estimating the underlying variability (i.e., the residual variance), without which there is no basis for valid statistical inference. The key property of a p-rep design is that there is adequate replication on a subset of treatments to enable reliable estimation of the residual variance, and thus valid statistical inference. For example, in the design above, the six replicated treatments (1,4,6,7,11,12) serve to provide an estimate of the residual variance, and to allow for block effect adjustments.

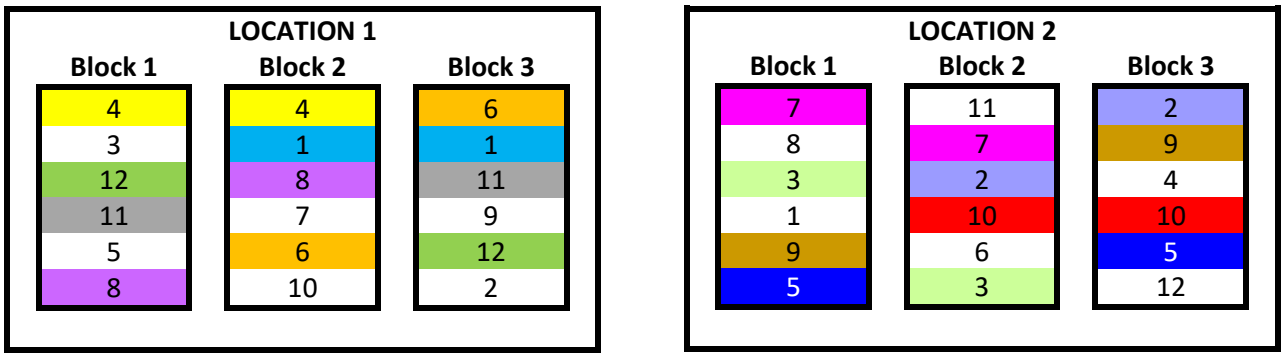
Similarly, here's an example of a single-location p-rep ***row-column*** design, again with 12 treatments but this time with 8 treatments duplicated within the 5 row by 4 column grid. The design is constructed such that the duplicated treatments can be used to estimate the residual variance, the row effects and the column effects.

	Column 1	Column 2	Column 3	Column 4
Row 1	3	8	1	7
Row 2	10	11	6	12
Row 3	1	4	7	5
Row 4	6	12	8	9
Row 5	5	3	2	11

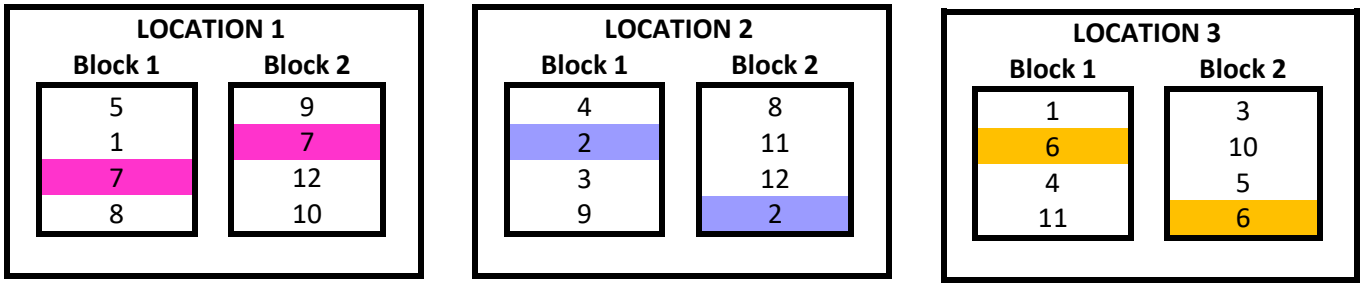
The p-rep design concept can also be extended across multiple locations. For a ***multi-location*** p-rep design, at each location some treatments occur twice, some occur only once, and others may not occur at all. Across the entire design, treatments are replicated either *r* or *r*+1 times. However, in

some situations, equally replicated designs are possible – that is, all treatments are replicated *r* times.

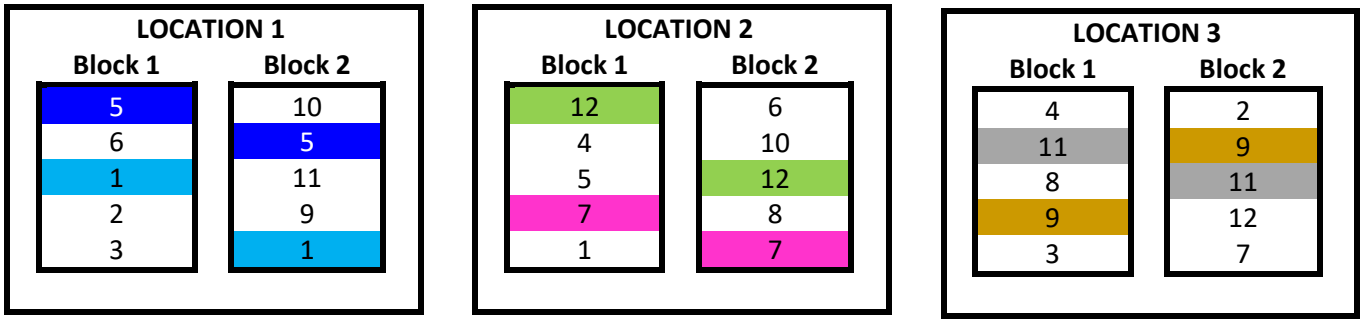
Here's an example of a two-location p-rep block design. Notice that in this example each treatment is equally replicated, with the unreplicated treatments at location 1 (2,3,5,7,9,10) replicated twice at location 2 (and vice-versa for those treatments duplicated at location 1).



And here's another example of a multi-location p-rep block design with equal replication. However, this time the 12 treatments are arranged across three locations but with only two blocks of size 4 per location. Thus, not all treatments can occur at all locations. And although each treatment is replicated twice over the whole design, some are replicated across different locations (e.g., treatment 5 occurs at location 1 and location 3) whereas others are duplicated within locations (e.g., treatment 7 occurs twice within location 1). This highlights a key challenge when generating multi-location p-rep designs: To even-out the partial replication within and between locations in such a way that the overall design has good efficiency.



In this final example of a multi-location p-rep block design, notice that some treatments are replicated *r*=2 times across the whole design, whilst the others are replicated *r*+1=3 times. And in particular, observe that the treatments replicated 3 times are those duplicated within a location (i.e., 1,5,7,9,11,12).



When are p-rep designs used?

p-rep designs are useful when practical constraints limit the amount of replication possible. For example, physical constraints may limit the block size or the row-column dimensions. Alternatively, other practical considerations, such as availability of experimental material (e.g., seeds in a plant-breeding trial) or cost may limit the number of times a treatment can feasibly be replicated.

In addition, p-rep designs offer an effective way of screening a large number of treatments (e.g., different genotypes in a plant-breeding programme).

Today, p-rep designs are commonly used in early generation plant-breeding trials. They offer a new and efficient alternative to the more traditional augmented design, with its replicated check entries (i.e., standards) and many unreplicated test entries (i.e., treatments).

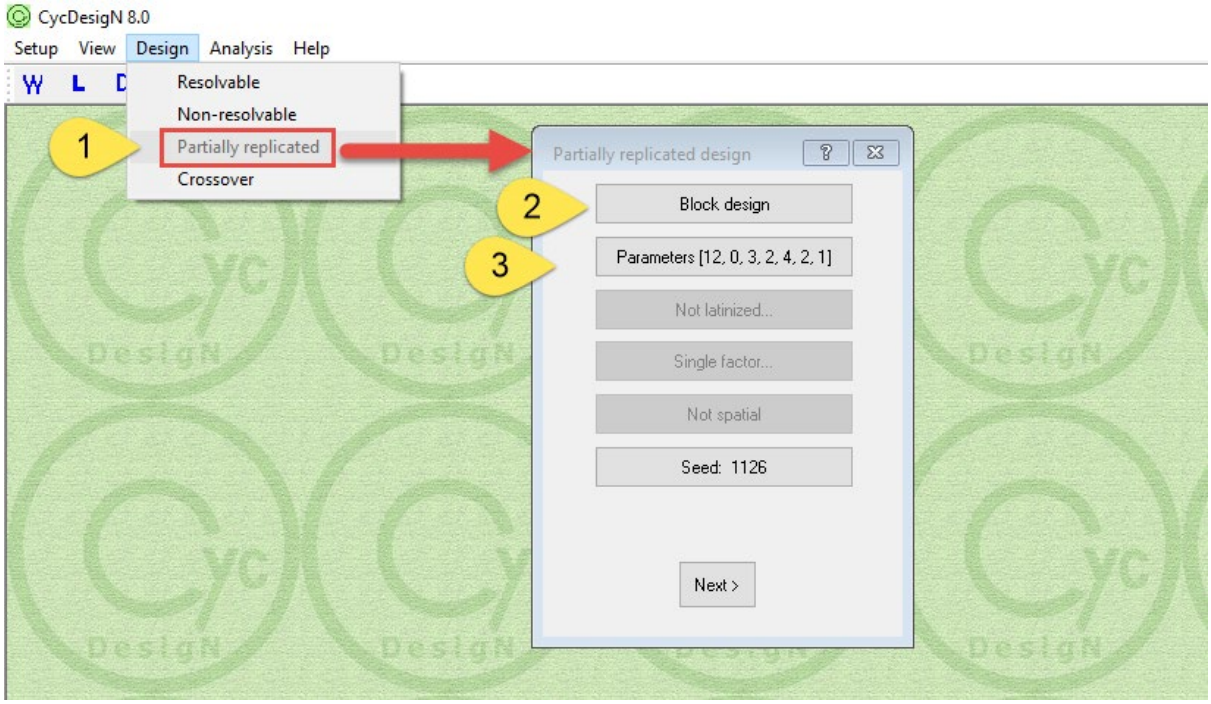
How do I generate a p-rep design in CycDesigN?

We'll begin by walking through a simple example, before looking at a couple of more complex examples, including p-rep designs with standards and spatial p-rep designs.

Example 1: A single-location p-rep block design

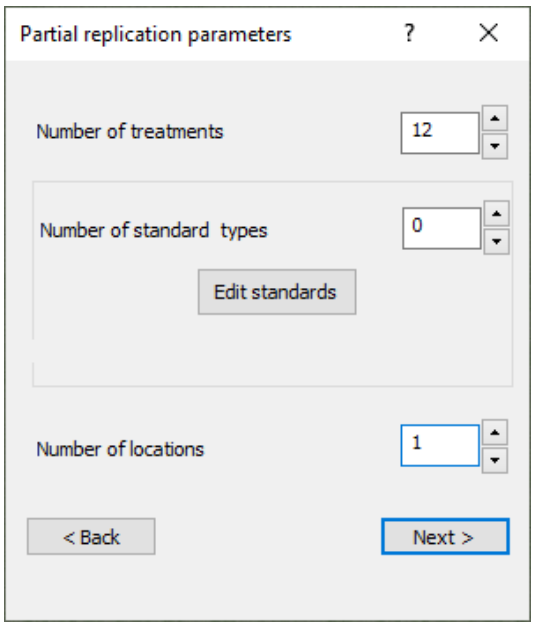
Let's generate a single-location p-rep block design, with 12 treatments and three blocks of size 6 (*cf* the single-location example above).

1. On the main menu of CycDesigN, under Design select Partially replicated to open the window dialog for generating p-rep designs.
2. The top button in this dialog toggles between a Block design and a Row-column design. We're going to generate a p-rep block design.
3. Clicking on the next button, Parameters [ ], opens another window dialog where we set the parameters of our design.



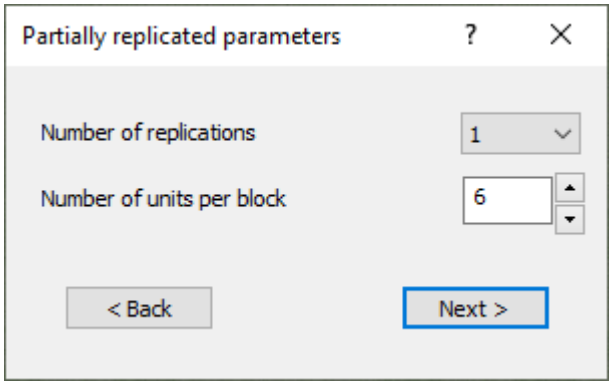
For our example, we set the Number of treatments to 12, the Number of standard types to 0, and the Number of locations to 1, then click Next to set further parameters.

Note: Standards (e.g., check entries) can be incorporated into p-rep design. These are replicated more frequently than the treatments. In an ***augmented p-rep design***, a design for a single-location, all treatments are unreplicated except for the standard types.



In the next window we set the minimum Number of replications of the treatments across the whole design. In p-rep designs, treatments are replicated either *r* or *r*+1 times across the whole design. Thus, this design parameter is set to the value given by *r*. For a single-location p-rep, this will be 1: recall that only a subset of treatments in a single-location p-rep are duplicated, with the remainder being unreplicated.

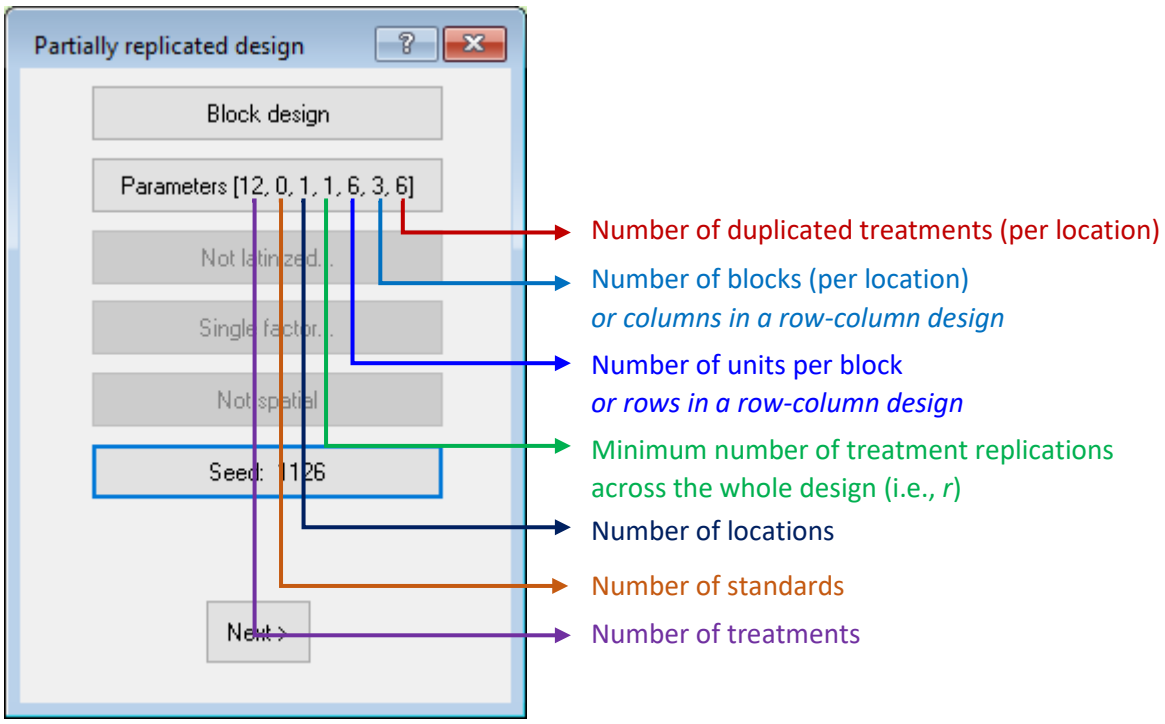
Next, we set the Number of units per block, which in our example is 6.



Click Next to open the final window for parameter setting. Set the Number of blocks per location to 3. Notice that the Number of duplicates has automatically changed to 6. This is because our design of three blocks of size 6 provides a total of 18 experimental units to which our 12 treatments are randomly assigned. This enables six of the 12 treatments to be duplicated (18-12=6).

Clicking Next closes the final parameter setting window.

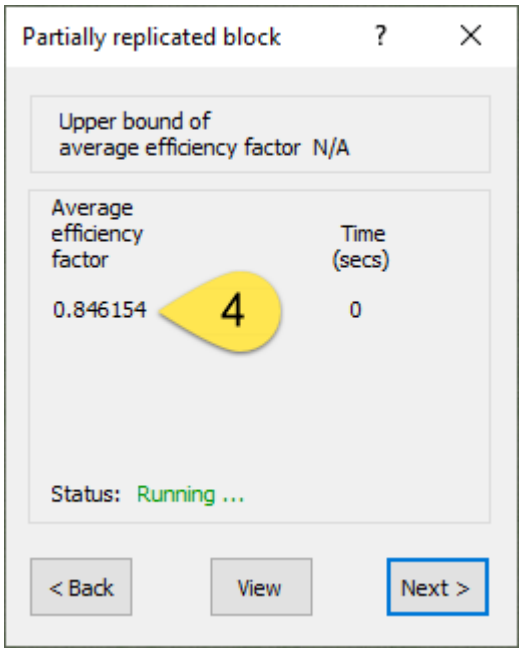
The numbers in the square brackets on Parameters [ ] button in the Partially replicated design window correspond to the values of the design parameters we just have set.



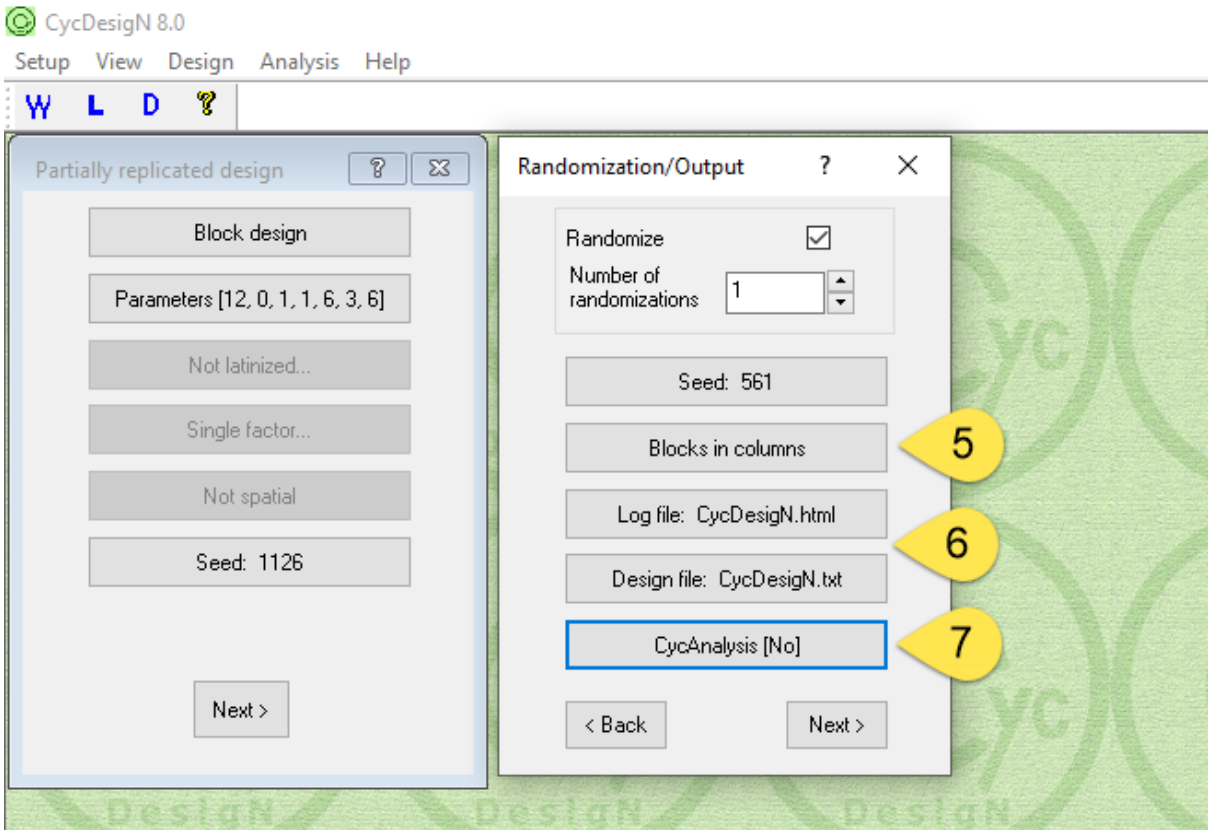
The Seed button allows us to set the seed used by CycDesigN's algorithm to generate the design. By default, the seed is taken from the computer clock.

Clicking Next on the Partially replicated design window starts the algorithm for generating the design.

4 When the updating process has stabilised, that is when there's no improvement to the average efficiency factor after a minute or two, we click Next to randomize the design and generate the output.



In the Randomization/Output window, we tick Randomize to randomize the design, and set the Number of randomizations to 1 to produce one randomization of our design. The Seed button allows us to set the seed of the random number generator used to randomize the design. If we use the same seed more than once, we will get the same random numbers, and hence the same randomization.



5 This button toggles between the blocks of the design being printed in columns or in rows in the .txt Design file.

6 For your design, CycDesigN generates a:

- ***Log file*** containing all the necessary information to reproduce the exact same design at a later date, and a,
- ***Design file*** giving details on the design type, parameters chosen and the final design layout.

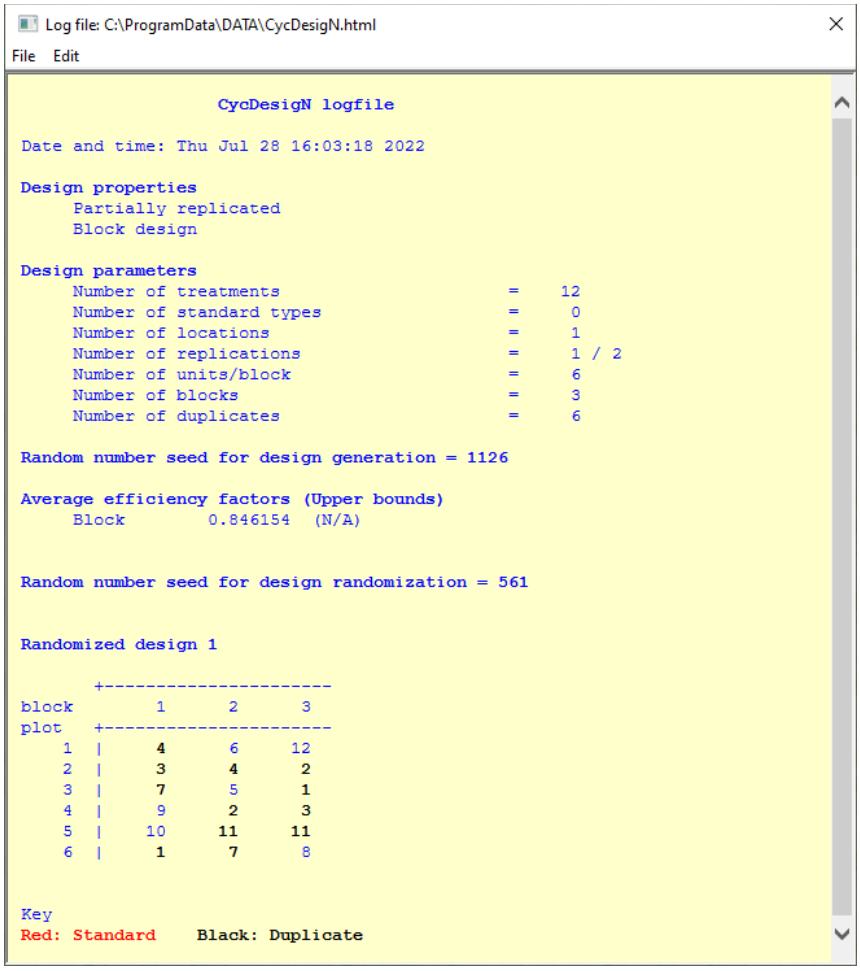
We can use the Log file and Design file buttons to specify the names of these files and where they are saved. Default names are automatically generated, and unless specified using these buttons, the files are saved in the working directory.
(Note: the working directory can changed be via Setup on the main menu bar.)

Two other files with the same name as the log file will also be generated:

- a .csv file containing the design in long format, and
- a .aux file for use by the CycAnalysis module (see below)

7 This button toggles between opening (yes) and not opening (no) the CycAnalysis window with the design just generated pre-loaded.
Note: In the CycAnalysis window we can generate template code for the analysis of our design using Genstat or SAS, save test data in a .xlsx file, and also make changes to the treatment and blocking factors, such as renaming them. The CycAnalysis window can also be accessed by clicking on Analysis in the main menu bar.

And finally, clicking Next on the Randomization/Output window will generate our design and the accompanying output.



Note: you can use View on the main menu bar to open your log and design file in CycDesigN.

Example 2: A multi-location p-rep block design

In this example, assume there are four locations, three blocks of size 5 per location and 16 treatments, each of which is to be replicated at least three times across the whole design. Therefore, in the design parameter windows, set:

- Number of treatments to 16
- Number of standard types to 0
- Number of locations to 4
- Number of replications to 3
- Number of units per block to 5
- Number of blocks per location to 3

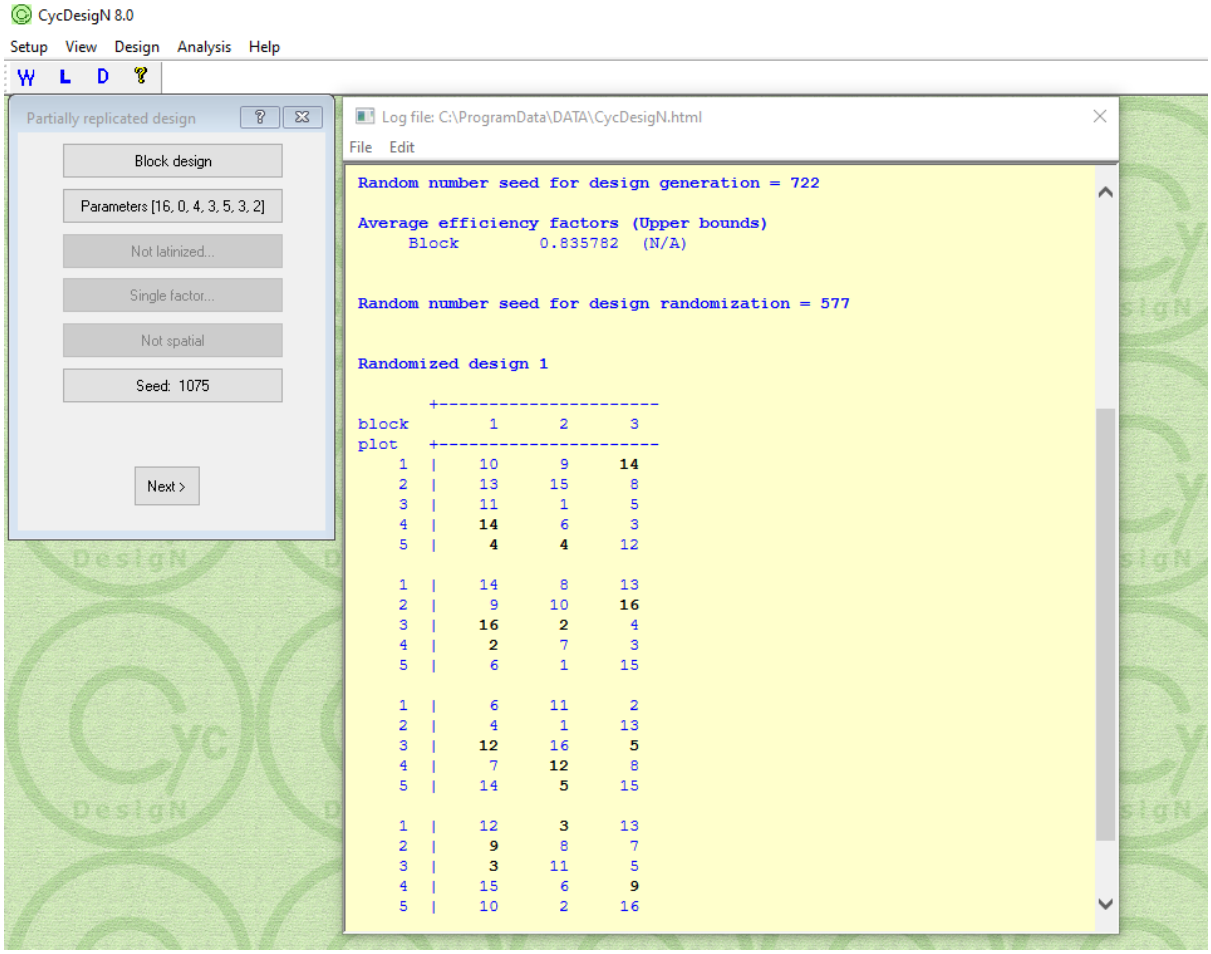
For the final parameter, Number of duplicates per location, we have a choice of settings: 0 to 4. Let's duplicate two treatments per location.

Note on the number treatment replications across the whole design, *r*

In a multi-location p-rep design, individual treatments are replicated either *r* or *r*+1 times. The range of possible *r* values depends on the number of locations (which we'll denote by *c*).

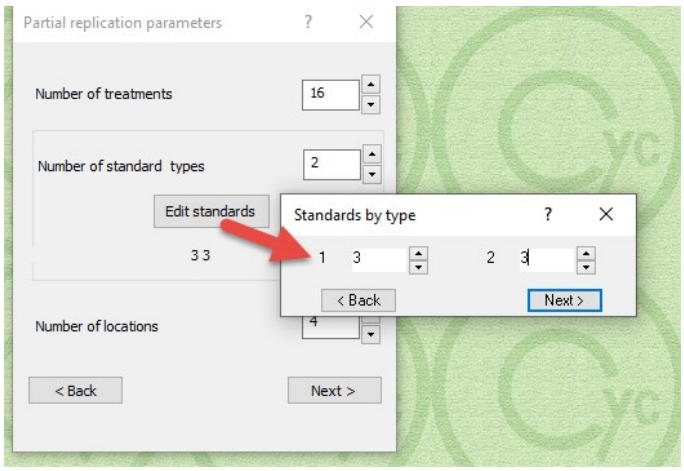
For *c* locations, *r* must be at least *c*-2, and no larger than 2*c*-1 (i.e., $c-2 \leq r \leq 2c-1$).

In our example, with *c* = 4 locations, *r* can thus range from 2 to 7.

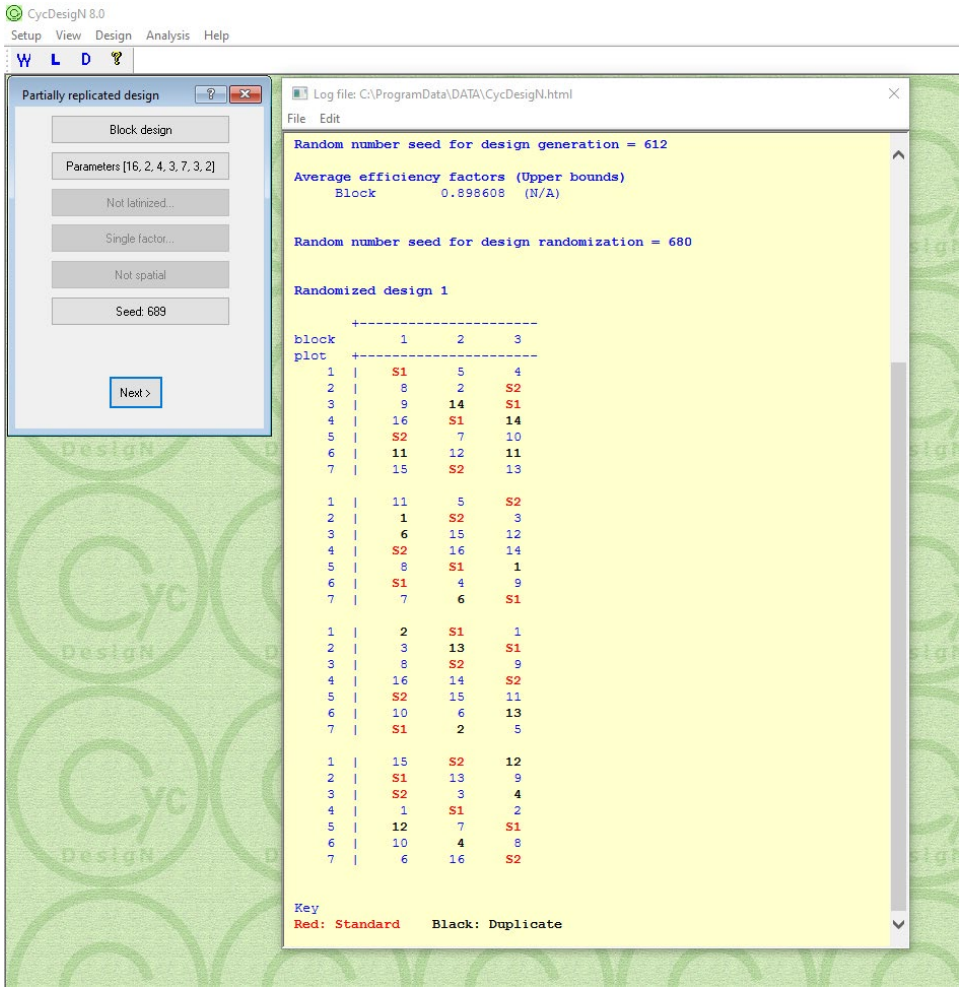


Example 3: A multi-location p-rep block design with standard types

Now let's modify the design above to include two standard types (denoted by S1 and S2), each of which is to occur three times per location. We do this by setting the Number of standard types to 2, then clicking in the Edit standards button and entering the required number of standards per location into the field for each standard type.



To accommodate these standards in our design, we either must change the dimensions of the block design (i.e., increase the number blocks and/or increase the number of units per block) OR decrease the number of times our treatments are replicated across the design. Reducing replication is never ideal, so, if possible, you should aim to increase the dimensions of the block design, preferably by increasing the units per block. Let's assume we can increase the units per block from 5 to 7 – giving 6 more plots per location. This will enable us to accommodate the 6 standards per location (2 standard types by 3 reps) without reducing the amount of replication possible for the treatments.



When generating a p-rep block design with standards, CycDesigN will set out the standards at each location such that they occur as evenly as possible within the blocks. In this particular example, each standard type thus appears exactly once in every block.

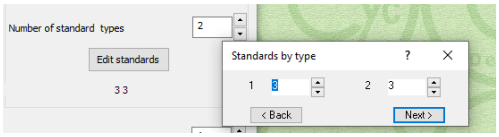
Example 4: A multi-location p-rep row-column design with standards

Until now we've only considered p-rep designs laid out in blocks. Let's look at one laid out in rows and columns at each location.

As above, let's assume there are four locations, 16 treatments (each of which is to be replicated at least three times across the whole design), and two standard types (each of which is to be replicated three times per location), but this time, within each location the design is to be arranged in a five row by four column grid.

Using the top button in the Partially replicated design menu dialog, toggle to Row-column design. In the design parameter windows, set:

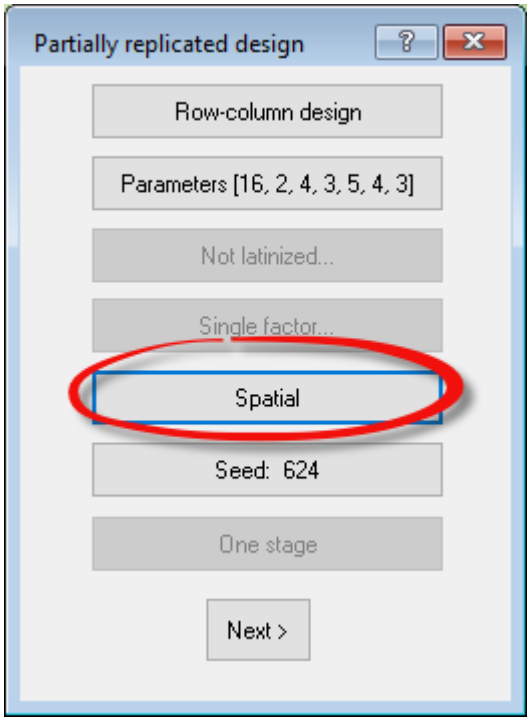
- Number of treatments to 16
- Number of standard types to 2
- Edit standards to 3 for each standard type



- Number of locations to 4
- Number of replications to 3
- Number of rows to 5
- Number of columns to 4

For the final parameter, Number of duplicates per location, we have a choice of settings: 0 to 3. Let's duplicate three treatments per location.

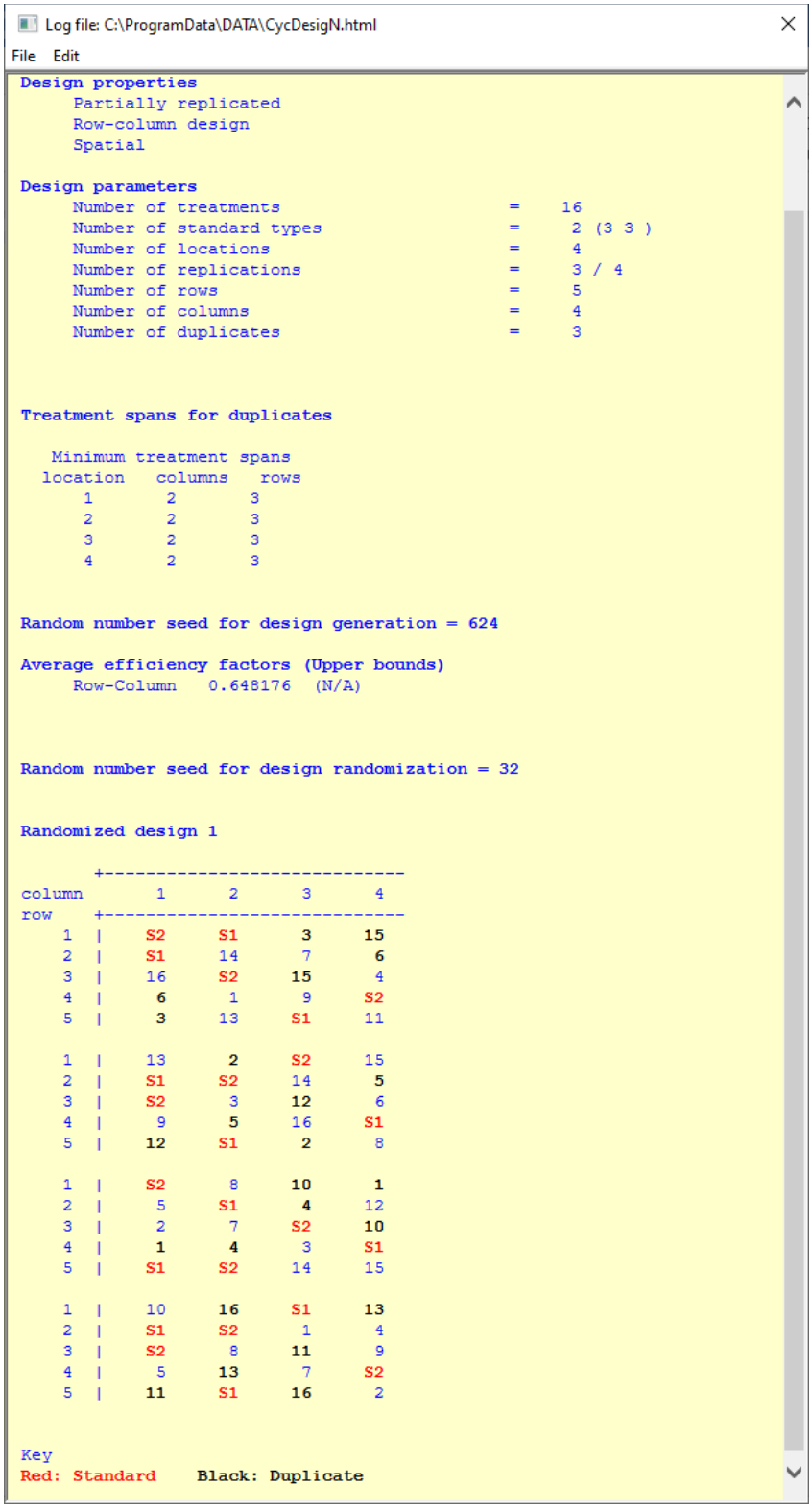
Notice that the 5th button in the Partially replicated design menu dialog is active. This button toggles between spatial and non-spatial designs. A common problem with classical row-column designs is that replications of the individual treatments may be clustered together. A ***spatial*** row-column p-rep design takes the distribution and separation of the treatments across the rows and columns into account, generating a design more robust to spatial trends and environmental heterogeneity across the row-column grid. Let's generate a spatial design.



To learn more about spatial design in CycDesigN, please check out the tutorial video: https://youtu.be/aZRzIBUNM14 and the reference paper: https://onlinelibrary.wiley.com/doi/epdf/10.1111/jac.12463

Click Next on the Partially replicated design menu dialog and CycDesigN will run through its iterative procedure to optimize the standards, ensuring they are evenly spread across the rows and columns. When this process has stabilized click Spatial. CycDesigN will now run through its algorithm to optimize the spatial criteria of the design.

When the updating process has stabilised, click Next, set your desired options in the Randomization/Output menu, and finally, click Next to generate your design and save the output.

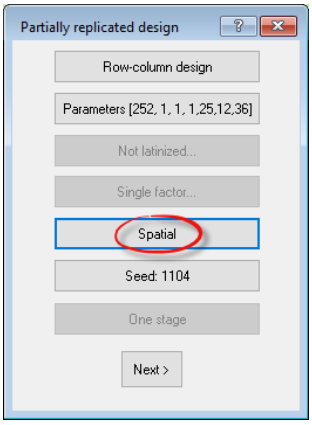


The "Treatment span for duplicates" table summaries how close the duplicated treatments within a location are. For example, a span of 3 for a treatment in the row direction means that its duplicates range over 3 rows. CycDesigN's spatial algorithm tries to maximize the overall span of the treatment replications. In our design, at each location, the duplicated treatments span at least 2 columns and at least 3 rows.

Real-world examples in plant-breeding

Today, p-rep designs are commonly used in early generation plant-breeding trials. Typically, such trials are laid out in a rectangular grid of experimental units (i.e., in rows and columns), rather than in blocks. A spatial design will be more robust to spatial trends and environmental heterogeneity across the row-column grid than the non-spatial variant, enabling more reliable estimates of the treatment effects. Piepho et al.'s 2020 paper in the Journal of Agronomy and Crop Science, entitled "*Generating row–column field experimental designs with good neighbour balance and even distribution of treatment replications*", provides three real-world examples of p-rep row-column designs, which they generate to be spatially robust. Let's also generate these three spatial p-rep row-column designs in CycDesigN.

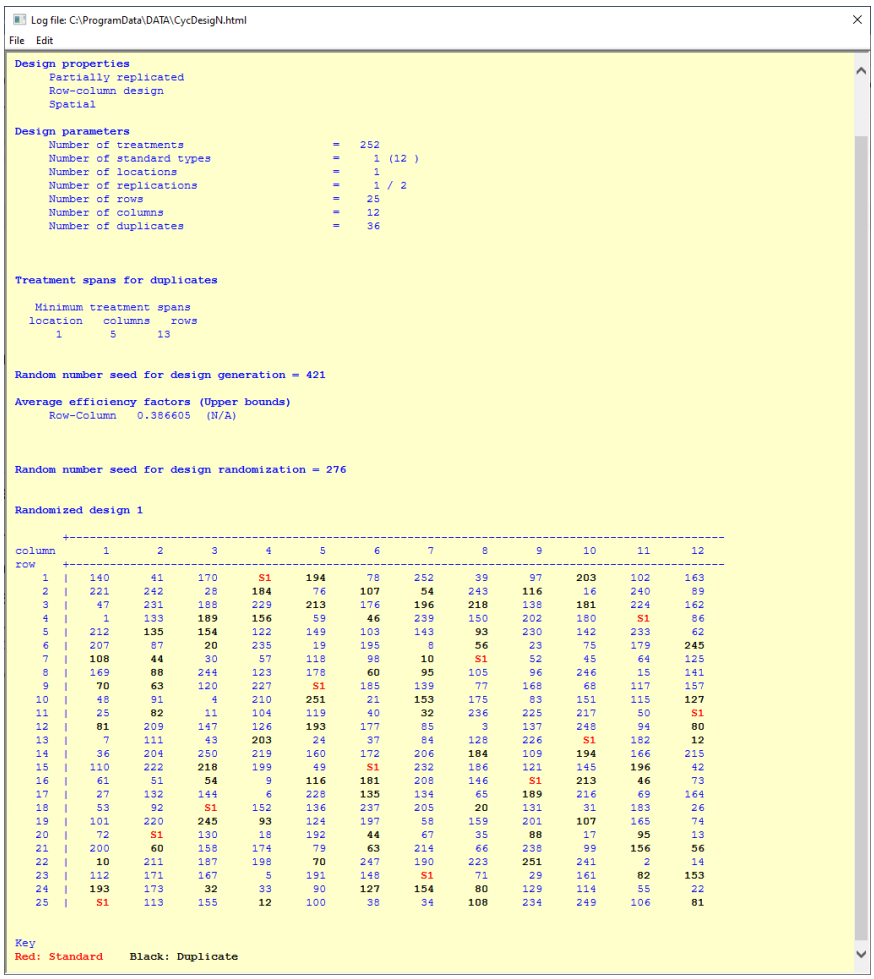
Recall that in CycDesigN, ***spatial*** row-column p-rep designs are generated by toggling the 5th button in the Partially replicated design menu to Spatial. A spatial design will be more robust to spatial trends and environmental heterogeneity across the row-column grid than the non-spatial variant, enabling more precise estimates of the treatment comparisons.



Example 3 in Piepho et al. (2020) : Single-location spatial p-rep row-column design augmented with a standard type

In this single-location trial, there are 252 treatment lines, of which 36 are duplicated. In addition, there is a single standard variety, replicated 12 times. The trial is laid out over 25 rows and 12 columns.

The Log file from generating a design for this trial in CycDesigN is below. Notice the settings of the design properties and parameters at the top of the Log file.



Log file: C:\ProgramData\DATA\CycDesigN.html

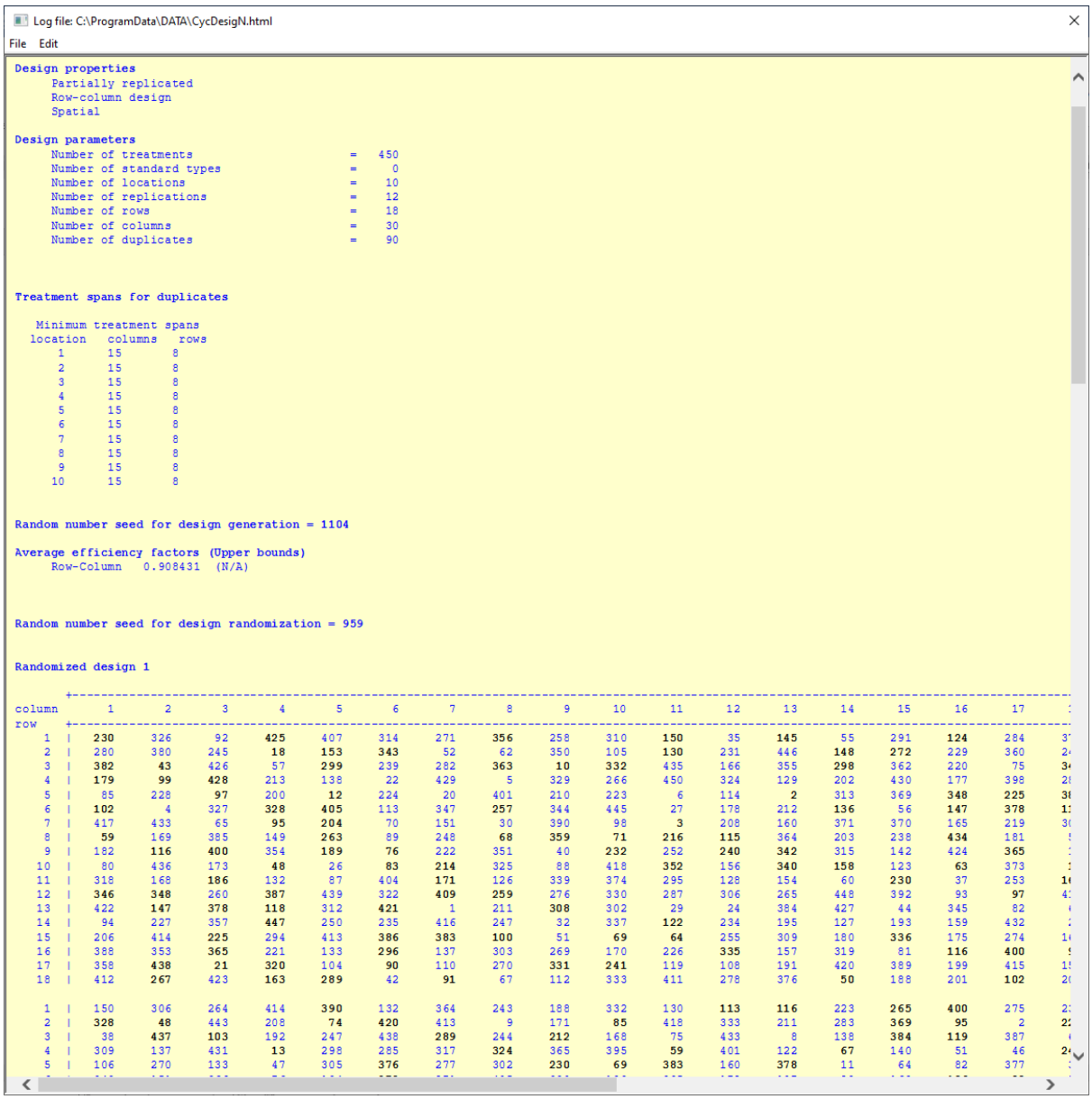
File Edit

```
Design properties
     Partially replicated
     Row-column design
     Spatial

Design parameters
     Number of treatments                =   252
     Number of standard types            =     1 (12 )
     Number of locations                 =     1
     Number of replications              =     1 / 2
     Number of rows                      =    25
     Number of columns                   =    12
     Number of duplicates                =    36


Treatment spans for duplicates

   Minimum treatment spans
  location   columns   rows
     1       5       13

Random number seed for design generation = 421

Average efficiency factors (Upper bounds)
     Row-Column   0.386605  (N/A)


Random number seed for design randomization = 276


Randomized design 1

       +-----------------------------------------------------------------------------------------------
column      1       2       3       4       5       6       7       8       9      10      11      12
row    +-----------------------------------------------------------------------------------------------
   1   |   140      41     170      S1     194      78     252      39      97     203     102     163
   2   |   221     242      28     184      76     107      54     243     116      16     240      89
   3   |    47     231     188     229     213     176     196     218     138     181     224     162
   4   |     1     133     189     156      59      46     239     150     202     180      S1      86
   5   |   212     135     154     122     149     103     143      93     230     142     233      62
   6   |   207      87      20     235      19     195       8      56      23      75     179     245
   7   |   108      44      30      57     118      98      10      S1      52      45      64     125
   8   |   169      88     244     123     178      60      95     105      96     246      15     141
   9   |    70      63     120     227      S1     185     139      77     168      68     117     157
  10   |    48      91       4     210     251      21     153     175      83     151     115     127
  11   |    25      82      11     104     119      40      32     236     225     217      50      S1
  12   |    81     209     147     126     193     177      85       3     137     248      94      80
  13   |     7     111      43     203      24      37      84     128     226      S1     182      12
  14   |    36     204     250     219     160     172     206     184     109     194     166     215
  15   |   110     222     218     199      49      S1     232     186     121     145     196      42
  16   |    61      51      54       9     116     181     208     146      S1     213      46      73
  17   |    27     132     144       6     228     135     134      65     189     216      69     164
  18   |    53      92      S1     152     136     237     205      20     131      31     183      26
  19   |   101     220     245      93     124     197      58     159     201     107     165      74
  20   |    72      S1     130      18     192      44      67      35      88      17      95      13
  21   |   200      60     158     174      79      63     214      66     238      99     156      56
  22   |    10     211     187     198      70     247     190     223     251     241       2      14
  23   |   112     171     167       5     191     148      S1      71      29     161      82     153
  24   |   193     173      32      33      90     127     154      80     129     114      55      22
  25   |    S1     113     155      12     100      38      34     108     234     249     106      81


Key
Red: Standard    Black: Duplicate
```

Example 4 in Piepho et al. (2020) : Multi-location spatial p-rep row-column design

In this multi-location p-rep trial, 450 breeding lines of triticale (a hybrid of wheat and rye) are tested at 10 locations. Each line is replicated 12 times across the entire trial. At each location, the layout has 18 rows and 30 columns, and 90 duplicated lines.

Below is the top section of the Log file (which includes the design properties and parameters) from generating a multi-location spatial p-rep row-column design for such a trial.



Log file: C:\ProgramData\DATA\CycDesigN.html

File Edit

Design properties

Partially replicated
Row-column design
Spatial

Design parameters

Number of treatments	450
Number of standard types	0
Number of locations	10
Number of replications	12
Number of rows	18
Number of columns	30
Number of duplicates	90

Treatment spans for duplicates

Minimum treatment spans

location	columns	rows
1	15	8
2	15	8
3	15	8
4	15	8
5	15	8
6	15	8
7	15	8
8	15	8
9	15	8
10	15	8

Random number seed for design generation = 1104

Average efficiency factors (Upper bounds)

Row-Column 0.908431 (N/A)

Random number seed for design randomization = 959

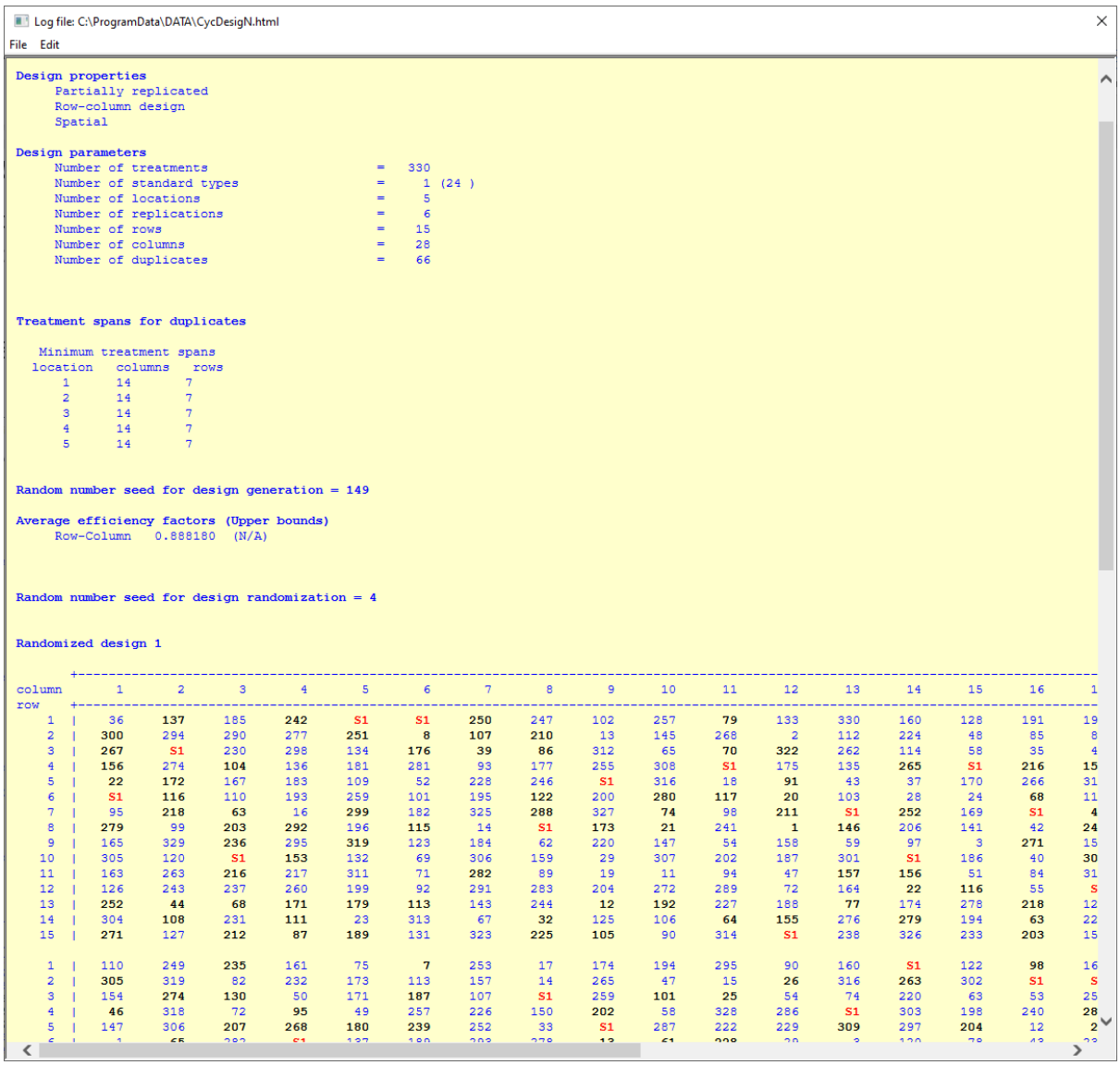
Randomized design 1

column	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
row																	
1	230	326	92	425	407	314	271	356	258	310	150	35	145	55	291	124	284
2	280	380	245	18	153	343	52	62	350	105	130	231	446	148	272	229	360
3	382	43	426	57	299	239	282	363	10	332	435	166	355	298	362	220	75
4	179	99	428	213	138	22	429	5	329	266	450	324	129	202	430	177	398
5	85	228	97	200	12	224	20	401	210	223	6	114	2	313	369	348	225
6	102	4	327	328	405	113	347	257	344	445	27	178	212	136	56	147	378
7	417	433	65	95	204	70	151	30	390	98	3	208	160	371	370	165	219
8	59	169	385	149	263	89	248	68	359	71	216	115	364	203	238	434	181
9	182	116	400	354	189	76	222	351	40	232	252	240	342	315	142	424	365
10	80	436	173	48	26	83	214	325	88	418	352	156	340	158	123	63	373
11	318	168	186	132	87	404	171	126	339	374	295	128	154	60	230	37	253
12	346	348	260	387	439	322	409	259	276	330	287	306	265	448	392	93	97
13	422	147	378	118	312	421	1	211	308	302	29	24	384	427	44	345	82
14	94	227	357	447	250	235	416	247	32	337	122	234	195	127	193	159	432
15	206	414	225	294	413	386	383	100	51	69	64	255	309	180	336	175	274
16	388	353	365	221	133	296	137	303	269	170	226	335	157	319	81	116	400
17	358	438	21	320	104	90	110	270	331	241	119	108	191	420	389	199	415
18	412	267	423	163	289	42	91	67	112	333	411	278	376	50	188	201	102
1	150	306	264	414	390	132	364	243	188	332	130	113	116	223	265	400	275
2	328	48	443	208	74	420	413	9	171	85	418	333	211	283	369	95	2
3	38	437	103	192	247	438	289	244	212	168	75	433	8	138	384	119	387
4	309	137	431	13	298	285	317	324	365	395	59	401	122	67	140	51	46
5	106	270	133	47	305	376	277	302	230	69	383	160	378	11	64	82	377

Example 5 in Piepho et al. (2020) : Multi-location spatial p-rep row-column design augmented with a standard type

In this trial, 330 lines of winter barley are tested across 5 locations. Each line is replicated 6 times across the entire trial. At each location, the design has 15 rows and 28 columns, and 66 duplicated lines. In addition, at each location, a standard variety is replicated 24 times.

Below is the top section of the Log file (including the design properties and parameters) from generating a design for this spatial p-rep trial.



Log file: C:\ProgramData\DATA\CycDesigN.html

File Edit

Design properties

Partially replicated
Row-column design
Spatial

Design parameters

Number of treatments	=	330
Number of standard types	=	1 (24)
Number of locations	=	5
Number of replications	=	6
Number of rows	=	15
Number of columns	=	28
Number of duplicates	=	66

Treatment spans for duplicates

Minimum treatment spans

location	columns	rows
1	14	7
2	14	7
3	14	7
4	14	7
5	14	7

Random number seed for design generation = 149

Average efficiency factors (Upper bounds)

Row-Column 0.888180 (N/A)

Random number seed for design randomization = 4

Randomized design 1

column / row	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	1
1	36	137	185	242	S1	S1	250	247	102	257	79	133	330	160	128	191	19
2	300	294	290	277	251	8	107	210	13	145	268	2	112	224	48	85	8
3	267	S1	230	298	134	176	39	86	312	65	70	322	262	114	58	35	4
4	156	274	104	136	181	281	93	177	255	308	S1	175	135	265	S1	216	15
5	22	172	167	183	109	52	228	246	S1	316	18	91	43	37	170	266	31
6	S1	116	110	193	259	101	195	122	200	280	117	20	103	28	24	68	11
7	95	218	63	16	299	182	325	288	327	74	98	211	S1	252	169	S1	4
8	279	99	203	292	196	115	14	S1	173	21	241	1	146	206	141	42	24
9	165	329	236	295	319	123	184	62	220	147	54	158	59	97	3	271	15
10	305	120	S1	153	132	69	306	159	29	307	202	187	301	S1	186	40	30
11	163	263	216	217	311	71	282	89	19	11	94	47	157	156	51	84	31
12	126	243	237	260	199	92	291	283	204	272	289	72	164	22	116	55	S
13	252	44	68	171	179	113	143	244	12	192	227	188	77	174	278	218	12
14	304	108	231	111	23	313	67	32	125	106	64	155	276	279	194	63	22
15	271	127	212	87	189	131	323	225	105	90	314	S1	238	326	233	203	15
1	110	249	235	161	75	7	253	17	174	194	295	90	160	S1	122	98	16
2	305	319	82	232	173	113	157	14	265	47	15	26	316	263	302	S1	S
3	154	274	130	50	171	187	107	S1	259	101	25	54	74	220	63	53	25
4	46	318	72	95	49	257	226	150	202	58	328	286	S1	303	198	240	28
5	147	306	207	268	180	239	252	33	S1	287	222	229	309	297	204	12	2

This brings us to the end of our quick-start guide to p-rep designs in CycDesigN. To learn more about using CycDesigN, including how to generate other types of designs, please check out the tutorial videos: https://cycdesign.kb.vsni.co.uk/videos/