

# AssimInfo.txt

## General

The *AssimInfo.txt* file contains additional model settings to [info.txt](#) and is therefore located in the same folder as *info.txt*. If assimilation is switched on (`assimilation Y`) in *info.txt*, the *AssimInfo.txt* file is used to define what kind of data assimilation is to be performed. The file information is divided into four groups:

1. general settings,
2. setting of control variables,
3. observation settings, and
4. meteorological forcing data settings.

The settings are usually kept in their four groups, but that is not necessary. The general settings all start with the letters `G_`, while the control variable settings start with `A_`, observations settings with `O_` and forcing settings with `F_`. The control variable settings determine the group of HYPE states to which the filtering is applied.

## File format

The basic format in the *AssimInfo.txt* file is simply a row-wise combination of codes and argument(s):

```
!! <comment>
<code 1.1> [<code 1.2>] <argument 1> [<argument 2>] ... [<argument n>]
<code 2.1> [<code 2.2>] <argument 1> [<argument 2>] ... [<argument m>]
...
```

Comment rows can be added anywhere and are marked with double exclamation marks, i.e. `!!`, or `!!!` followed by a space. For other rows, the first code string decides what information is to be read. The code can be written within or without apostrophes (`'...'`). Codes are not case sensitive. Maximum 18000 characters can be read on a single line.

## General settings

Code	Argument	Description
G_NE	<i>Integer</i>	Ensemble size, number of ensemble members (default=100)
G_MV	<i>Real</i>	Missing value for the assimilation routine (default=-9999) (not useful for HYPE, must be -9999?)
G_MEANOUT	<i>0/1</i>	mean(1) or median(0) value printed in ordinary output files (default=mean)

Code	Argument	Description
G_STATOUT	0/1/2/3/4/5/..	Extra output files for statistics (0-5) and ensemble members (6-). 1 give minimum as _002, 2 give min and max (as _003), 6 and up give ensemble members 1 and up to maximum 5+NE (as _007 and up). Note: 3-5 is not implemented. They should in the future give: 3 give previous and 0.025-perc (as _004), 4 give previous and median (as _005), 5 give previous and 0.975-perc (as _006) (default=0)
G_XYLOC	Real	Horizontal length scale [m] for covariance localization (distance with ~90% covariance reduction) (default=1000000)
G_ZLOC	Real	Vertical length scale [m] for covariance localization (default=100000)
G_USEBINX	0/1/2	Use bin-files to hold state ensembles (0=no, 1=one bin-file, 2=several bin-files) (default is no)
G_USEBINFA	0/1/2	Use bin-files to hold forcing and auxiliary ensembles (0=no, 1=one bin-file, 2=several bin-files) (default is no)
G_STOP	0/1	Stop simulation when cholesky factorisation fails (0/1) (default=0=not stop)
G_CNC	0/1	Collapse non-controlled states to ensemble mean(or median) (0=no, 1=yes) (default=0)
G_TRANSTAT	0/1	Transform state variables (and some outvar) before the EnKF analysis (0=no, 1=yes) (default=0). If yes, then statevariables with physical range [0,+inf] will be log-transformed, and variables with range [0,1] will be logit transformed. Once implemented, Yeo-Johnson transform will be used on variables with unbounded physical limits (such as temperatres)
G_TRANEPS	Real	epsilon = minimum value used for log and logit transforms (used for state variables with physical range [0,+inf] and [0,1], respectively) (default=0.000001)

## Control variable settings

These are the variables controlled by assimilation. They are set as a group by category or separate by name. The categories and names are specific to the HYPE model. If they are turned off, the analysis is NOT applied to variables in this category. Instead, they are re-initialized to the ensemble mean (or median depending on G\_MEANOUT) after each time step. If they are turned on, the analysis IS APPLIED whenever there are observations available. No re-initialization.

Format of control variable lines: They start with A\_ followed by include\_ or exclude\_ followed by bycategory or byname followed by [category] or [name]. To identify the variable by name, the category of the variable need to be set on a line directly before the variable. Example:

```
A_INCLUDE_BYCATEGORY SNOW
A_INCLUDE_BYCATEGORY SOIL
A_EXCLUDE_BYCATEGORY GLACIER
```

Code	Argument	Description
A_INCLUDE_BYCATEGORY	Category	Category is defined in the HYPE code (see table below)
A_EXCLUDE_BYCATEGORY	Category	Category as above
A_INCLUDE_BYNAME	Name	Name is state variable name in HYPE code (see table below).
A_EXCLUDE_BYNAME	Name	Name as above
Category	Names	
SNOW	snow csnow snowage snowdepth snowcov snowmax snowheat snowliq	

Category	Names
GLACIER	glacvol
LAKEICE	lakesnow lakesnowage lakesnowdepth lakeice lakebice lakeicecov lakeicepor
RIVERICE	riversnow riversnowage riversnowdepth riverice riverbice rivericecov rivericepor
SOIL	water temp deeptemp conc humusN fastN partP fastP humusP fastC humusC PPrelpool Srelpool oldgrw partT1 surface icelens
AQUIFER	water conc lastrecharge clastrecharge nextoutflow cnextoutflow
RIVERWT	water temp conc TPmean temp10 temp20 Psed qqueue cqueue cwetland Qdayacc Q365 Qmean T1sed Ssed
LAKEWT	water temp conc TPmean temp10 temp20 uppertemp lowertemp volfrac
MISC	temp5 temp30 temp10 temp20 gdd gsbegin nextirrigation cnextirrigation updatestationsarcorr floodwater cfloodwater partT1sf nexttransfer cnexttransfer

## Observation settings

The observation settings determine which observations should be assimilated. The observations settings are given as a table with one observation variable per line. The settings include which HYPE outvar variables to compare, the ensemble generation model, minimum and maximum values allowed, standard deviation parameters, and parameters for generation of spatially correlated perturbations. The columns are in the order given in the table below.

Column number	Column name	Type	Value range	Description
1	Observation	Character	0_nnn	Beginning with the code for observation setting ("O_"), the following characters (nnn) are a description for the user
2	IDobs	4 characters	HYPE variable ID	The 4 letter code for the observation as used by HYPE
3	IDmod	4 characters	HYPE variable ID	The 4 letter code for the corresponding simulated variable as used by HYPE for output
4	EnsType	Integer	0-4	Ensemble generation model (following Turner et al). EnsType definition: 0=not used, 1=unrestricted, 2=semi-restricted(minimum), 3=semi-restricted(maximum), 4=constrained (max and min)
5	Min	Real	-	Minimum value allowed (EnsType 2,3,4). Perturbations outside this range will be truncated to the min value.
6	Max	Real	-	Maximum value allowed (EnsType 2,3,4). Perturbations outside this range will be truncated to the max value.
7	Minsigma	Real	-	Standard deviation parameter. Minsigma is minimum allowed standard deviation

Column number	Column name	Type	Value range	Description
8	Sigma	Real	-	Standard deviation parameter. Sigma is constant standard deviation used for EnsType = 1, also used as minimum allowed standard deviation for EnsType = 2-4
9	SemiMeta	Real	-	Standard deviation parameter. SemiMeta is relative standard deviation used for EnsType = 2 & 3
10	RestMeta	Real	-	Standard deviation parameter. RestMeta is relative standard deviation for EnsType = 4
11	Lscale	Real	-	correlation length (horizontal)
12	GridSize	Real	-	cellsize (x and y dir) in the 2D grid used for the 2D spatially correlated random fields (interpolated to the model coordinates)
13	CorrType	Integer	1-3	correlation function: 0 none 1 Gaussian, 2 Compact 5th degree polynomial, 3 Power law
14	Coordid	Integer	1-4	spatial domain of observation (1=subbasin, 2=upstream area (ie. COUT), 3=aquifer, 4=outregions)
15	Transform	Integer	0-3	kind of transformation to be applied to the variable before filtering (0=none, 1=log, 2=Yeo-Johnson (not implemented yet), 3=logit
16	epsilon	Real	-	minimum value used to avoid 0 in log or logit transform
17	ClassGroup	Character	-	Optional. If a class group variable is used, the class group name (as defined in info.txt) is given.

## Forcing data settings

The meteorological forcing data settings determine which forcing data should be perturbed and included in assimilation. The settings are given as a table with one forcing variable per line. The settings include the ensemble generation model, minimum and maximum values allowed, standard deviation parameters, and parameters for generation of spatially correlated perturbations. The columns are in the order given in the table below.

Column number	Column name	Type	Value range	Description
1	Forcing	Character	F_nnn	Beginning with a code for forcing data setting (F_), the following characters (nnn) are a description for the user
2	IDobs	4 characters	-	A letter code for the forcing as used by HYPE. It is the filename without the file ending, e.g. Pobs
3	EnsType	Integer	0-4	Ensemble generation model (following Turner et al). EnsType definition: 0=not used, 1=unrestricted, 2=semi-restricted(minimum), 3=semi-restricted(maximum), 4=constrained (max and min)
4	Min	Real	-	Minimum value allowed (EnsType 2,3,4). Perturbations outside this range will be truncated to the min value. Note: TMIN and TMAX is handled as deviations from Tobs in the DA, thus their range is the range of the deviation (negative for TMIN, positive for TMAX).

Column number	Column name	Type	Value range	Description
5	Max	Real	-	Maximum value allowed (EnsType 2,3,4). Perturbations outside this range will be truncated to the max value. Note: TMIN and TMAX is handled as deviations from Tobs in the DA, thus their range is the range of the deviation (negative for TMIN, positive for TMAX).
6	Minsigma	Real	-	Standard deviation parameter. Minsigma is minimum allowed standard deviation
7	Sigma	Real	-	Standard deviation parameter. Sigma is constant standard deviation used for EnsType = 1, also used as minimum allowed standard deviation for EnsType = 2-4
8	SemiMeta	Real	-	Standard deviation parameter. SemiMeta is relative standard deviation used for EnsType = 2 & 3
9	RestMeta	Real	-	Standard deviation parameter. RestMeta is relative standard deviation for EnsType = 4
10	Lscale	Real	-	correlation length (horizontal)
11	GridSize	Real	-	cellsize (x and y dir) in the 2D grid used for the 2D spatially correlated random fields (interpolated to the model coordinates)
12	CorrType	Integer	1-3	correlation function: 0 none 1 Gaussian, 2 Compact 5th degree polynomial, 3 Power law
13	Tau	Real	-	perturbation memory coefficient (fraction of perturbation propagated from previous timestep)