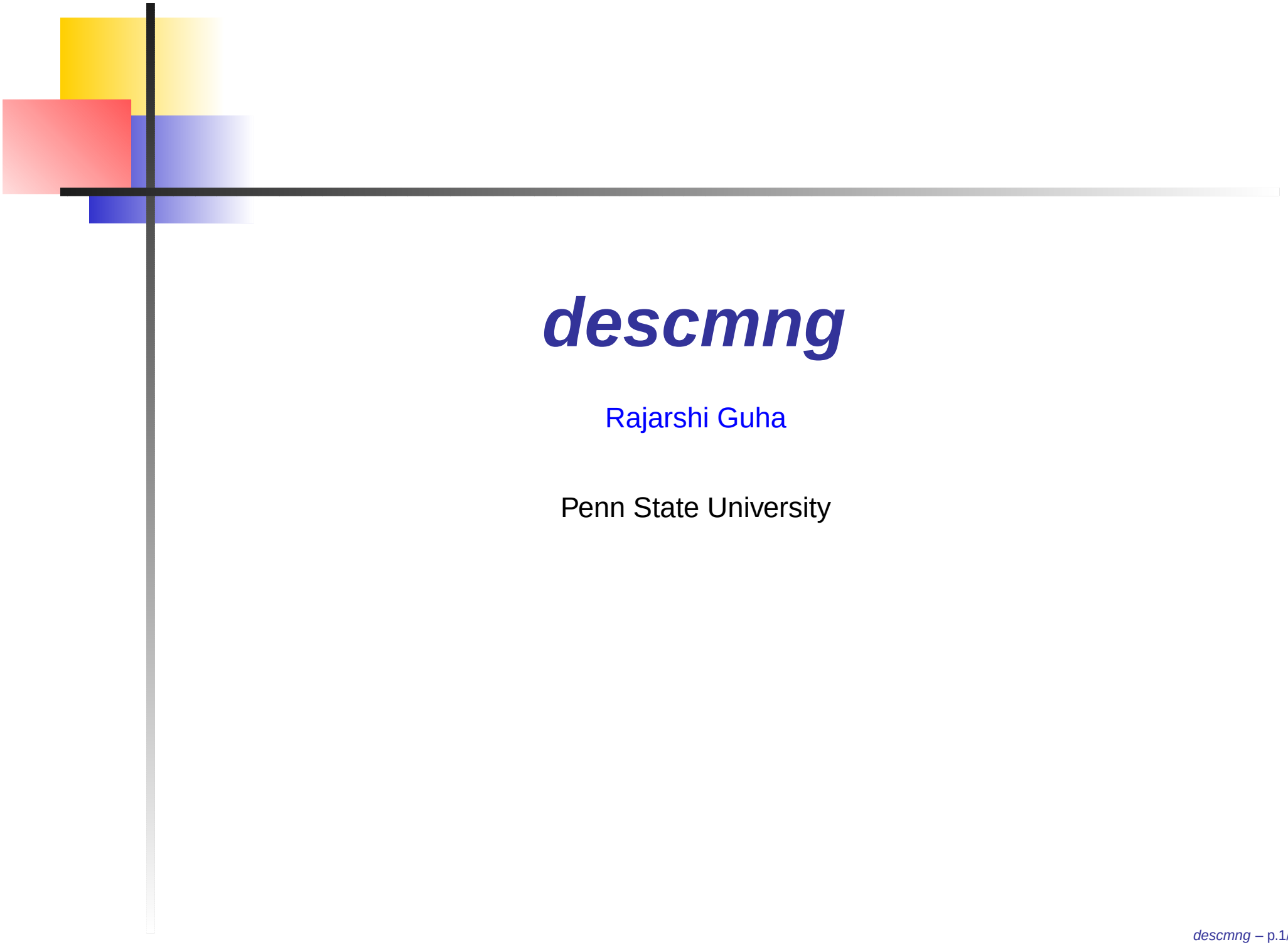


descmng

Rajarshi Guha

Penn State University

ADAPT Limitations?

- Fixed number of atoms.
- Storage space for descriptors is limited.
- Separate workareas are requred for individual problems.
- Lack of automation.
- Stuck to *ares* & *hera*.

ADAPT Limitations

- Storage Space
  - Fixed upper limit.
  - We need to delete descriptors as we go along.
  - Descriptor storage is linked to workareas.
- Each type of model involves several steps.

Descriptor Manager (descmng)

- Carries out a number of ADAPT functions.
- No limits on descriptor number.
- No limits on number of molecules.
- Implements a number of recent items:
  - Tropsha set generation
  - Diversity Indices
  - KNN averaged predictions

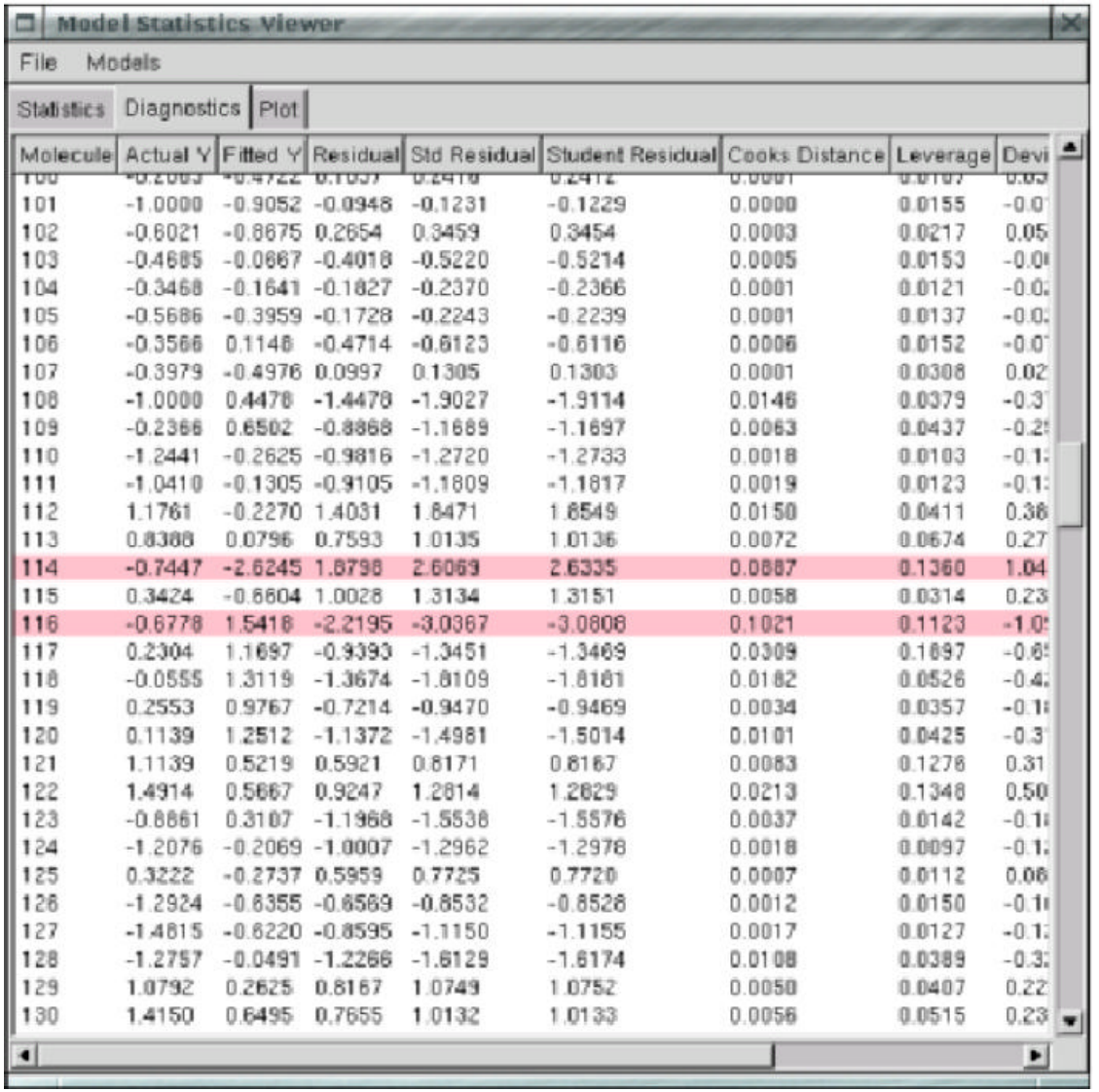
descmng - Storage

- It can store all calculated descriptors.
- Implements descriptor reduction using correlation and identical testing.
- Can generate output files in *annlin*, *qnetin*, *dragon* or *pnn* formats.
- It can generate scrambled sets.

descmng - Analysis

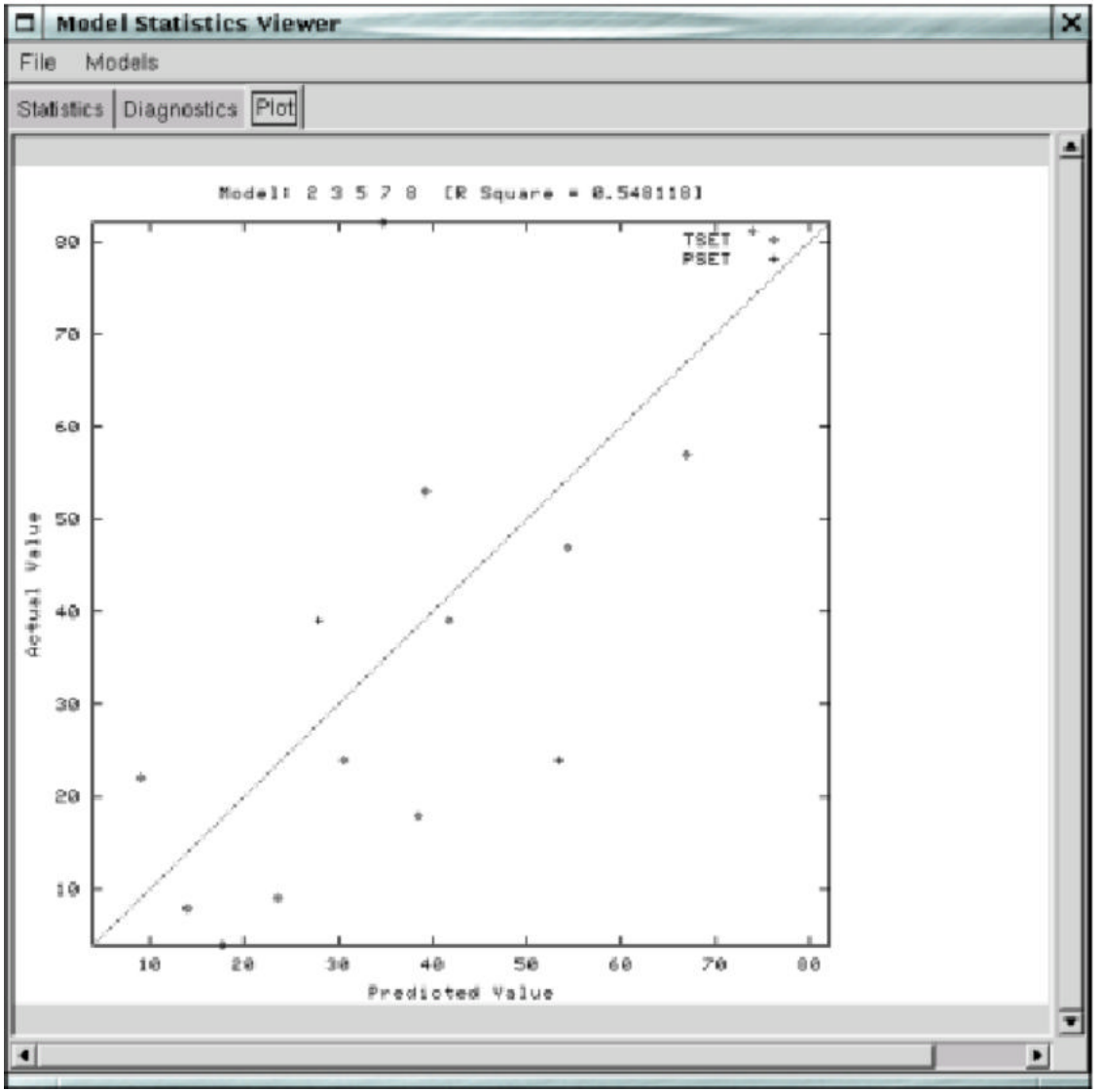
- Performs multiple linear regression (multiple inputs possible).
- Generate diagnostics for outlier detection.
- Generates plots of predicted versus actual values.

Viewing Type I Models

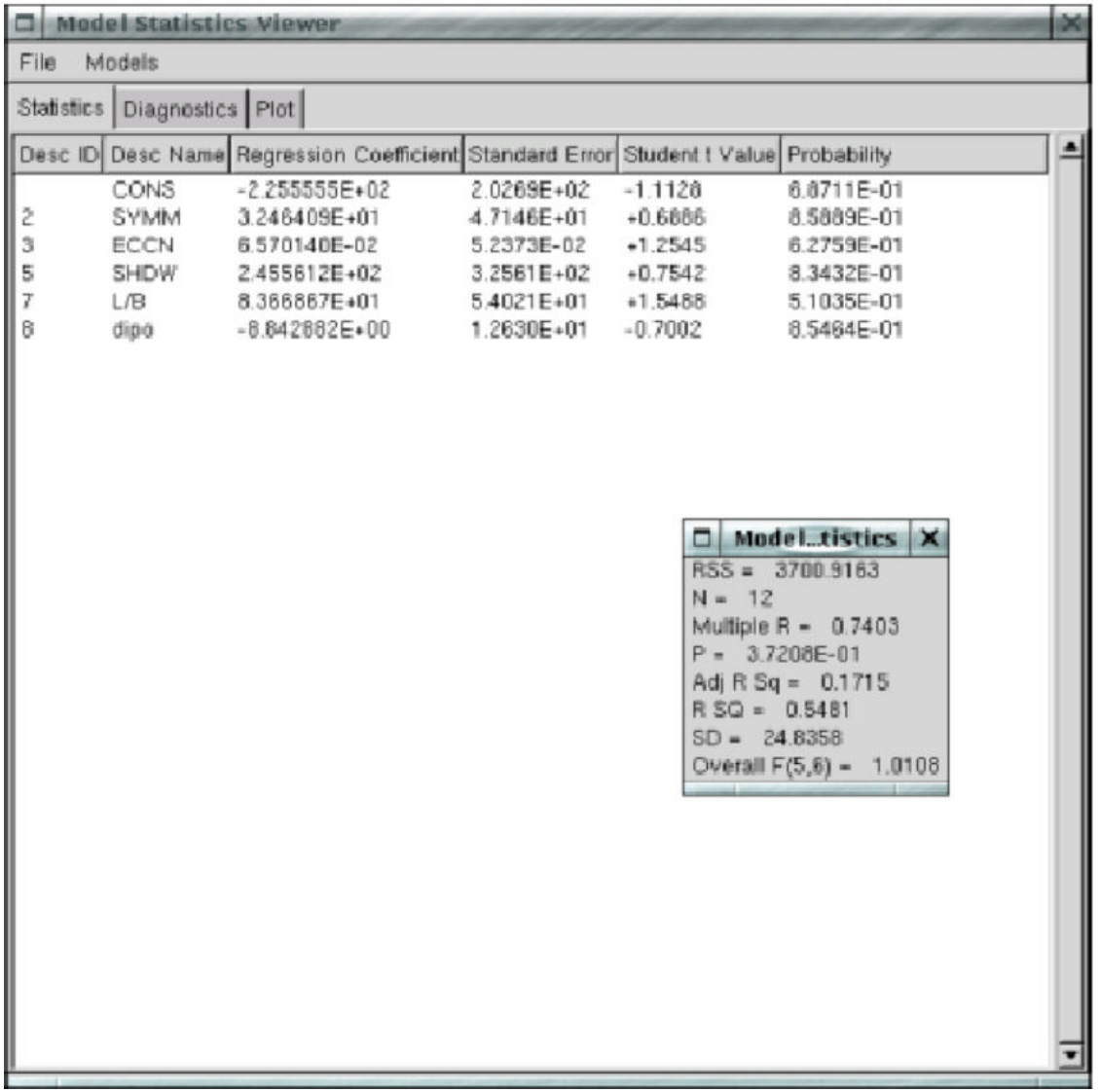


Molecule	Actual Y	Fitted Y	Residual	Std Residual	Student Residual	Cooks Distance	Leverage	Devi
101	-1.0000	-0.9052	-0.0948	-0.1231	-0.1229	0.0000	0.0155	-0.0
102	-0.6021	-0.8675	0.2654	0.3459	0.3454	0.0003	0.0217	0.05
103	-0.4685	-0.0667	-0.4018	-0.5220	-0.5214	0.0005	0.0153	-0.0
104	-0.3468	-0.1641	-0.1827	-0.2370	-0.2366	0.0001	0.0121	-0.0
105	-0.5686	-0.3959	-0.1728	-0.2243	-0.2239	0.0001	0.0137	-0.0
106	-0.3566	0.1148	-0.4714	-0.6123	-0.6116	0.0006	0.0152	-0.0
107	-0.3979	-0.4976	0.0997	0.1305	0.1303	0.0001	0.0308	0.02
108	-1.0000	0.4478	-1.4478	-1.9027	-1.9114	0.0146	0.0379	-0.3
109	-0.2366	0.6502	-0.8868	-1.1689	-1.1697	0.0063	0.0437	-0.2
110	-1.2441	-0.2625	-0.9816	-1.2720	-1.2733	0.0018	0.0103	-0.1
111	-1.0410	-0.1305	-0.9105	-1.1809	-1.1817	0.0019	0.0123	-0.1
112	1.1761	-0.2270	1.4031	1.8471	1.8549	0.0150	0.0411	0.36
113	0.8388	0.0796	0.7593	1.0135	1.0136	0.0072	0.0674	0.27
114	-0.7447	-2.6245	1.8798	2.6069	2.6335	0.0887	0.1360	1.04
115	0.3424	-0.6604	1.0028	1.3134	1.3151	0.0058	0.0314	0.23
116	-0.6778	1.5416	-2.2195	-3.0367	-3.0808	0.1021	0.1123	-1.0
117	0.2304	1.1697	-0.9393	-1.3451	-1.3469	0.0309	0.1697	-0.6
118	-0.0555	1.3119	-1.3674	-1.8109	-1.8181	0.0182	0.0526	-0.4
119	0.2553	0.9767	-0.7214	-0.9470	-0.9469	0.0034	0.0357	-0.1
120	0.1139	1.2512	-1.1372	-1.4981	-1.5014	0.0101	0.0425	-0.3
121	1.1139	0.5219	0.5921	0.8171	0.8167	0.0083	0.1276	0.31
122	1.4914	0.5667	0.9247	1.2814	1.2829	0.0213	0.1348	0.50
123	-0.8861	0.3107	-1.1968	-1.5538	-1.5576	0.0037	0.0142	-0.1
124	-1.2076	-0.2069	-1.0007	-1.2962	-1.2978	0.0018	0.0097	-0.1
125	0.3222	-0.2737	0.5959	0.7725	0.7720	0.0007	0.0112	0.06
126	-1.2924	-0.6355	-0.6569	-0.8532	-0.8526	0.0012	0.0150	-0.1
127	-1.4815	-0.6220	-0.8595	-1.1150	-1.1155	0.0017	0.0127	-0.1
128	-1.2757	-0.0491	-1.2266	-1.6129	-1.6174	0.0108	0.0389	-0.3
129	1.0792	0.2625	0.8167	1.0749	1.0752	0.0050	0.0407	0.22
130	1.4150	0.6495	0.7655	1.0132	1.0133	0.0056	0.0515	0.23

Viewing Type I Models



Viewing Type I Models



Automation

- Type I
  - Will run for varying descriptor lengths in one run.
  - Can automatically set the required validation number.
  - Runs committes and reports averaged RMS errors.
  - GUI available to review statistics, outliers and plots.

Automation

- Type II
  - Automatically process multiple CNN architectures.
  - For N neuron input layer, it will process architectures from N-(N-1)-1 to N-2-1.
  - Runs committees for each architecture and reports averaged RMS errors and averaged prediction values.