

Exceptional service in the national interest



```
import sys
from h5py import *
import numpy

filename = 'gridded_dataset.h5'
h5file = File(filename, mode='w')

h5grp = h5file.create_group('surface')
nx = 3; ny = 3; nz = 1; nt = 50
h5grp.attrs['Dimension'] = numpy.string_('XYZ')
h5grp.attrs['Discretization'] = [2, 5, 2, 5]
h5grp.attrs['Origin'] = [0, 0]
h5grp.attrs['Max Buffer Size'] = [4]
tarray = numpy.zeros(nt, dtype='f8')
for t in range(nt):
    tarray[t] = 10.*float(t)
h5dsset = h5grp.create_dataset('Times', data=tarray)
darray = numpy.zeros((nx, ny, nt), dtype='f8')
for t in range(nt):
    for j in range(ny):
        for i in range(nx):
            darray[i][j][t] = float(i) + 1.e-2*float(j) + 2. + \
                1.e2*float(t)
h5grp.create_dataset('Data', data=darray)
h5file.close()
```

surface Attributes

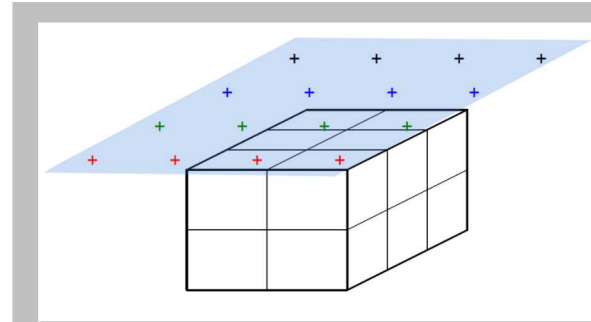
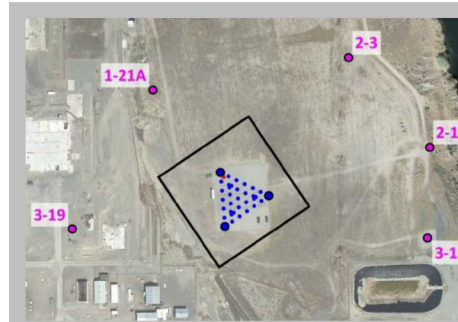
Name	Type	Array Size	Value50(...)
Dimension	String length=2: string	Scalar	XYZ
Discretization	64-bit floating-point	2	2.5, 2.5
Max Buffer Size	32-bit integer	1	4
Origin	64-bit floating-point	2	0.0, 0.0

Data (Slices)

	0	1	2
0	2.0	2.01	2.02
1	3.0	3.01	3.02
2	4.0	4.01	4.02

Times

	0	1	2	3	4
0	100.0	100.01	100.02	100.03	100.04
1	101.0	101.01	101.02	101.03	101.04
2	102.0	102.01	102.02	102.03	102.04



PFLOTRAN HDF5 Datasets

HDF5 Datasets in Input Decks

PFLOTRAN_DIR/shortcourse/exercises/regional_doublet/stochastic_regional_doublet.in

```
DATASET perm_field
  FILENAME ./perm_fields.h5
  HDF5_DATASET_NAME Permeability
  REALIZATION_DEPENDENT
/
...
MATERIAL_PROPERTY soil2
  ID 2
  POROSITY 0.35d0
  TORTUOSITY 0.5d0
  CHARACTERISTIC_CURVES cc1
  PERMEABILITY
    DATASET perm_field
    VERTICAL_ANISOTROPY_RATIO 0.1
  /
END
```

Type of HDF5 Datasets

- CELL_INDEXED – Data values are specified for each grid cell in the domain.
 - Typically use for initial conditions and material properties; i.e. not transient
- GRIDDED – Data values are defined on a regular 1D, 2D or 3D grid.

Not covered:

- GLOBAL – For a specific degree of freedom (dof), data values are specified for each grid cell in the domain.
- MAPPED – Heterogeneous approach where M data values are mapped to N cells.
 - $N \geq M$

CELL_INDEXED Datasets

Within the HDF5 File, create two top-level HDF5 Datasets (not within an HDF5 Group)

- **Cell Ids** – IDs of cells 1-N, where N is the maximum cell ID
- **<string>** – An integer or double precision dataset with a user-defined name (spaces are not allowed). This dataset holds parameters values, one for each grid cell (e.g. material ID, porosity, permeability, concentration, etc.).

Example CELL_INDEXED Datasets

```
import sys
from h5py import *
import numpy

filename = 'cell_indexed_dataset.h5'
h5file = File(filename, mode='w')
n = 10
# create integer array for cell ids
iarray = numpy.arange(n, dtype='i4')
# convert to 1-based
iarray[:] += 1
dataset_name = 'Cell Ids'
h5dset = h5file.create_dataset(dataset_name, data=iarray)
# create double array for porosities
rarray = numpy.zeros(n, dtype='f8')
rarray[:] = 0.25
rarray[4:7] = 0.3
dataset_name = 'poros'
h5dset = h5file.create_dataset(dataset_name, data=rarray)
h5file.close()
```

Cell Ids

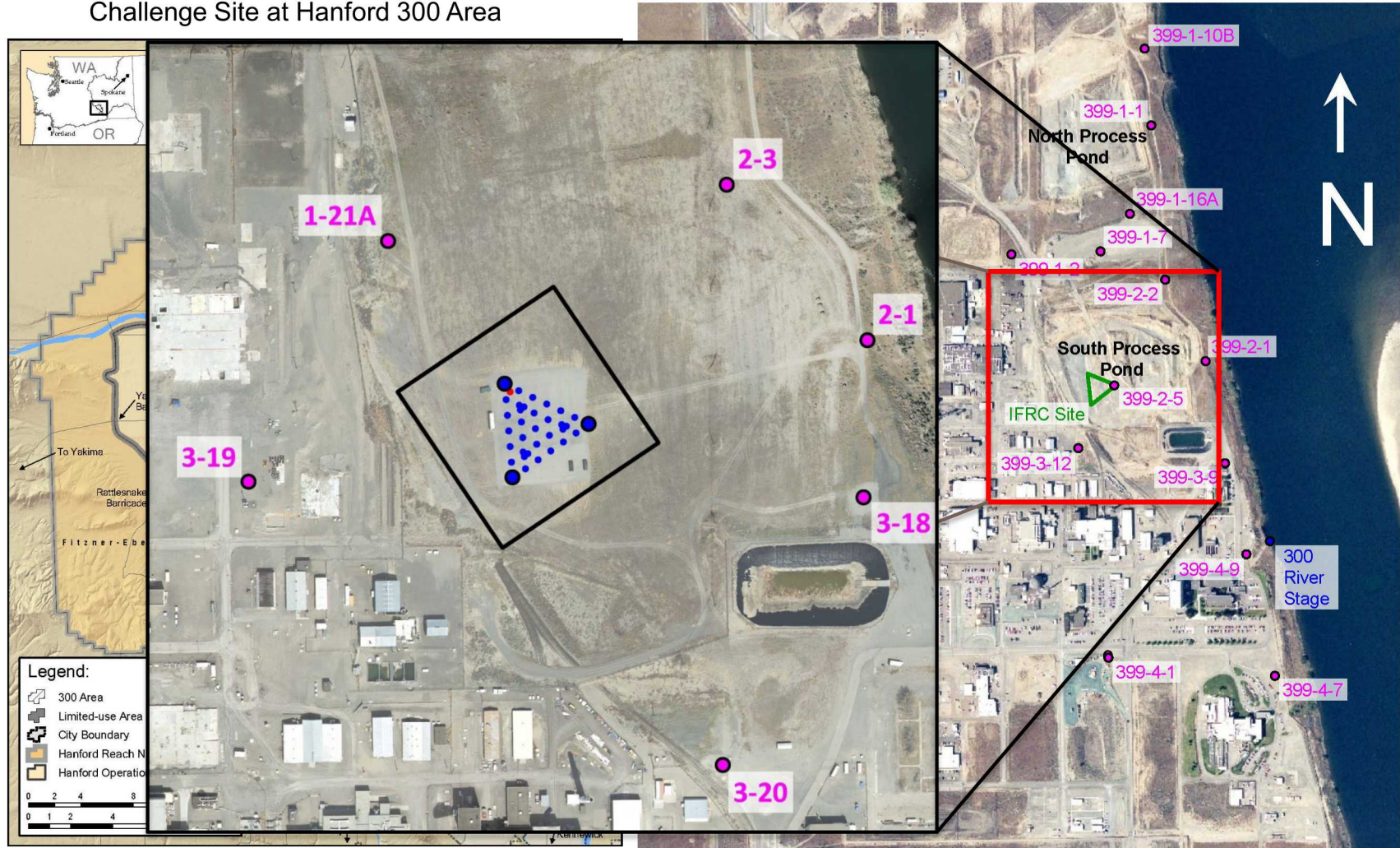
0	1
1	2
2	3
3	4
4	5
5	6
6	7
7	8
8	9
9	10

poros

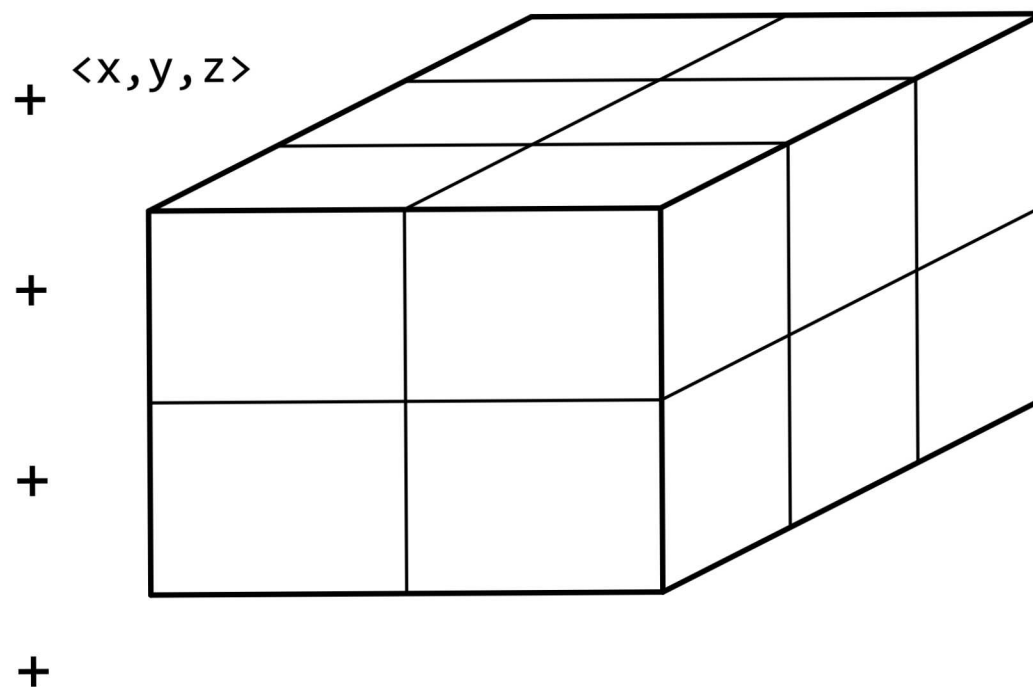
0	0.25
1	0.25
2	0.25
3	0.25
4	0.3
5	0.3
6	0.3
7	0.25
8	0.25
9	0.25

Motivation for GRIDDED Datasets

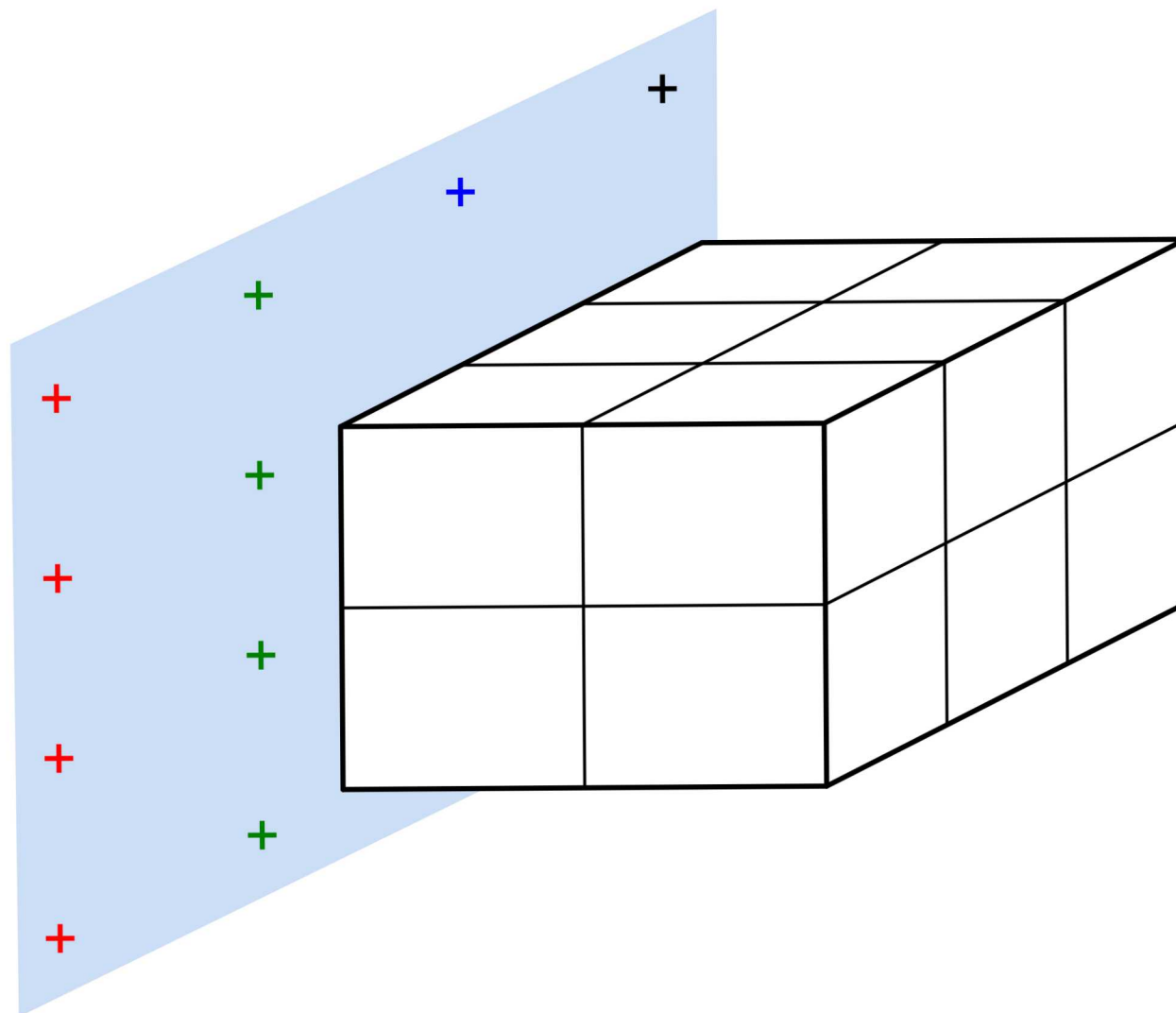
Integrated Field-Scale Subsurface Research
Challenge Site at Hanford 300 Area



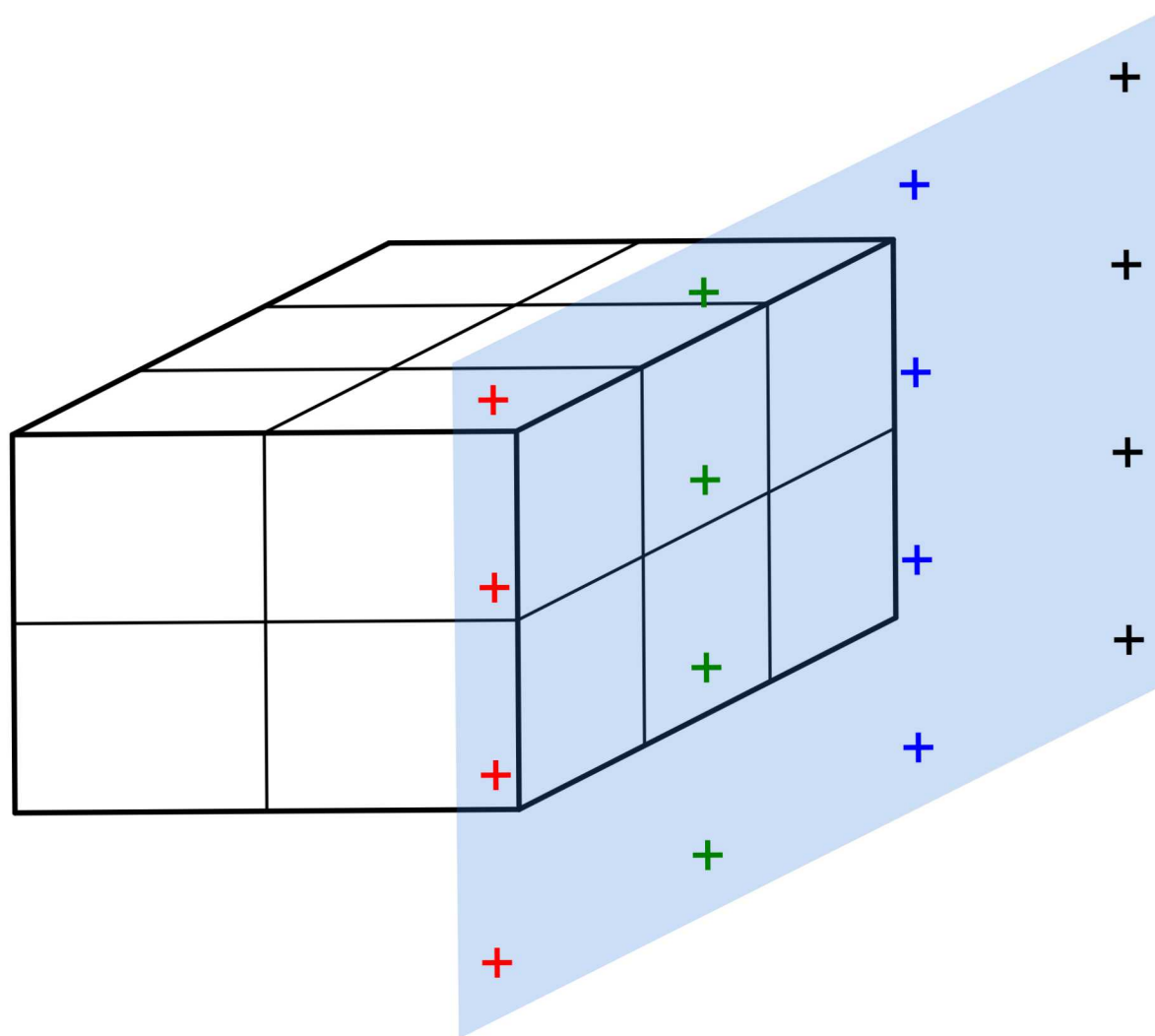
1D GRIDDED Datasets



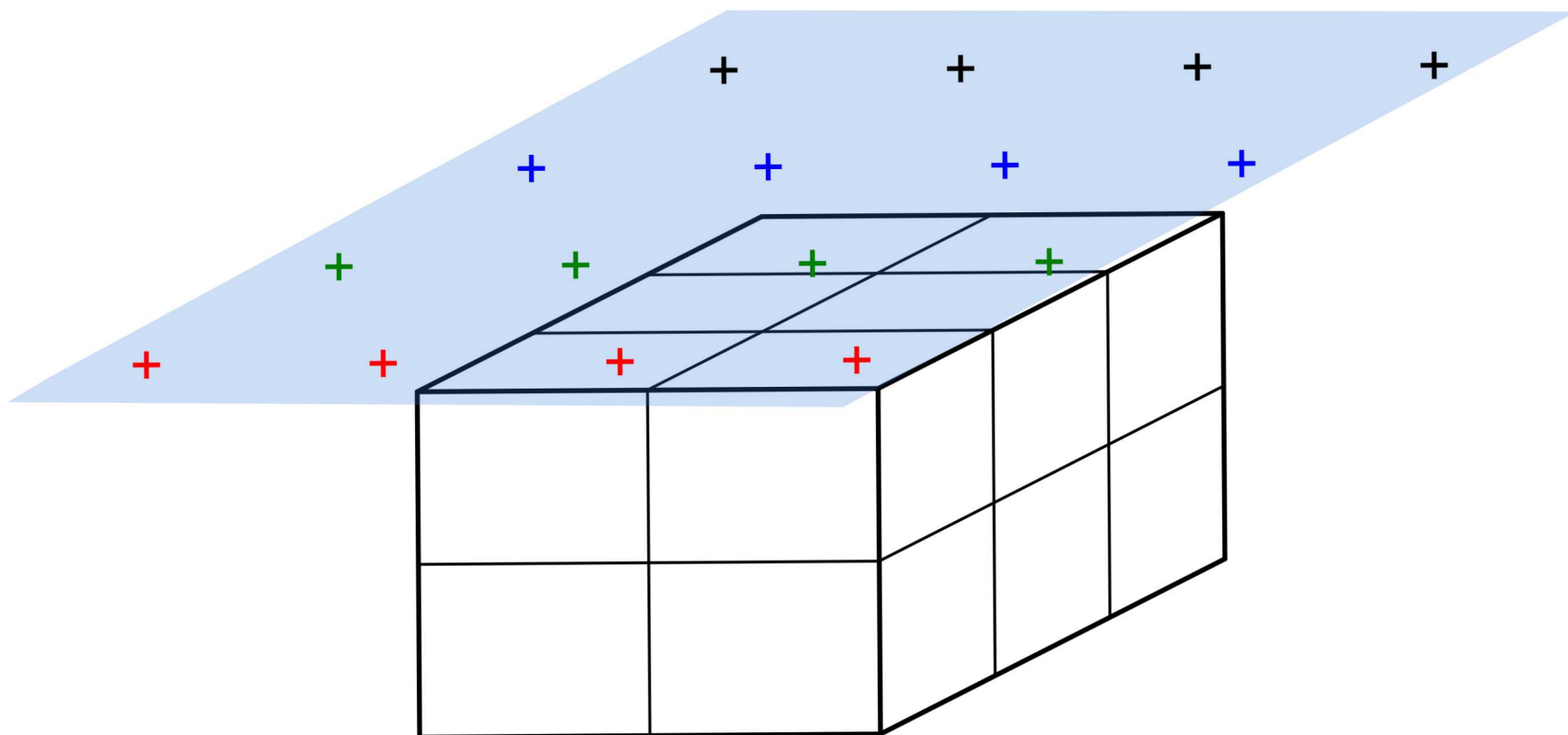
2D GRIDDED Datasets



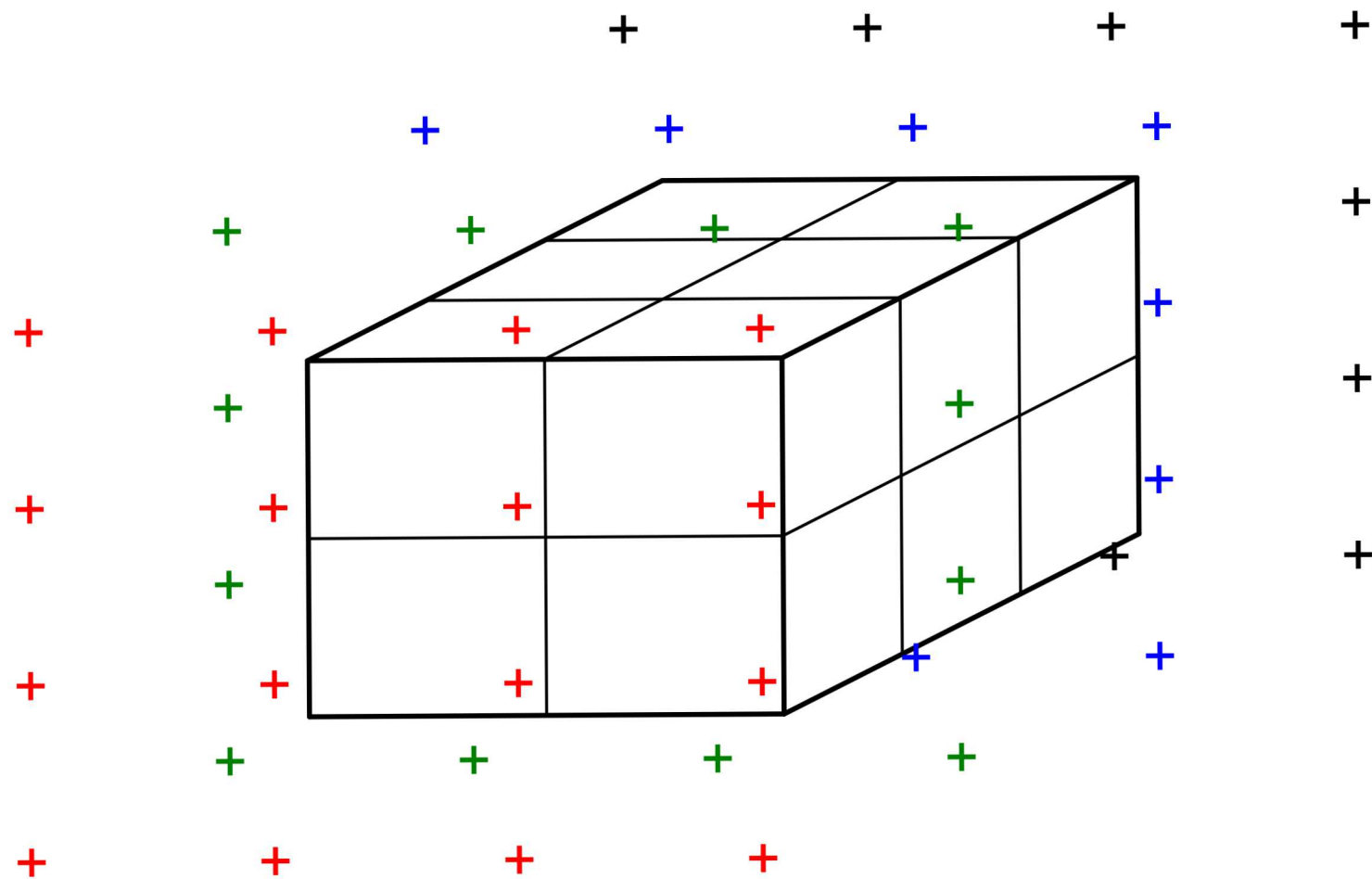
2D GRIDDED Datasets



2D GRIDDED Datasets



3D GRIDDED Datasets



GRIDDED Datasets

- Within the HDF5 File, create a top level HDF5 Group with a user-defined name (e.g. **surface**) and the following Attributes:
 - **Dimension** – Dimensions of dataset: X, Y, Z, XY, XZ, YZ, XYZ
 - **Discretization** – Resolution of regular grid associated with dataset. The number of values must be consistent and aligned with the dimension.
 - **Origin** – $\langle x, y, z \rangle$ coordinate of the grid origin
 - **Cell Centered** – Optional attribute specifying the cell-centered location of data values. Otherwise, data values are defined at cell corners.
 - **Interpolation Method** – Optional interpolation method: LINEAR or STEP (default = LINEAR)
 - **Max Buffer Size** – Optional attribute specifying the maximum number of time slices to load into the buffer during each file access. Buffering helps balance file access with memory use. Smaller buffers require more file accesses and less memory.
- Within the HDF5 Group, create up to two double precision HDF5 Datasets
 - **Data** – 1D, 2D, or 3D Dataset associated with dataset grid
 - **Times** – Optional 1D Dataset associated with dataset time slices. Required if transient.

Example GRIDDED Dataset

```
import sys
from h5py import *
import numpy

filename = 'gridded_dataset.h5'
h5file = File(filename,mode='w')
h5grp = h5file.create_group('surface')
nx = 3; ny = 3; nz = 1; nt = 50
h5grp.attrs['Dimension'] = numpy.string_('XY')
h5grp.attrs['Discretization'] = [2.5,2.5]
h5grp.attrs['Origin'] = [0.,0.]
h5grp.attrs['Max Buffer Size'] = [4]
tarray = numpy.zeros((nt),'=f8')
for t in range(nt):
    tarray[t] = 10.*float(t)
h5dset = h5grp.create_dataset('Times', data=tarray)
darray = numpy.zeros((nx,ny,nt),'=f8')
for t in range(nt):
    for j in range(ny):
        for i in range(nx):
            darray[i][j][t] = float(i) + 1.e-2*float(j) + 2. + \
                               1.e2*float(t)
h5grp.create_dataset('Data', data=darray)
h5file.close()
```

surface Attributes

Name	Type	Array Size	Value[50](...)
Dimension	String, length = 2, string ...	Scalar	XY
Discretization	64-bit floating-point	2	2.5, 2.5
Max Buffer Size	32-bit integer	1	4
Origin	64-bit floating-point	2	0.0, 0.0

Data (Slices)

	0	1	2
0	2.0	2.01	2.02
1	3.0	3.01	3.02
2	4.0	4.01	4.02

1	103.0	103.01	103.02
2	104.0	104.01	104.02

1	203.0	203.01	203.02
2	204.0	204.01	204.02

1	303.0	303.01	303.02
2	304.0	304.01	304.02

1	403.0	403.01	403.02
2	404.0	404.01	404.02

Times

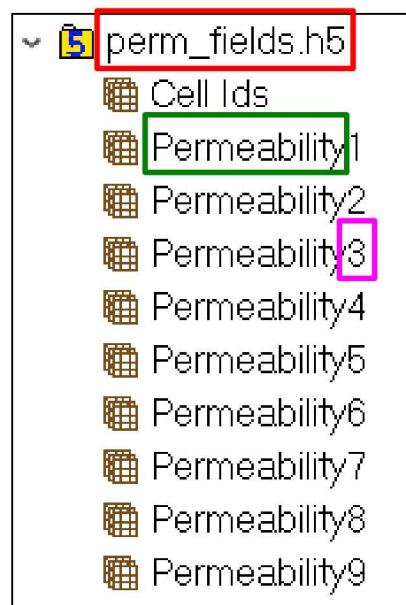
0	0.0
1	10.0
2	20.0
3	30.0
4	40.0
5	50.0
6	60.0
7	70.0
8	80.0
9	90.0
10	100.0
11	110.0

REALIZATION_DEPENDENT Datasets

- CELL_INDEXED – Realization IDs are appended to the HDF5 Dataset name.
- GRIDDED – Realization IDs are appended to the HDF5 Group name.

```
DATASET perm_field
  FILENAME ./perm_fields.h5
  HDF5_DATASET_NAME Permeability
  REALIZATION_DEPENDENT
/
...
MATERIAL_PROPERTY soil2
  ID 2
  POROSITY 0.35d0
  TORTUOSITY 0.5d0
  CHARACTERISTIC_CURVES cc1
  PERMEABILITY
    DATASET perm_field
    VERTICAL_ANISOTROPY_RATIO 0.1
  /
END
```

As seen in HDFView



HDF5 Datasets in Input Decks

PFLOTRAN_DIR/shortcourse/exercises/regional_doublet/stochastic_regional_doublet.in

```
DATASET perm_field
  FILENAME ./perm_fields.h5
  HDF5_DATASET_NAME Permeability
  REALIZATION_DEPENDENT
/
...
MATERIAL_PROPERTY soil2
  ID 2
  POROSITY 0.35d0
  TORTUOSITY 0.5d0
  CHARACTERISTIC_CURVES cc1
  PERMEABILITY
    DATASET perm_field
    VERTICAL_ANISOTROPY_RATIO 0.1
  /
END
```