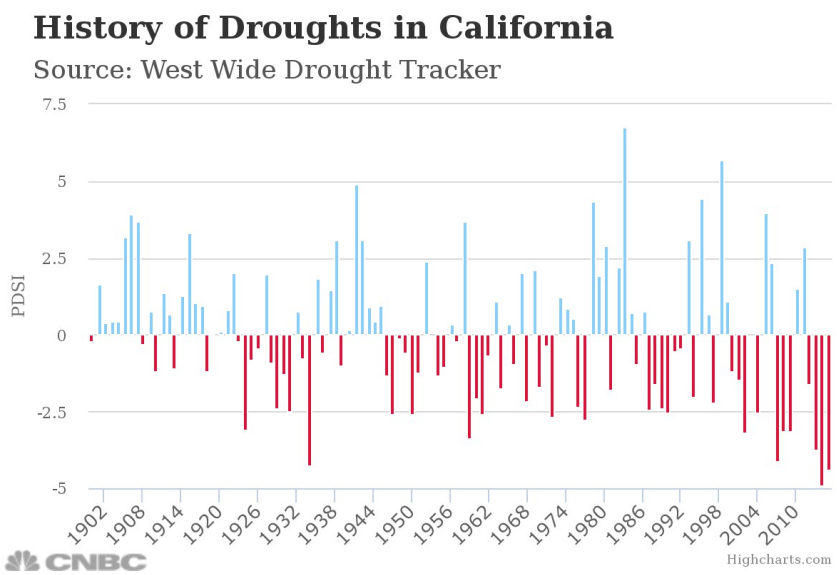
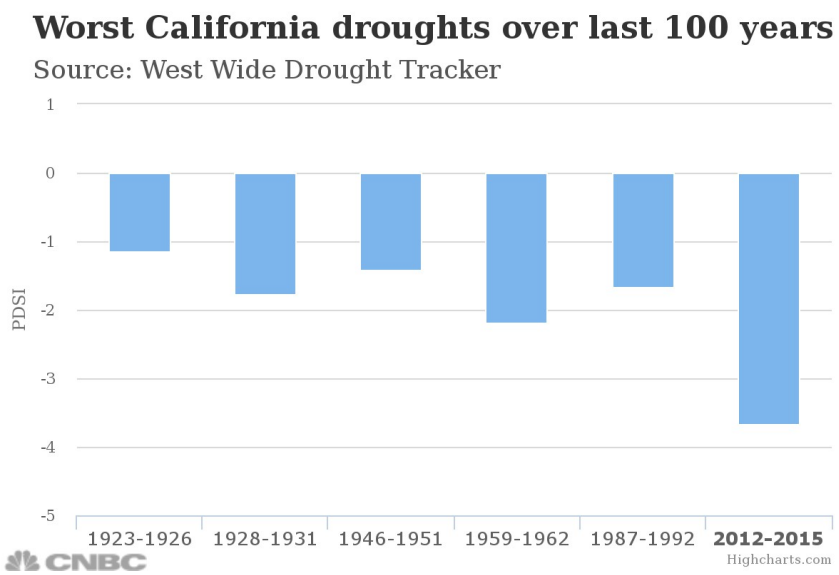
Medium-term Precipitation Projections with Neural Networks

THE SEARCH FOR INSIGHTS INTO THE PHENOMENOLOGY OF DROUGHT AND EL NINO CONDITIONS

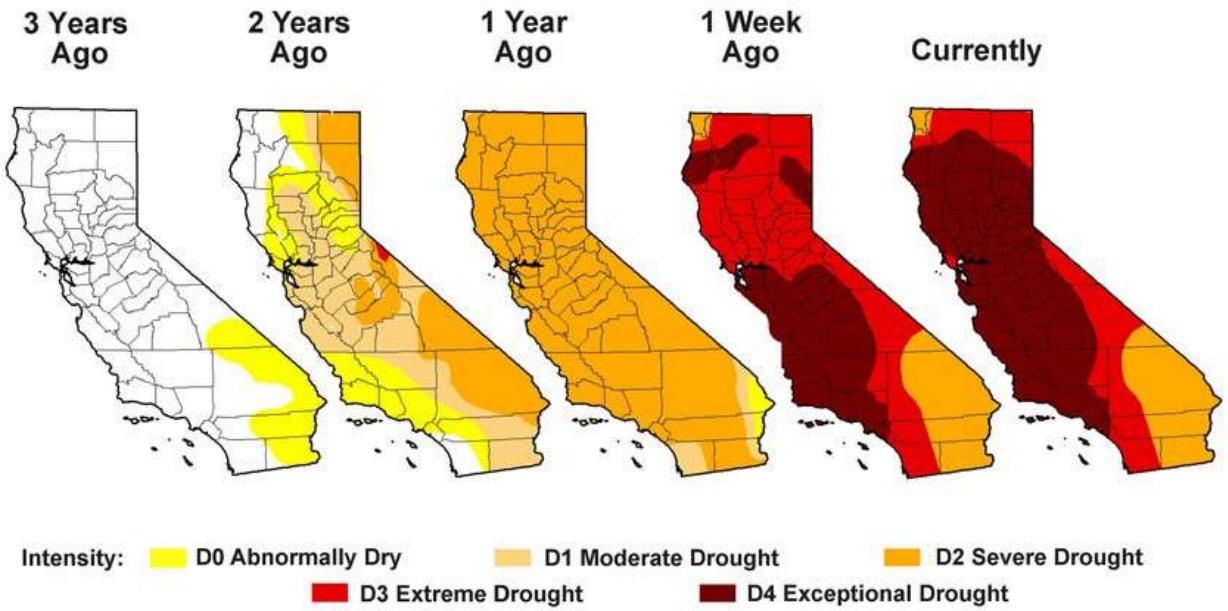
JUSTIN LE, SCHMID COLLEGE OF SCIENCE AND TECHNOLOGY

California Drought





California Drought



Environmental Impacts of Drought

Short-term:

- Detriments to Hydropower production and recreation
- Low farm yields

Long-term:

- Groundwater storage lost
- Food and water shortages
- Over-pumping can cause permanent groundwater loss, land elevation sinking, and seawater intrusion
- Wildfire risks

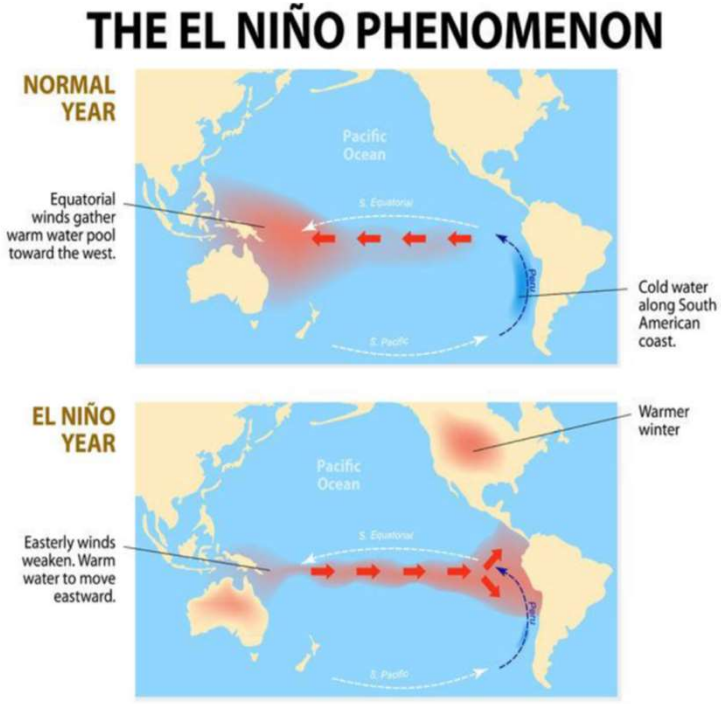
2015 drought season estimated to have cost more than **$2.2 billion** in economic loss.

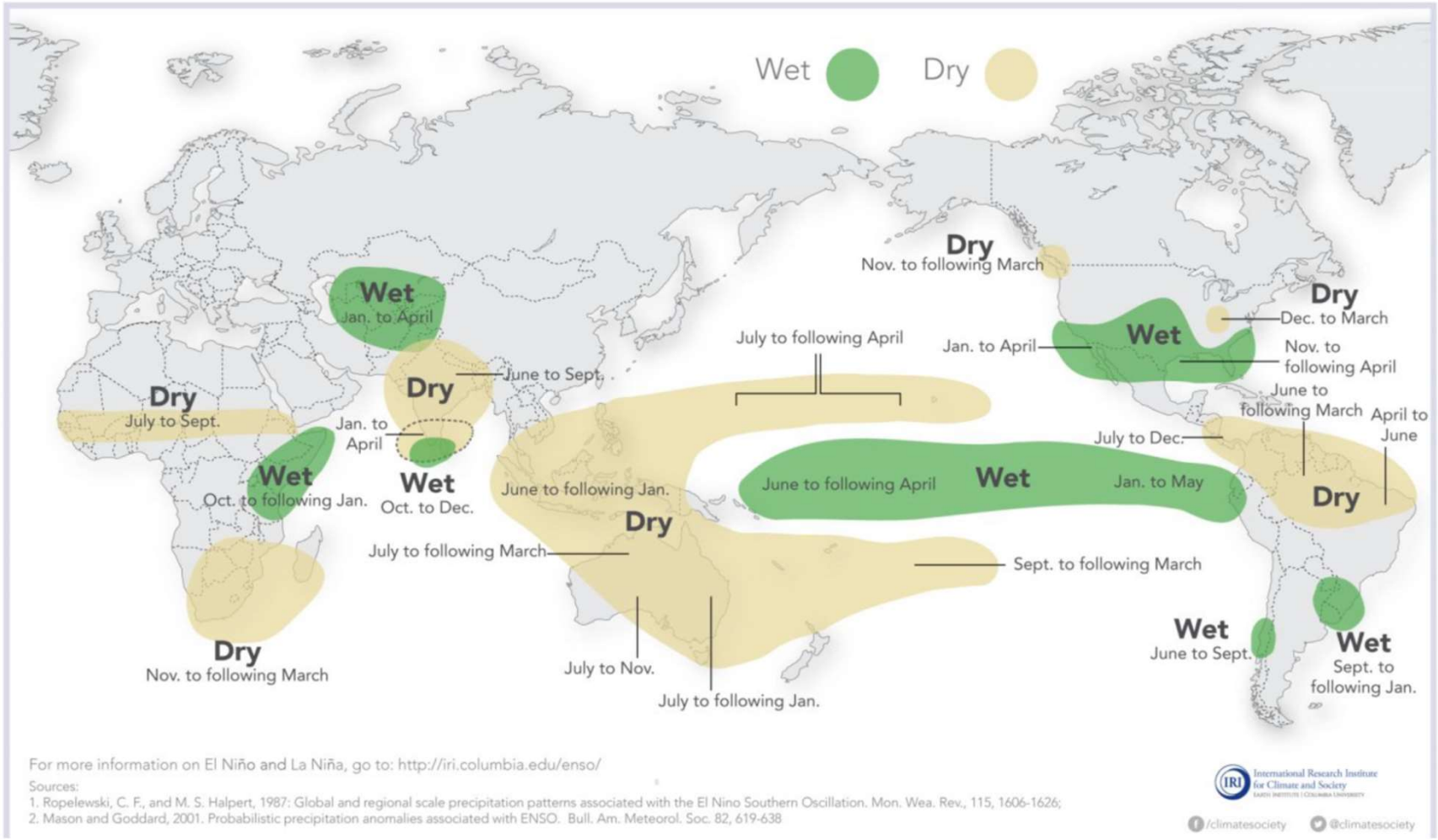
El Nino

Global periodic phenomenon linked to high ocean temperatures near South American coast

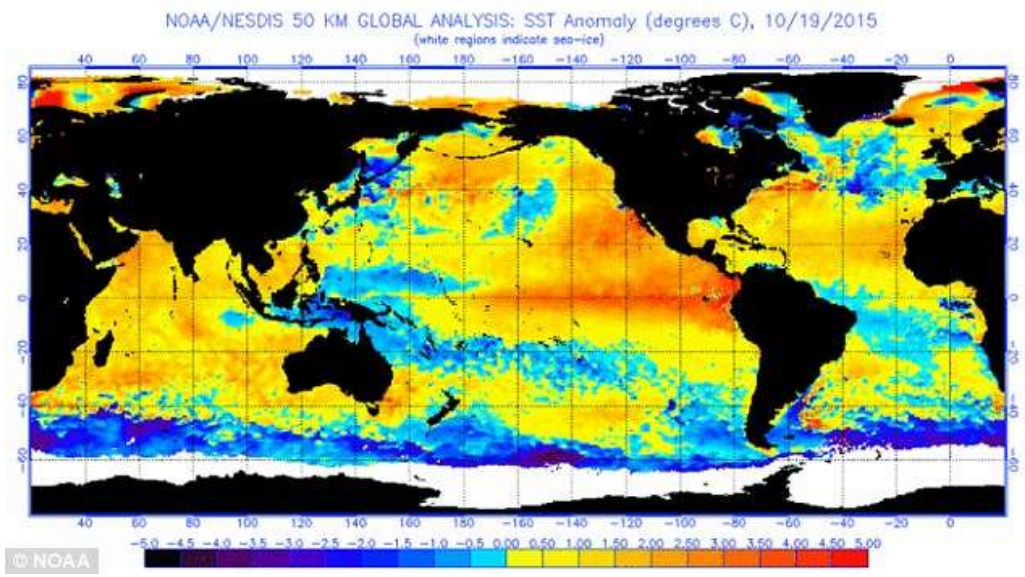
In California, manifests as a season of extreme rainfall, flooding, and warm temperatures.

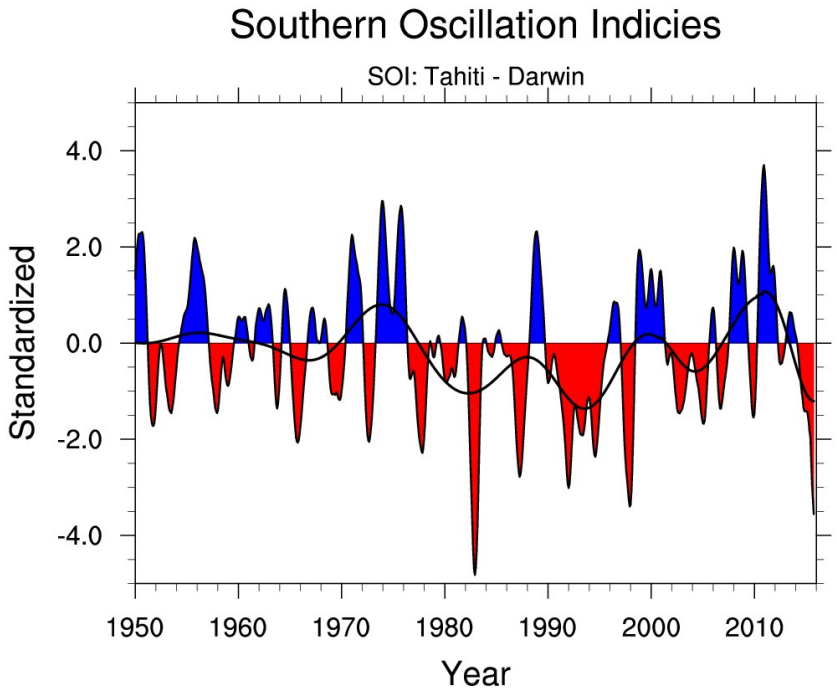
1997-1998 El Nino Season is estimated to have caused **$25 billion** in damages to the entire country.

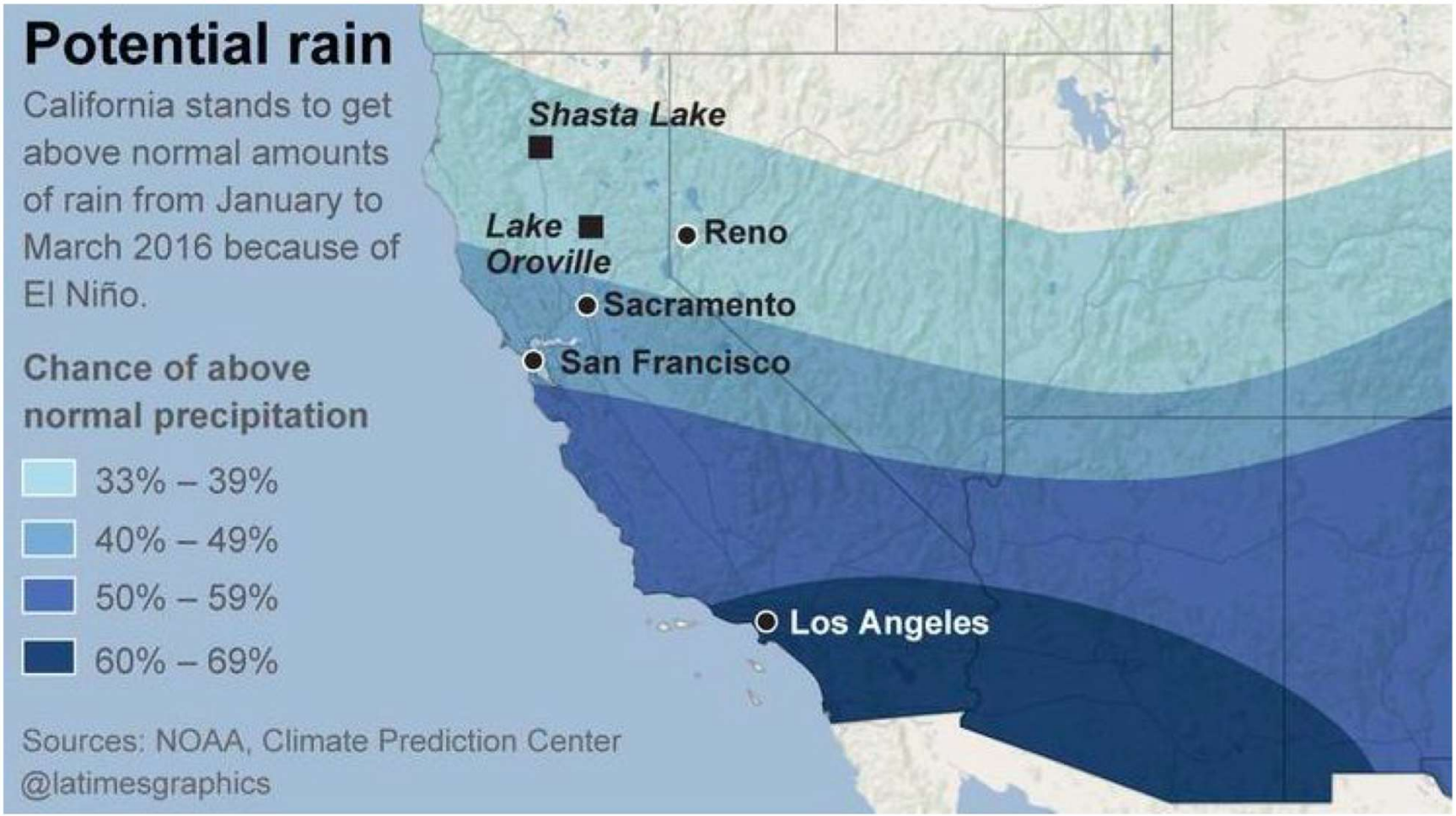



Wet
Dry
Dry
Nov. to following March
Wet
Jan. to April
Dry
Dec. to March
Wet
Jan. to April
Nov. to following April
July to following April
Dry
June to Sept.
Dry
July to Sept.
Jan. to April
Wet
Oct. to following Jan.
Wet
Oct. to Dec.
June to following March
April to June
July to Dec.
June to following Jan.
Dry
June to following April
Wet
Jan. to May
Dry
July to following March
Sept. to following March
Dry
Nov. to following March
July to Nov.
July to following Jan.
Wet
June to Sept.
Wet
Sept. to following Jan.
For more information on El Niño and La Niña, go to: http://iri.columbia.edu/enso/
Sources:
1. Ropelewski, C. F., and M. S. Halpert, 1987: Global and regional scale precipitation patterns associated with the El Nino Southern Oscillation. Mon. Wea. Rev., 115, 1606-1626;
2. Mason and Goddard, 2001. Probabilistic precipitation anomalies associated with ENSO. Bull. Am. Meteorol. Soc. 82, 619-638
International Research Institute for Climate and Society
/climatesociety
@climatesociety

2015-2016 El Nino Season






Potential rain
California stands to get above normal amounts of rain from January to March 2016 because of El Niño.
Chance of above normal precipitation
33% – 39%
40% – 49%
50% – 59%
60% – 69%
Sources: NOAA, Climate Prediction Center
@latimesgraphics
Shasta Lake
Lake Oroville
Reno
Sacramento
San Francisco
Los Angeles

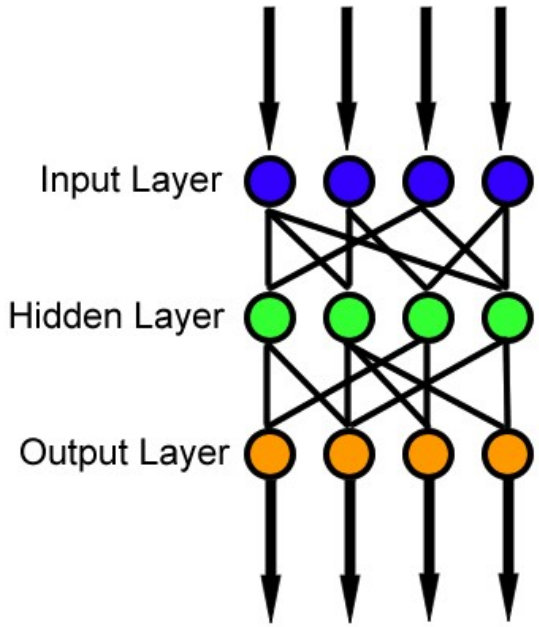
Questions

- Can machine learning techniques give us:
  - Autocorrelative patterns?
  - Insight on physical phenomenon?

Neural networks

- "Universal Approximators"
- Inspired by topology of human brain
- Actively explored space of training algorithms to configure a network to approximate any function to arbitrary accuracy

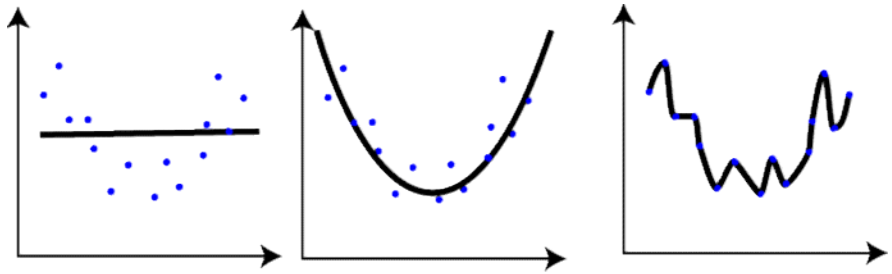
Feed-Forward Neural Networks



- Input comes in through input layer, and information proceeds through to the output.
- Connections from layer to layer have different strength and transfer functions
- "Training" involves selecting how strong each connection should be.

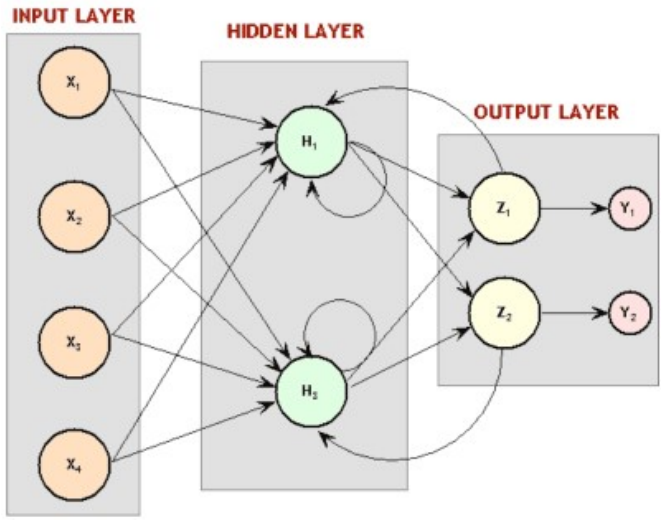
Issues

- Search space very large and erratic compared to what common sense dictates
- Strong potential of overfitting



Recurrent Neural Networks

- Trainable state machines
- Universal approximation of **processes**, not functions
- Nodes receive delayed signals from their past outputs
- Much more suited for time series, and topology is much closer to reality.
- Much less parameters; “recentness” built in.
- Training and prediction done over a **series**, not over a **data point**.



Ability to look at physical phenomenon

Cell sensitive to position in line:

```
The sole importance of the crossing of the Berezina lies in the fact
that it plainly and indubitably proved the fallacy of all the plans for
cutting off the enemy's retreat and the soundness of the only possible
line of action--the one Kutuzov and the general mass of the army
demanded--namely, simply to follow the enemy up. The French crowd fled
at a continually increasing speed and all its energy was directed to
reaching its goal. It fled like a wounded animal and it was impossible
to block its path. This was shown not so much by the arrangements it
made for crossing as by what took place at the bridges. When the bridges
broke down, unarmed soldiers, people from Moscow and women with children
who were with the French transport, all--carried on by vis inertiae--
pressed forward into boats and into the ice-covered water and did not,
surrender.
```

Cell that turns on inside quotes:

```
"You mean to imply that I have nothing to eat out of.... On the
contrary, I can supply you with everything even if you want to give
dinner parties," warmly replied Chichagov, who tried by every word he
spoke to prove his own rectitude and therefore imagined Kutuzov to be
animated by the same desire.

Kutuzov, shrugging his shoulders, replied with his subtle penetrating
smile: "I meant merely to say what I said."
```

Cell that robustly activates inside if statements:

```
static int __dequeue_signal(struct sigpending *pending, sigset_t *mask,
        siginfo_t *info)
{
  int sig = next_signal(pending, mask);
  if (sig) {
    if (current->notifier) {
      if (sigismember(current->notifier_mask, sig)) {
        if (!(current->notifier)(current->notifier_data)) {
          clear_thread_flag(TIF_SIGPENDING);
          return 0;
        }
      }
    }
    collect_signal(sig, pending, info);
  }
  return sig;
}
```

A large portion of cells are not easily interpretable. Here is a typical example:

```
/* Unpack a filter field's string representation from user-space
 * buffer. */
char *audit_unpack_string(void **bufp, size_t *remain, size_t len)
{
  char *str;
  if (!*bufp || (len == 0) || (len > *remain))
    return ERR_PTR(-EINVAL);
  /* Of the currently implemented string fields, PATH_MAX
   * defines the longest valid length.
   */
```

Cell that turns on inside comments and quotes:

```
/* Duplicate LSM field information.  The lsm_rule is opaque, so
 * re-initialized. */
static inline int audit_dupe_lsm_field(struct audit_field *df,
                                       struct audit_field *sf)
{
  int ret = 0;
  char *lsm_str;
  /* our own copy of lsm_str */
  lsm_str = kstrdup(sf->lsm_str, GFP_KERNEL);
  if (unlikely(!lsm_str))
    return -ENOMEM;
  df->lsm_str = lsm_str;
  /* our own (refreshed) copy of lsm_rule */
  ret = security_audit_rule_init(df->type, df->op, df->lsm_str,
                                 (void **)&df->lsm_rule);
  /* Keep currently invalid fields around in case they
   * become valid after a policy reload. */
  if (ret == -EINVAL) {
    pr_warn("audit rule for LSM \'%s\' is invalid\n",
            df->lsm_str);
    ret = 0;
  }
  return ret;
}
```

Cell that is sensitive to the depth of an expression:

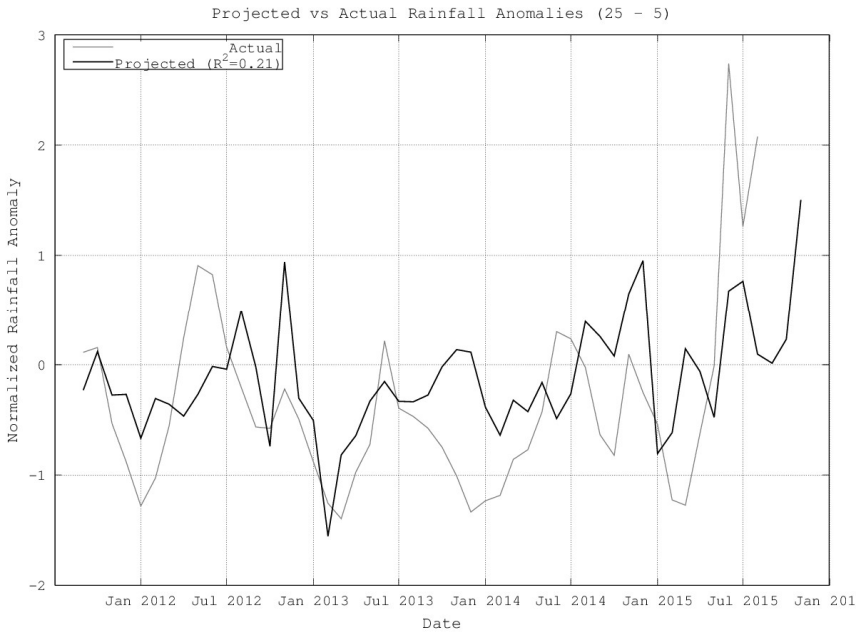
```
#ifdef CONFIG_AUDITSYSCALL
static inline int audit_match_class_bits(int class, u32 *mask)
{
  int i;
  if (classes[class]) {
    for (i = 0; i < AUDIT_BITMASK_SIZE; i++)
      if (mask[i] & classes[class][i])
        return 0;
  }
  return 1;
}
```

Cell that might be helpful in predicting a new line. Note that it only turns on for some ")":

```
char *audit_unpack_string(void **bufp, size_t *remain, si
{
  char *str;
  if (!*bufp || (len == 0) || (len > *remain))
    return ERR_PTR(-EINVAL);
  /* Of the currently implemented string fields, PATH_MAX
   * defines the longest valid length.
   */
  if (len > PATH_MAX)
    return ERR_PTR(-ENAMETOOLONG);
  str = kmalloc(len + 1, GFP_KERNEL);
  if (unlikely(!str))
    return ERR_PTR(-ENOMEM);
  memcpy(str, *bufp, len);
  str[len] = 0;
  *bufp += len;
  *remain -= len;
  return str;
}
```

Andrej Karpathy, "The Unreasonable Effectiveness of Recurrent Neural Networks"
May 21 2015 http://karpathy.github.io/2015/05/21/rnn-effectiveness/

Initial Results



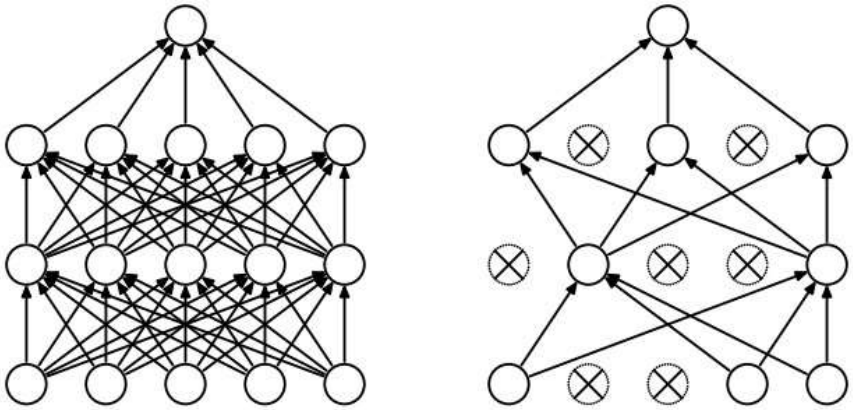
- Initial results demonstrated correlation, making this avenue apparently worth pursuing.
- Strong signs of overfitting.
- Parameter reduction (and removing hidden layers) tended to improve results, but hit a wall.

Combatting Overfitting with Noise

- Realization: we are trying to model a **distribution**, not a function.
- Train with noise! Forces **robustness**.
- **Instant** decreases in overfitting observed, and higher correlations.

Combatting overfitting with Dropout

- Train handicapped versions of networks in hopes that they find different features, and a way to work together despite missing neurons.
- Inspired by sexual selection in evolution. Power in variation.
- Overall effect is similar to **running several networks** at once, within only **one network**!

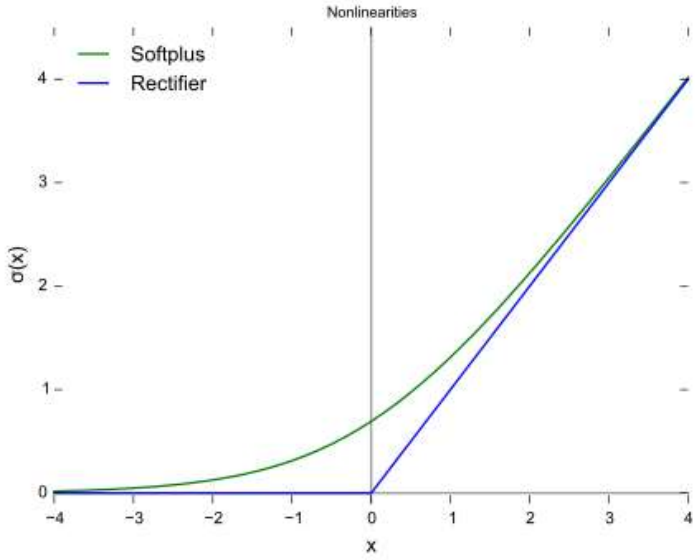


(a) Standard Neural Net

(b) After applying dropout.

Increasing Expressivity with Rectified Linear Unit Activation Functions

- "Activation functions" determine neuron firing behavior
- Originally began with the classic logistic function:
  - $\frac{1}{1+e^{-x}}$
  - Outputs range between 0 and 1
  - Gradient vanishes dramatically at high inputs
- Switched according to new research showing success of **rectifiers** as activation functions
  - Outputs are unbounded and positive.
  - Gradient expressed at high inputs
  - Allows neurons to be more expressive in what information they pass down.

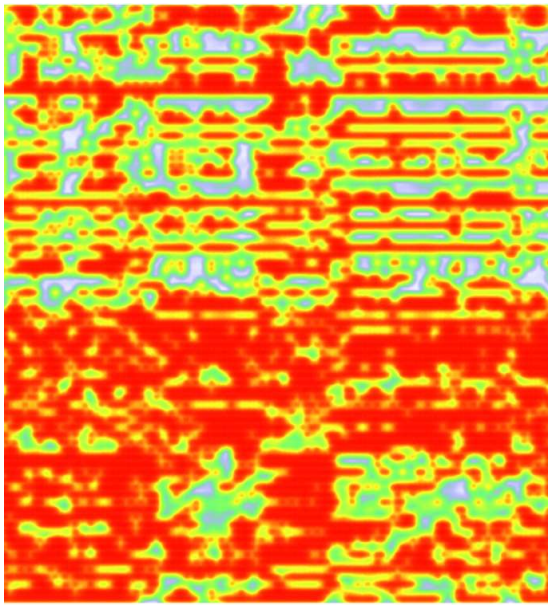


Predicting the Future

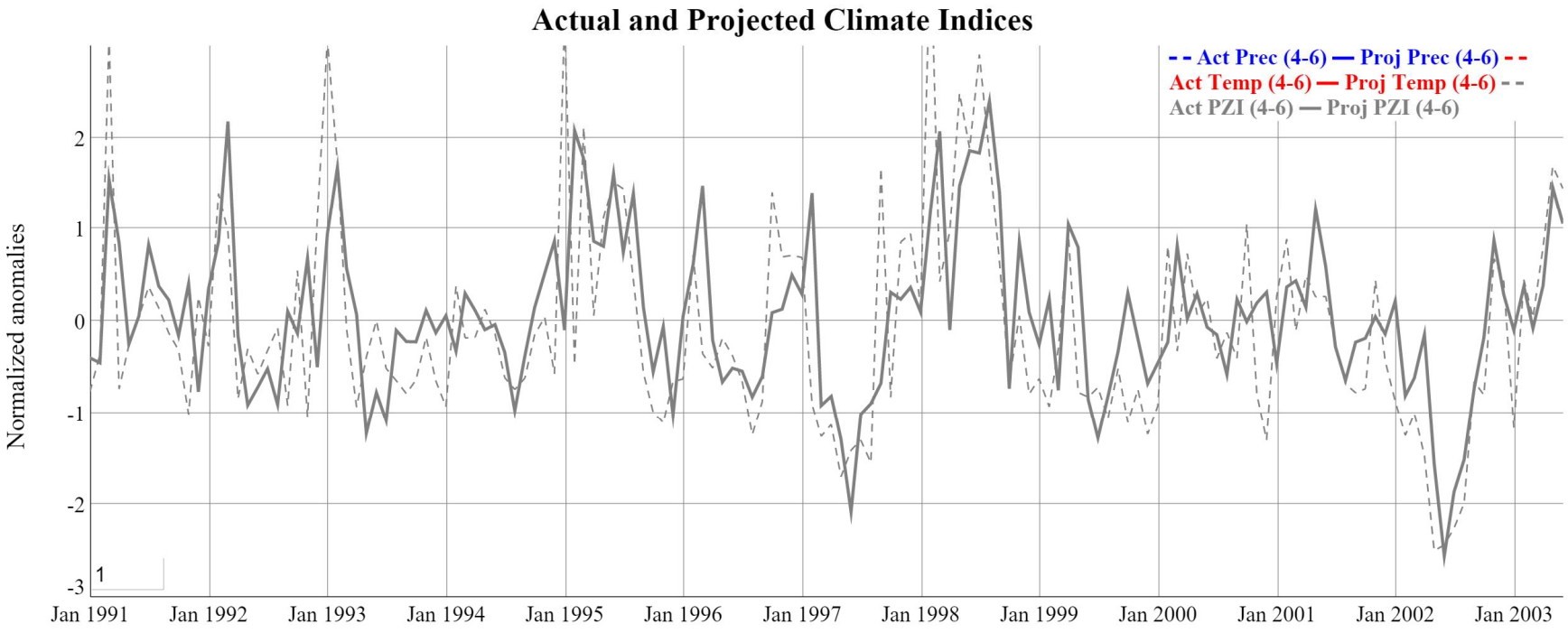
- Network is shaped to predict the next month of climate data/indices. But can it predict long-term?
- "**Leapfrogging into the future**".
- Add noise to get confidence intervals.

Node Activation Maps

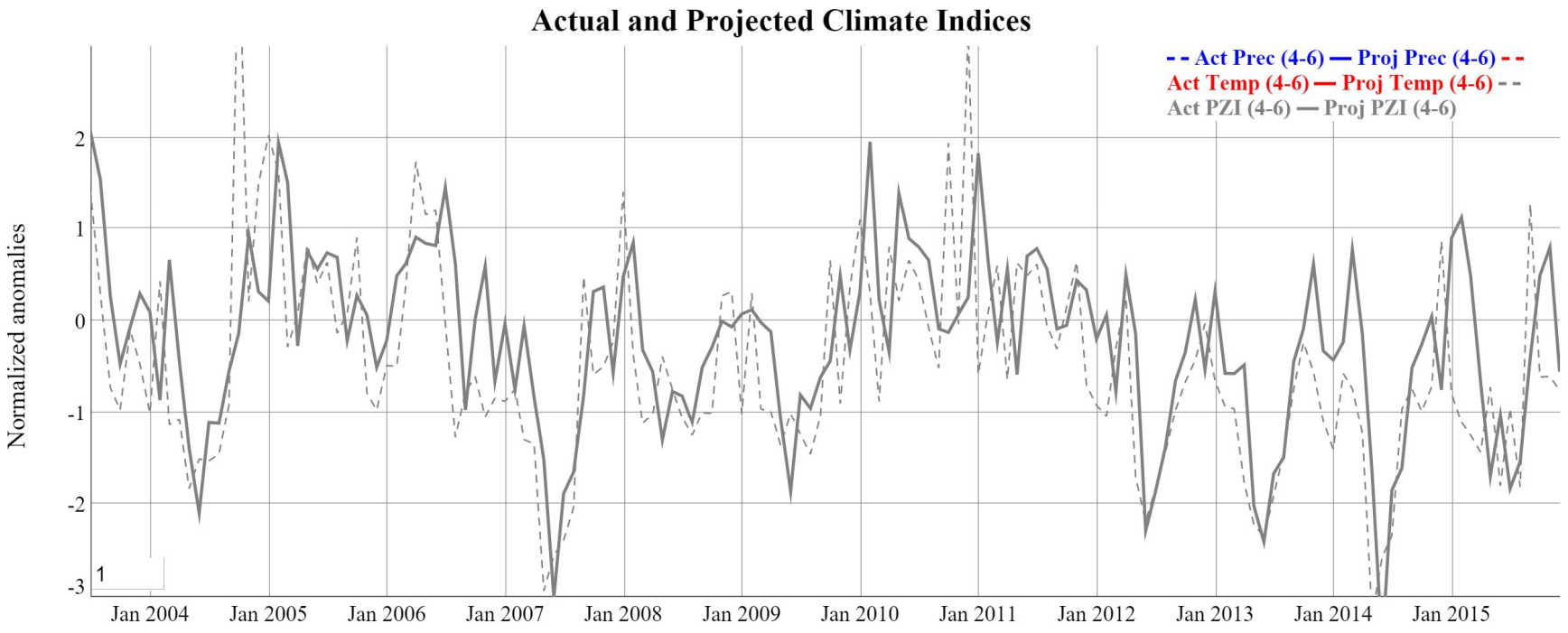
- Developing visualizations to peek into the mind of networks to find new potential phenomenon
- Might yield:
  - Features that correspond to known or unknown physical phenomenon
    - Correlate with outside data and find any links
  - The ability to identify weak nodes or overly correlated nodes, and delete them: “brain surgery”
  - A way to gauge the health of a neural network.



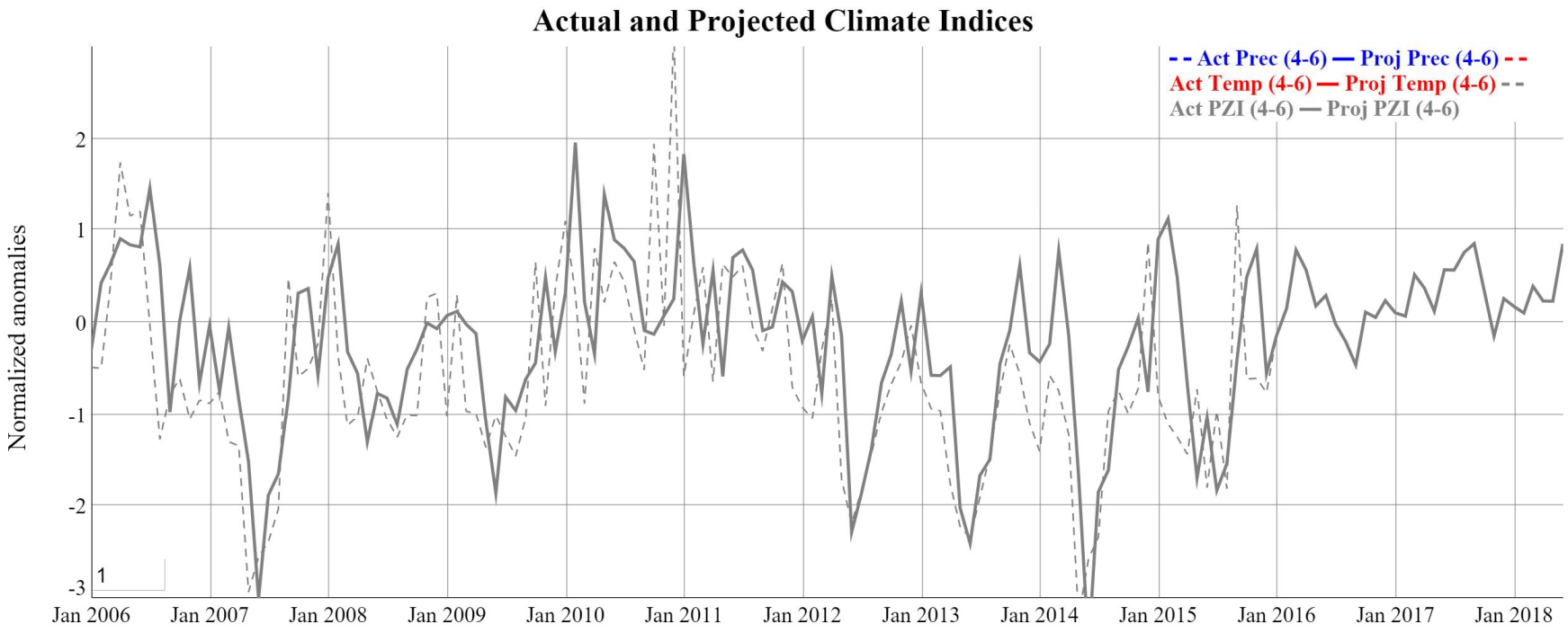
Training


Actual and Projected Climate Indices
Act Prec (4-6) Proj Prec (4-6)
Act Temp (4-6) Proj Temp (4-6)
Act PZI (4-6) Proj PZI (4-6)
Normalized anomalies
2
1
0
-1
-2
-3
1
Jan 1991
Jan 1992
Jan 1993
Jan 1994
Jan 1995
Jan 1996
Jan 1997
Jan 1998
Jan 1999
Jan 2000
Jan 2001
Jan 2002
Jan 2003

Validation


Actual and Projected Climate Indices
Act Prec (4-6) — Proj Prec (4-6)
Act Temp (4-6) — Proj Temp (4-6)
Act PZI (4-6) — Proj PZI (4-6)
Normalized anomalies
2
1
0
-1
-2
-3
Jan 2004
Jan 2005
Jan 2006
Jan 2007
Jan 2008
Jan 2009
Jan 2010
Jan 2011
Jan 2012
Jan 2013
Jan 2014
Jan 2015

Prediction


Actual and Projected Climate Indices
Act Prec (4-6) Proj Prec (4-6)
Act Temp (4-6) Proj Temp (4-6)
Act PZI (4-6) Proj PZI (4-6)
Normalized anomalies
2
1
0
-1
-2
-3
1
Jan 2006
Jan 2007
Jan 2008
Jan 2009
Jan 2010
Jan 2011
Jan 2012
Jan 2013
Jan 2014
Jan 2015
Jan 2016
Jan 2017
Jan 2018

Historical El Nino Z Indices

Season	Peak Anomaly	Peak Anomaly Month	Annual Anomaly
1957-1958	3.03	April	0.50
1982-1983	2.22	April	1.15
1997-1998	4.13	February	1.10
2009-2010	1.10	January	0.45
2015-2016*	0.72	March	0.24

* Projected

2012-2016 Annual Z-Index Anomalies

Year	Annual Anomaly
2012	-0.93
2013	-1.30
2014	-1.17
2015	-0.85
2016*	0.24

* Projected

Answers?

- Definitely evidence of predictive power with high correlation (.5-.7 depending on the index being predicted).
- Drought conditions always predicted accurately.
- What's in the future?
  - Looks like milder, average-level precipitation with drought indices moving towards non-drought.
  - Predicting a steady climb out of drought