**<Sources of variability in electrophysiology data of mammalian neurons>**

**Introduction**

Neurophysiology as a field is moving towards larger kinds of data analyses trying to not only understand one neuron type in isolation, but to study many kinds of neurons simultaneously (Kandel et al., 2013). For example, the first goal of the US NIH BRAIN project is to generate a “census of cell types” (Jorgenson et al., 2015), involving neuron comparison using genetic, morphological and electrophysiological characteristics. One way to address this is by aggregating vast amounts of already published neuroscientific data. However, combining and comparing electrophysiology data across labs directly and on a large scale is challenging, because such data is often collected under different experimental conditions. It is generally thought that subtle variation in experimental conditions introduces certain variation into the corresponding ephys measurements. Therefore, comparing data across differently designed experiments without accounting for variability introduced by experimental conditions could lead to incorrect or inconsistent results.

The common practice among neurophysiologists is to only analyze data that they have collected themselves. This is largely because it is generally thought that subtle variation in experimental conditions introduces certain variation into the corresponding measurements.

Many neurophysiologists address the task of exploring the effects of experimental conditions on neuron ephys properties using experimental electrophysiology techniques (Armentia and Sah, 2004; Kim and Connors, 2012; Lee et al., 2005). In the past, ephys data has been shown to be sensitive to experimental conditions. For example, differences in animal ages, especially during development (Suter et al., 2013); or varying extracellular Ca2+ concentrations (Aivar et al., 2014) result in changes in electrophysiological properties of neurons. However, this experimental approach is limited to varying a single condition and studying one or several neuron types at a time. Therefore, it is unclear how well the discovered relationships between electrophysiology properties and experimental conditions would generalize to other neuron types, animal species, ages and other confounding factors that typically remained fixed throughout each experiment.

To the best of our knowledge, there are no comprehensive and systematic analysis (or meta-analyses) of the effects of experimental conditions on electrophysiological measurements. However, there are papers that explore the effects of specific experimental setups (Aghajanian and Rasmussen, 1989; Moyer and Brown, 1998). Additionally, the effects of specific experimental solution components on neuron survival rates were discussed by several papers (MacGregor et al., 2001; Richerson and Messer, 1995; Tanaka et al., 2008).

Previously, we designed and created NeuroElectro, an online database that contains text-mined and curated population mean electrophysiological measurements, neuron type and experimental setup information from normal control samples of published neuroscientific studies (Tripathy et al., 2015). Using a large-scale meta-analysis method, he showed that animal age, recording temperature, electrode type choices significantly explain the study-to-study variance in reported ephys values.

Since a typical electrophysiological experiment uses carefully designed solutions inside and outside the measured neurons, we hypothesized that study-to-study ephys variability could be partially explained by the experimental setup (metadata) differences, focusing on the recording and pipette solution compositions. To test that hypothesis, we employed a combination of text-mining and curation approaches to extract experimental solutions used in published neurophysiological articles. Then, we integrated the solution extraction algorithms into the NeuroElectro database. After exploring the external and internal solution recipes commonly used by electrophysiologists, we applied multiple regression models to uncover the effects of solutions on the measured ephys values.

**Materials and Methods**

We used the existing NeuroElectro (www.neuroelectro.org) database as a starting point for determining metadata factors that could influence the variability in reported electrophysiological measurements (Tripathy et al. 2015). Briefly, NeuroElectro stores and data-mines thousands of Neuroscience articles that may contain electrophysiology data (resting membrane potential, input resistance, action potential spike half-width and amplitude, etc.). Text-mining tables of articles for ephys properties and methods sections for experimental conditions is done in C-Python using HTML-parsing tools and libraries. The process of text-mining for ephys properties and previously analyzed metadata (animal age, weight, species, strain; recording temperature, electrode type, junction potential) has not been significantly adjusted since the previous NeuroElectro paper (Tripathy et. al, 2015). However, here we have implemented a robust text-mining approach that can handle complicated metadata extraction such as solution compositions used during electrophysiological experiments. We assume that all the relevant metadata information can be found in the Methods section of each article.

The text-mined ephys properties and metadata get verified by at least two trained curators. Neuron types are assigned based on an expert-defined list of neuron types provided by NeuroLex.org (Larson et al. 2013). Neuron instances reported in articles that could not be curated unambiguously to a single type were curated to the general neuron type “other”. Besides the NeuroLex definition, NeuroElectro preserves the author-defined neuron types capturing as many details as the original article’s authors provided. The curators are responsible for checking the correctness of external and internal solution sentences, but they do not validate the extracted chemical component concentration values.

**Statistical analysis**

**Preprocessing**

The curated and standardized ephys values, neuron names and metadata are aggregated into a single CSV file of the following format: each line of the file corresponds to a unique combination of an ephys table in a neuroscientific article and a neuron type reported in that table. For example, an article with 1 ephys table that provide ephys information about 4 neuron types will have 4 data rows in the CSV data spreadsheet. Each row of data contains information about the article (PubMed ID, title, year published, authors, etc.), NeuroElectro and author-defined neuron types, ephys properties found in the table and all metadata we could gather from the Methods section of the article. Each ephys property is stored exactly as it is reported in each article: mean +/- standard deviation or standard error and number of measurements.

Little processing is performed on the metadata entries since some of them are pre-defined categorical variables (species, strain, electrode type, preparation type and junction potential correction status). Continuous variables are checked against possible value thresholds (age and weight cannot be negative, recording temperature should be within a reasonable range). In the text-mining and curation stage we record the sentences containing external and internal solutions, but little analysis can be done on solutions described in text. We define ‘compounds of interest’ that are very commonly used in electrophysiological solutions (Ions: Ca, Mg, Na, K, Cl, Cs; Other: glucose, ATP, GTP, HEPES, EGTA, EDTA, BAPTA). We programmatically extract the reported concentrations of these compounds and disregard the rest of the solution text. Importantly, we also disregard each individual compound’s dissociation constants since we do not possess that information for every chemical used in electrophysiological experiments at the recording temperature values.

**Post-processing of the data spreadsheet in RStudio and reasons for each assumption / correction.**

The CSV data spreadsheet was imported into a local installation of RStudio (R version 3.3.0). Several filtering and processing steps were required to clean up the data since this article focuses on the effect of metadata and, more specifically, solutions on the resulting ephys values. Subsequently, we have filtered out articles that did not have any solutions associated with them in our database. Possible reasons include: solutions used were described in another paper and only cited in the article of interest (more curation work is required to deal with these cases), solutions were missed by both text-mining and curation efforts and the least likely explanation is that solutions were simply not reported by the authors.

Here, we use 11 most commonly reported electrophysiological properties (in order of abundant to sparse: input resistance, RMP, AP threshold, AP amplitude, AP half-width, membrane time constant, AHP amplitude, rheobase, maximum firing frequency, cell capacitance, adaptation ratio) and ignore the rest due to their sparsity. Our statistical pipeline does not depend on the number of ephys properties analyzed, but it does require a reasonable number of articles reporting them, otherwise the resulting models would have insignificant explanatory power. With more articles being added to NeuroElectro, our analysis can be applied to more ephys properties. For a full and most up-to-date list of ephys properties visit [neuroelectro.org/ephys\_prop/index/](http://neuroelectro.org/ephys_prop/index/).

The most important data preparation steps include: assigning a default compound concentration and reversing appropriate ephys values correction for junction potential. The first step involves assigning an arbitrary small default concentration of 10-6 millimoles (mM) for each ion concentration that was not mentioned in the solution sentence. We calculate reversal potentials in our analysis and the data would otherwise be too sparse. Ideally, the second preparation step would be done in the other direction – correcting resting membrane potential and action potential threshold values for junction potentials in the articles that did not do so themselves. Liquid junction potential is the voltage difference between two different solutions that are in contact with each other. Unfortunately, we discovered that only 46% of articles report junction potential value and 24% correct for it, thus we decided that it would be easier to reverse the corrections (since the articles that correct for junction potential are likely to report its value) than attempt to impute unreported junction potentials based on solutions used. It was important to address the junction potential problem to standardize the ephys properties like resting membrane potential and AP threshold.

**Model creation**

To predict electrophysiological property quantitative values from experimental metadata attributes, we built models for predicting each of the 11 most abundant ephys properties using RandomForest package in R, adapting a previously developed approach (Tripathy, 2015).

It implements a supervised machine learning approach that is based on creating decision trees from samples of our dataset. Each decision tree gives its best prediction of what the ephys value should be given a set of metadata values, then these predictions are combined across all decision trees to produce a final predicted value.

To evaluate model performance, we examined the predictive power of each of the 30 metadata features using 10-fold cross-validation utilizing a leave-one-feature-out and use-one-feature approaches. For each of the 10 folds we ensured that unique articles were assigned to each fold, otherwise the metadata would be predicting PubMed ID instead of ephys properties. We have also tested several thematic models, examples include only using chemical compound concentrations, neuron types, other metadata or combinations of them. We then select the most valuable features using the internal feature importance (varimp) of the cforest (party R package) implementation of random forest. We then use AICc to compare using various number of top features for ephys property prediction. AICc assigns a score to each model based on its performance, adjusted for the number of features used and the amount of data that is available.

Based on the results we propose a model for each of the commonly reported ephys properties that allows to reduce the effects of experimental conditions on the ephys values.

**Data and code availability**

The python code used for text-mining and preprocessing is incorporated into the NeuroElectro codebase and can be found on <https://github.com/neuroelectro/neuroelectro_org> in assign\_metadata.py file. The most up-to-date CSV data spreadsheet can be found at <http://neuroelectro.org/static/src/article_ephys_metadata_curated.csv>. The R files with the data wrangling, analysis and model creation are located in <https://github.com/dtebaykin/neuronephys>.

**Results**

The main goal was to measure the impact of experimental solution recipes on the results of electrophysiological experiments. We performed an analysis of 882 published intracellular neurophysiology articles to explore common experimental solution compositions and to discover the general effects of experimental solutions on neuronal electrophysiology. We then extend the analysis to include previously known sources of ephys variability (Examples: animal species, age, type of electrode used, recording temperature) and compare their relative impact. Finally, we propose models for several commonly reported ephys properties that allow adjusting the ephys values from one set of experimental conditions to another. We validate these models using a new dataset provided by Allen Institute for Brain Science.

**Data overview**

In NeuroElectro, we have gathered electrophysiology, neuron type data and experimental conditions (metadata) from text-mined and manually curated neuroscience articles. NeuroElectro does not have access to the original raw experimental measurements (i.e. voltage traces), instead the ephys values are curated as population means with standard errors and number of samples (Tripathy et al., 2014). The dataset primarily contains ephys data reported under normal control conditions (control samples, defined by the original paper), enabling comparisons across articles. The NeuroElectro curation team has identified 1588 neuron type mentions in the collected articles. For each of these mentions, ephys data and metadata information can be reported or missing from the article’s methods section. Each data entry in NeuroElectro is annotated with one of 120 neuron types that are defined by the extended dictionary, originally provided by NeuroLex.org. The full list of NeuroElectro neuron types can be found here: <http://neuroelectro.org/neuron/index/>.

|  |  |  |  |
| --- | --- | --- | --- |
| **Entity name** | **Quantity**  **(rows of data)** | **Entity name** | **Quantity**  **(rows of data)** |
| **Unique PubMed ID** | 882 | **Solutions metadata:** | |
|  | | **External [Na]** | 1471 |
| **Electrophysiological properties:** | | **External [K]** | 1466 |
| Input Resistance (Rin, rin) | 1435 | **External [Cl]** | 1486 |
| Resting Membrane Potential (rmp) | 1314 | **External [Mg]** | 1478 |
| Action Potential Threshold (apthr) | 935 | **External [Ca]** | 1479 |
| Action Potential Amplitude (apamp) | 990 | **External [Cs]** | 2 |
| Action Potential Half-Width (aphw) | 980 | **External [glucose]** | 1446 |
| AfterHyperPolarization Amp. (ahpamp) | 687 | **External [HEPES]** | 84 |
| Membrane Time Constant (τ, tau) | 682 | **External [EGTA]** | 6 |
| Adaptation Ratio (adratio) | 308 | **External [EDTA]** | 0 |
| Rheobase (rheo) | 303 | **External [BAPTA]** | 0 |
| CellCapacitance (cap) | 258 | **External [ATP]** | 4 |
| Maximum AP Frequency (maxfreq) | 229 | **External [GTP]** | 4 |
|  | | **Internal [Na]** | 1119 |
| **Neuron Type** | 1588 | **Internal [K]** | 1466 |
|  | | **Internal [Cl]** | 1340 |
| **Basic metadata:** | | **Internal [Mg]** | 1217 |
| **Species** | 943 | **Internal [Ca]** | 244 |
| **Strain** | 887 | **Internal [Cs]** | 60 |
| **Electrode Type** | 943 | **Internal [glucose]** | 46 |
| **Preparation Type** | 943 | **Internal [HEPES]** | 1241 |
| **Recording Temperature** | 1511 | **Internal [EGTA]** | 797 |
| **Animal Age** | 1388 | **Internal [EDTA]** | 3 |
| **Animal Weight** | 272 | **Internal [BAPTA]** | 30 |
| **Junction Potential** | 1588 | **Internal [ATP]** | 1193 |
| **Junction Offset** | 551 | **Internal [GTP]** | 1083 |

**Table 1: Summary of data stored in NeuroElectro database. Color highlights: green – 11 most commonly reported ephys properties, yellow – neuron type mentions defined by NeuroLex, orange – basic metadata, blue – recording (external) and pipette (internal) solutions metadata. Data extracted on: 25.09.2016**

An ephys property can be reported multiple times in the same article (once per measured neuron type), resulting in the total number of measured properties (Examples: Rin, RMP) exceeding the number of articles. NeuroElectro only contains data that authors choose to publish, therefore

some ephys properties are not reported as consistently (Examples: rheobase, capacitance, maximum firing frequency). My analysis focuses on the top 11 commonly reported ephys properties (EPs), shown in table 1. The full up-to-date list of ephys properties can be found here: http://neuroelectro.org/ephys\_prop/index/.

One of the challenges of comparing values of ephys properties across studies stems from inconsistent definitions. For example, action potential spike amplitude can be measured from resting membrane potential (Cui et al., 2011; Perkowski and Murphy, 2011) or AP threshold (Novkovic et al., 2015). The adaptation ratio, defined in NeuroElectro as the ratio of durations between early and late APs inter-spike intervals (ISI) in an AP train, is even less standardized. It can be reported as a ratio of first / last ISI (Nassar et al., 2015; Scorza et al., 2011), a ratio of last / first ISI (Novkovic et al., 2015; Zhou et al., 2015), a percentage (Fujiwara-Tsukamoto et al., 2004; Zaitsev et al., 2009), or 1 – ratio first / last ISI (Derchansky et al., 2008; Lamsa et al., 2007). In each of these cases, NeuroElectro curators have standardized these ephys measurements for the different baselines. However, there are many other reporting inconsistencies that the curation team has not been able to address. These examples simply outline the types of problems in attempting to aggregate electrophysiological data that go above and beyond the effects of experimental conditions metadata.

Here, we distinguish experimental conditions (metadata) stored in NeuroElectro into two types: basic (recording temperature, animal age, species, etc.) and solutions metadata (pipette and extracellular concentrations of ions and compounds). Typically, all metadata is curated once per article and then copied into all rows of data extracted from that article. The exceptions are articles that alter experimental conditions between measurements. There are 4 basic metadata types in NeuroElectro (preparation type, animal weight, junction potential and junction offset) that are not used directly in my analysis. We only use *in vitro* studies for modeling ephys properties, animal weights get converted to animal age, junction potentials and junction offsets are used to standardize RMP and AP threshold values before the analysis (see Methods for more details).

### Analysis of experimental solution recipes used by neurophysiologists

The first step to understanding the effects of solutions on electrophysiological variance is determining the magnitude of variance within solutions themselves. We text-mined concentration values of ions and compounds commonly used in ACSF and pipette solutions during ephys recordings.

We observed the following general trends throughout the literature: external solutions use ~150 mM of sodium and ~130 mM of chloride with small amounts (1-3 mM) of magnesium and calcium (Figure 4). The potassium concentration is commonly kept very close to 0 mM; however, we identified a subset (~10%) of articles that include 5-6 mM of K into their artificial cerebrospinal fluid (ACSF) composition. On the other hand, there is a clear distinction between internal solutions used by patch-clamp and sharp electrodes: the former commonly uses ~140 mM of potassium with a wide variety of chloride concentrations (0-200 mM), magnesium (1-8 mM) and sodium (0-50 mM); while the latter tends to contain several moles of potassium, typically paired up with acetate, methylsulfate or chloride.

Next, we examined the distributions of the 26 extracted major ions and other compounds concentrations in experimental solution recipes. We excluded chemical compounds that were used less than ten times from this recipe analysis. The concentration values of major ions (Na, K, Cl, Ca, Mg) used in extracellular solutions generally follow approximately normal distributions (Figure 4A). However, the other common compounds concentrations are not normally distributed across the recipes: glucose is primarily used at a concentration of 10 mM, with the rest of the recipes increasing it up to 40 mM; and HEPES is usually not included into ACSF, but in ~5% of the recipes it was present at a concentration of 10 mM. In many cases, HEPES was included into external solutions of cell culture ephys experiments, however, it was also included into dissociated neuron experiments (Gittis and Lac, 2007) and two-photon guided *in vivo* whole-cell recordings (Chen et al., 2015).

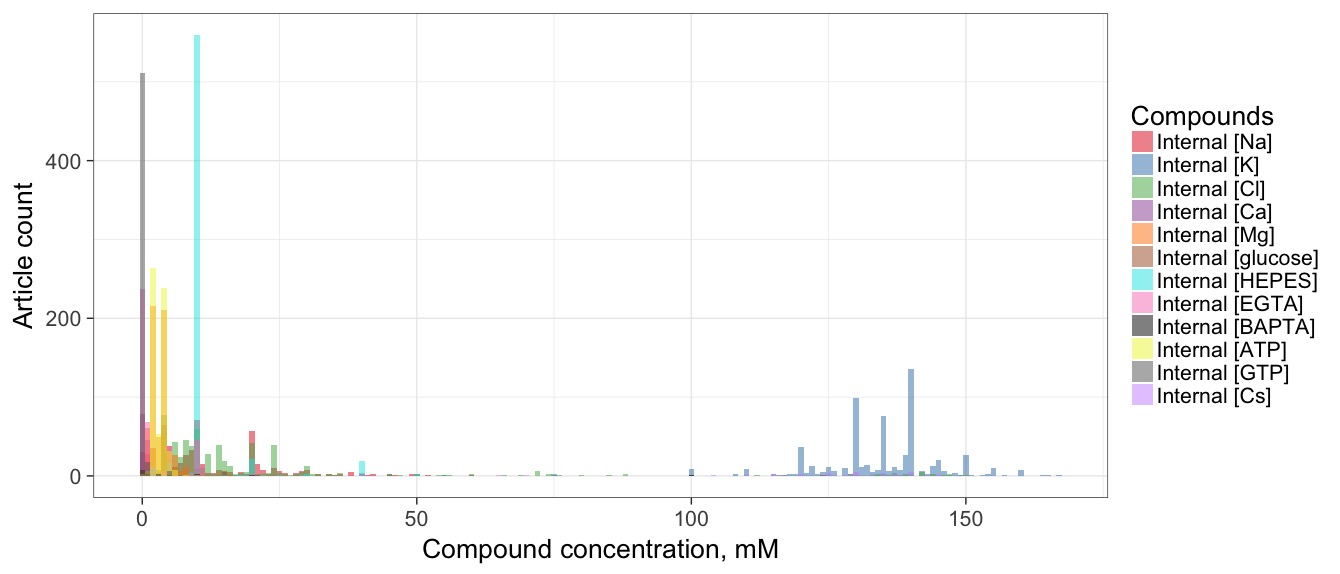
The recipes of intracellular solutions vary more than the compositions of extracellular solutions. Potassium concentration values are almost uniformly distributed, except for the large peaks at

130, 135 and 140 mM (Figure 4B). Cesium concentration values closely follow potassium’s distribution, since voltage-clamp recordings utilize cesium-based recipes to block K channels.



B

A



**Figure 1: Chemical compositions of experimental solutions. Data from 731 curated Patch-clamp solutions. Histograms of compounds that are commonly found in: A) External (extracellular, ACSF) and B) internal (pipette, electrode) solutions. The ion concentrations were calculated by summing concentrations of their respective compounds, assuming complete dissociation. Histogram bin width is set to 1 mM on the main plots and to 0.5 mM on the 0-15 mM histograms. Arrows denote CSF composition as described in medical literature: “142 Na+, 2.5 K+, 1.3 Ca++, 0.8 Mg++, 124 Cl-, 3.9 glucose” (Hall, 2015).**

There is a lot of variation in internal sodium and chloride concentrations, though both distributions are skewed towards the 0-30 mM range. Magnesium and ATP are predominantly used at 2 mM and 4 mM concentrations. Additionally, HEPES, GTP and EGTA are consistently included into pipette solutions at the respective concentrations of 10 mM, 0-1 mM and 0-11 mM. Finally, calcium, glucose and BAPTA are rarely included into the intracellular solutions at small concentrations (Gall et al., 2003; Goldfarb et al., 2007; Prestori et al., 2008). It is important to remember that the concentration values extraction algorithm performs at roughly 90% accuracy, and while the above concentration value distributions represent the true summary of recipes used by electrophysiologists, some of the edge cases (internal Ca, glucose and BAPTA concentrations) have slightly inflated numbers.

The above exploration gives the impression of a great deal of at least minor variability in solution makeups. A quantitative analysis confirms this: considering only the five major ions, there are 358 (49%) different external and 482 (66%) different internal solutions in my data, with 603 (82%) papers using unique combinations of the two – out of 731 possible patch-clamp solution recipes. The most frequent ACSF recipe was used 62 times, (in mM): 151.25 Na, 2.5 K, 133.5 Cl, 1 Mg, 2 Ca. This recipe was most commonly used by the Spruston lab: 6 times over the course of almost 20 years: “ACSF consisted of 125 mM NaCl, 2.5 mM KCl, 25 mM NaHCO3, 1.25 mM NaH2PO4, 1 mM MgCl2, 2 mM CaCl2, and 25 mM dextrose” (Cembrowski et al., 2016; Cooper et al., 2003; Golding et al., 2005; Graves et al., 2012; Lübke et al., 1998; Staff et al., 2000). While there was little consistency overall, the differences between recipes were generally minor. For example, external Na varied from 150 mM to 153 mM (interquartile range; see Figure 4).

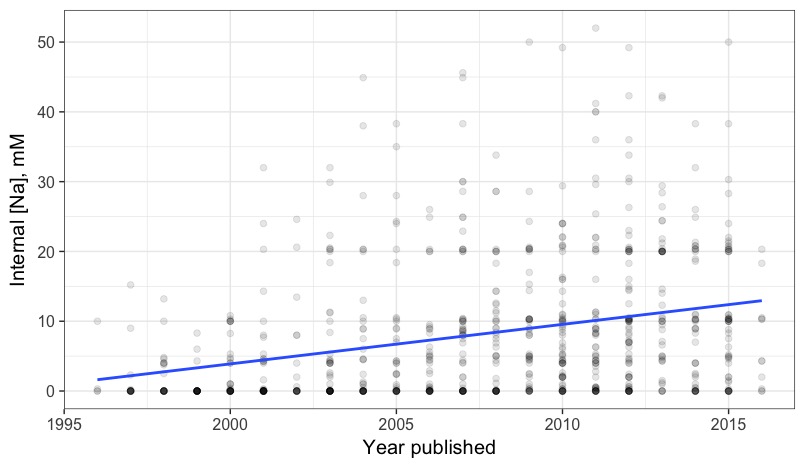
Patch-clamp intracellular solutions are even more diverse: out of N recipes from 731 papers, only 55 were used twice, 24 – 3 times, 2 – 4 times, 4 – 5 times, 2 – 6 times and single recipes were used 7 and 8 times, the other 481 recipes were unique. The pipette solution that was used 7 times contained, (in mM): 120 K, 6 Cl, 4 Mg; and the one that was used 8 times, (in mM): 140 K, 14 Cl, 4 Mg. Among the patch-clamp articles, 41 recipes for both solutions were shared between 2 articles, 8 recipes – 3 articles, 4 recipes – 4 articles and 1 recipe was the same in 6 articles.

Out of the 128 curated Sharp electrode articles: 5 ACSF recipes were used 2 times, 5 – 3 times, 3 – 4 times, 1 – 5 times. This most common recipe was, (in mM): 151.25 Na, 3 K, 131 Cl, 2 Mg, 2 Ca. The pipette solutions of Sharp electrodes are less diverse: 3 recipes were used twice, 1 – 4 times, 2 – 5 times, 2 – 14 times (1 M and 4 M of K), 1 – 19 times (3 M of K) and 1 – 34 times (2 M of K).

To explore possible patterns in recipe creation for recording and pipette solution, we used principal component analysis supplemented by hierarchical clustering to identify trends in recipe creation for recording and pipette solutions. No obviously distinct clusters presented themselves, suggesting that electrophysiologists use similar recipes with slight variations, within biologically reasonable concentration values. However, two trends were identifiable among the Patch-clamp solution recipes: internal solutions could be separated into those with low Na, Cl concentrations and high Na, Cl (Table 2); external solutions can be split by their relatively low and high Mg concentrations. It is important to note that cesium-based solutions (associated with voltage-clamp experiments) were generally avoided during the curation process, because NeuroElectro curation heavily prioritizes current-clamp experiments, as it is focused on ephys properties such as action potential characteristics.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Chemical** | **Internal Solution, mM** | | **External Solution, mM** | | |
| **Low Na, Cl**  **(N = 353)** | **High Na, Cl**  **(N = 183)** | **High Mg (N = 276)** | | **Low Mg (N = 371)** |
| **Na** | **0 - 10** | **15 - 50** | **140 – 160** | | |
| **K** | **120 - 150** | | **1 - 5** | | |
| **Cl** | **0 - 30** | **15 - 50** | **125 - 140** | | |
| **Cs** | **0** | | **0** | | |
| **Mg** | **0 - 10 (81% in 2-4 range)** | | **2 – 2.5** | **1 - 1.5** | |
| **Ca** | **0 - 1 (95% use 0 mM)** | | **2 - 3** | | |
| **HEPES** | **5 - 15 (96% use 10 mM)** | | **0-10 (89% use 0 mM)** | | |
| **EGTA** | **0 - 10 (87% use 0 mM)** | | **0** | | |
| **ATP** | **5 - 10** | | **0** | | |
| **GTP** | **0 - 10 (~95% use 0 mM)** | | **0** | | |
| **glucose** | **0** | | **0 - 25** | | |

**Table 2: Summary of trends in electrode and recording solution designs. In this general trend analysis, outlier recipes were not considered. Number of articles analyzed: 703 Patch-clamp, in vitro studies performed on rats, mice or guinea pigs. N is the number of articles with the specified solution composition.**



r = 0.29

p < 0.001

**Figure 2: Internal sodium concentration increased in 2003. Boxes represent solution concentrations from articles published in the corresponding year (X-axis). The blue line is a linear fit between internal sodium concentration and publication year. Internal sodium concentration significantly increases throughout the years (r = 0.29, p < 0.001).**

An interesting variation in Patch-clamp pipette solution recipes is internal sodium concentrations increase throughout the years (Figure 5). It seems to be caused by the introduction of 10-20 millimoles per litre of Na2-phosphocreatine to internal solutions, which became popular in mid-2000s. The implication is that recipes do change over time and it is entirely possible that a single lab or a small set of labs can start new trends in the designs of solution recipes.

This exploration of the distributions of solution components and verifying that neurophysiologists tend to use at least slightly different solutions, we proceeded to the task of determining whether these solution component variations help to further explain the variability of commonly reported ephys properties.

### Assessing study-to-study electrophysiological variability

Electrophysiology values might have relatively large intrinsic cell-to-cell variability, which could conceal the effects of experimental conditions on the results (Tripathy et al., 2015). To assess this, we considered whether between-experimental variance for a single neuron type is greater than within-study variance. If the within-study variance was higher, the meta-analysis approach would likely yield inconclusive results.

In the context of a single experiment, the scientist measuring RMPs of hippocampus CA1 pyramidal neurons expects to observe values that are approximately normally distributed, with a sample mean providing an estimate of the population mean. If experimental conditions do not introduce significant variance when comparing ephys properties across studies, then multiple electrophysiology studies should report similar ranges of values while measuring from one neuron type. Figure 1 shows mean +/- standard error of the mean for three relatively common neuron types in NeuroElectro, after correcting for junction offset. Disregarding several outliers, SEMs do not cover the whole range of reported RMP means. In the case of hippocampus CA1 pyramidal cells the mean RMPs range from -73 mV to -53 mV with an average SEM of 1.7 mV. There is an even greater spread in the reported mean resting membrane potentials in Martinotti cells (from -73 mV to -48 mV with a mean SEM of 2.6 mV) and medium spiny neurons (-95 mV to -61 mV, mean SEM of 2.8 mV), still with relatively small standard errors. These data do not support the hypothesis that different electrophysiology experiments report ephys values from the same normal distribution (ANOVA P-value of 4.04\*10-15 for RMPs of hippocampus CA1 pyramidal neurons). We found that, other ephys properties behave very similarly to RMP (not shown). Thus, the hypothesis that ephys measurements are unaffected by experimental

../../Neuroelectro%20documents/Plots/exampleRMP.pdf

Article index

**Figure 3: Electrophysiological variability is higher between experiments than within experiments. Resting potentials of hippocampal CA1 pyramidal neurons, neocortex Martinotti cells and Striatum medium spiny neurons, across articles in NeuroElectro. Each point and line is an RMP mean +/- SEM, reported by an article.**

conditions must be false. These inter-study differences must be partially due to differences in experimental procedures.

Since the RMP means for a single neuron type reported in different articles are highly unlikely to originate from the same normal distribution, we hypothesize that there are factors contributing to the high variability of resting membrane potentials when compared across labs. This argument holds for other electrophysiological properties. In fact, certain experimental conditions (animal species, age, electrode type, recording temperature) have been previously shown to be systematically correlated with variance in ephys measurements (Tripathy et al., 2015). This analysis motivated my consideration of experimental solution compositions as a potential explanation for inter-study variance.

### Modeling electrophysiological properties with experimental metadata

B

A

../../Neuroelectro%20documents/Plots/rinVsNaInt.pdf../../Neuroelectro%20documents/Plots/rmpVsMgExt.pdf

r = 0.35

p < 0.001

r = -0.16

p > 0.05

**Figure 4: Univariate relationships between electrophysiological properties and solution concentrations. Each point is a mean ephys value reported by an article for Hippocampus CA1 pyramidal neurons. Blue line is the best univariate linear fit for the data, grey area shows 95% confidence interval for the linear fit. A) Input resistance increases with internal sodium concentration (r = 0.35, p < 0.001). B) RMP linear model is driven by 3 outliers in the 7-7.5 mM range of external magnesium concentration, rendering its results insignificant.**

The simplest approach to studying the effects of major ion concentrations on ephys properties is to model the ions one at a time. This approach ignores the possibility of more complicated interactions, and cannot handle multiple cell types. However, it can identify strong correlations between specific solution components and ephys properties in a single cell type. Two examples of this univariate approach, when applied to hippocampus CA1 pyramidal neuron type: input resistance significantly correlates with internal sodium concentration (Figure 7A). As an example of a negative finding, the relationship between resting membrane potential and external magnesium concentration is driven by outliers (Figure 7B).

Systematically applying this univariate linear modeling approach to each neuron type, we did not find significant relationships between ephys data and individual ion or compound concentrations (FDR < 0.05). Confounding effects of other factors (age, species, electrode type, other solution components) likely mask true correlations if any exist. Searching for articles that have the same methods apart from a single solution component was not a feasible approach due to sparsity of the dataset. There are too few articles that use the same experimental conditions except one, at that point the limited sample size would render the analysis statistically underpowered. Subsequently, we considered a multiple regression approach that incorporates the influence of several experimental parameters on the same ephys property simultaneously.

Building on the regression approach developed previously (Tripathy, 2015), we hypothesized that it should be possible to model the effects of solution parameters on the resulting ephys measurements. To that purpose, we used a Random Forest machine learning algorithm to construct regression models relating ephys properties to metadata features (described in detail in Methods). The models were designed to simultaneously capture the effects of neuron type, solution composition information and basic metadata like species, age, temperature, electrode type (Table 3). We chose Random Forest over the classic linear regression approach because it is a non-linear model that empirically better handles statistical overfitting (which would cause a failure to generalize well to unseen test data) when using datasets with many features relative to sample size (Breiman, 2001). All models were trained and tested using 10-fold cross-validation, with performance summarized by R2. An R2 value of 1 means that the model was 100% correct in all predictions (which is essentially unattainable). An R2 value of 0 means that the predictions are as accurate as using the mean value of the training ephys data for the predictions of test samples. A negative R2 means that the model performs worse than the mean because of overfitting to the training data. Additionally, we define a ‘baseline’ R2 value (calculated to be -0.30), which is generated by randomly shuffling ephys values. It serves as a lower bound for the worst predictions that could be made when the model is essentially predicting noise. When my models consistently achieve positive R2 values, they should be used for predicting ephys values (instead of using the mean value). We consider models with negative (but above the lower bound, figure 9) R2 values capable of explaining a small amount of ephys variability, however they should not be used for predicting ephys values.

To compare the effect of solutions metadata to basic metadata when modeling the variability in ephys properties, we designed several models: neuron type only, neuron type + basic metadata, solutions only, neuron type + solutions, basic metadata only and all three sets of features combined. We expected the all features model to have the best performance since it has access to the information other models lack.

Applying the Random Forest algorithm to the data, we used a model that related all metadata features to input resistance (Figure 3A). The predicted values are the model’s best estimates of what the observed ephys values should be given the experimental conditions from each article. In general, the predicted ephys values have less variance than the observed ones. That behaviour is expected, because the models can only partially predict the ephys variance.

The next step was to evaluate the relative contributions of neuron type, basic metadata and solutions when predicting ephys properties, starting with input resistance (Figure 3B, metadata details listed in table 1). Since the folds are assigned to articles in a random fashion, the performance of the models in each fold is slightly different. However, all 6 models are run using the data from the same 10 folds, there is no reshuffling of data between different models. Judging by the model performances, solution features help to predict input resistance.

B

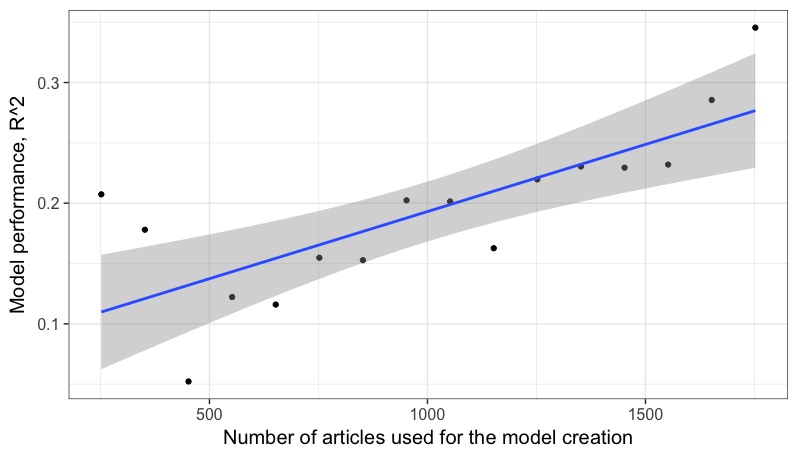
A



**Figure 5: Multivariate regression models can predict ephys properties. Predictions are performed on held-out data (10x cross-validation). A) Each point is an input resistance value, reported by an article and predicted by a model using all metadata features (1 fold). B) Comparison of 6 different models for input resistance, each model uses a different set of features. Briefly, Neuron Type (NT) indicates a model using neuron type information only, basic metadata refers to information like animal age, recording temperature, etc., solutions refer to the use of internal and external solution concentrations, and all features refers to the combined set of metadata.**

All features and the neuron type + solution features perform very similarly and better than neuron type + basic metadata. However, solutions on their own perform worse than basic metadata. It could mean that neuron type and basic metadata features provide similar information to the models, whereas solutions explain additional variance in input resistance. Neuron types alone cannot predict input resistance values as well as in conjunction with basic and solution features.

To formalize the effect of available data on a model’s performance, we quantified the effect of varying the number of data points that a model can use to predict an ephys property. There is a strong correlation between the number of articles for an ephys property and the R2 values of a model that predicts it (Figure 5).

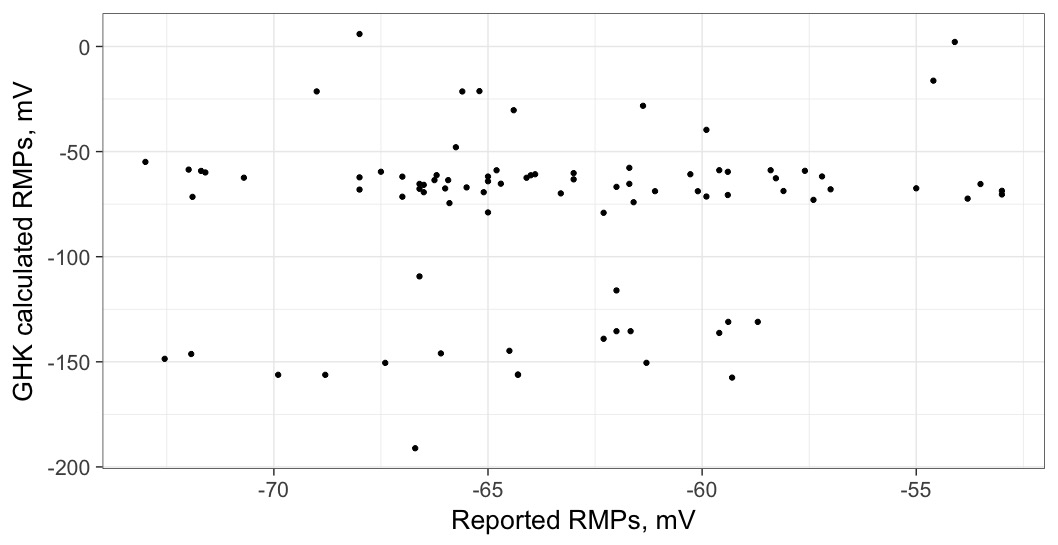


r = 0.76

p < 0.001

**Figure 6: Multivariate model performance improves with N. R2 performance for predicting input resistance with N rows of data. Blue line is the linear model best fit; grey region represents 95% confidence interval for the fitted line.**

Additionally, we evaluated the performance of the GHK equation for modeling membrane potentials of neurons at rest by comparing its predictions based on recording temperature and Na, Cl, K concentrations and text-book permeability values (Hille, 1984) to the observed ephys values. Strikingly, the GHK model essentially failed to predict the RMP values of hippocampal CA1 neurons (R2 < 0).

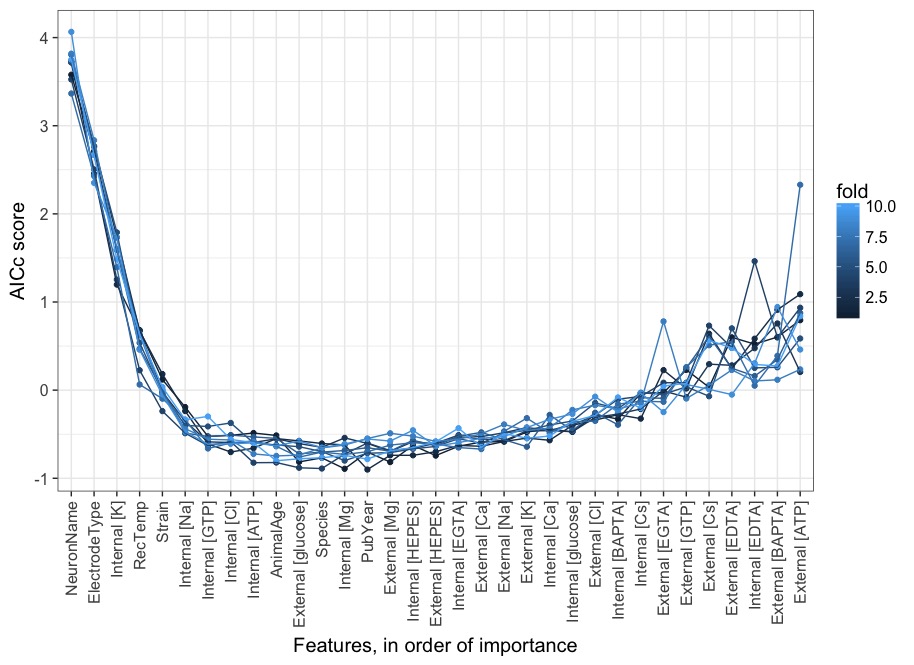


R2 = -0.06

**Figure 7: Predicting RMPs of hippocampus CA1 pyramidal cells with the GHK equation. Each point is a mean RMP reported by a single article in NeuroElectro. GHK calculated RMPs refer to the usage of experimental metadata stored in NeuroElectro for the prediction of resting membrane potentials.**

**Optimizing multiple regression models for predicting specific electrophysiological properties**

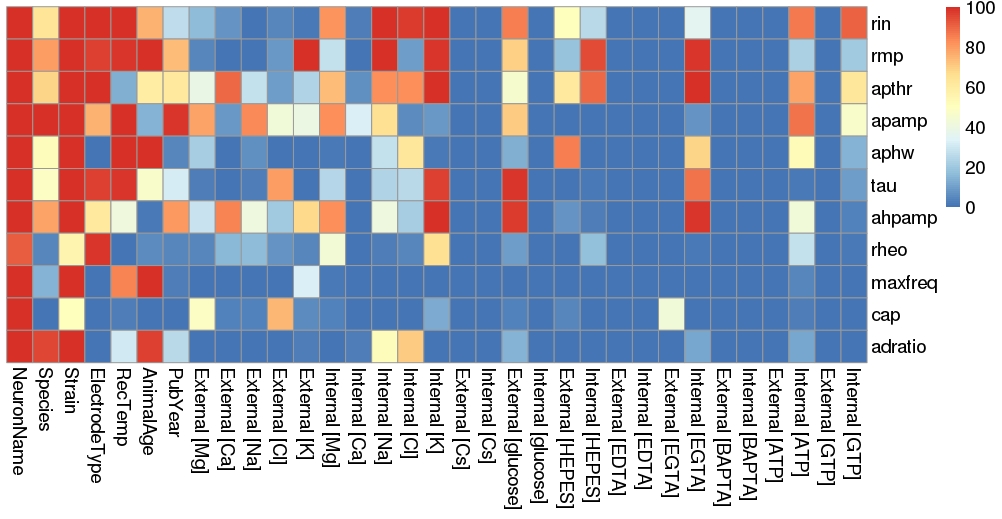
The final step of the project was to create models using only the important features, which might be different for each ephys property. We used Random Forest internal variable importance and AICc to create and evaluate models for each ephys property. The lower AICc is, the better the model.



**Figure 8: Model comparison using AICc score. One run of 10-fold cross-validation, each line is an AICc curve (per fold) calculated by adding top X (from 1 to all 33) features to the model that predicts input resistance. Model with the lowest AICc score is the best performing one. Metadata features are ordered from high to low based on their importance, calculated by cforest (X-axis).**

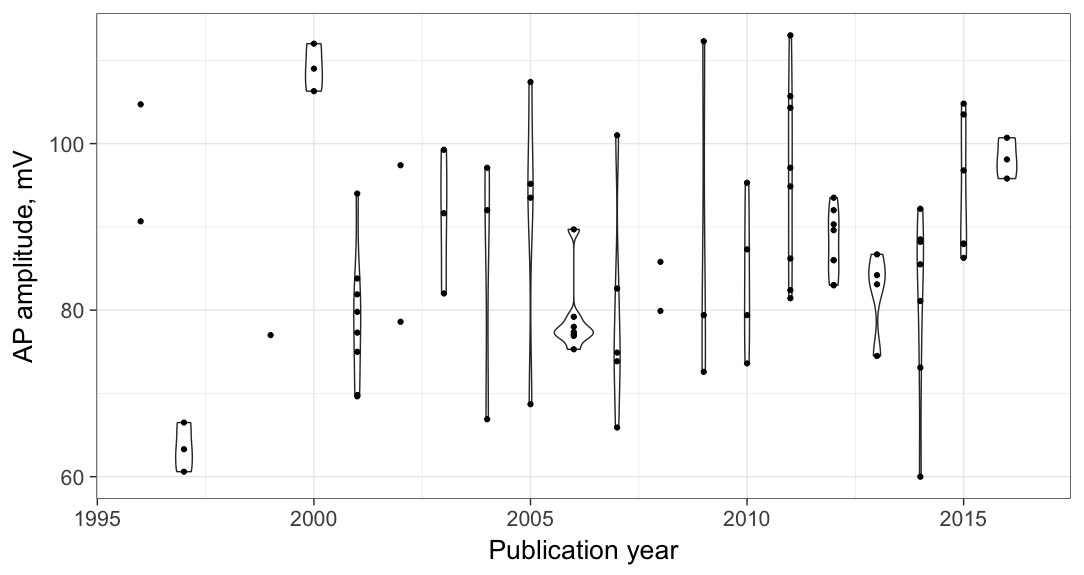
Adding the first 6 features results in a big drop off, meaning that those features should always be included into the best input resistance model. After that, the AICc curve shifts up and down, depending on the fold. Finally, adding any features after external calcium and sodium would hurt the model rather than help it (at least at the current amount of data, this might change when more articles are added to NeuroElectro). The last 7 features illustrate the amount of instability and noise bad features can add to a model (the effect of overfitting), making its performance vary greatly between different folds.

Generalizing the above Random Forest variable importance ranking complemented by AICc approach for input resistance model selection, we applied the same algorithm to the 11 of the most abundant electrophysiological properties in NeuroElectro. We have also performed this entire procedure 10 times for each ephys property to ensure we were getting stable results. Ten runs of the 10-fold cross-validation are summarized in Figure 8.



**Figure 9: Feature importance, based on the frequency of inclusion into the models. Ephys properties and metadata features are listed vertically and horizontally, respectively. Color represents the number of times the feature was chosen for the ephys property’s model (from 0 to 100 times). NeuronName stands for neuron type.**

Neuron type and strain features are almost always chosen for modeling each ephys property. The species feature is included less frequently because most of its contribution is covered by strain, which is also more informative, because 93% of the data in NeuroElectro comes from the experiments performed on mice or rats. A few solution features get consistently included for at least 1 ephys property as well: external K for RMP, internal Na for Rin and RMP, etc. It reinforces the earlier observation that certain solution components are important when predicting specific electrophysiological properties. Surprisingly, the publication year was an important feature when modeling AP amplitude (Figure 14). We hypothesize that this effect can be explained by changes in the AP amplitude calculation or measurement protocols.



**Figure 10: Reported action potential amplitudes of CA1 pyramidal cells vary with time. Each point is a population mean APamp value of Hippocampus CA1 pyramidal cells, reported by a single article published in the corresponding year. Violins outline the distributions of values for each year.**

Another important aspect of the feature selection heatmap is that very few cells are colored yellow. It means that most features are either robustly good or robustly poor when predicting the corresponding ephys property. The poor performance of certain solutions features (BAPTA, EDTA, Cs, etc.) can be linked to their sparsity. If a solution component is included into <5% of recipes, it is unlikely to be predictive of electrophysiological variance between studies stored in NeuroElectro. However, NeuroElectro curation generally targets current-clamp experiments, naturally causing low Cesium (a common voltage-clamp compound) inclusion rates. When a feature gets included into the best model <50% of the times, its performance is unstable, likely due to overfitting. Between 50% and 90% inclusion is the uncertain area where the feature might not be important enough to be included all the time but it does provide some useful information. The final models were created using features that are included in >90% of the best models to minimize overfitting.

##### Validating optimized models with NeuroElectro and Allen Institute Cell Types data

Expanding input resistance modeling to 11 commonly reported ephys properties, we evaluated the effectiveness of each model type in predicting them (Figure 10).

Here, the null-model (using ephys property means as predictions) has an R2 of 0. The figure can be interpreted as: the models that perform better than the null-model in all 10 folds can capture a significant amount of the ephys variance; the models with R2 values between baseline (-0.3) and 0 can explain some amount of variance in the ephys property, but not enough to outperform the null-model. Input resistance and resting membrane potential models can explain a significant amount of their respective variance, however, in most cases the models are only slightly better (if at all) than simply taking an average of the observed values and using that as an estimate for the ephys property. Interestingly, the 4 out of 5 properties on the right-hand side of the plot (AHP amplitude, rheobase, maximum firing frequency and adaptation ratio) get the best predictions out of neuron type only model. Additionally, ephys properties with less available data (Ordered from left to right: abundant to sparse) have much less stable performance levels. These effects are likely to be artifacts of not having enough data to sufficiently train the multiple regression Random Forest models. We observed a general increase of 0.2-0.5 in the predictive power of my models when comparing to the baseline, implying that, in most cases, the models can partially explain the variability in ephys values. On average, solutions contribute less to the overall model predictive power than neuron name or basic metadata, however, in some cases they substantially increase all features models performances (Rin, APthr, APamp). Therefore, solutions can contribute different information than neuron name or basic metadata when modeling ephys properties.

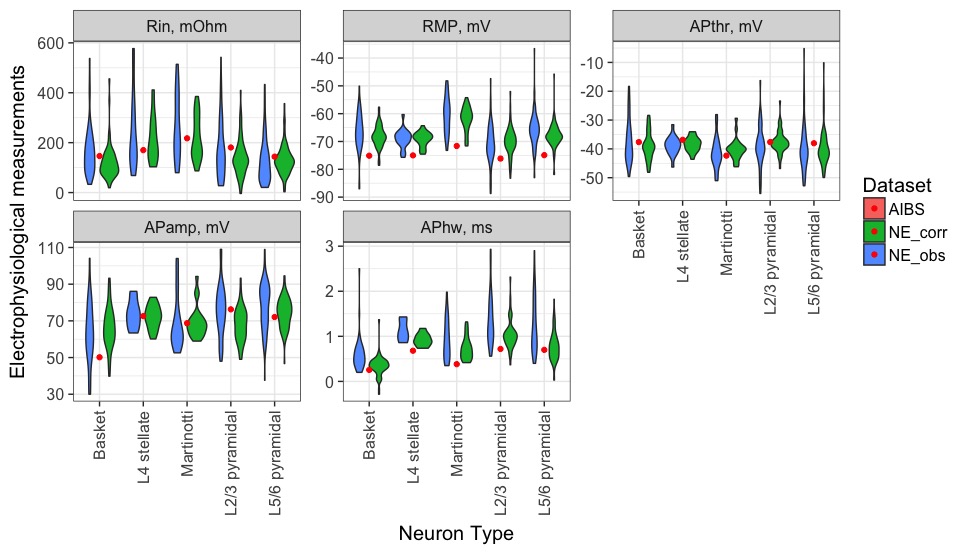
Having created the models for all ephys properties, we needed to compare their performance to the models used previously (Figure 10, ‘best features’ models). To achieve that, we once again employ 10-fold cross-validation and calculate R2 values for each model, substituting the neuron type + solutions model for the best features model. The optimized models achieve the highest performance levels (or on par with other models) when predicting Rin, APthr, APamp. However, they fall short of “no solutions” model swhen predicting RMP and APhw.

After comparing the feature-selected models to the basic ones, we applied these new models to data unused in the fitting process or cross-validation, from the Allen Institute for Brain Science. Only AIBS neuron types that could be definitively assigned to a NeuroLex cell type were used in this analysis. Since AIBS data was produced during a set of experiments in a single lab – all ephys property measurements were aggregated into mean values per neuron type, because NeuroElectro stores reported means, not individual measurements.



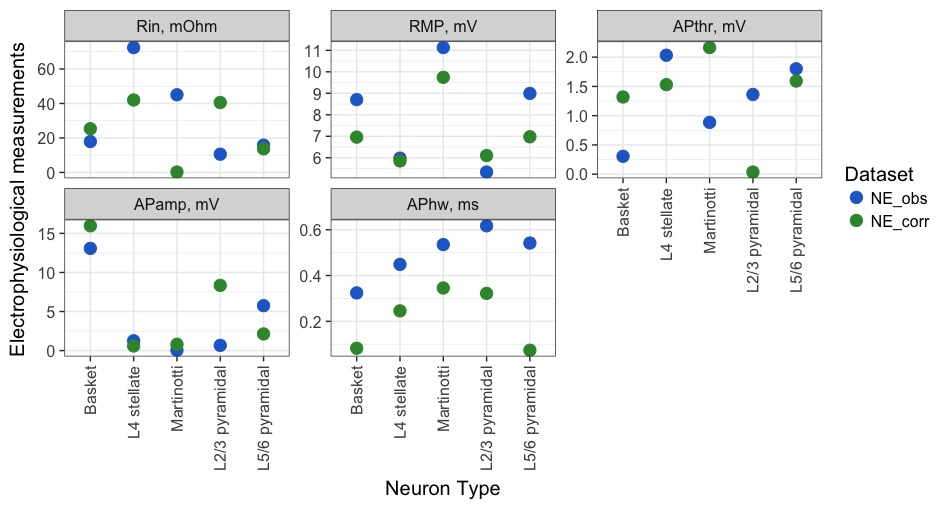
**Figure 11: Comparison of feature-selected models to basic models. The best model (per AICc) for each property is shown in green color. Variable refers to the list of commonly reported ephys properties. R2 value on the y-axis represents each model’s performance.**

To evaluate the feature-selected models performance, we used them to ‘shift’ NeuroElectro ephys data (NE) to AIBS experimental conditions baseline (See Methods for details). Briefly, we first predicted and removed ephys variance from NE data that could be explained by the models, then we added ephys variance introduced by AIBS experimental conditions.



B

A



**Figure 12: Adjusting NeuroElectro data to AIBS conditions. All AIBS neuron types come from neocortex. A) Violin plots of NeuroElectro data (blue), Allen Institute for Brain Science data (red), adjusted NeuroElectro data using the feature-selected models (green). Each point is a mean ephys property value reported by an experiment. B) Absolute differences between NeuroElectro raw and adjusted ephys values when compared to AIBS ephys means. The model correction tends to squeeze NeuroElectro data around the mean and bring it closer to AIBS value.**

If the models work, the adjusted NE data should be more closely distributed around AIBS electrophysiological measurements. Figures 11A and 11B support that claim. Generally, the corrected NeuroElectro ephys values have less variance, which is the result of removing the explained variance from each reported ephys value, and their means are in most cases closer to the AIBS mean ephys values than raw NeuroElectro data.