In [1]:

```
import re
import json
import os
import datetime
import pickle
from itertools import product
from cycler import cycler
import pandas as pd
import numpy as np
import matplotlib.pylab as plt
import matplotlib
from sklearn.linear_model import LinearRegression, Ridge, BayesianRidge
from sklearn.kernel_ridge import KernelRidge
from sklearn.neighbors import KNeighborsRegressor
from sklearn.gaussian_process import GaussianProcessRegressor
from sklearn.svm import SVR
from sklearn.naive_bayes import GaussianNB, BernoulliNB
from sklearn.ensemble import RandomForestRegressor, GradientBoostingRegressor
from scipy.stats import pearsonr
from scipy.optimize import minimize, Bounds, NonlinearConstraint
from itertools import combinations


ELEMS = ['Ag', 'Al', 'Cd', 'Ga', 'In', 'Mg', 'Sb', 'Sn', 'Si', 'Ti', 'Zn', 'Zr']

pd.set_option('display.max_rows', 500)
%matplotlib notebook
```

In [2]:

```
def composition_string_to_vector(name, get_max_el=False):
    elems = ELEMS + ['O']
    vec = [0]*len(elems)
    reg_exp = re.compile("([a-zA-Z]+)([0-9]+)")
    res = reg_exp.findall(name)
    for group in res:
        vec[elems.index(group[0])] = int(group[1])
    vec = vec[:-1]
    if get_max_el:
        return elems[vec.index(max(vec))]
    vec = np.array(vec) / sum(vec)
    return vec


def get_model(df, key, model, direct=False, for_cv=False, **kwargs):
    if not direct:
        model = model(**kwargs)
    name = str(model.__class__).split('.')[-1].replace('>', '').replace("'", "")
    model.fit(list(df['descriptor']), df[key])

    if for_cv:
        return model

    data_org = df[key]
    data_predicted = model.predict(list(df['descriptor']))
    diff = abs(data_predicted - data_org)
    max_val = max(max(data_org), max(data_predicted))
    min_val = min(min(data_org), min(data_predicted))
    corr, _ = pearsonr(data_org, data_predicted)


    fig, ax = plt.subplots()
    ax.plot([min_val,max_val], [min_val,max_val], c='red')
```

```
    ax.scatter(data_org, data_predicted, s=1)

    key_name = key.replace('_', ' ').replace('iswo', 'ISWO').replace('val', 'Valence
    key_name = key_name.replace('dissociation', 'formation')
    title = f"{name} model for {key_name}"
    ax.set_title(title, size=14)
    ax.set_xlabel('Origial Data', size=12)
    ax.set_ylabel('Predicted Data', size=12)
    fig = ax.get_figure()
    fig.savefig(title.replace(' ', '_') + '.eps')
    print(f'Mean Absolute diviation: {np.mean(diff):2.3f}')
    print(f'Pearsons correlation: {corr:0.3f}')
    print(f"R2: {model.score(list(df['descriptor']), df[key]):0.3f}")
    return model
```

Load the dataframe with all data on the 200 atom systems

In [3]:
```
df = pd.read_json('full_200_dataframe.json')
df
```

Out[3]:

	Ag	Al	Ba	Ca	Cd	Ga	Ge	Hf	In	Mg	...
Mg40Si40O120-8	0.000000	0.0	0.0	0.0	0.000000	0.000000	0.0	NaN	0.000000	0.500000	...
Mg29Ga57O114-7	0.000000	0.0	0.0	0.0	0.000000	0.662791	0.0	NaN	0.000000	0.337209	...
Ag56Sn28O112-3	0.666667	0.0	0.0	0.0	0.000000	0.000000	0.0	NaN	0.000000	0.000000	...
Ag80O120-1	1.000000	0.0	0.0	0.0	0.000000	0.000000	0.0	NaN	0.000000	0.000000	...
Zn40Sn40O120-3	0.000000	0.0	0.0	0.0	0.000000	0.000000	0.0	NaN	0.000000	0.000000	...
...	...	...	...	...	...	...	...	...	...	...	...
Si5Sb50O135-9.cif	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
In14Ag70O126-5.cif	0.833333	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.166667	NaN	...
Cd50Ga20O130-7.cif	NaN	NaN	NaN	NaN	0.714286	0.285714	NaN	NaN	NaN	NaN	...
Ti60Cd3O126-8.cif	NaN	NaN	NaN	NaN	0.047619	NaN	NaN	NaN	NaN	NaN	...
Ga60Sn30O126-8.cif	NaN	NaN	NaN	NaN	NaN	0.666667	NaN	NaN	NaN	NaN	...

4648 rows × 49 columns

Reduce to 3 most stable and add descriptor and main element

In [4]:

```
groups = []

for name, group in df.groupby('composition'):
    std = group['total_energy'].std()
    group = group.sort_values(by='total_energy').head(3)
    groups.append(group)

df = pd.concat(groups)

df['descriptor'] = [composition_string_to_vector(i) for i in df.index]

main_elements = []
for entry in df.itertuples():
    main_elements.append(composition_string_to_vector(entry.composition, get_max_el=

df['main_elements'] = main_elements
df
```

Out[4]:

	Ag	Al	Ba	Ca	Cd	Ga	Ge	Hf	In	Mg	...	skip_val	std_is
Ag10Sb50O140-5.cif	0.166667	0.0	0.0	0.0	0.0	0.0	0.0	NaN	0.0	0.0	...	4.0	
Ag10Sb50O140-2.cif	0.166667	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	1.0	
Ag10Sb50O140-1.cif	0.166667	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	3.0	
Ag16Sn80O104-5	0.166667	0.0	0.0	0.0	0.0	0.0	0.0	NaN	0.0	0.0	...	6.0	
Ag16Sn80O104-7	0.166667	0.0	0.0	0.0	0.0	0.0	0.0	NaN	0.0	0.0	...	2.0	
...	...	...	...	...	...	...	...	...	...	...	...	...	
Zr9Sn90O108-8	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	NaN	0.0	0.0	...	2.0	
Zr9Sn90O108-5	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	NaN	0.0	0.0	...	3.0	
Zr9Zn90O108-4	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	NaN	0.0	0.0	...	3.0	
Zr9Zn90O108-7	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	NaN	0.0	0.0	...	4.0	
Zr9Zn90O108-2	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	NaN	0.0	0.0	...	7.0	

1434 rows × 51 columns

In [5]:

```
df.describe()
```

Out[5]:

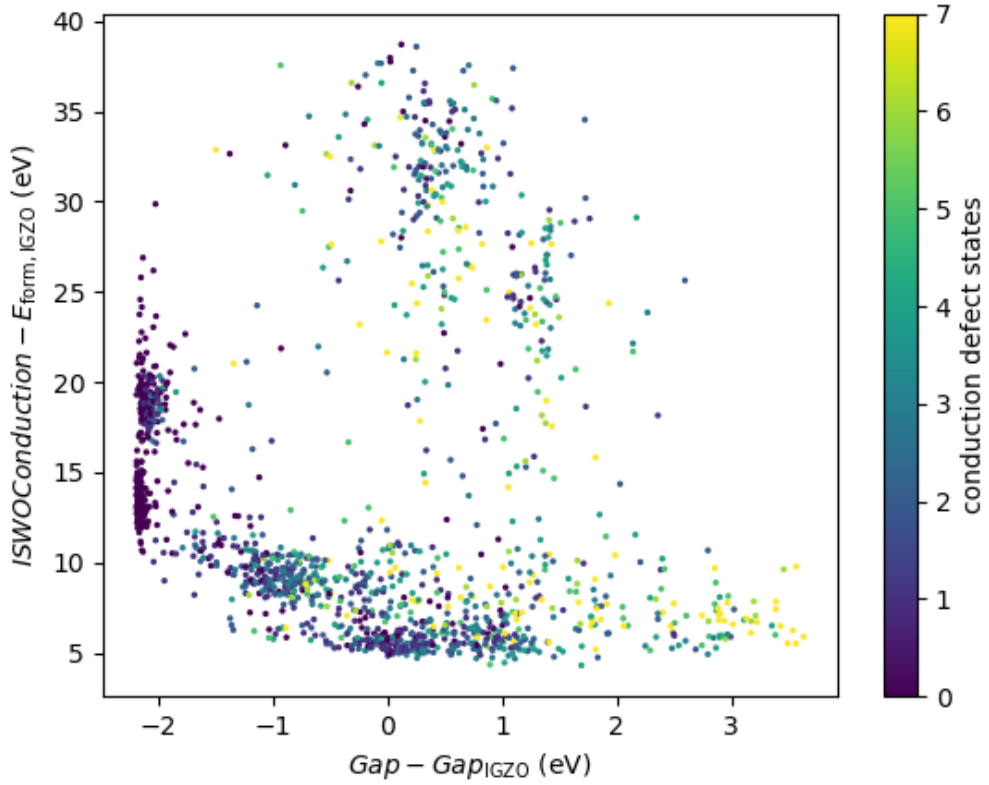
	Ag	Al	Ba	Ca	Cd	Ga	Ge	Hf	I
count	1260.000000	1258.000000	1222.0	1222.0	1255.000000	1260.000000	1222.0	0.0	1258.00000
mean	0.105044	0.099009	0.0	0.0	0.085841	0.100456	0.0	NaN	0.10043
std	0.269985	0.262365	0.0	0.0	0.242851	0.261704	0.0	NaN	0.26459

	Ag	Al	Ba	Ca	Cd	Ga	Ge	Hf	I
min	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.0	NaN	0.00000
25%	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.0	NaN	0.00000
50%	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.0	NaN	0.00000
75%	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.0	NaN	0.00000
max	1.000000	1.000000	0.0	0.0	1.000000	1.000000	0.0	NaN	1.00000

8 rows × 44 columns

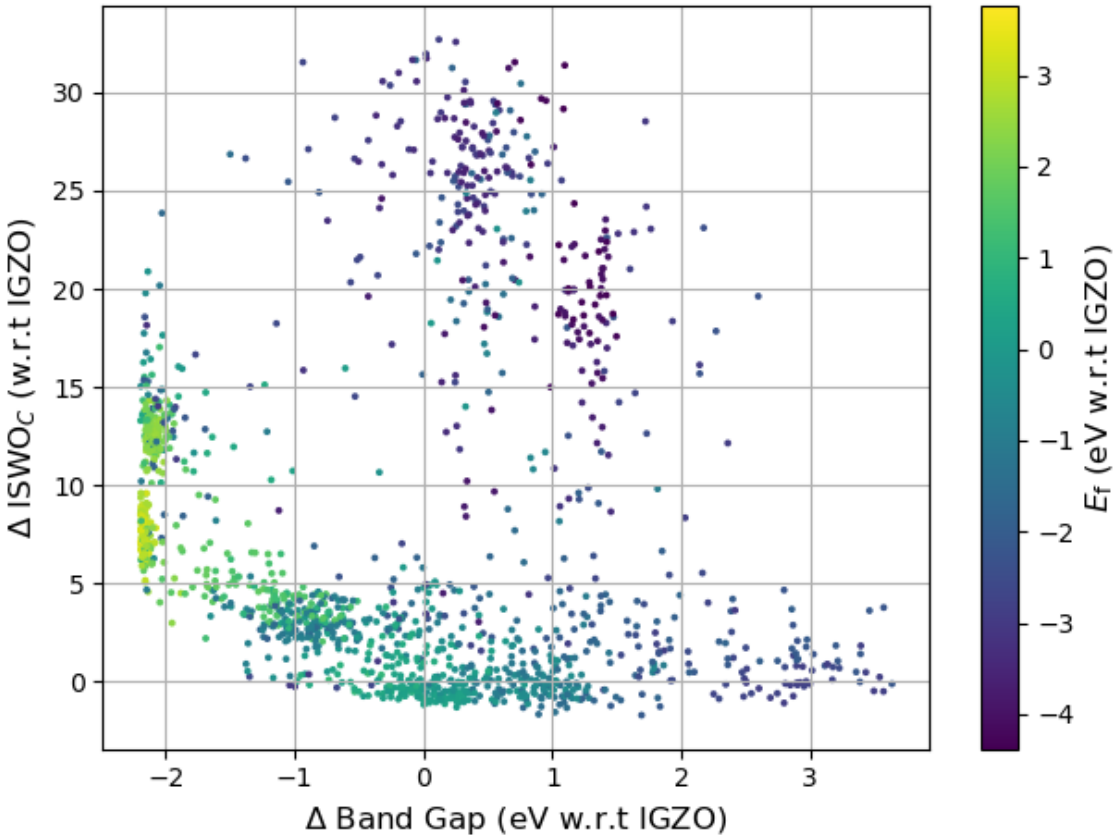
Exploring data

In [6]:
```
fig, ax = plt.subplots()

figure = ax.scatter(df['transport_gap']-2.2, df['mean_iswo_cond'], s=2, c=df['skip_c
ax.set_xlabel(r'$Gap - Gap_\mathrm{IGZO}$ (eV)')
ax.set_ylabel(r'$ISWO Conduction - E_\mathrm{form, IGZO}$ (eV)')
cb = plt.colorbar(figure)
cb.set_label('conduction defect states')
```



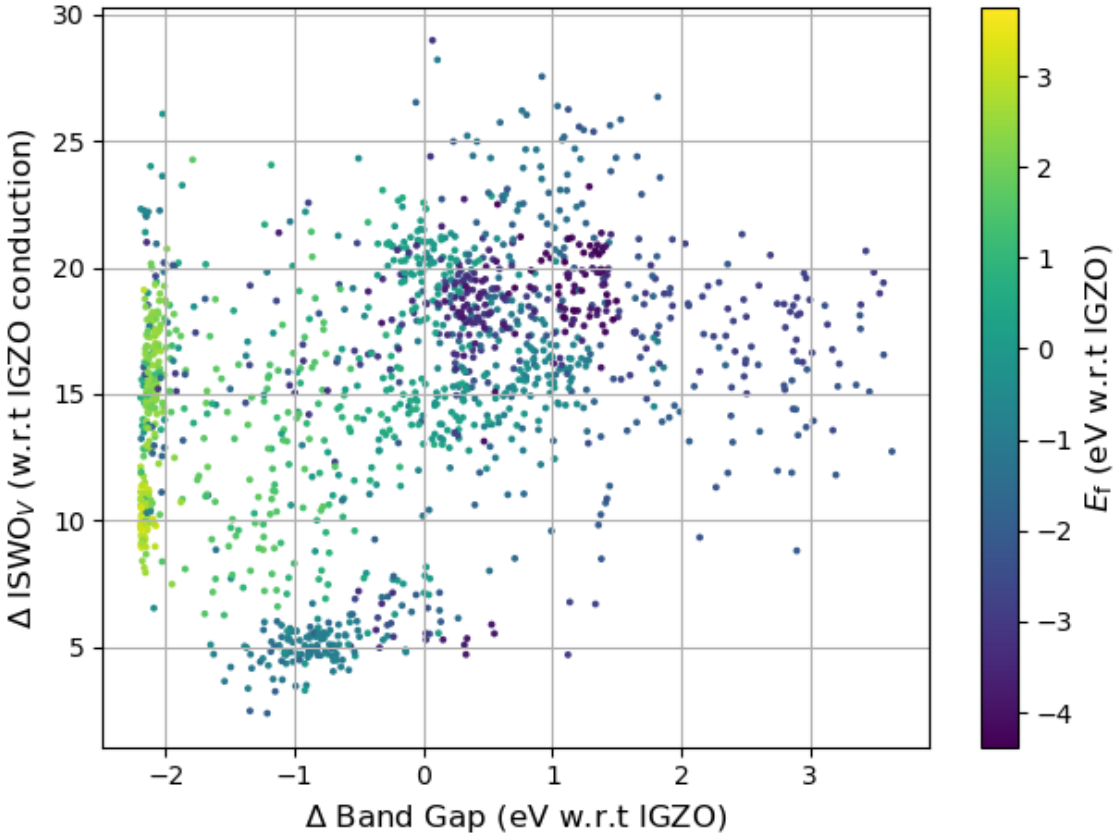
In [7]:
```
fig, ax = plt.subplots()

figure = ax.scatter(df['transport_gap']-2.2, df['mean_iswo_cond'] - 6.0, s=3, c=df['
ax.set_xlabel(r'$\Delta$ Band Gap (eV w.r.t IGZO)', size=12)
ax.set_ylabel(r'$\Delta$ ISWO$_C$ (w.r.t IGZO)', size=12)
ax.grid()
```

```
cb = plt.colorbar(figure, cmap='jet')
cb.set_label(r'$E_\mathrm{f}$ (eV w.r.t IGZO)', size=12)

plt.tight_layout()
fig.savefig('training_set_energy.eps')
fig.savefig('training_set_energy.jpeg')
```



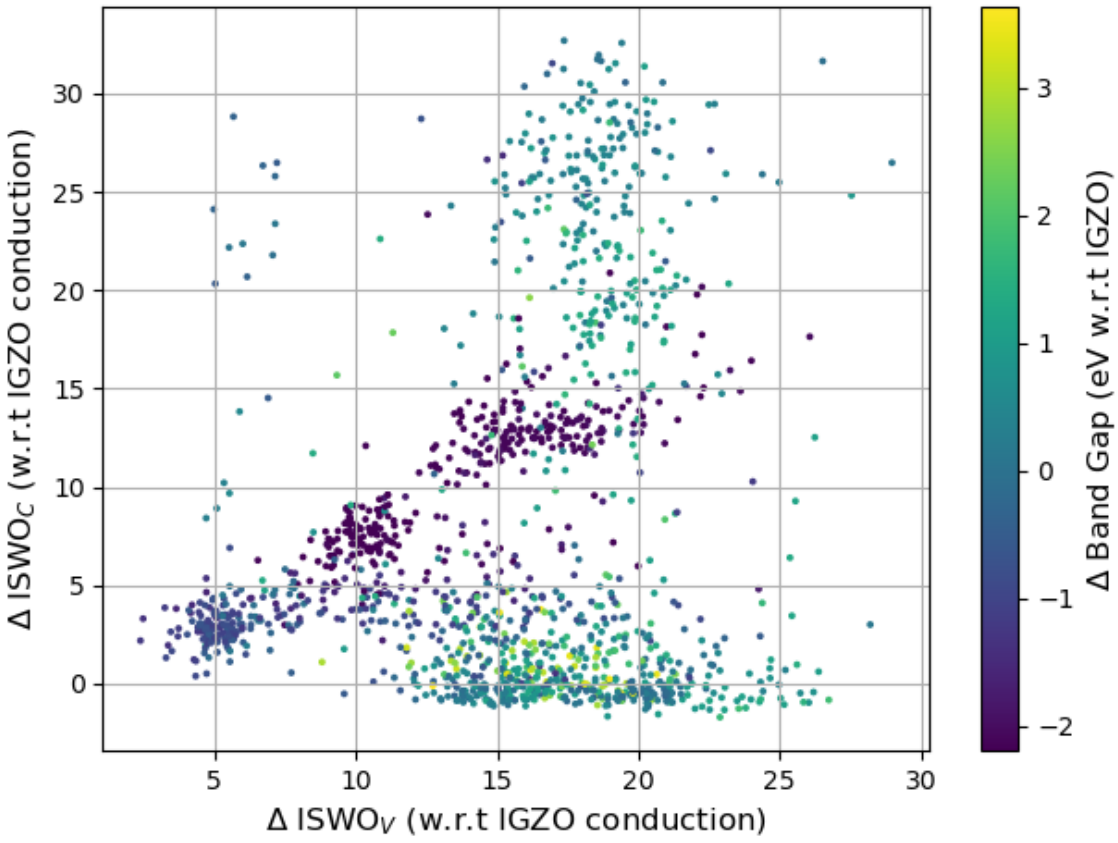
In [8]:

```
fig, ax = plt.subplots()

figure = ax.scatter(df['transport_gap']-2.2, df['mean_iswo_val'] - 6.0, s=3, c=df['d
ax.set_xlabel(r'$\Delta$ Band Gap (eV w.r.t IGZO)', size=12)
ax.set_ylabel(r'$\Delta$ ISWO$_V$ (w.r.t IGZO conduction)', size=12)
ax.grid()
cb = plt.colorbar(figure, cmap='jet')
cb.set_label(r'$E_\mathrm{f}$ (eV w.r.t IGZO)', size=12)

plt.tight_layout()
fig.savefig('training_set_energy.eps')
fig.savefig('training_set_energy.jpeg')
```



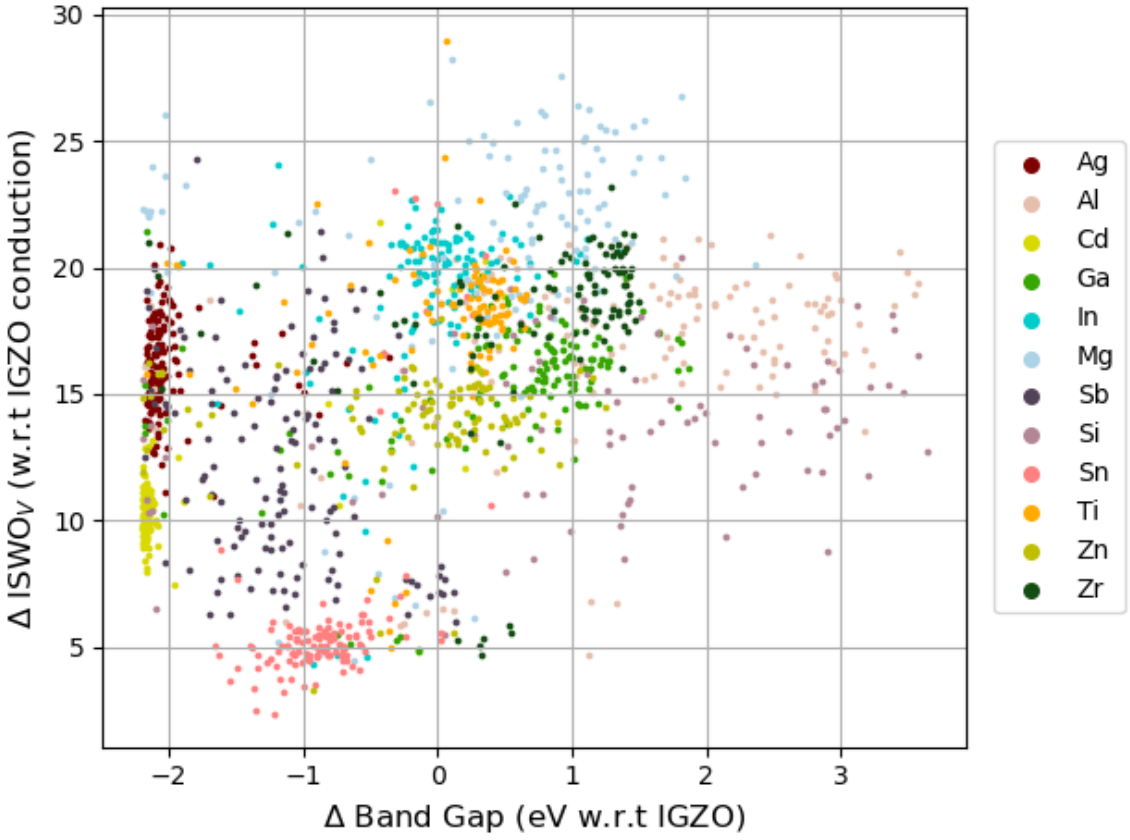
In [9]:

```
fig, ax = plt.subplots()

figure = ax.scatter(df['mean_iswo_val']-6.0, df['mean_iswo_cond'] - 6.0, s=3, c=df['
ax.set_xlabel(r'$\Delta$ ISWO$_V$ (w.r.t IGZO conduction)', size=12)
ax.set_ylabel(r'$\Delta$ ISWO$_C$ (w.r.t IGZO conduction)', size=12)
ax.grid()
cb = plt.colorbar(figure, cmap='jet')
cb.set_label(r'$\Delta$ Band Gap (eV w.r.t IGZO)', size=12)

plt.tight_layout()
fig.savefig('training_set_energy.eps')
fig.savefig('training_set_energy.jpeg')
```



In [10]:

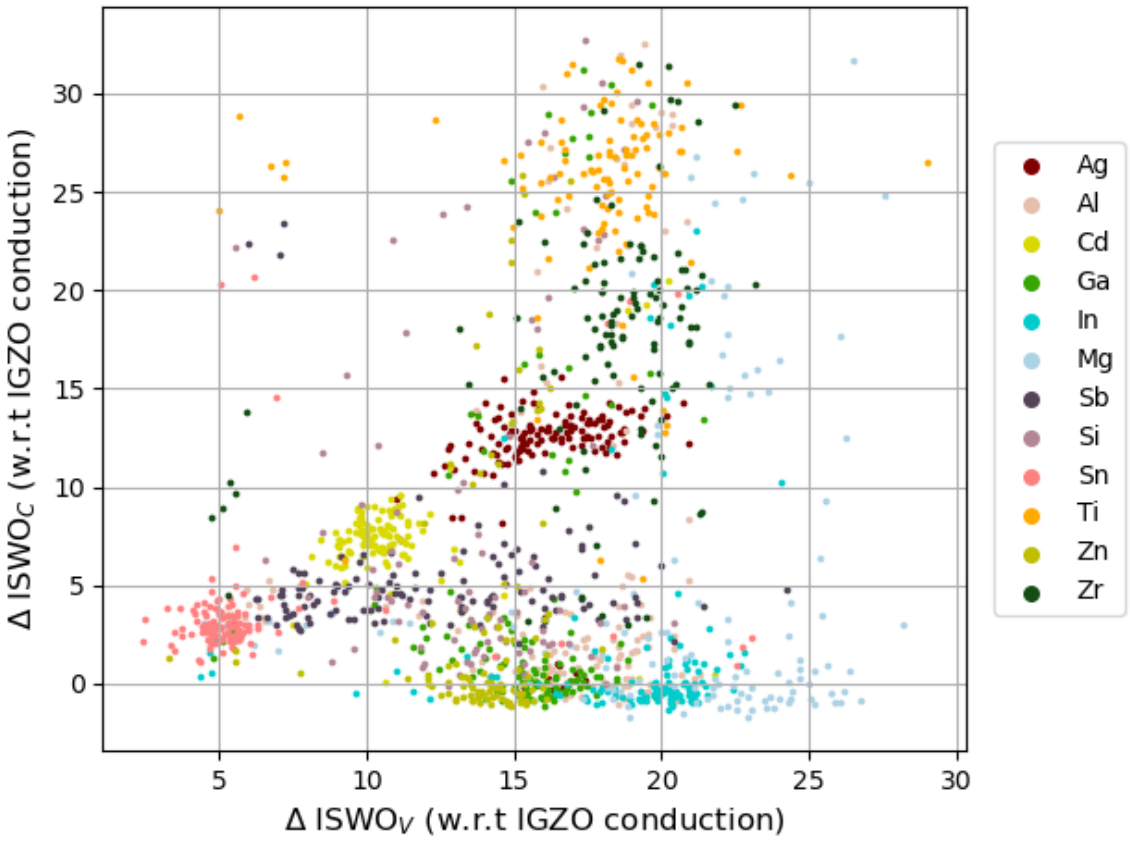
```
fig, ax = plt.subplots()


colors = ['#7f0000', '#e6bfac', '#d9d900', '#37a600', '#00cccc', '#acd2e6', '#524359
          '#ff8080', '#ffaa00', '#bfbf00', '#134d13', '#008080', '#001540', '#800080
          '#734d00', '#7f8020', '#ace6bf', '#004040', '#202080', '#4d134d', '#ff5500

custom_cycler = cycler(color=colors)
ax.set_prop_cycle(custom_cycler)


for name, group in df.groupby(by='main_elements'):
    ax.scatter(group['transport_gap']-2.2, group['mean_iswo_val'] - 6.0, s=3, label=
    ax.set_xlabel(r'$\Delta$ Band Gap (eV w.r.t IGZO)', size=12)
    ax.set_ylabel(r'$\Delta$ ISWO$_V$ (w.r.t IGZO conduction)', size=12)

ax.grid()

legend = ax.legend(loc="right", bbox_to_anchor=(1.2, 0.5), ncol=1)

for hh in legend.legendHandles:
    hh._sizes = [30]
plt.tight_layout()
fig.savefig('training_set_composition.eps')
```




```
The PostScript backend does not support transparency; partially transparent artists
will be rendered opaque.
```

In [11]:

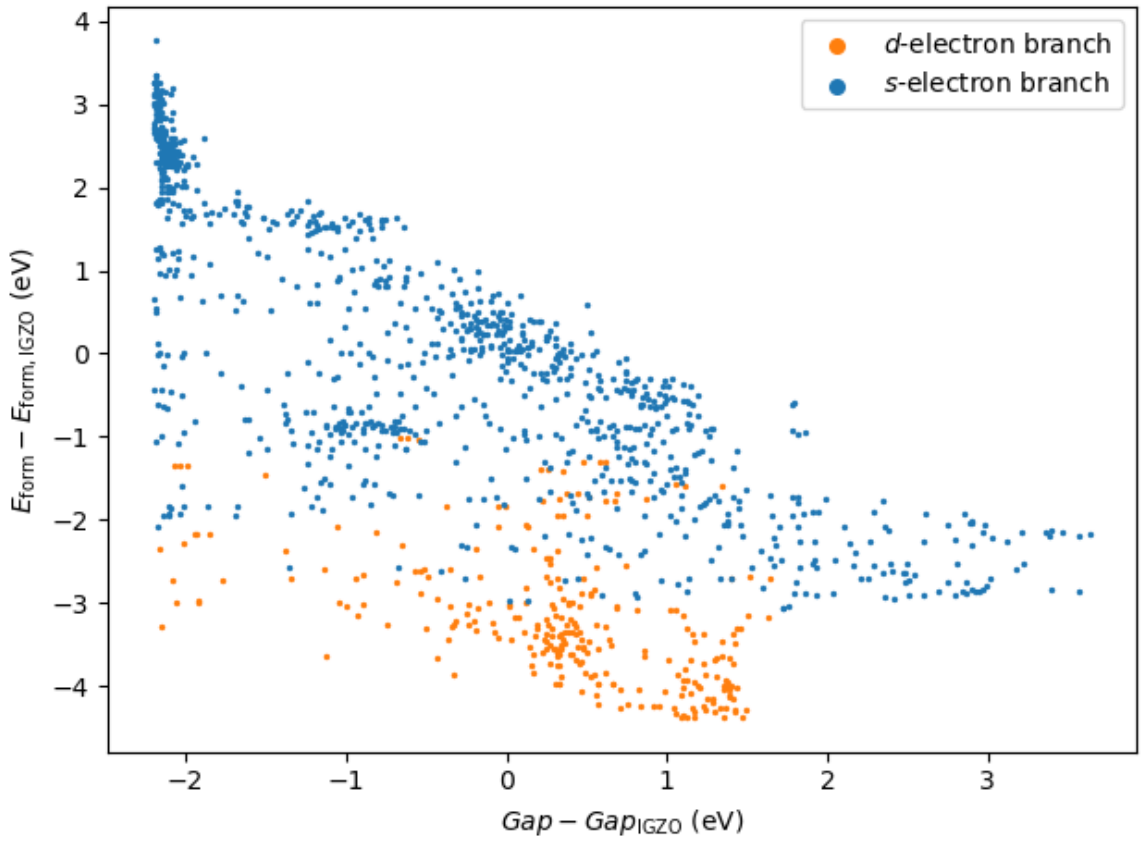
```
fig, ax = plt.subplots()

from cycler import cycler


colors = ['#7f0000', '#e6bfac', '#d9d900', '#37a600', '#00cccc', '#acd2e6', '#524359
          '#ff8080', '#ffaa00', '#bfbf00', '#134d13', '#008080', '#001540', '#800080
          '#734d00', '#7f8020', '#ace6bf', '#004040', '#202080', '#4d134d', '#ff5500

custom_cycler = cycler(color=colors)
ax.set_prop_cycle(custom_cycler)


for name, group in df.groupby(by='main_elements'):
    ax.scatter(group['mean_iswo_val']-6.0, group['mean_iswo_cond'] - 6.0, s=3, label
    ax.set_xlabel(r'$\Delta$ ISWO$_V$ (w.r.t IGZO conduction)', size=12)
    ax.set_ylabel(r'$\Delta$ ISWO$_C$ (w.r.t IGZO conduction)', size=12)

ax.grid()

legend = ax.legend(loc="right", bbox_to_anchor=(1.2, 0.5), ncol=1)

for hh in legend.legendHandles:
    hh._sizes = [30]
plt.tight_layout()
fig.savefig('training_set_composition.eps')
```




```
The PostScript backend does not support transparency; partially transparent artists
will be rendered opaque.
```

In [12]:

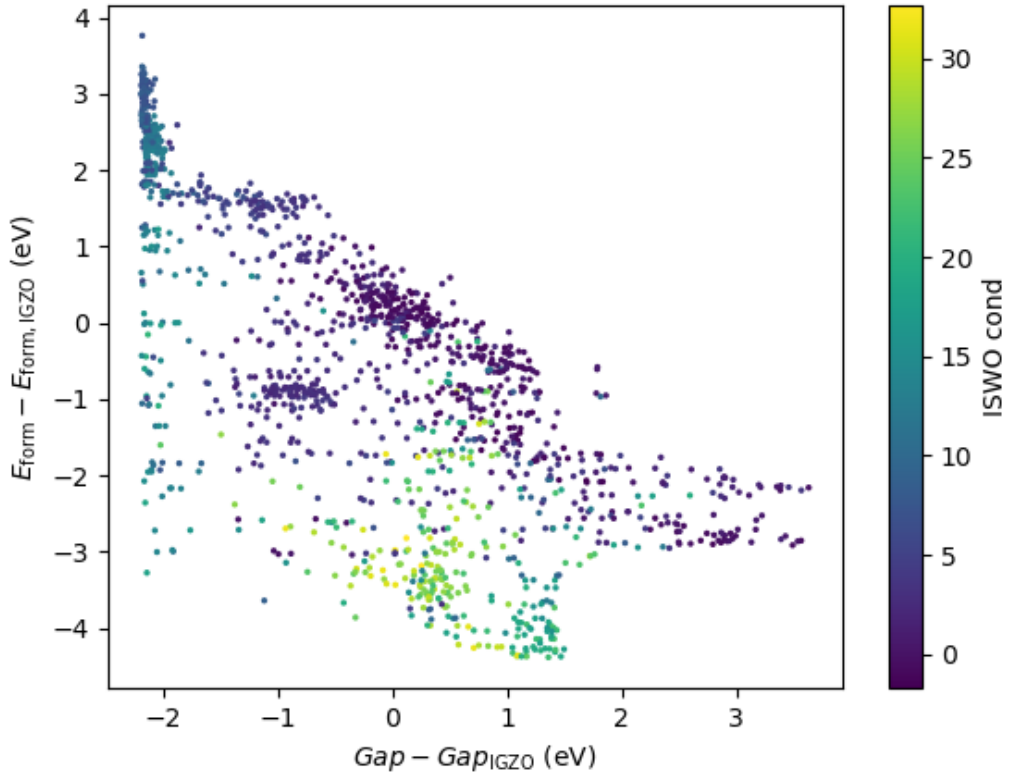
```
fig, ax = plt.subplots()
x= 0.25
df_ti = df[df['Ti'] > x]
df_zr = df[df['Zr'] > x]
df_s = df[df['Ti'] < x]
df_s = df_s[df['Zr'] < x]

ax.scatter(df_ti['transport_gap']-2.2, df_ti['dissociation_energy_per_O'], s=2, c='t
ax.scatter(df_zr['transport_gap']-2.2, df_zr['dissociation_energy_per_O'], s=2, c='t
ax.scatter(df_s['transport_gap']-2.2, df_s['dissociation_energy_per_O'], s=2, c='tab
ax.set_xlabel(r'$Gap - Gap_\mathrm{IGZO}$ (eV)')
ax.set_ylabel(r'$E_\mathrm{form} - E_\mathrm{form, IGZO}$ (eV)')
legend = ax.legend()
for hh in legend.legendHandles:
    hh._sizes = [30]
plt.tight_layout()
```




```
/home/vsette63/anaconda3/envs/2022.4/lib/python3.7/site-packages/ipykernel_launcher.
py:6: UserWarning: Boolean Series key will be reindexed to match DataFrame index.
```

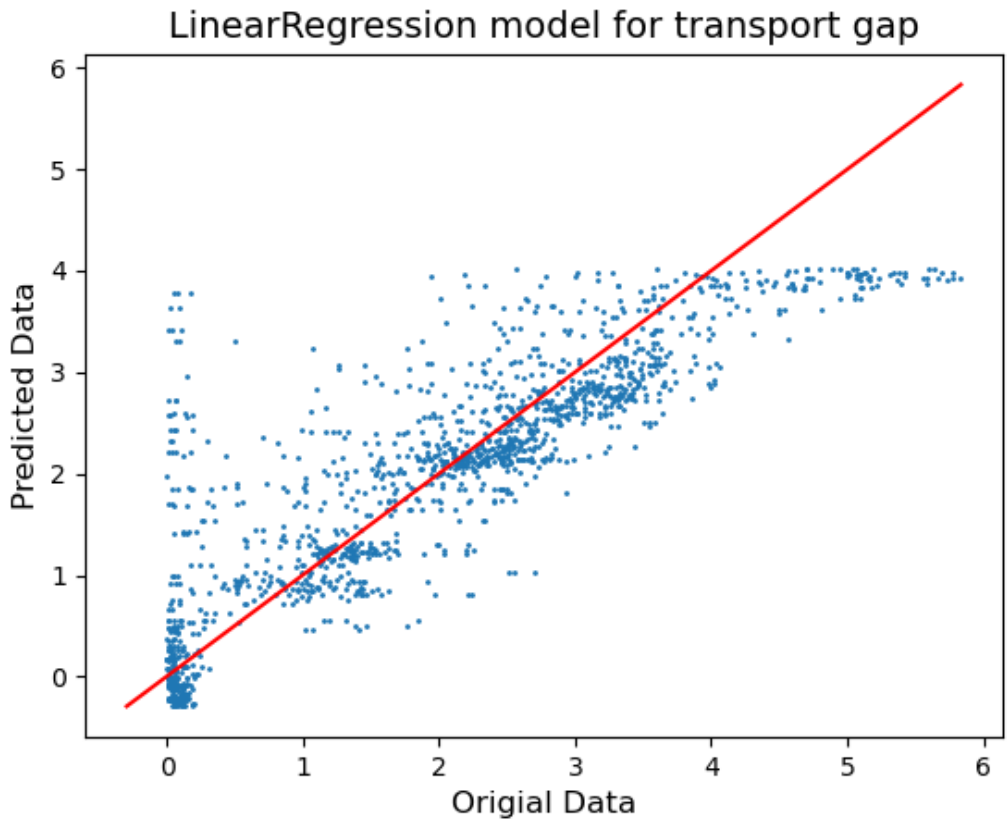
In [13]:

```
fig, ax = plt.subplots()

figure = ax.scatter(df['transport_gap']-2.2, df['dissociation_energy_per_O'], s=2, c
ax.set_xlabel(r'$Gap - Gap_\mathrm{IGZO}$ (eV)')
ax.set_ylabel(r'$E_\mathrm{form} - E_\mathrm{form, IGZO}$ (eV)')
cb = plt.colorbar(figure)
cb.set_label('ISWO cond')
```



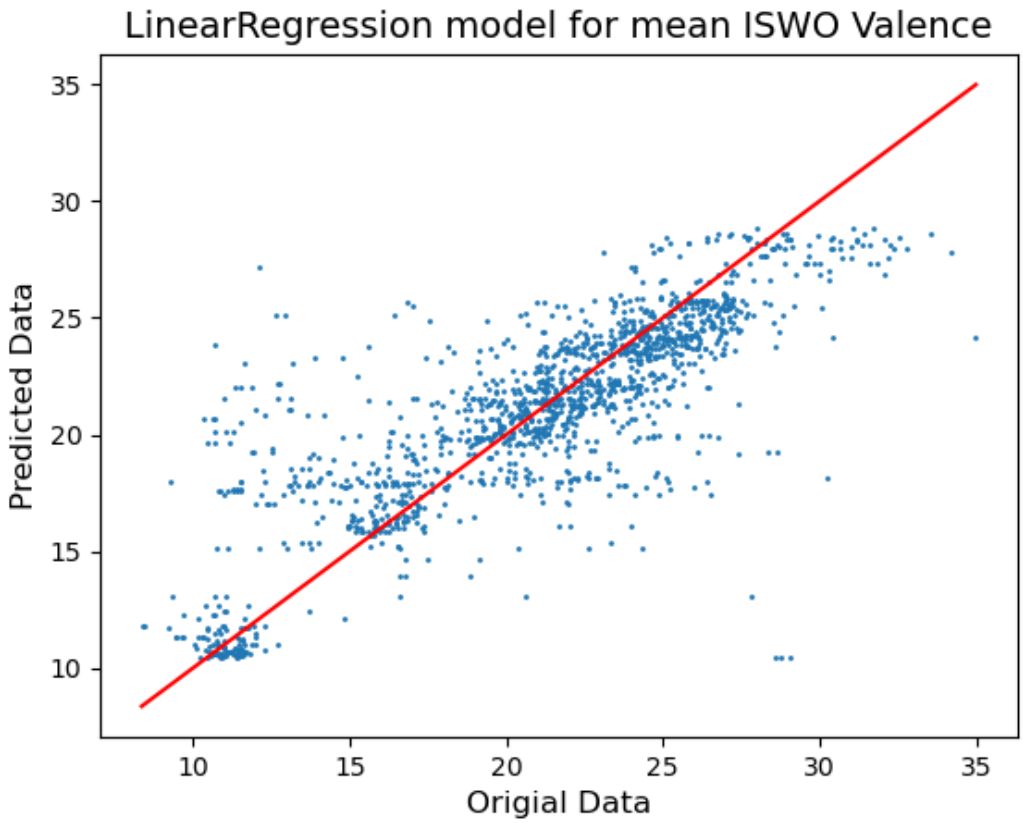
linear model

In [14]:
```
gap_model = get_model(df, 'transport_gap', LinearRegression)
```



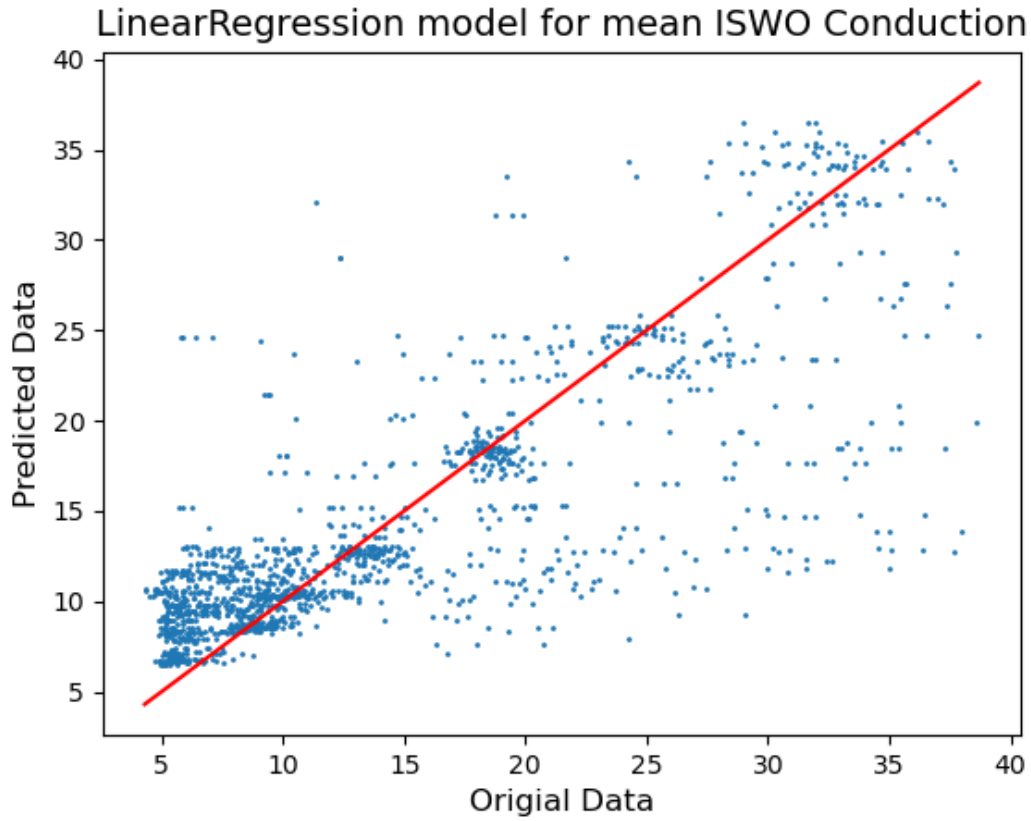
```
Mean Absolute diviation: 0.486
Pearsons correlation: 0.858
R2: 0.736
```

In [15]:
```
iswo_val_model = get_model(df, 'mean_iswo_val', LinearRegression)
```




```
Mean Absolute diviation: 1.990
Pearsons correlation: 0.815
R2: 0.664
```

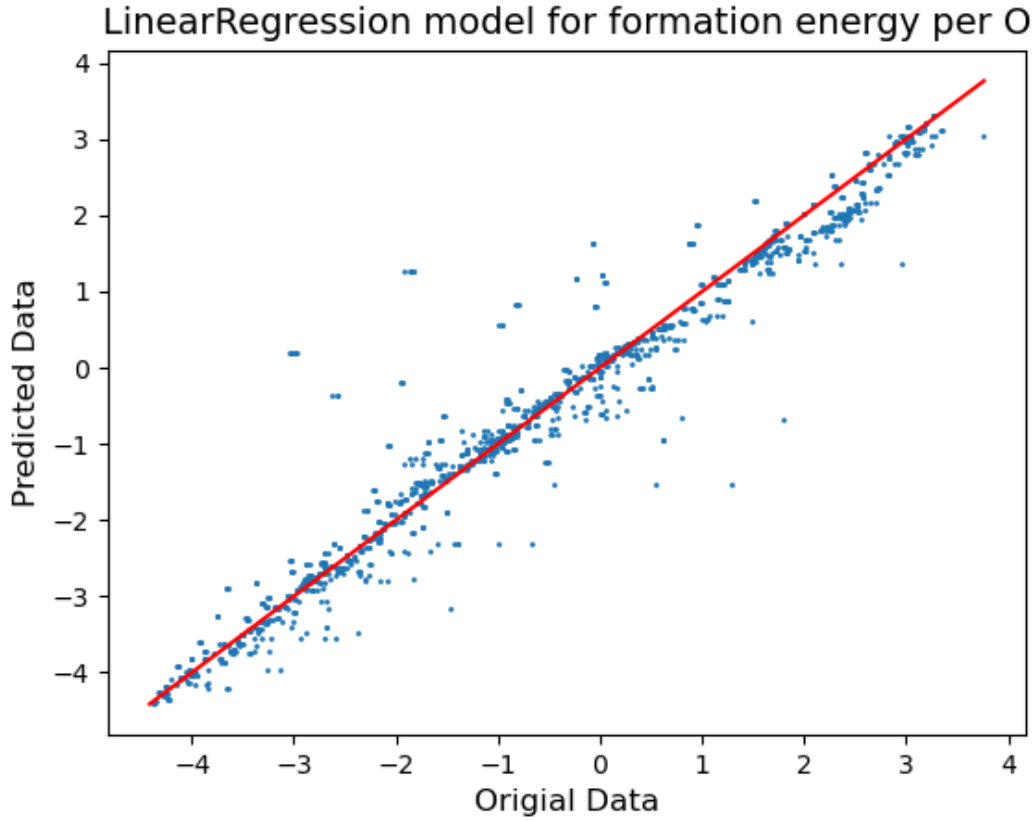
In [16]:
```
iswo_cond_model = get_model(df, 'mean_iswo_cond', LinearRegression)
```




```
Mean Absolute diviation: 3.420
Pearsons correlation: 0.802
R2: 0.643
```

In [17]:

```
formation_model = get_model(df, 'dissociation_energy_per_O', LinearRegression)
```



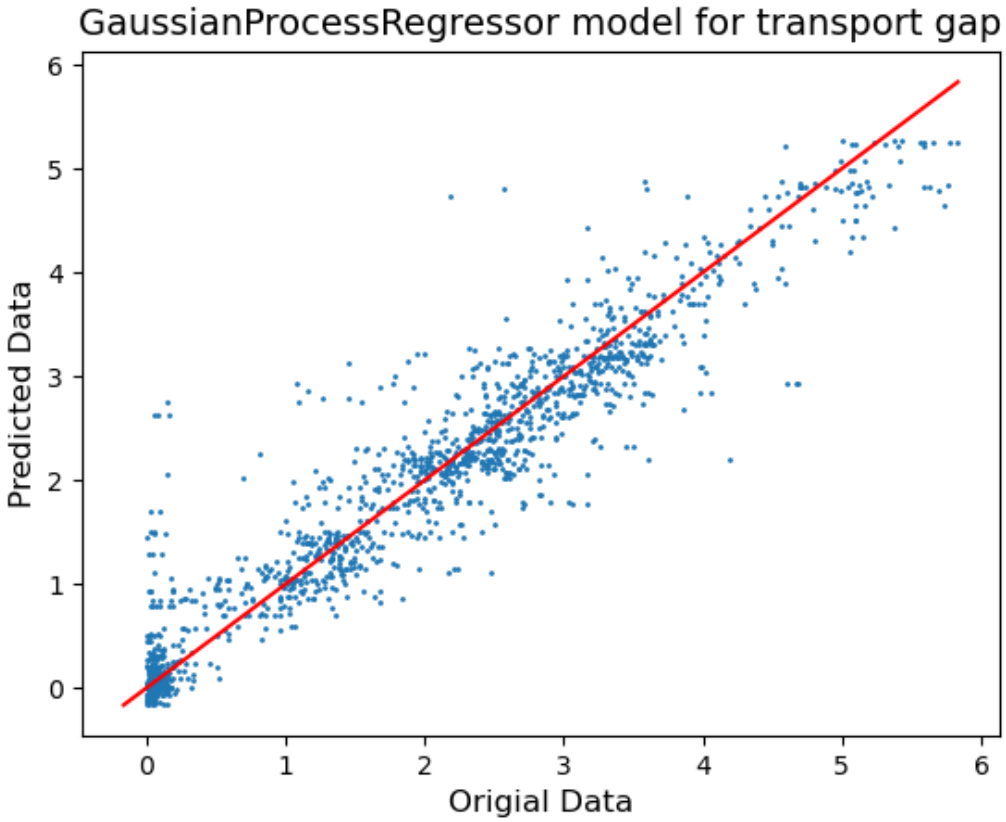

```
Mean Absolute diviation: 0.248
```

```
Pearsons correlation: 0.971
R2: 0.944
```

Gaussian process

In [18]:

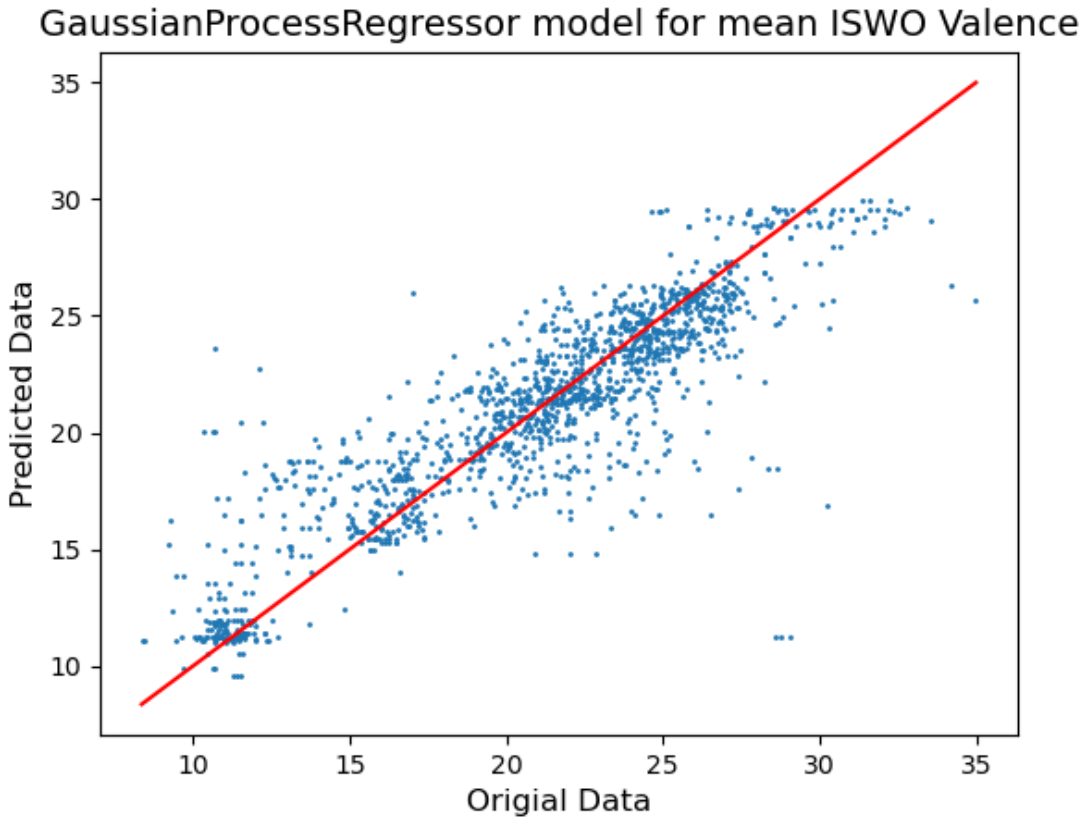
```
from sklearn.gaussian_process.kernels import DotProduct, WhiteKernel, RBF, RationalQ
kernel = RBF(length_scale=5, length_scale_bounds='fixed')
gap_model = get_model(df, 'transport_gap', GaussianProcessRegressor, kernel=kernel)
```




```
Mean Absolute diviation: 0.292
Pearsons correlation: 0.949
R2: 0.900
```

In [19]:

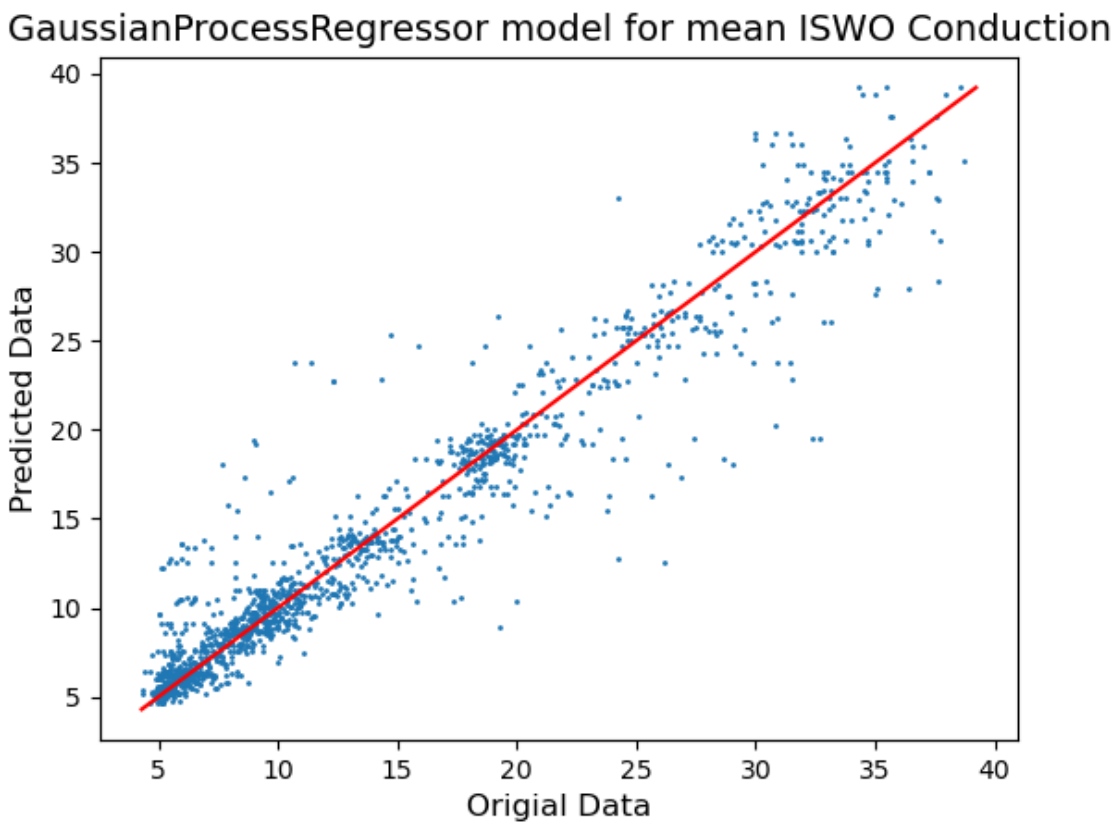
```
kernel = RBF(length_scale=10, length_scale_bounds='fixed')
iswo_val_model = get_model(df, 'mean_iswo_val', GaussianProcessRegressor, kernel = k
```




```
Mean Absolute diviation: 1.651
Pearsons correlation: 0.878
R2: 0.771
```

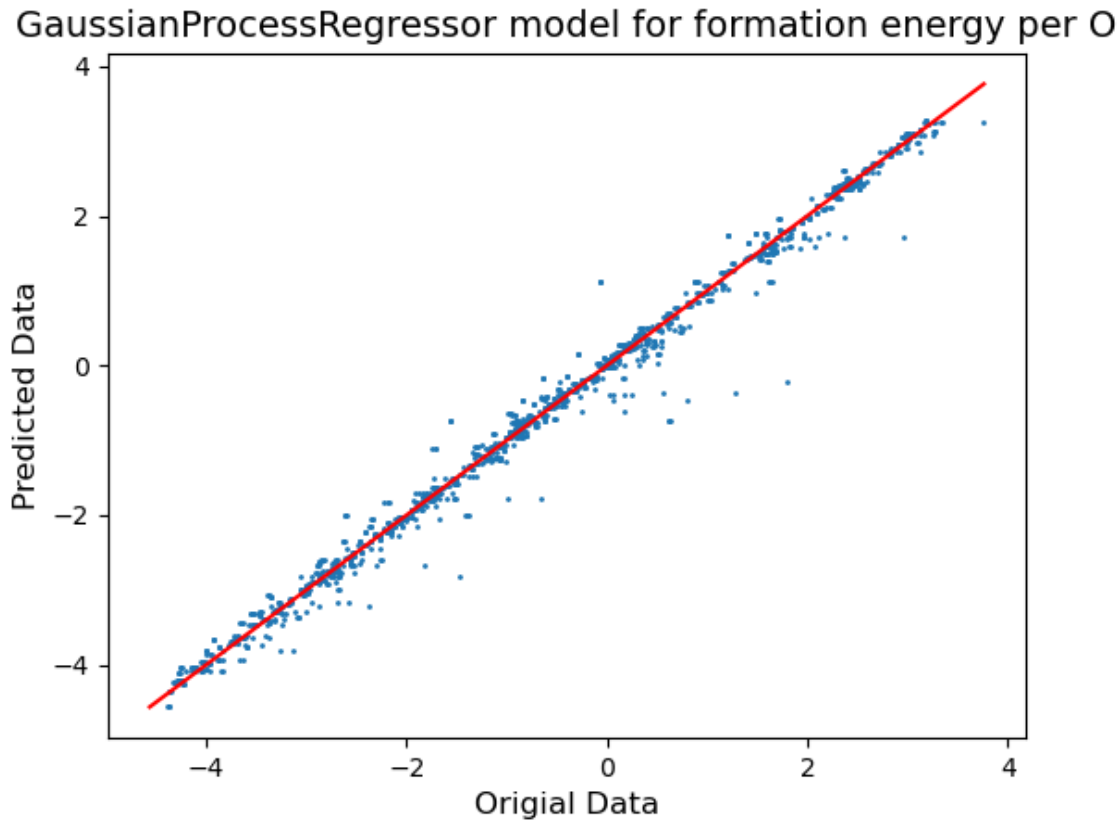
In [20]:

```
kernel = RBF(length_scale=20, length_scale_bounds='fixed')
kernel = RationalQuadratic(length_scale=4, alpha=200, length_scale_bounds='fixed', a
iswo_cond_model = get_model(df, 'mean_iswo_cond', GaussianProcessRegressor, kernel=k
```



```
Mean Absolute diviation: 1.423
Pearsons correlation: 0.963
R2: 0.927
```

In [21]:
```
kernel = RBF(length_scale=2, length_scale_bounds='fixed')
formation_model = get_model(df, 'dissociation_energy_per_O', GaussianProcessRegresso
```

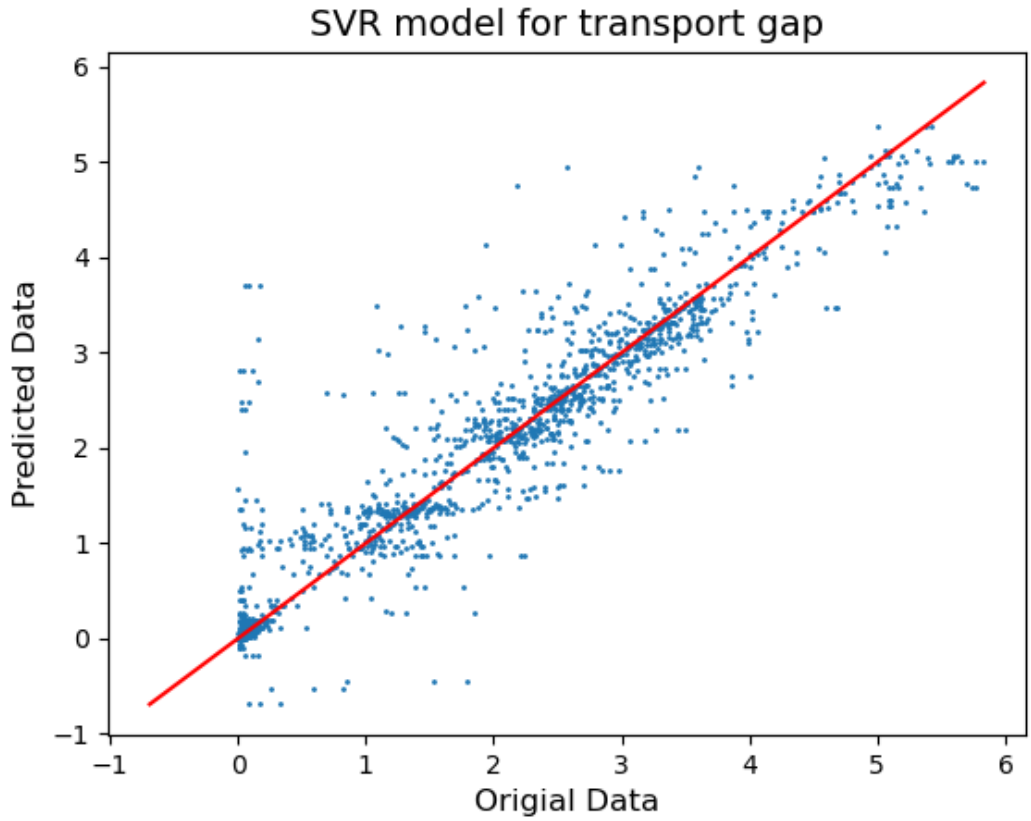



```
Mean Absolute diviation: 0.110
Pearsons correlation: 0.995
R2: 0.990
```

svm

In [22]:
```
degree = 10
epsilon = 0.001
C=400
kernel='poly'
```

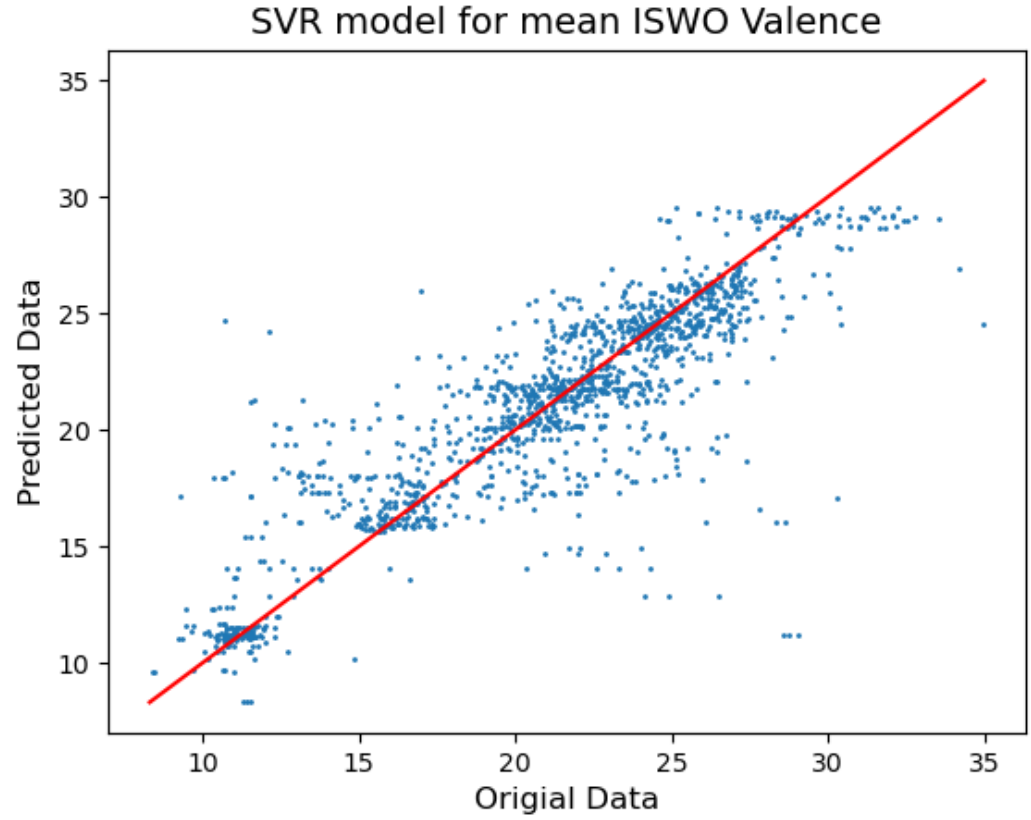
In [23]:
```
gap_model = get_model(df, 'transport_gap', SVR, kernel=kernel, C=C, gamma='auto', de
                              epsilon=epsilon, coef0=1)
```




```
Mean Absolute diviation: 0.301
Pearsons correlation: 0.928
R2: 0.857
```

In [24]:

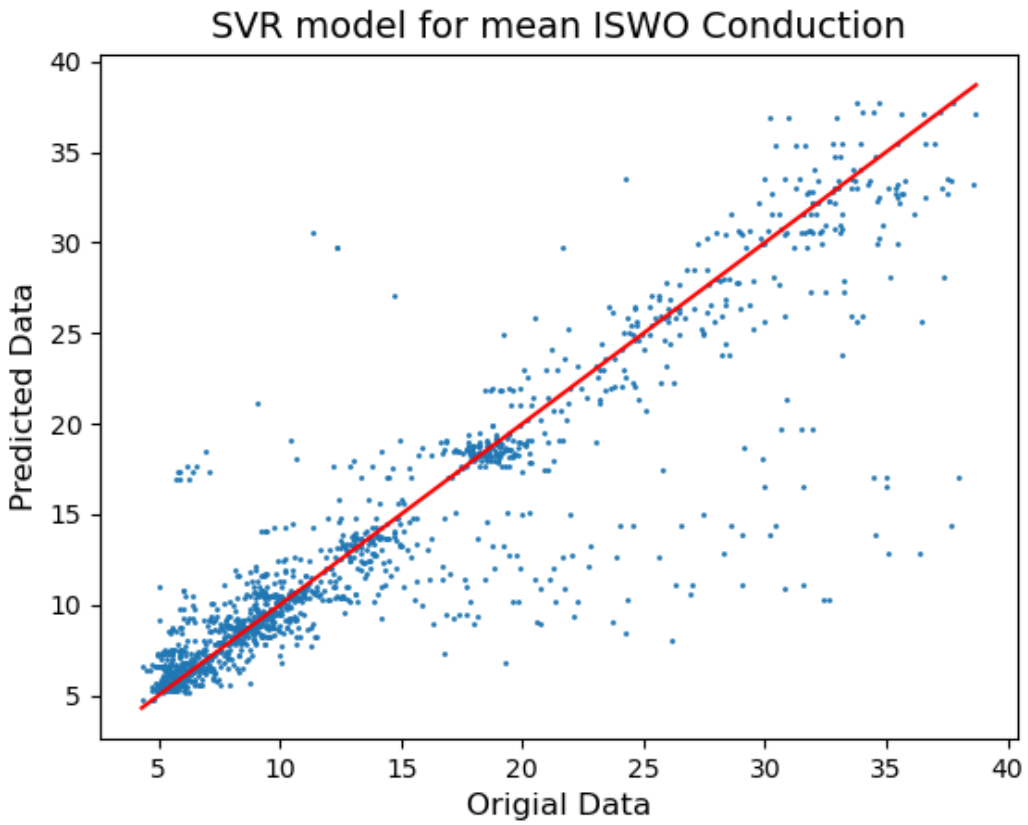
```
iswo_val_model = get_model(df, 'mean_iswo_val', SVR, kernel='poly', C=100, gamma='au
                           epsilon=epsilon, coef0=1)
```



```
Mean Absolute diviation: 1.654
Pearsons correlation: 0.861
R2: 0.740
```

In [25]:

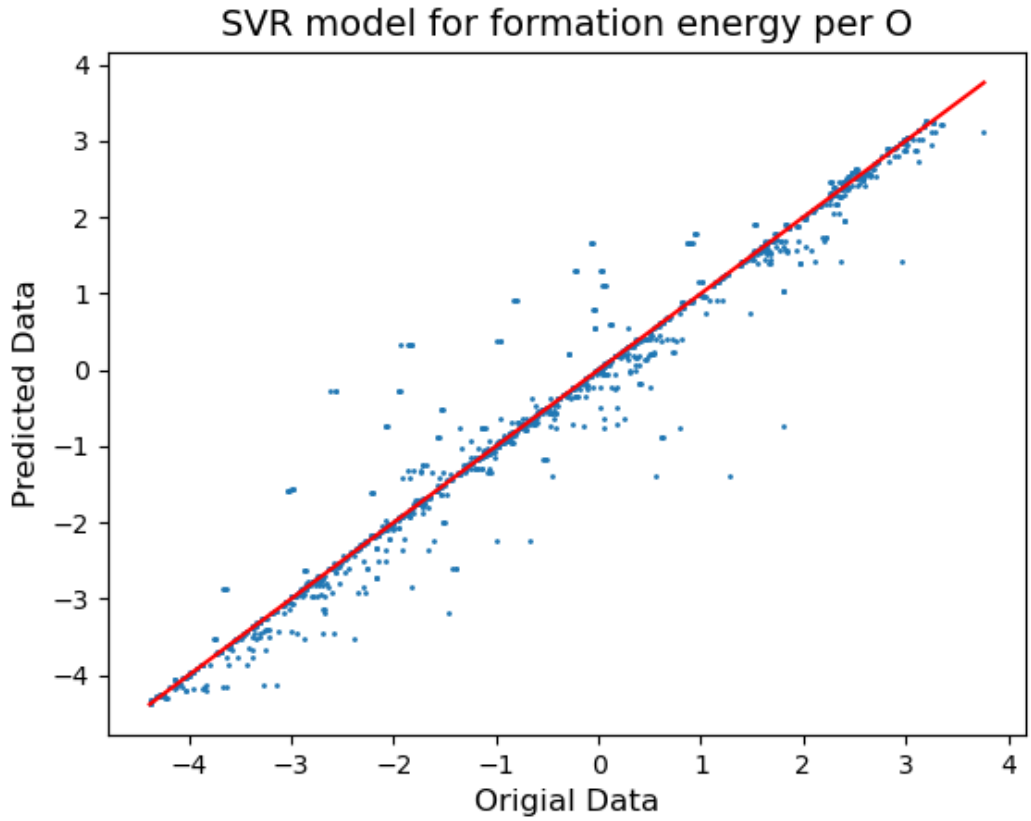
```
iswo_cond_model = get_model(df, 'mean_iswo_cond', SVR, kernel='poly', C=100, gamma='
                            epsilon=epsilon, coef0=1)
```




```
Mean Absolute diviation: 1.903
Pearsons correlation: 0.906
R2: 0.818
```

In [26]:

```
formation_model = get_model(df, 'dissociation_energy_per_O', SVR, kernel='poly', C=1
                            epsilon=epsilon, coef0=1)
```




```
Mean Absolute diviation: 0.155
Pearsons correlation: 0.982
R2: 0.964
```

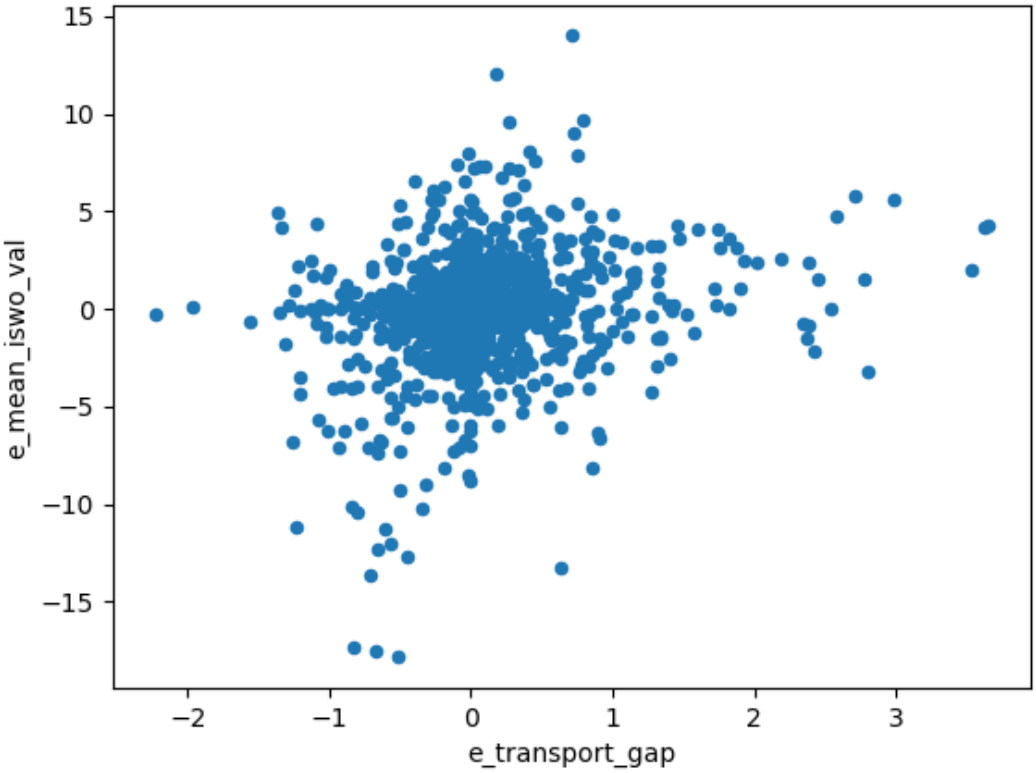
Correlations between model errors

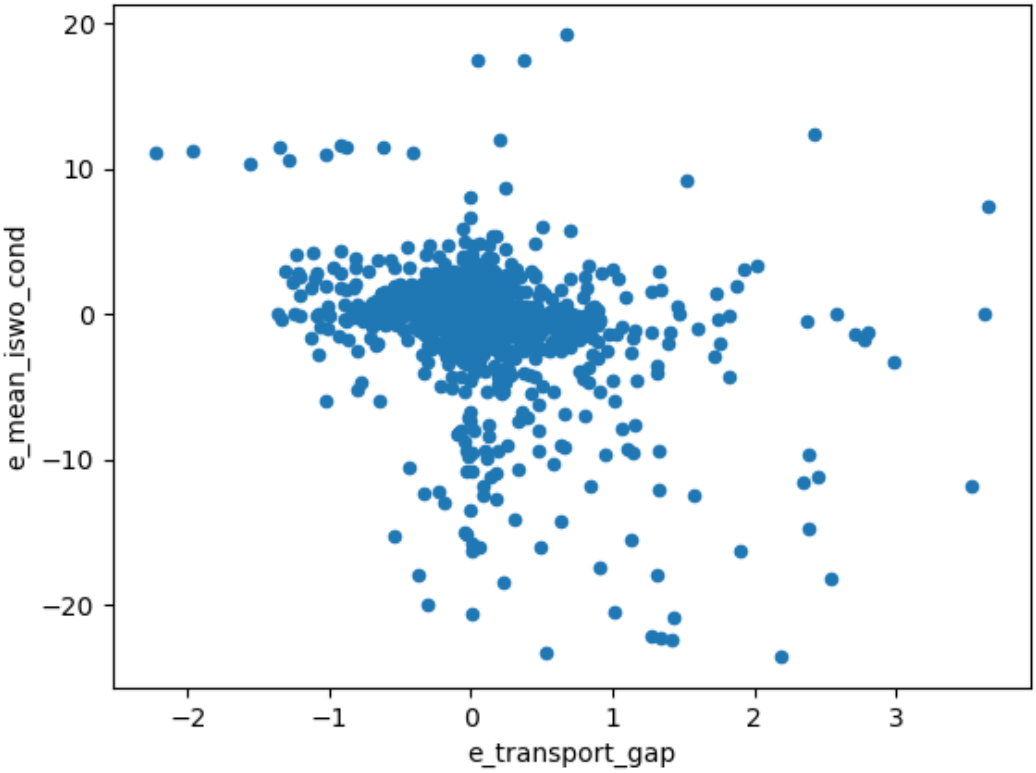
In predictions for the ISWO and gap values we observe several severe outliers. These points do not show any correlation. There are only some individual systems that have serious outlier errors for two observables at the same time. These cases, however, do not form a specific class of materials.

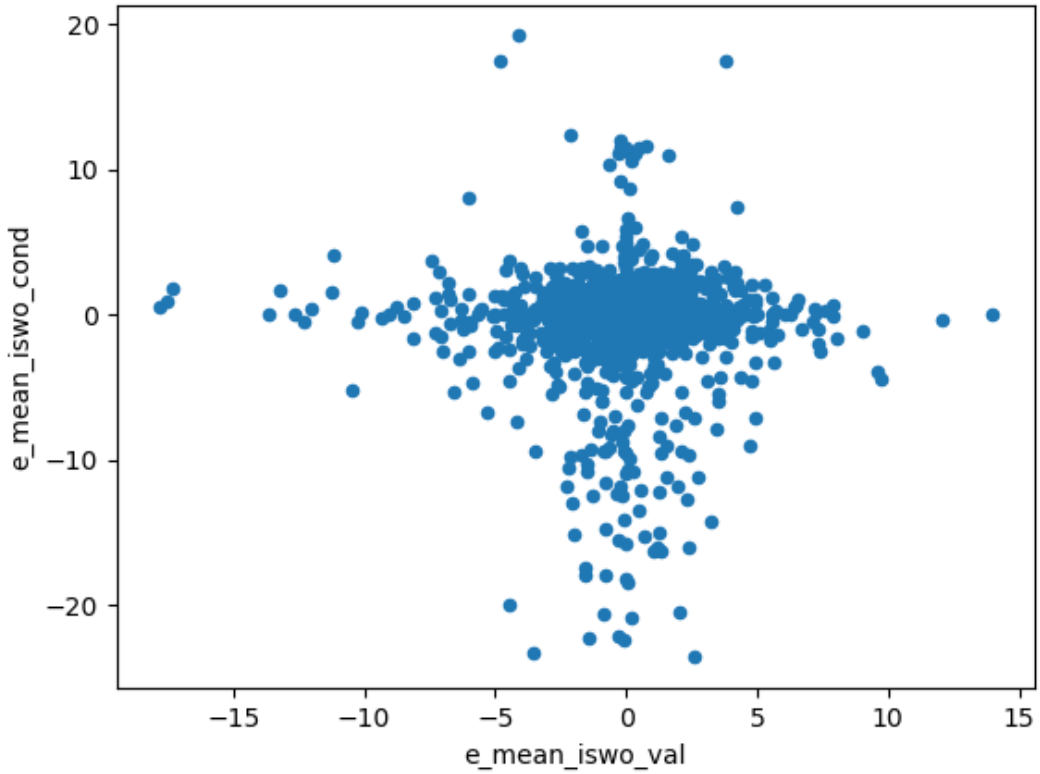
In [27]:

```
import copy
df_compare = copy.deepcopy(df[['composition', 'descriptor', 'dissociation_energy_per
df_compare['p_dissociation_energy_per_O'] = formation_model.predict(list(df_compare[
df_compare['e_dissociation_energy_per_O'] = df_compare['p_dissociation_energy_per_O'
df_compare['p_mean_iswo_cond'] = iswo_cond_model.predict(list(df_compare['descriptor
df_compare['e_mean_iswo_cond'] = df_compare['p_mean_iswo_cond'] - df_compare['mean_i
df_compare['p_mean_iswo_val'] = iswo_val_model.predict(list(df_compare['descriptor']
df_compare['e_mean_iswo_val'] = df_compare['p_mean_iswo_val'] - df_compare['mean_isw
df_compare['p_transport_gap'] = gap_model.predict(list(df_compare['descriptor']))
df_compare['e_transport_gap'] = df_compare['p_transport_gap'] - df_compare['transpor
df_compare.plot.scatter('e_transport_gap','e_mean_iswo_val' )
df_compare.plot.scatter('e_transport_gap','e_mean_iswo_cond' )
df_compare.plot.scatter('e_mean_iswo_val','e_mean_iswo_cond' )
```

```
/home/vsette63/anaconda3/envs/2022.4/lib/python3.7/site-packages/pandas/plotting/_ma
tplotlib/core.py:345: RuntimeWarning: More than 20 figures have been opened. Figures
created through the pyplot interface (`matplotlib.pyplot.figure`) are retained until
explicitly closed and may consume too much memory. (To control this warning, see the
rcParam `figure.max_open_warning`).
  fig = self.plt.figure(figsize=self.figsize)
```


15
10
5
0
−5
−10
−15
e_mean_iswo_val
−2
−1
0
1
2
3
e_transport_gap


20
10
0
−10
−20
e_mean_iswo_cond
−2
−1
0
1
2
3
e_transport_gap



Out[27]: <AxesSubplot:xlabel='e_mean_iswo_val', ylabel='e_mean_iswo_cond'>

corrrelations

In [28]:

```
stabilities = []
pd.set_option("precision", 2)


def _color_blue_or_red(val):
    color = 'red' if val < 0 else 'blue'
    return 'color: %s' % color

for r in df.itertuples():
    x = r.descriptor
    e = 0
    for i, y in enumerate(x):
        v = [0]*len(x)
        v[i] = 1
        e += y * formation_model.predict([v])

    stability = formation_model.predict([x]) - e

    stabilities.append(stability[0])

df['stability'] = stabilities

elements = ['Ag', 'Al', 'Cd', 'Ga', 'In', 'Mg', 'Sb', 'Si', 'Sn', 'Ti', 'Zn', 'Zr']
corr = df[['Ag', 'Al', 'Cd', 'Ga', 'In', 'Mg', 'Sb', 'Si', 'Sn', 'Ti', 'Zn', 'Zr',
           'skip_cond', 'skip_val', 'mean_iswo_cond','mean_iswo_val',
           'transport_gap', 'dissociation_energy_per_O', 'stability']].corr()
corr_sort = corr[['dissociation_energy_per_O', 'skip_cond', 'skip_val', 'mean_iswo_c
           'transport_gap','stability']].sort_values(by='dissociation_energy_per_O')
corr_sort[corr_sort.index.isin(elements)].style.applymap(_color_blue_or_red)
```

Out[28]:

	dissociation_energy_per_O	skip_cond	skip_val	mean_iswo_cond	mean_iswo_val	transport_gap
Zr	-0.51	0.17	0.10	0.35	0.20	0.24
Ti	-0.41	0.14	0.08	0.64	0.16	0.07
Al	-0.33	0.17	0.02	-0.08	0.14	0.42
Si	-0.19	0.15	0.10	-0.02	-0.06	0.33
Mg	-0.10	0.06	0.04	-0.12	0.39	0.14
Sn	-0.06	0.06	-0.07	-0.20	-0.61	-0.16
Ga	0.01	0.03	0.14	-0.14	0.03	0.19
In	0.12	-0.05	0.13	-0.25	0.26	0.04
Zn	0.12	-0.06	0.00	-0.18	-0.06	0.03
Sb	0.34	-0.08	0.05	-0.13	-0.19	-0.25
Ag	0.40	-0.24	-0.26	0.16	0.07	-0.48
Cd	0.50	-0.32	-0.33	-0.03	-0.27	-0.41

In [29]:

```
from itertools import combinations
import copy

corrs = []

for c in combinations(elements, 2):
    not_c = [e for e in elements if e not in c]
    #print(c, not_c)
    df_work = copy.deepcopy(df)
    for e in not_c:
        #print(e)
        df_work = df_work[df_work[e] ==  0]
    prop_list = ['dissociation_energy_per_O','mean_iswo_cond','mean_iswo_val', 'tran
    corr_list = list(c) + prop_list
    corr = df_work[corr_list].corr()
    corr = corr[corr.index.isin(c)]
    corrs.append(corr[prop_list])

corrs_df = pd.concat(corrs).reset_index()
corrs_df = corrs_df.groupby(by='index').agg('mean')
corrs_df.sort_values(by='dissociation_energy_per_O').style.applymap(_color_blue_or_r
print(corrs_df.sort_values(by='dissociation_energy_per_O').to_latex())
corrs_df.sort_values(by='dissociation_energy_per_O').style.applymap(_color_blue_or_r
```

```
\begin{tabular}{lrrrrrrr}
\toprule
{} &  dissociation\_energy\_per\_O &  mean\_iswo\_cond &  mean\_iswo\_val &  transpo
rt\_gap &  stability &  skip\_cond &  skip\_val \\
index &                              &                 &                &
&            &            &           \\
\midrule
Zr    &                       -0.98 &            0.74 &           0.44 &           0.
37 &      -0.06 &       0.21 &      0.12 \\
Ti    &                       -0.75 &            0.65 &           0.19 &           0.
07 &       0.07 &       0.06 &      0.19 \\
Al    &                       -0.59 &           -0.18 &           0.23 &           0.
65 &       0.18 &       0.21 &     -0.06 \\
Si    &                       -0.40 &           -0.03 &          -0.14 &           0.
```

```
81 &       0.28 &       0.55 &      0.37 \\
Mg    &                     -0.21 &           -0.41 &           0.80 &           0.
27 &       0.09 &       0.08 &     -0.02 \\
Sn    &                     -0.03 &           -0.14 &          -0.32 &          -0.
36 &      -0.09 &       0.10 &      0.04 \\
Ga    &                      0.04 &           -0.29 &          -0.06 &           0.
30 &      -0.11 &      -0.05 &      0.24 \\
In    &                      0.34 &           -0.58 &           0.48 &          -0.
16 &       0.05 &      -0.08 &      0.15 \\
Zn    &                      0.43 &           -0.51 &          -0.34 &          -0.
11 &       0.15 &      -0.09 &     -0.15 \\
Sb    &                      0.47 &            0.11 &          -0.49 &          -0.
37 &      -0.35 &      -0.18 &      0.11 \\
Ag    &                      0.73 &            0.37 &          -0.13 &          -0.
61 &      -0.36 &      -0.19 &     -0.39 \\
Cd    &                      0.96 &            0.28 &          -0.69 &          -0.
85 &       0.16 &      -0.62 &     -0.61 \\
\bottomrule
\end{tabular}
```

Out[29]:

index	dissociation_energy_per_O	mean_iswo_cond	mean_iswo_val	transport_gap	stability	skip_con
Zr	-0.98	0.74	0.44	0.37	-0.06	0.2
Ti	-0.75	0.65	0.19	0.07	0.07	0.0
Al	-0.59	-0.18	0.23	0.65	0.18	0.2
Si	-0.40	-0.03	-0.14	0.81	0.28	0.5
Mg	-0.21	-0.41	0.80	0.27	0.09	0.0
Sn	-0.03	-0.14	-0.32	-0.36	-0.09	0.1
Ga	0.04	-0.29	-0.06	0.30	-0.11	-0.0
In	0.34	-0.58	0.48	-0.16	0.05	-0.0
Zn	0.43	-0.51	-0.34	-0.11	0.15	-0.0
Sb	0.47	0.11	-0.49	-0.37	-0.35	-0.1
Ag	0.73	0.37	-0.13	-0.61	-0.36	-0.1
Cd	0.96	0.28	-0.69	-0.85	0.16	-0.6

finding optimal compositions

In [30]:

```python
def metric(element_vector):
    gap_delta_aim = 1
    iswo_cond_factor = 1
    iswo_val_factor = 0.2
    formation_model_factor = 1
    metric = (gap_model.predict([element_vector])[0] - 2.0 - gap_delta_aim)**2
    metric += iswo_cond_factor * iswo_cond_model.predict([element_vector])[0]
    metric += - iswo_val_factor * iswo_val_model.predict([element_vector])[0]
    return metric
```

In [31]:

```python
igzo_gap = gap_model.predict([composition_string_to_vector('In50Ga50Zn50O200')])[0]
igzo_iswo_c = iswo_cond_model.predict([composition_string_to_vector('In50Ga50Zn50O20
```

```
igzo_iswo_v = iswo_val_model.predict([composition_string_to_vector('In50Ga50Zn50O200
igzo_formation = formation_model.predict([composition_string_to_vector('In50Ga50Zn50
print(igzo_gap)
print(igzo_iswo_c)
print(igzo_iswo_v)
print(igzo_formation)
```

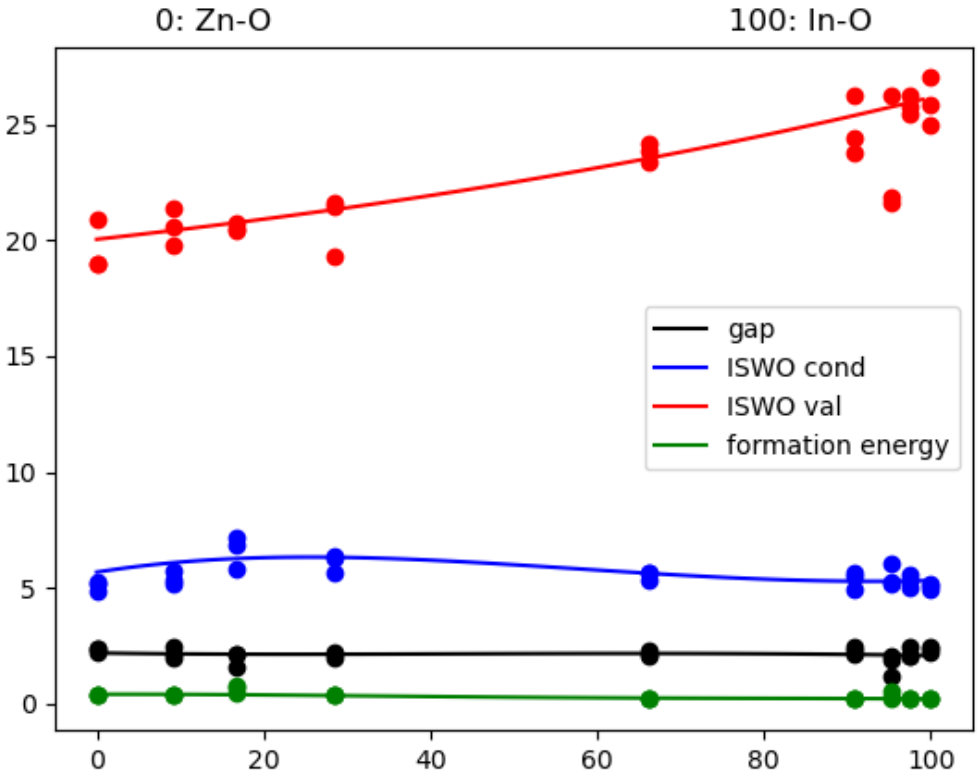
```
2.5579790682542263
5.8702958759611015
22.6086060861648
-0.025452122791994825
```

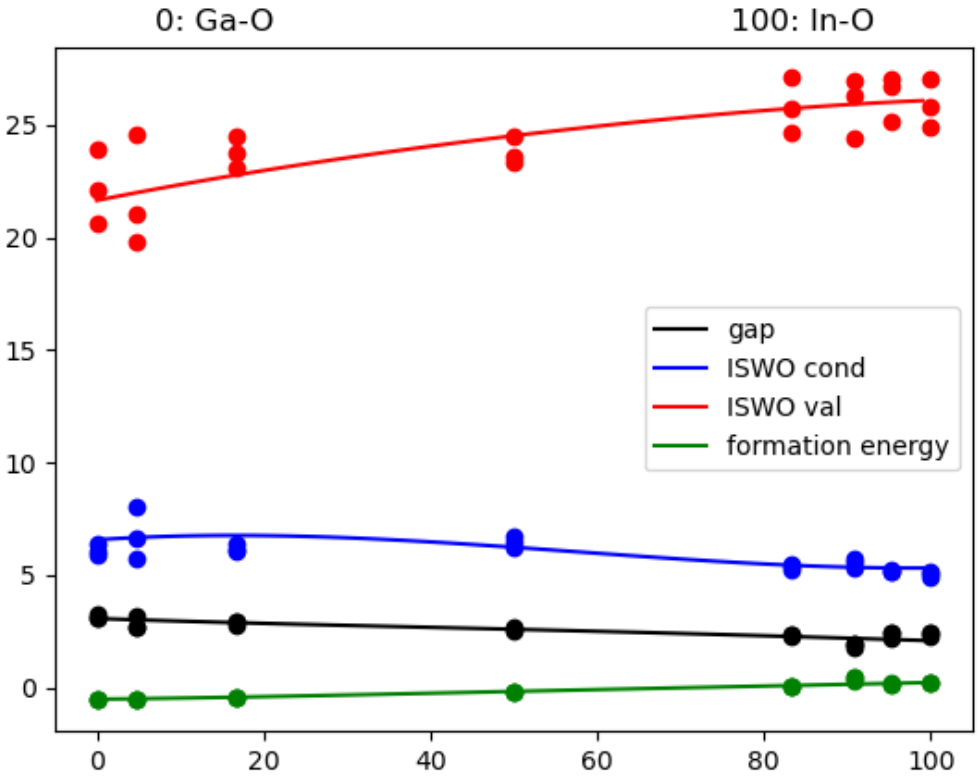
pairwise data comparison

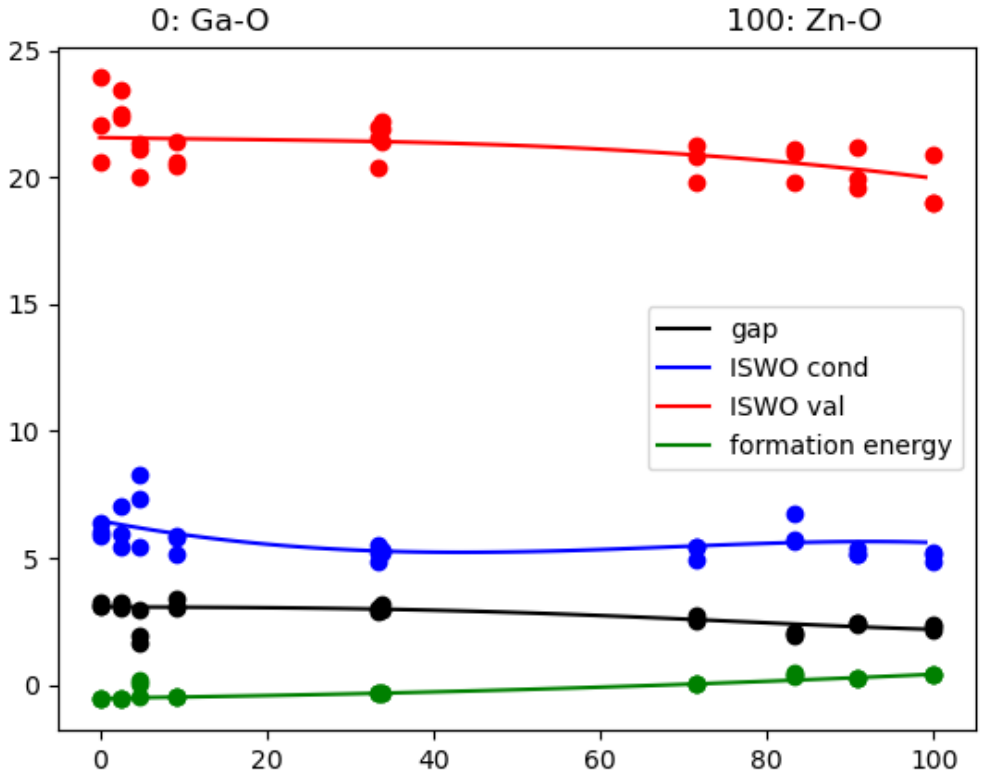
Comparison of the predictions to the actual data in the one dimensional subspaces of biniary oxides.

In [32]:

```
elems = ['Ag', 'Al', 'Cd', 'Ga', 'In', 'Mg', 'Sb', 'Sn', 'Si', 'Ti', 'Zn', 'Zr']
from itertools import combinations
n = 20
count = 0
for c in combinations(['In', 'Zn', 'Ga'], 2):
    fig, ax = plt.subplots(1)
    y = [gap_model.predict([composition_string_to_vector(f'{c[0]}{n}{c[1]}{100-n}')]
    ax.plot(y, label='gap', color = 'k')
    y = [iswo_cond_model.predict([composition_string_to_vector(f'{c[0]}{n}{c[1]}{100
    ax.plot(y, label='ISWO cond', color = 'b')
    y = [iswo_val_model.predict([composition_string_to_vector(f'{c[0]}{n}{c[1]}{100-
    ax.plot(y, label='ISWO val', color = 'r')
    y = [formation_model.predict([composition_string_to_vector(f'{c[0]}{n}{c[1]}{100
    ax.plot(y, label='formation energy', color = 'g')
    ax.set_title(f'0: {c[1]}-O                                                    100: {c
    test = np.ones(len(elems))
    x_data = []
    iswo_c_data = []
    iswo_v_data = []
    gap_data = []
    formation_data = []
    indices = [elems.index(e) for e in c]
    for i in indices:
        test[i] = 0
    for i in df.itertuples():
        if np.dot(np.array(i.descriptor), test) == 0:
            x = 100 * i.descriptor[indices[0]] / (i.descriptor[indices[0]] + i.descr
            x_data.append(x)
            gap_data.append(i.transport_gap)
            iswo_c_data.append(i.mean_iswo_cond)
            iswo_v_data.append(i.mean_iswo_val)
            formation_data.append(i.dissociation_energy_per_O)
    ax.scatter(x_data, iswo_c_data, color='b')
    ax.scatter(x_data, iswo_v_data, color='r')
    ax.scatter(x_data, gap_data, color='k')
    ax.scatter(x_data, formation_data, color='g')
    ax.legend()
    count += 1
    if count > n:
        break
```


0: Zn-O
100: In-O
gap
ISWO cond
ISWO val
formation energy
0
5
10
15
20
25
0
20
40
60
80
100


0: Ga-O
100: In-O
gap
ISWO cond
ISWO val
formation energy
0
5
10
15
20
25
0
20
40
60
80
100



store models

In [33]:

```
models = {'gap_model': gap_model, 'iswo_cond_model': iswo_cond_model,
          'iswo_val_model': iswo_val_model, 'formation_model': formation_model}

t = datetime.datetime.now()

path = f'models_{t.year}-{t.month}-{t.day}_{t.hour}-{t.minute}'
models_path = os.path.join(os.getcwd(), path)
os.makedirs(models_path)

for model in models:
    with open(os.path.join(models_path, f'{model}.pickle'), 'wb') as f:
        print(model)
        pickle.dump(models[model], f)
```

```
gap_model
iswo_cond_model
iswo_val_model
formation_model
```

predict optima

In [34]:

```
gap_delta_aims = [0, 0.25, 0.5, 0.75, 1]
iswo_cond_factors = [0.1, 0.2]
iswo_val_factors = [0.01, 0.02]
formation_factors = [0.0, 0.5, 1.0, 1.5]
elems = ['Ag', 'Al', 'Cd', 'Ga', 'In', 'Mg', 'Sb', 'Sn', 'Si', 'Ti', 'Zn', 'Zr']
active = elems

def get_formula(vec):
```

```
    form = ''
    for i, val in enumerate(vec):
        if val > 0.01:
            form += f'{elems[i]}{val:0.2f}'
    return form

def full_from_active(vec):
    element_vector = [0]*len(elems)
    for i, v in enumerate(vec):
        element_vector[elems.index(active[i])] = v
    return element_vector

solutions = []

for gap_delta_aim, iswo_cond_factor, iswo_val_factor, formation_factor in product(ga
    print(gap_delta_aim, iswo_cond_factor, iswo_val_factor, formation_factor)

    def metric(element_vector_in):
        element_vector = full_from_active(element_vector_in)
        metric = (gap_model.predict([element_vector])[0] - 2.0 - gap_delta_aim)**2
        metric += iswo_cond_factor * iswo_cond_model.predict([element_vector])[0]
        metric -= iswo_val_factor * iswo_val_model.predict([element_vector])[0]
        metric += formation_factor * formation_model.predict([element_vector])[0]
        return metric

    def constraint_fun(x):
        return(sum(x))

    import random

    best = None

    constraint = NonlinearConstraint(constraint_fun, 1, 1)

    for c in range(0,50):
        x0 = [random.random() for i in range(0,len(active))]
        x0 = [x / sum(x0) for x in x0]
        bounds = len(active)*[(0,1)]
        res = minimize(metric, x0=x0, bounds=bounds, constraints=constraint, method=
        if not res.success:
            continue
        if best is None:
            best = res
        elif res.fun < best.fun:
            best = res
            print(get_formula(full_from_active(res.x)), res.fun)

    best = dict(best)
    x = full_from_active(best['x'])
    best['jac'] = list(best['jac'])
    best['x'] = list(x)
    best['success'] = bool(best['success'])

    solution = {'formula': get_formula(x),
                'gap': round(float(gap_model.predict([x])), 2),
                'iswo_c': round(float(iswo_cond_model.predict([x])), 2),
                'iswo_v': round(float(iswo_val_model.predict([x])), 2),
                'formation': round(float(formation_model.predict([x])), 2),
                'res' : best,
                'target' : {'gap_delta_aim': gap_delta_aim, 'iswo_cond_factor': iswo
                            'iswo_val_factor': iswo_val_factor, 'formation_factor':
                'models' : {'gap_model': str(gap_model), 'iswo_cond_model': str(iswo
                            'iswo_val_model': str(iswo_val_model), 'formation_model'
                'models_path': models_path}
```

```
    e = 0
    for i, y in enumerate(x):
        v = [0]*len(x)
        v[i] = 1
        e += y * formation_model.predict([v])

    stability = formation_model.predict([x]) - e

    solution['stability'] = stability[0]

    solutions.append(solution)

    print(f"gap             : {gap_model.predict([x])[0]:1.2f}")
    print(f"iswo conduction : {iswo_cond_model.predict([x])[0]:1.2f}")
    print(f"iswo valence    : {iswo_val_model.predict([x])[0]:1.2f}")
    print(f"formation       : {formation_model.predict([x])[0]:1.2f}")
    print(f"stability       : {stability[0]:1.2f}")

    get_formula(x)

with open(os.path.join(models_path, 'solutions.json'), 'w') as f:
    json.dump(solutions, f, indent=2)
```

```
0 0.1 0.01 0.0
In1.00 0.2784521816610798
In1.00 0.27845218165706664
In1.00 0.2784521816516457
In1.00 0.27845218154610224
In1.00 0.27845218151787665
In1.00 0.2784521808420297
gap             : 2.09
iswo conduction : 5.31
iswo valence    : 26.08
formation       : 0.21
stability       : 0.00
0 0.1 0.01 0.5
Al0.71Sn0.29 -0.3687735482185699
Sn0.53Zr0.47 -0.4972320965481041
Sn0.53Zr0.47 -0.49723217022844113
gap             : 2.04
iswo conduction : 10.46
iswo valence    : 8.01
formation       : -2.93
stability       : -0.41
0 0.1 0.01 1.0
Sn0.43Zr0.57 -2.039420566221399
Sn0.43Zr0.57 -2.0394213800621896
gap             : 2.26
iswo conduction : 12.21
iswo valence    : 9.25
formation       : -3.23
stability       : -0.37
0 0.1 0.01 1.5
Sn0.33Zr0.67 -3.7279221282635753
Sn0.33Zr0.67 -3.7279222340308253
Sn0.33Zr0.67 -3.7279222897136126
Sn0.33Zr0.67 -3.7279222984798164
gap             : 2.46
iswo conduction : 14.40
iswo valence    : 11.26
formation       : -3.51
stability       : -0.30
0 0.1 0.02 0.0
```

```
In1.00 0.017634655617461248
In1.00 0.017634655566571844
gap             : 2.09
iswo conduction : 5.31
iswo valence    : 26.08
formation       : 0.21
stability       : 0.00
0 0.1 0.02 0.5
Al0.71Sn0.29 -0.5216052745434137
Sn0.53Zr0.47 -0.5775710610194147
Sn0.53Zr0.47 -0.5775710782792106
Sn0.53Zr0.47 -0.5775710859803733
gap             : 2.05
iswo conduction : 10.55
iswo valence    : 8.06
formation       : -2.95
stability       : -0.40
0 0.1 0.02 1.0
Sn0.42Zr0.58 -2.132949639187236
Sn0.42Zr0.58 -2.132949639189418
Sn0.42Zr0.58 -2.1329497959325416
gap             : 2.28
iswo conduction : 12.45
iswo valence    : 9.46
formation       : -3.27
stability       : -0.36
0 0.1 0.02 1.5
Ag0.05Sn0.13Zr0.82 -3.8734842091862536
Ag0.05Sn0.13Zr0.82 -3.873484222882431
gap             : 2.37
iswo conduction : 20.14
iswo valence    : 18.52
formation       : -3.77
stability       : -0.18
0 0.2 0.01 0.0
In1.00 0.8094632507141805
In1.00 0.8094632341011374
In1.00 0.8094632199947489
gap             : 2.09
iswo conduction : 5.31
iswo valence    : 26.08
formation       : 0.21
stability       : 0.00
0 0.2 0.01 0.5
Al0.46Mg0.16Sn0.38 0.5297321157012564
Al0.46Mg0.16Sn0.38 0.5297319518049157
Al0.46Mg0.16Sn0.38 0.5297319296177474
Al0.46Mg0.16Sn0.38 0.5297316546156188
Sn0.60Zr0.40 0.4956308597081598
Sn0.60Zr0.40 0.495630759275838
Sn0.60Zr0.40 0.4956307519747707
Sn0.60Zr0.40 0.4956307317557429
gap             : 1.88
iswo conduction : 9.53
iswo valence    : 7.58
formation       : -2.70
stability       : -0.42
0 0.2 0.01 1.0
Al0.73Sn0.27 -0.6137240036206699
Al0.73Sn0.27 -0.6137271567760956
Sn0.53Zr0.47 -0.9161585670839645
Sn0.53Zr0.47 -0.9161588817135984
Sn0.53Zr0.47 -0.9161590874093015
gap             : 2.05
```

```
iswo conduction : 10.52
iswo valence    : 8.05
formation       : -2.94
stability       : -0.41
0 0.2 0.01 1.5
Sn0.46Zr0.54 -2.442792273097615
Sn0.46Zr0.54 -2.4427923278928803
gap             : 2.21
iswo conduction : 11.72
iswo valence    : 8.87
formation       : -3.16
stability       : -0.38
0 0.2 0.02 0.0
gap             : 2.09
iswo conduction : 5.31
iswo valence    : 26.08
formation       : 0.21
stability       : 0.00
0 0.2 0.02 0.5
Al0.48Mg0.15Sn0.37 0.3864259513320296
Al0.48Mg0.15Sn0.37 0.3864257862911311
Al0.48Mg0.15Sn0.37 0.38642574645608097
Al0.48Mg0.15Sn0.37 0.3864254276857422
gap             : 2.21
iswo conduction : 8.27
iswo valence    : 14.43
formation       : -2.05
stability       : -0.11
0 0.2 0.02 1.0
Sn0.52Zr0.48 -0.9968099375122856
Sn0.52Zr0.48 -0.9968099429877377
Sn0.52Zr0.48 -0.9968104750753095
Sn0.52Zr0.48 -0.9968113417558391
gap             : 2.06
iswo conduction : 10.59
iswo valence    : 8.08
formation       : -2.96
stability       : -0.40
0 0.2 0.02 1.5
Sn0.45Zr0.55 -2.5319768363684614
Sn0.45Zr0.55 -2.5319768719761506
Sn0.45Zr0.55 -2.5319769090455964
Sn0.45Zr0.55 -2.5319769358481583
Sn0.45Zr0.55 -2.5319795347315663
gap             : 2.22
iswo conduction : 11.85
iswo valence    : 8.97
formation       : -3.18
stability       : -0.37
0.25 0.1 0.01 0.0
Ga0.09In0.91 0.2790827740236828
Ga0.09In0.91 0.2790827688678003
Ga0.09In0.91 0.2790827625528427
gap             : 2.19
iswo conduction : 5.35
iswo valence    : 25.90
formation       : 0.14
stability       : -0.01
0.25 0.1 0.01 0.5
Al0.75Sn0.25 -0.43212883301513716
Al0.75Sn0.25 -0.43212883531110347
Al0.75Sn0.25 -0.43212883650022427
Sn0.46Zr0.54 -0.4948058211794286
Sn0.46Zr0.54 -0.4948060604687252
```

```
gap             : 2.20
iswo conduction : 11.67
iswo valence    : 8.83
formation       : -3.15
stability       : -0.38
0.25 0.1 0.01 1.0
Sn0.36Zr0.64 -2.1446379458023532
Sn0.36Zr0.64 -2.1446379458555533
Sn0.36Zr0.64 -2.144638319111033
gap             : 2.41
iswo conduction : 13.79
iswo valence    : 10.66
formation       : -3.44
stability       : -0.32
0.25 0.1 0.01 1.5
Sn0.27Zr0.73 -3.9313693055717125
Sn0.27Zr0.73 -3.9313693130363525
Sn0.27Zr0.73 -3.9313693173788353
Sn0.27Zr0.73 -3.9313693189970635
gap             : 2.60
iswo conduction : 16.19
iswo valence    : 13.12
formation       : -3.69
stability       : -0.25
0.25 0.1 0.02 0.0
Ga0.08In0.92 0.019964605014389014
Ga0.08In0.92 0.019964587441563486
Ga0.08In0.92 0.019964586362792747
Ga0.08In0.92 0.019964585636698895
Ga0.08In0.92 0.019964564568322407
gap             : 2.19
iswo conduction : 5.35
iswo valence    : 25.94
formation       : 0.14
stability       : -0.00
0.25 0.1 0.02 0.5
Al0.76Sn0.24 -0.595044141651533
Al0.76Sn0.24 -0.5950441448049077
Al0.76Sn0.24 -0.5950441449617965
gap             : 2.39
iswo conduction : 9.42
iswo valence    : 16.33
formation       : -2.46
stability       : -0.06
0.25 0.1 0.02 1.0
Sn0.34Zr0.66 -2.2532225086274584
Sn0.34Zr0.66 -2.253222510290489
Sn0.34Zr0.66 -2.2532225103281145
Sn0.34Zr0.66 -2.253222514046051
Sn0.34Zr0.66 -2.2532225141478053
gap             : 2.45
iswo conduction : 14.20
iswo valence    : 11.06
formation       : -3.49
stability       : -0.31
0.25 0.1 0.02 1.5
Sn0.25Zr0.75 -4.065664117384116
gap             : 2.64
iswo conduction : 16.76
iswo valence    : 13.75
formation       : -3.75
stability       : -0.24
0.25 0.2 0.01 0.0
Ga0.06In0.91Zn0.03 0.8120989303401116
```

```
gap             : 2.17
iswo conduction : 5.32
iswo valence    : 25.71
formation       : 0.17
stability       : -0.01
0.25 0.2 0.01 0.5
Al0.48Mg0.19Sn0.33 0.43401841826176746
gap             : 2.45
iswo conduction : 7.90
iswo valence    : 15.44
formation       : -2.06
stability       : -0.11
0.25 0.2 0.01 1.0
Al0.77Sn0.23 -0.7263937523264665
Sn0.48Zr0.52 -0.906886818519645
Sn0.48Zr0.52 -0.9068868434218302
Sn0.48Zr0.52 -0.9068868455668158
Sn0.48Zr0.52 -0.906886883670484
gap             : 2.16
iswo conduction : 11.36
iswo valence    : 8.60
formation       : -3.10
stability       : -0.39
0.25 0.2 0.01 1.5
Al0.78Sn0.22 -1.9728578190561779
Al0.78Sn0.22 -1.9728580276948804
Sn0.41Zr0.59 -2.5099684812610725
Sn0.41Zr0.59 -2.5099685183439773
Sn0.41Zr0.59 -2.509968519151654
Sn0.41Zr0.59 -2.509968519739983
Sn0.41Zr0.59 -2.509968692591556
gap             : 2.31
iswo conduction : 12.73
iswo valence    : 9.70
formation       : -3.31
stability       : -0.35
0.25 0.2 0.02 0.0
Ga0.07In0.93 0.5535789550954417
Ga0.07In0.93 0.5535787547197056
gap             : 2.17
iswo conduction : 5.33
iswo valence    : 25.95
formation       : 0.16
stability       : -0.00
0.25 0.2 0.02 0.5
Al0.49Mg0.19Sn0.32 0.27884709718466505
Al0.49Mg0.19Sn0.32 0.2788470558448397
Al0.49Mg0.19Sn0.32 0.2788470539016854
Al0.49Mg0.19Sn0.32 0.2788467381681059
gap             : 2.47
iswo conduction : 7.93
iswo valence    : 15.60
formation       : -2.09
stability       : -0.11
0.25 0.2 0.02 1.0
Sn0.47Zr0.53 -0.9933037808804754
Sn0.47Zr0.53 -0.993303807135058
Sn0.47Zr0.53 -0.9933038173616997
Sn0.47Zr0.53 -0.9933040643934081
gap             : 2.18
iswo conduction : 11.48
iswo valence    : 8.68
formation       : -3.12
stability       : -0.38
```

```
0.25 0.2 0.02 1.5
Sn0.40Zr0.60 -2.6077958713418767
Sn0.40Zr0.60 -2.6077958735514613
Sn0.40Zr0.60 -2.6077959361647625
Sn0.40Zr0.60 -2.6077988353641723
gap             : 2.33
iswo conduction : 12.93
iswo valence    : 9.87
formation       : -3.34
stability       : -0.35
0.5 0.1 0.01 0.0
Ga0.10In0.66Mg0.15Zn0.09 0.294268367165692
Ga0.10In0.66Mg0.15Zn0.09 0.2942683597253882
Mg0.41Zn0.59 0.29101389540195266
Mg0.41Zn0.59 0.29101328938717674
gap             : 2.56
iswo conduction : 5.05
iswo valence    : 21.76
formation       : -0.40
stability       : -0.03
0.5 0.1 0.01 0.5
Al0.79Sn0.21 -0.4912948144069258
Al0.79Sn0.21 -0.49129491670156045
gap             : 2.62
iswo conduction : 9.30
iswo valence    : 17.21
formation       : -2.52
stability       : -0.05
0.5 0.1 0.01 1.0
Sn0.29Zr0.71 -2.1996567427089353
Sn0.29Zr0.71 -2.199656774373863
Sn0.29Zr0.71 -2.1996567871823736
Sn0.29Zr0.71 -2.1996567891989094
Sn0.29Zr0.71 -2.1996568032290735
Sn0.29Zr0.71 -2.199656817301824
gap             : 2.56
iswo conduction : 15.64
iswo valence    : 12.53
formation       : -3.64
stability       : -0.27
0.5 0.1 0.01 1.5
Sn0.20Zr0.80 -4.07861238026275
Sn0.20Zr0.80 -4.078612429710845
gap             : 2.74
iswo conduction : 18.11
iswo valence    : 15.29
formation       : -3.86
stability       : -0.20
0.5 0.1 0.02 0.0
Al0.02Ga0.10In0.71Mg0.16 0.035981894636108236
Al0.02Ga0.10In0.71Mg0.16 0.03598186211759946
Al0.02Ga0.10In0.71Mg0.16 0.03598185148259159
Al0.02Ga0.10In0.71Mg0.16 0.03598183050693837
Al0.02Ga0.10In0.71Mg0.16 0.03598182590400567
Al0.02Ga0.10In0.71Mg0.16 0.03598181312645188
gap             : 2.50
iswo conduction : 5.61
iswo valence    : 26.24
formation       : -0.14
stability       : 0.04
0.5 0.1 0.02 0.5
Al0.79Sn0.21 -0.6637734897142203
Al0.79Sn0.21 -0.663773523639877
Al0.79Sn0.21 -0.6637735238810398
```

```
Al0.79Sn0.21 -0.66377363486548
gap             : 2.63
iswo conduction : 9.28
iswo valence    : 17.28
formation       : -2.53
stability       : -0.05
0.5 0.1 0.02 1.0
Sn0.27Zr0.73 -2.3281079630941814
Sn0.27Zr0.73 -2.3281079655348176
Sn0.27Zr0.73 -2.328107996387274
gap             : 2.60
iswo conduction : 16.24
iswo valence    : 13.18
formation       : -3.70
stability       : -0.25
0.5 0.1 0.02 1.5
Sn0.18Zr0.82 -4.2353030379170775
Sn0.18Zr0.82 -4.23530305612128
Sn0.18Zr0.82 -4.235303058028784
Sn0.18Zr0.82 -4.235303058779923
Sn0.18Zr0.82 -4.235303059303547
Sn0.18Zr0.82 -4.235305640633477
gap             : 2.79
iswo conduction : 18.77
iswo valence    : 16.07
formation       : -3.92
stability       : -0.18
0.5 0.2 0.01 0.0
Ga0.10In0.64Mg0.14Zn0.13 0.8409778702329669
Al0.02Mg0.37Zn0.60 0.7923520037536482
Al0.02Mg0.37Zn0.60 0.792351879589624
Al0.02Mg0.37Zn0.60 0.7923517966949499
gap             : 2.61
iswo conduction : 4.98
iswo valence    : 21.60
formation       : -0.41
stability       : -0.05
0.5 0.2 0.01 0.5
Al0.48Mg0.24Sn0.28 0.33287588097326926
Al0.48Mg0.24Sn0.28 0.3328756963825823
Al0.48Mg0.24Sn0.28 0.3328756599611671
Al0.48Mg0.24Sn0.28 0.3328756452058641
Al0.48Mg0.24Sn0.28 0.33287525948478813
gap             : 2.71
iswo conduction : 7.51
iswo valence    : 16.75
formation       : -2.09
stability       : -0.11
0.5 0.2 0.01 1.0
Al0.80Sn0.20 -0.8325502514590333
Al0.80Sn0.20 -0.8325502568760241
Al0.80Sn0.20 -0.8325505118963801
gap             : 2.71
iswo conduction : 9.18
iswo valence    : 17.55
formation       : -2.54
stability       : -0.05
0.5 0.2 0.01 1.5
Sn0.36Zr0.64 -2.5049630883926324
gap             : 2.42
iswo conduction : 13.86
iswo valence    : 10.74
formation       : -3.45
stability       : -0.32
```

```
0.5 0.2 0.02 0.0
Al0.02Mg0.41Zn0.57 0.5749019667834305
Al0.02Mg0.41Zn0.57 0.5749019611317607
gap             : 2.62
iswo conduction : 4.99
iswo valence    : 21.90
formation       : -0.46
stability       : -0.04
0.5 0.2 0.02 0.5
Al0.13Mg0.81Sb0.06 0.12566894204163215
gap             : 2.89
iswo conduction : 6.20
iswo valence    : 27.75
formation       : -1.42
stability       : 0.04
0.5 0.2 0.02 1.0
Al0.80Sn0.19 -1.008428867902223
Al0.80Sn0.19 -1.008428893683202
Al0.80Sn0.19 -1.0084288945755147
Al0.80Sn0.19 -1.0084289014306123
Al0.80Sn0.19 -1.0084292621393858
Al0.80Sn0.19 -1.008429600065921
Al0.80Sn0.19 -1.0084296884381954
gap             : 2.73
iswo conduction : 9.19
iswo valence    : 17.62
formation       : -2.55
stability       : -0.05
0.5 0.2 0.02 1.5
Sn0.34Zr0.66 -2.613671743065199
gap             : 2.44
iswo conduction : 14.15
iswo valence    : 11.01
formation       : -3.48
stability       : -0.31
0.75 0.1 0.01 0.0
Ga0.11In0.45Mg0.27Zn0.16 0.28882937233138317
Al0.03Ga0.04Mg0.42Zn0.51 0.2618658659933507
Al0.03Ga0.04Mg0.42Zn0.51 0.2618657282186357
gap             : 2.81
iswo conduction : 4.82
iswo valence    : 22.39
formation       : -0.56
stability       : -0.06
0.75 0.1 0.01 0.5
Al0.79Mg0.03Sn0.18 -0.5483122772101232
Al0.79Mg0.03Sn0.18 -0.5483122868247011
Al0.79Mg0.03Sn0.18 -0.548312289716524
gap             : 2.87
iswo conduction : 8.86
iswo valence    : 18.09
formation       : -2.53
stability       : -0.06
0.75 0.1 0.01 1.0
Sn0.21Zr0.79 -2.205454382323347
Sn0.21Zr0.79 -2.205454382365028
Sn0.21Zr0.79 -2.205454383026409
Sn0.21Zr0.79 -2.2054543873556205
gap             : 2.72
iswo conduction : 17.78
iswo valence    : 14.91
formation       : -3.84
stability       : -0.21
0.75 0.1 0.01 1.5
```

```
Sn0.14Zr0.86 -4.172863739959609
Sn0.14Zr0.86 -4.172863747025374
gap             : 2.89
iswo conduction : 20.17
iswo valence    : 17.77
formation       : -4.02
stability       : -0.14
0.75 0.1 0.02 0.0
Al0.04Ga0.08In0.54Mg0.30Zn0.04 0.03089708491781007
Al0.04Ga0.08In0.54Mg0.30Zn0.04 0.030897066147116403
Al0.04Ga0.08In0.54Mg0.30Zn0.04 0.030897063389623725
Al0.04Ga0.08In0.54Mg0.30Zn0.04 0.03089705991558256
Al0.04Ga0.08In0.54Mg0.30Zn0.04 0.0308970509053641
gap             : 2.77
iswo conduction : 5.60
iswo valence    : 26.49
formation       : -0.40
stability       : 0.05
0.75 0.1 0.02 0.5
Ag0.07Al0.93 -0.7184344128217588
Al0.79Mg0.03Sn0.18 -0.7296550054388857
Al0.79Mg0.02Sn0.18 -0.7296552757835442
Al0.79Mg0.03Sn0.18 -0.7296554417667455
Al0.79Mg0.03Sn0.18 -0.7296554805660265
Al0.79Mg0.03Sn0.18 -0.7296555486045032
Al0.79Mg0.02Sn0.18 -0.7296557747319453
gap             : 2.88
iswo conduction : 8.90
iswo valence    : 18.17
formation       : -2.55
stability       : -0.05
0.75 0.1 0.02 1.0
Sn0.19Zr0.81 -2.358992188811368
gap             : 2.77
iswo conduction : 18.56
iswo valence    : 15.81
formation       : -3.90
stability       : -0.19
0.75 0.1 0.02 1.5
Sn0.12Zr0.88 -4.3549859178793
gap             : 2.94
iswo conduction : 20.88
iswo valence    : 18.65
formation       : -4.07
stability       : -0.12
0.75 0.2 0.01 0.0
Al0.03Ga0.06Mg0.38Zn0.52 0.740824118731755
Al0.03Ga0.06Mg0.38Zn0.52 0.7408240747209744
Al0.03Ga0.06Mg0.38Zn0.52 0.7408227897342732
Al0.03Ga0.06Mg0.38Zn0.52 0.7408207959976588
gap             : 2.85
iswo conduction : 4.76
iswo valence    : 22.19
formation       : -0.54
stability       : -0.08
0.75 0.2 0.01 0.5
Al0.46Mg0.31Sn0.23 0.22349423440531546
Al0.17Mg0.79Sb0.04 0.21408777093826892
Al0.17Mg0.79Sb0.04 0.21408776850990807
Al0.17Mg0.79Sb0.04 0.2140876928827553
Al0.17Mg0.79Sb0.04 0.21408721258777175
gap             : 3.12
iswo conduction : 5.76
iswo valence    : 28.11
```

```
formation       : -1.59
stability       : -0.02
0.75 0.2 0.01 1.0
Al0.77Mg0.05Sn0.18 -0.9390987755445199
Al0.77Mg0.05Sn0.18 -0.9390989021015665
Al0.77Mg0.05Sn0.18 -0.9390992625692312
Al0.77Mg0.05Sn0.18 -0.9390993273596215
gap             : 2.97
iswo conduction : 8.59
iswo valence    : 18.45
formation       : -2.52
stability       : -0.06
0.75 0.2 0.01 1.5
Sn0.31Zr0.69 -2.4271127360229765
Sn0.31Zr0.69 -2.42711274165798
Sn0.31Zr0.69 -2.4271127417363663
gap             : 2.52
iswo conduction : 15.15
iswo valence    : 12.02
formation       : -3.59
stability       : -0.28
0.75 0.2 0.02 0.0
Ga0.12In0.40Mg0.30Zn0.19 0.5759544922245876
Ga0.12In0.40Mg0.30Zn0.19 0.5759544236496
Al0.02Ga0.06Mg0.42Zn0.49 0.5172917871596776
gap             : 2.86
iswo conduction : 4.77
iswo valence    : 22.52
formation       : -0.58
stability       : -0.07
0.75 0.2 0.02 0.5
Al0.16Mg0.80Sb0.04 -0.06769201567358862
Al0.16Mg0.80Sb0.04 -0.06769203804591573
Al0.16Mg0.80Sb0.04 -0.06769211965770605
gap             : 3.13
iswo conduction : 5.75
iswo valence    : 28.24
formation       : -1.60
stability       : -0.03
0.75 0.2 0.02 1.0
Al0.77Mg0.05Sn0.18 -1.1240095891389414
Al0.77Mg0.05Sn0.18 -1.1240098887766579
Al0.77Mg0.05Sn0.18 -1.1240109454237222
Al0.77Mg0.05Sn0.18 -1.124011626135546
gap             : 2.99
iswo conduction : 8.57
iswo valence    : 18.53
formation       : -2.52
stability       : -0.06
0.75 0.2 0.02 1.5
Al0.84Sn0.16 -2.4248659209372807
Sn0.29Zr0.71 -2.5493369802913897
Sn0.29Zr0.71 -2.5493370144170897
Sn0.29Zr0.71 -2.5493370174739565
Sn0.29Zr0.71 -2.5493370175530137
Sn0.29Zr0.71 -2.549337110986845
Sn0.29Zr0.71 -2.549340523202282
gap             : 2.55
iswo conduction : 15.54
iswo valence    : 12.43
formation       : -3.63
stability       : -0.27
1 0.1 0.01 0.0
Al0.04Ga0.09Mg0.43Zn0.43 0.2342907514393039
```

```
Al0.04Ga0.09Mg0.43Zn0.43 0.2342901272609481
gap             : 3.05
iswo conduction : 4.62
iswo valence    : 23.02
formation       : -0.72
stability       : -0.09
1 0.1 0.01 0.5
Al0.75Mg0.08Sn0.17 -0.6076572091293975
gap             : 3.12
iswo conduction : 8.23
iswo valence    : 19.05
formation       : -2.51
stability       : -0.07
1 0.1 0.01 1.0
Sn0.14Zr0.86 -2.1699349109992614
Sn0.14Zr0.86 -2.169934995493538
Sn0.14Zr0.86 -2.1699349962065053
Sn0.14Zr0.86 -2.1699350228877052
Sn0.14Zr0.86 -2.1699350290023394
gap             : 2.90
iswo conduction : 20.26
iswo valence    : 17.88
formation       : -4.03
stability       : -0.14
1 0.1 0.01 1.5
Sn0.08Zr0.92 -4.222775773079165
Sn0.08Zr0.92 -4.222775836077663
Sn0.08Zr0.92 -4.22277591866991
Sn0.08Zr0.92 -4.2227759247561325
Sn0.08Zr0.92 -4.222775925050442
Sn0.08Zr0.92 -4.22277592648236
gap             : 3.06
iswo conduction : 22.38
iswo valence    : 20.57
formation       : -4.17
stability       : -0.08
1 0.1 0.02 0.0
Al0.07Mg0.59Zn0.34 -0.0015791270918306677
Al0.07Mg0.59Zn0.34 -0.0015791643413713974
gap             : 3.07
iswo conduction : 4.79
iswo valence    : 24.30
formation       : -1.02
stability       : -0.10
1 0.1 0.02 0.5
Ag0.02Al0.23Mg0.75 -0.8176065808997205
Ag0.02Al0.23Mg0.75 -0.81760659517433
gap             : 3.26
iswo conduction : 6.12
iswo valence    : 29.01
formation       : -1.83
stability       : -0.13
1 0.1 0.02 1.0
gap             : 2.96
iswo conduction : 21.13
iswo valence    : 18.96
formation       : -4.09
stability       : -0.12
1 0.1 0.02 1.5
Sn0.06Zr0.94 -4.433086899340425
Sn0.06Zr0.94 -4.433086912527972
Sn0.06Zr0.94 -4.4330869212628015
Sn0.06Zr0.94 -4.4330869213843345
Sn0.06Zr0.94 -4.43308692378341
```

```
Sn0.06Zr0.94 -4.433087542957352
Sn0.06Zr0.94 -4.433087930942776
gap             : 3.12
iswo conduction : 23.08
iswo valence    : 21.49
formation       : -4.22
stability       : -0.07
1 0.2 0.01 0.0
Al0.05Ga0.11Mg0.39Zn0.45 0.6926440492881674
Al0.05Ga0.11Mg0.39Zn0.45 0.6926432777728493
Al0.05Ga0.11Mg0.39Zn0.45 0.6926424166867793
Al0.05Ga0.11Mg0.39Zn0.45 0.6926424000737751
gap             : 3.09
iswo conduction : 4.56
iswo valence    : 22.77
formation       : -0.69
stability       : -0.10
1 0.2 0.01 0.5
Al0.20Mg0.78Sb0.02 0.02931249777228606
Al0.20Mg0.78Sb0.02 0.02931243888994217
Al0.20Mg0.78Sb0.02 0.029310740874692875
gap             : 3.37
iswo conduction : 5.33
iswo valence    : 28.62
formation       : -1.77
stability       : -0.09
1 0.2 0.01 1.0
gap             : 3.23
iswo conduction : 7.95
iswo valence    : 19.50
formation       : -2.50
stability       : -0.08
1 0.2 0.01 1.5
gap             : 3.24
iswo conduction : 8.87
iswo valence    : 19.46
formation       : -2.66
stability       : -0.03
1 0.2 0.02 0.0
Al0.05Ga0.09Mg0.44Zn0.42 0.46312796556663044
Al0.05Ga0.09Mg0.44Zn0.42 0.46312786492991453
Al0.05Ga0.09Mg0.44Zn0.42 0.4631275228447868
Al0.05Ga0.09Mg0.44Zn0.42 0.4631260820110994
Al0.05Ga0.09Mg0.44Zn0.42 0.46312593479601954
gap             : 3.10
iswo conduction : 4.58
iswo valence    : 23.15
formation       : -0.76
stability       : -0.10
1 0.2 0.02 0.5
Al0.19Mg0.79Sb0.01 -0.25745078017532363
Al0.19Mg0.79Sb0.01 -0.25745252284531417
gap             : 3.38
iswo conduction : 5.33
iswo valence    : 28.73
formation       : -1.78
stability       : -0.10
1 0.2 0.02 1.0
Al0.73Mg0.12Sn0.15 -1.2461569486910222
Al0.73Mg0.12Sn0.15 -1.2461572537639263
Al0.73Mg0.12Sn0.15 -1.2461572711104596
Al0.73Mg0.12Sn0.15 -1.246158937955896
gap             : 3.25
iswo conduction : 7.88
```

```
iswo valence    : 19.60
formation       : -2.49
stability       : -0.08
1 0.2 0.02 1.5
Al0.87Sn0.13 -2.5547182205272305
Al0.87Sn0.13 -2.554718221911564
Al0.87Sn0.13 -2.5547182227016543
Al0.87Sn0.13 -2.5547189770038448
Al0.87Sn0.13 -2.5547215952955247
gap             : 3.25
iswo conduction : 8.86
iswo valence    : 19.51
formation       : -2.67
stability       : -0.03
```

In [35]:

```
from itertools import combinations
#elems_for_test = ['In', 'Ga', 'Zn', 'Mg', 'Al']
elems_for_test = ['Ag', 'Al', 'Cd', 'Ga', 'In', 'Mg', 'Sb', 'Sn', 'Si', 'Ti', 'Zn',
#elems_for_test = ['Al', 'Cd', 'Ga', 'In', 'Mg', 'Sb', 'Sn', 'Si', 'Zn']

degree = 10
epsilon = 0.001
C=400
kernel='poly'

data = {}

for pair in combinations(elems_for_test, 2):
    df_pair = df[1 - abs(df[pair[0]] + df[pair[1]]) < 0.0001 ]
    if pair[0] == 0 or pair[1] == 0:
        continue
    data[pair] = {}
    for observable in ['transport_gap', 'dissociation_energy_per_O', 'mean_iswo_cond
        mads = []
        for c, df_test in df_pair.groupby('composition'):
            df_training = df_pair[df_pair['composition'] != c]
            model = get_model(df_training, observable, SVR, kernel=kernel, C=C, gamm
                              epsilon=epsilon, coef0=1, for_cv=True)
            data_org = df_test[observable]
            data_predicted = model.predict(list(df_test['descriptor']))
            diff = abs(data_predicted - data_org)
#            print(c, np.mean(diff))
            mads.append(np.mean(diff))
        data[pair][observable] =np.mean(mads)
cv_df = pd.DataFrame(data).T
cv_df
```

Out[35]:

		transport_gap	dissociation_energy_per_O	mean_iswo_cond	mean_iswo_val
Ag	**Al**	1.33	0.52	4.98	1.39
	Cd	0.05	0.23	1.30	1.91
	Ga	1.11	0.43	6.74	1.85
	In	0.39	0.61	4.01	2.07
	Mg	0.66	0.04	3.39	2.32
	Sb	0.29	0.58	1.09	2.24
	Sn	0.16	0.27	1.17	2.82
	Si	0.37	1.25	1.64	1.66

		transport_gap	dissociation_energy_per_O	mean_iswo_cond	mean_iswo_val
	Ti	0.69	1.22	4.48	2.73
	Zn	0.41	0.06	2.67	1.08
	Zr	0.56	0.07	0.89	1.71
Al	**Cd**	0.51	0.31	1.96	2.13
	Ga	0.55	0.16	1.13	1.52
	In	0.40	0.14	0.72	1.66
	Mg	0.61	0.12	0.54	1.55
	Sb	0.52	0.47	1.85	2.79
	Sn	0.38	0.39	1.24	4.28
	Si	0.34	0.05	0.92	1.81
	Ti	0.73	0.05	6.22	0.95
	Zn	0.41	0.08	0.61	1.06
	Zr	0.31	0.06	6.61	1.49
Cd	**Ga**	0.26	0.41	1.92	1.89
	In	0.24	0.09	1.42	1.82
	Mg	0.48	0.05	3.29	0.92
	Sb	0.47	0.65	1.23	2.21
	Sn	0.36	0.31	1.15	2.85
	Si	0.87	0.43	2.57	1.68
	Ti	0.49	0.32	2.53	2.74
	Zn	0.44	0.06	1.71	0.85
	Zr	0.28	0.19	2.13	1.53
Ga	**In**	0.31	0.10	0.45	1.16
	Mg	0.35	0.05	0.53	1.75
	Sb	0.47	0.46	0.89	2.17
	Sn	0.34	0.29	0.88	2.90
	Si	0.34	0.08	1.30	2.03
	Ti	0.38	0.15	3.79	1.41
	Zn	0.24	0.09	0.44	0.96
	Zr	0.34	0.16	1.57	1.55
In	**Mg**	0.24	0.06	0.42	1.36
	Sb	0.65	0.85	0.98	4.56
	Sn	0.50	0.39	0.58	5.79
	Si	0.36	0.08	1.71	2.27
	Ti	0.20	0.05	5.90	1.05
	Zn	0.18	0.07	0.42	1.24

		transport_gap	dissociation_energy_per_O	mean_iswo_cond	mean_iswo_val
	Zr	0.44	0.10	1.51	0.83
Mg	Sb	0.39	0.49	1.12	3.30
	Sn	0.41	0.34	0.96	4.79
	Si	0.27	0.06	0.87	2.12
	Ti	0.36	0.12	6.22	1.36
	Zn	0.39	0.11	0.57	1.42
	Zr	0.31	0.08	1.62	2.48
Sb	Sn	0.41	0.53	0.79	3.20
	Si	0.76	0.75	1.75	2.37
	Ti	0.59	0.71	4.39	3.53
	Zn	0.58	0.78	1.45	2.78
	Zr	0.52	0.62	2.12	2.67
Sn	Si	0.27	0.32	1.65	2.83
	Ti	0.45	0.45	1.29	5.40
	Zn	0.44	0.49	0.52	6.24
	Zr	0.37	0.33	1.43	3.43
Si	Ti	1.00	0.19	5.67	1.78
	Zn	0.29	0.06	1.01	1.30
	Zr	0.51	0.09	3.89	1.86
Ti	Zn	0.20	0.08	4.72	1.02
	Zr	0.18	0.04	2.91	1.30
Zn	Zr	0.39	0.11	1.99	1.08

In [36]:
```
cv_df.describe()
```

Out[36]:

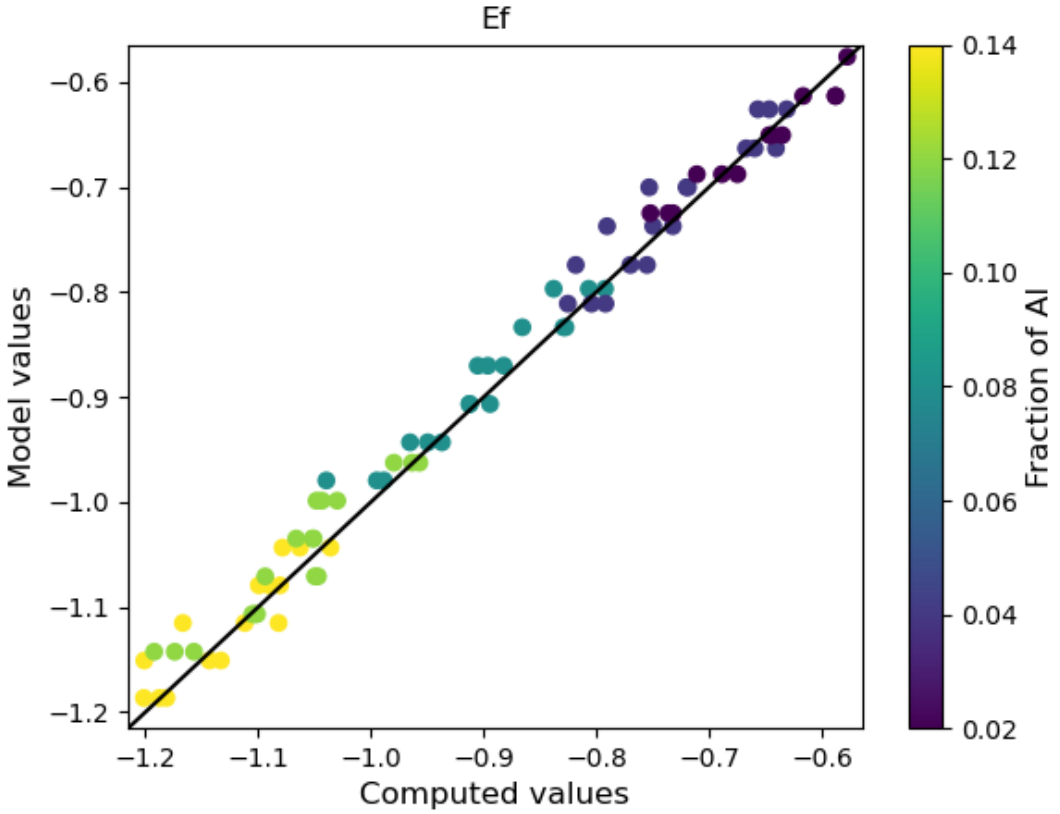
	transport_gap	dissociation_energy_per_O	mean_iswo_cond	mean_iswo_val
count	66.00	66.00	66.00	66.00
mean	0.44	0.29	2.13	2.19
std	0.22	0.28	1.73	1.18
min	0.05	0.04	0.42	0.83
25%	0.31	0.08	0.93	1.39
50%	0.39	0.17	1.48	1.86
75%	0.51	0.45	2.64	2.74
max	1.33	1.25	6.74	6.24

validation of the Mg-Zn-Al system

In [37]:
```
with open('MgZnAlO.json', 'r') as f:
    MgZnAl_data = json.load(f)

ens = []

for x in MgZnAl_data:
    energy = MgZnAl_data[x].get('total_energy_per_O', None)
    comp = MgZnAl_data[x].get('composition')
    ens.append([x, energy, comp])

selected = []

for name, group in pd.DataFrame(ens).groupby(by=2):
    selected += list(group.sort_values(by=1).head(3)[0])
```

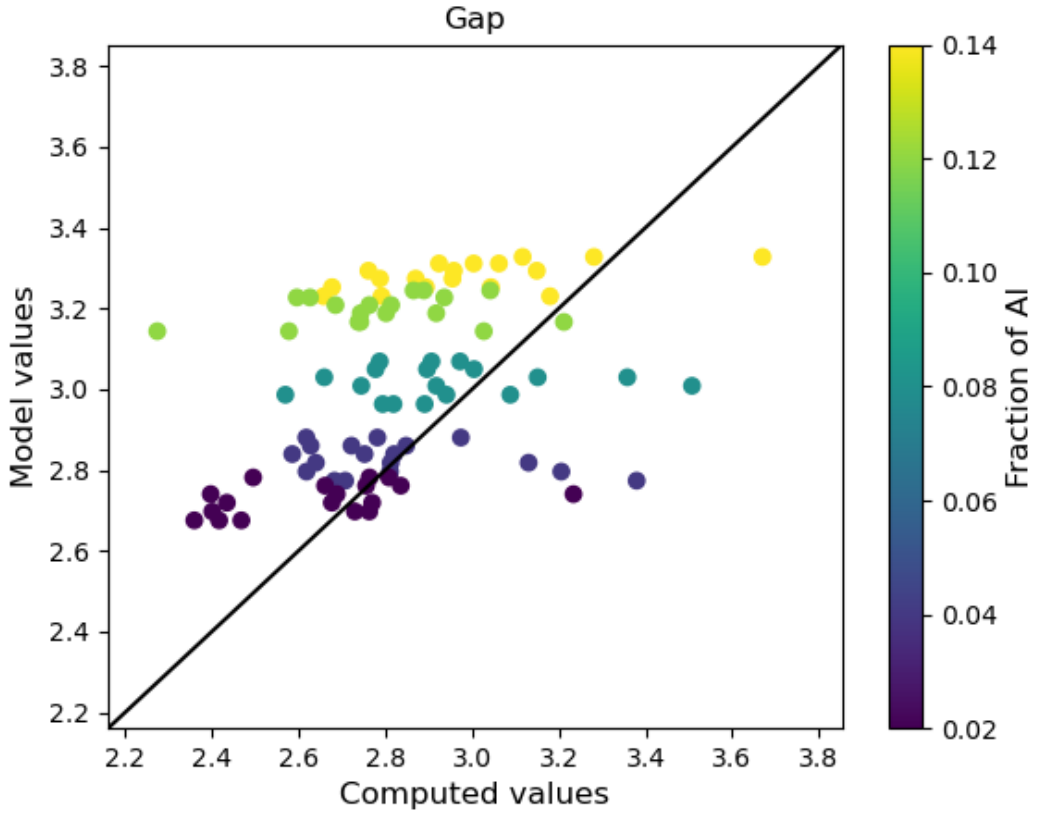
In [38]:
```
model_ef, model_ic, model_iv, model_g = [], [], [], []
calc_ef, calc_ic, calc_iv, calc_g = [], [], [], []
al, mg, zn = [], [], []

for x in selected:
#    print(MgZnAl_data[x]['Zn'], MgZnAl_data[x]['Mg'], MgZnAl_data[x]['Al'])
    v = [0]*len(elems)
    for el in ['Zn', 'Mg', 'Al']:
        v[elems.index(el)] = MgZnAl_data[x][el]

#    print(v)
#    print(formation_model.predict([v])[0], iswo_cond_model.predict([v])[0], iswo_va
#    print(MgZnAl_data[x]['dissociation_energy_per_O'], MgZnAl_data[x]['mean_iswo_co
#          MgZnAl_data[x]['mean_iswo_val'], MgZnAl_data[x]['transport_gap'])
    model_ef.append(formation_model.predict([v])[0])
    model_ic.append(iswo_cond_model.predict([v])[0])
    model_iv.append(iswo_val_model.predict([v])[0])
    model_g.append(gap_model.predict([v])[0])
    al.append(MgZnAl_data[x]['Al'])
    mg.append(MgZnAl_data[x]['Mg'])
    zn.append(MgZnAl_data[x]['Zn'])
    calc_ef.append(MgZnAl_data[x]['dissociation_energy_per_O'])
    calc_ic.append(MgZnAl_data[x]['mean_iswo_cond'])
    calc_iv.append(MgZnAl_data[x]['mean_iswo_val'])
    calc_g.append(MgZnAl_data[x]['transport_gap'])
```

In [39]:
```
fig, ax = plt.subplots()
p = ax.scatter(calc_ef, model_ef, c=al)
mi = min(calc_ef + model_ef) * 0.95
ma = max(calc_ef + model_ef) * 1.05
ax.plot([-5,30], [-5,30], c = 'k')
ax.set_xlim(mi, ma)
ax.set_ylim(mi, ma)
ax.set_title('Ef')
ax.set_xlabel('Computed values', size=12)
ax.set_ylabel('Model values', size=12)
cb = fig.colorbar(p, ax=ax)
cb.set_label('Fraction of Al', size=12)
```



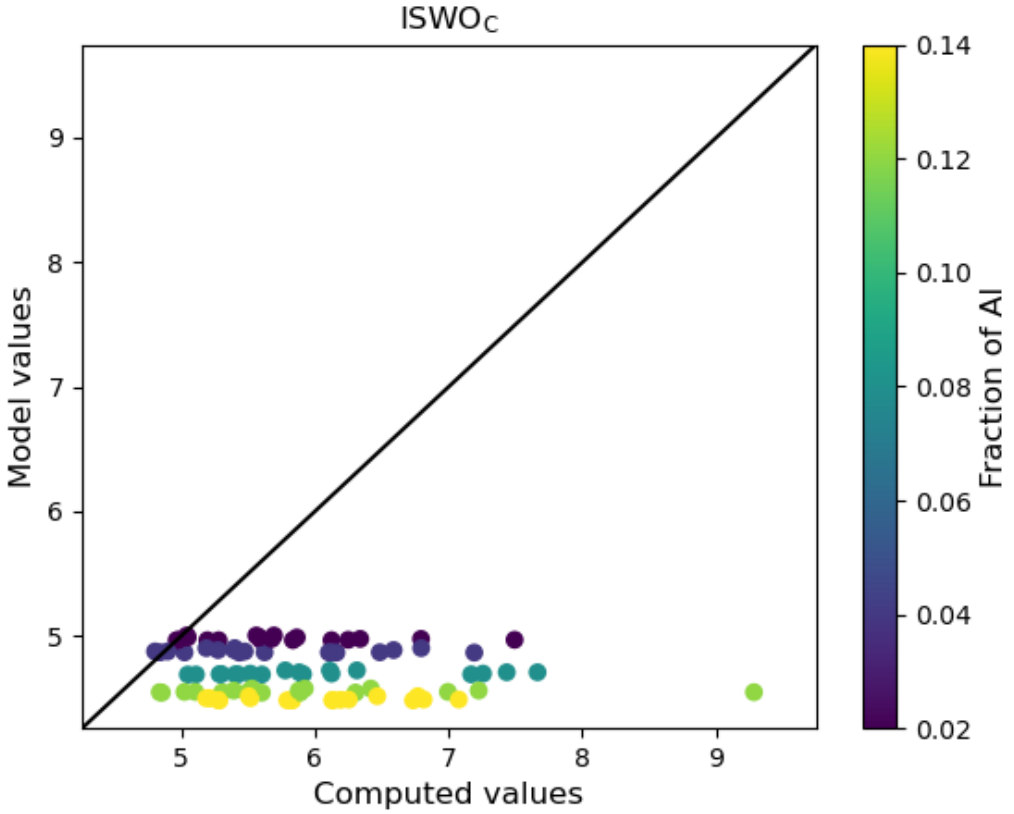
In [40]:

```
fig, ax = plt.subplots()
p = ax.scatter(calc_g, model_g, c=al)

mi = min(calc_g + model_g) * 0.95
ma = max(calc_g + model_g) * 1.05
ax.plot([-5,30], [-5,30], c = 'k')
ax.set_xlim(mi, ma)
ax.set_ylim(mi, ma)
ax.set_title('Gap')
ax.set_xlabel('Computed values', size=12)
ax.set_ylabel('Model values', size=12)
cb = fig.colorbar(p, ax=ax)
cb.set_label('Fraction of Al', size=12)
```



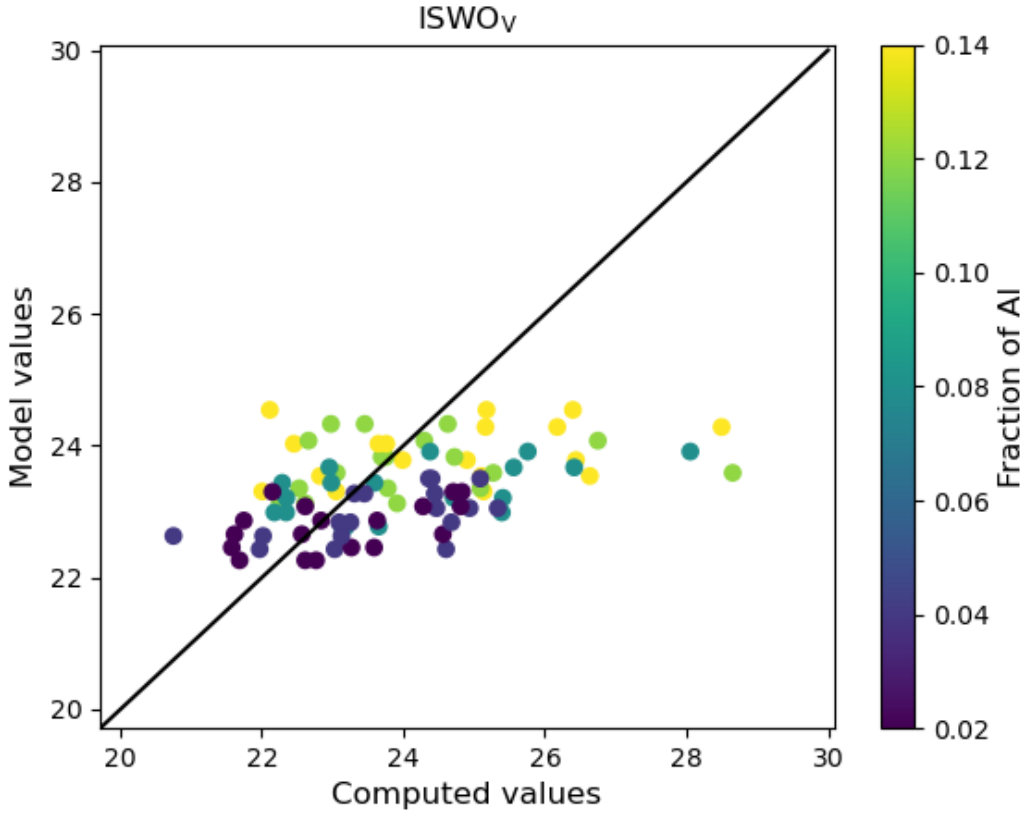
In [41]:

```
fig, ax = plt.subplots()
p = ax.scatter(calc_ic, model_ic, c=al)
mi = min(calc_ic + model_ic) * 0.95
ma = max(calc_ic + model_ic) * 1.05
ax.plot([-5,30], [-5,30], c = 'k')
ax.set_xlim(mi, ma)
ax.set_ylim(mi, ma)
ax.set_title('ISWO$_\mathrm{C}$')
ax.set_xlabel('Computed values', size=12)
ax.set_ylabel('Model values', size=12)
cb = fig.colorbar(p, ax=ax)
cb.set_label('Fraction of Al', size=12)
```



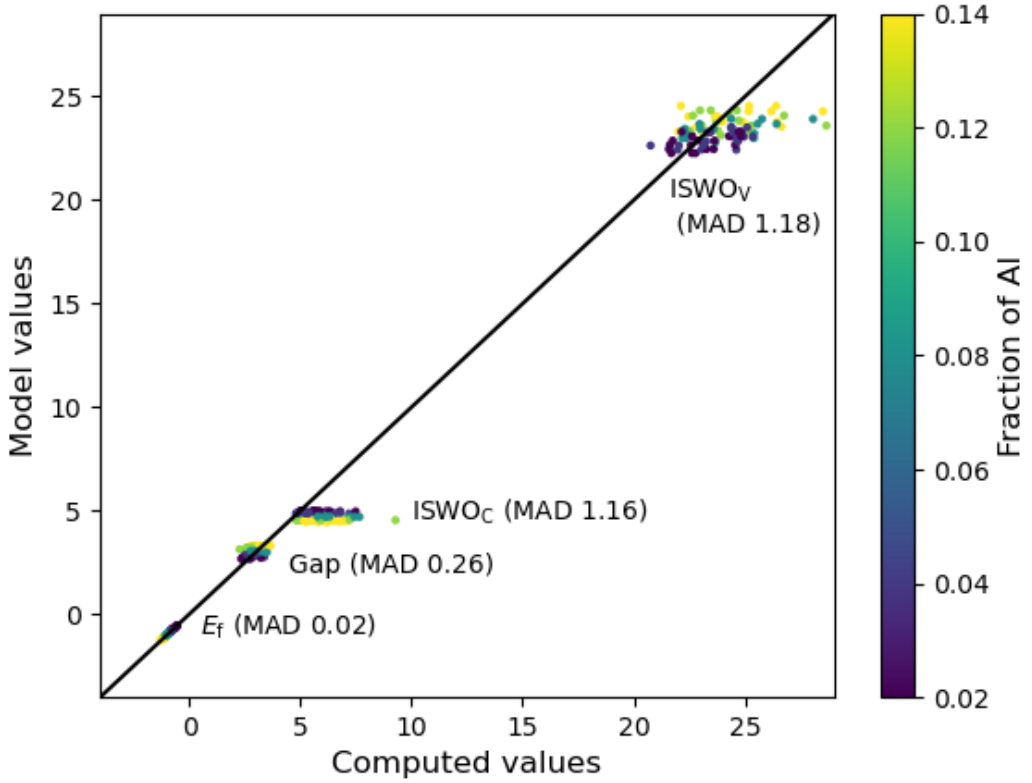
In [42]:

```
fig, ax = plt.subplots()
p = ax.scatter(calc_iv, model_iv, c=al)
mi = min(calc_iv + model_iv) * 0.95
ma = max(calc_iv + model_iv) * 1.05
ax.plot([-5,30], [-5,30], c = 'k')
ax.set_xlim(mi, ma)
ax.set_ylim(mi, ma)
ax.set_title('ISWO$_\mathrm{V}$')
ax.set_xlabel('Computed values', size=12)
ax.set_ylabel('Model values', size=12)
cb = fig.colorbar(p, ax=ax)
cb.set_label('Fraction of Al', size=12)
```



In [43]:

```
fig, ax = plt.subplots()
ax.plot([-5,30], [-5,30], c = 'k')
ax.scatter(calc_iv, model_iv, c=al, s=5)
ax.scatter(calc_ic, model_ic, c=al, s=5)
ax.scatter(calc_g, model_g, c=al, s=5)
p = ax.scatter(calc_ef, model_ef, c=al, s=5)
ax.set_xlabel('Computed values', size=12)
ax.set_ylabel('Model values', size=12)
ax.set_xlim(-4,29)
ax.set_ylim(-4,29)

mad_g = np.mean(np.abs(np.array(calc_g) - np.array(model_g)))
mad_iv = np.mean(np.abs(np.array(calc_iv) - np.array(model_iv)))
mad_ic = np.mean(np.abs(np.array(calc_ic) - np.array(model_ic)))
mad_ef = np.mean(np.abs(np.array(calc_ef) - np.array(model_ef)))

ax.annotate('ISWO$_\mathrm{V}$' + f'\n (MAD {mad_iv:1.2f})', [21.5,18.5])
ax.annotate('ISWO$_\mathrm{C}$' + f' (MAD {mad_ic:1.2f})', [10, 4.5])
ax.annotate('Gap' + f' (MAD {mad_g:1.2f})', [4.5, 2.0])
ax.annotate('$E_\mathrm{f}$' + f' (MAD {mad_ef:1.2f})', [0.5,-1.0])
cb = fig.colorbar(p, ax=ax)
cb.set_label('Fraction of Al', size=12)
fig.savefig('validation.eps')
```


Model values
Computed values
Fraction of Al
ISWO$_V$
(MAD 1.18)
ISWO$_C$ (MAD 1.16)
Gap (MAD 0.26)
E_f (MAD 0.02)
0
5
10
15
20
25
0.02
0.04
0.06
0.08
0.10
0.12
0.14

In [ ]:

In [ ]: