

動力的因果モデリングにおける 神経スパイク交差コレログラム解析について

大羽 成征^{1,2,a)} 中江 健¹ 吉本 潤一郎³

概要: スパイク統計に基づいて神経細胞間の機能的結合を知るための手法として、一般化線形モデル (GLM) に基づく応答関数推定法が提案されている。GLM によれば古典的な 2 ニューロン間クロスコレログラム解析で生じる擬似相関起源の偽陽性を排除できる。しかし、GLM によっても原理的に避けられない偽陽性要因が存在する。本発表ではそのうち 2 つを指摘する。第一は、スパイク検出時刻決定の誤差、第二は時間方向に非線形性の強いダイナミクスを GLM でフィッティングすることによって生じる歪みである。本発表ではまた、これらの偽陽性を避けるために、応答関数可視化法に関する小さなトリック「GLM クロスコレログラム」を提案する。これとクラスタリング法と組み合わせて、機能的結合パターンの典型例を見つけることで、通常手法を使った場合に導かれたであろう解釈誤りの可能性を減らせることを示す。

1. はじめに

神経細胞間の機能的結合の有無や、その詳しい様子を
知るために、神経スパイク計測データを用いるスパイク
統計の手法発展が顕著である。現時点で広く使われてい
る代表的手法として、一般化線形モデリング (generalized
linear modeling; GLM) によるスパイク応答関数解析が挙
げられる [1], [2], [3], [4], [5], [6]。GLM によるスパイク応
答関数解析は、動力的因果モデリング (dynamic causal
modeling; DCM) による実効的接続関係推定 [7] の思想を
スパイク統計に応用したものであり、多数ニューロンを経
由する信号通信ネットワークの動力的挙動を全体的にモ
デリングすることにより、古典的な 2 ニューロン間相互相
関解析 (cross correlogram) など生じる擬似相関起源の偽
陽性を避けることができる。

一般に、スパイク統計に基づいて推定される機能的結合
と、シナプス結合などによる物理構造的結合とは別のもの
であり、機能的結合が物理的結合を直接的に反映している
保証は無く、その検証は難しい。そこで、近年では機能的
結合のみを対象とすることで成立するデコーディング研究
が進んでいる。例えば [8][9] は、GLM による応答関数推定
を用いてニューロンのスパイク統計に基づく視覚的刺激デ

コーディングを行い視覚野ニューロンの受容野を調べてい
る。こうしたデコーディング研究では、ネットワーク全体
としての機能に各ニューロンがどのように関わるかが重要
であり、個々接続の有無に関する推定はさほど重要でない
ため、物理的・機能的結合の関係に関する議論を省略して
も影響が少ない。一方で、メゾスケール神経回路を構造レ
ベルで理解するために、依然として物理的結合の有無をク
リアに推定する必要がある。そのために、機能的結合推定
と物理的結合の関連について間接的な証拠を探す試みが続
いている。[10] はカルシウムイメージングに基づくスパイク
統計で推定された機能的結合が、物理的結合を良く反映
していることを間接的証拠から示した。一方で推定された
結合が、実際よりも多くの双方向接続を示す傾向も指摘し
ており、その原因については示していない。

本発表では、GLM による応答関数推定において生じて
いる可能性の高い偽陽性要因を二つ示す。第一は、スパイク
検出時刻決定の誤差、第二は時間方向に非線形性の強い
ダイナミクスを GLM でフィッティングすることによって
生じる歪みである。両者とも、GLM によるスパイク応答
関数推定の結果として、実際には単方向の結合であったも
のが双方向結合のように見えてしまう原因となる。したが
って [10] が示した偽陽性原因の一部は、我々が本発表
で示す要因から生じている可能性があり、これらの排除に
よって推定結果改善の余地がある。こうした偽陽性の可能
性は筆者らの知る限りこれまでに指摘されたことはなく、
これらを避ける方法も知られていない。

本発表ではさらに、解析者がこれらに起因する偽陽性の

¹ 京都大学情報学研究科
Gokasho, Uji, Kyoto 611-0011, Japan

² 日本科学技術振興機構

³ 沖縄先端科学技術大学院大学
1919-1 Tancha, Onna-son, Kunigami-gun, Okinawa 904-
0495, Japan

a) oba@i.kyoto-u.ac.jp

可能性を回避するために、GLM モデルの推定結果として得られた応答関数の可視化における簡単な工夫（GLM クロスコレログラムと呼ぶ）が有効であることを示す。

2. スパイク応答解析と GLM クロスコレログラム

スパイク応答解析の対象は、複数ニューロン $i = 1, \dots, C$ の各々について、その第 t 微小時間幅 ($t = 1, \dots, T$) におけるスパイク信号の有無 $N_i(t) = 1, 0$ をまとめたスパイク時系列データである。GLM に基づく確率的スパイク応答モデルでは、第 t 微小時間幅における各ニューロン i のスパイク信号が解析対象ニューロンの過去のスパイク活動 $N_c(t-1), \dots, N_c(t-M)$, $c = 1, \dots, C$ に依存した確率で生じると仮定する。

$$\Pr(N_i(t) = 1) = f(\lambda_i(t)), \quad (1)$$

ここで $\lambda_i(t)$ はニューロン i の全ニューロン $c = 1, \dots, C$ へのスパイク応答の合計であり、以下で表される。

$$\lambda_i(t) = R_{i0} + \sum_{c=1}^C \sum_{s=1}^M R_{ic}(s) N_c(t-s). \quad (2)$$

ここで $R_{ic}(s)$ はニューロン i のニューロン c に対するスパイク応答、 R_{i0} は定数項である。自然数 M は想定される応答遅れの最大値を表す。関数 $f(x)$ は、スパイク応答の合計 $\lambda_i(t)$ とスパイク発火確率との間の非線形関係を規定する任意の関数であり、以下の議論やシミュレーションでは $f(x) = 1/(1 + \exp(-x))$ のロジスティック回帰の場合を用いる。スパイク応答解析では、 $f(x) = 1 - \exp(-\Delta t \exp(x))$ のポアソン回帰を使うことも多いが、本発表の全ての結論はこれを含め他の場合でも同様に成り立つ。ここで Δt は時間フレームの大きさである。上記モデルとスパイク時系列データ $N_i(t)$, $i = 1, \dots, C$, $t = 1, \dots, T$ のもとで、スパイク応答 $R_{ic}(s)$, $i = 1, \dots, C$, $c = 1, \dots, C$, $s = 1, \dots, M$ を推定するとき、その推定結果はニューロン c からニューロン i への機能的結合を反映する。例えば結合が存在しないとき $R_{ic}(s)$ は全ての s においてゼロに近い値となることが期待できる。

スパイク応答関数 $R_{ic}(s)$ の推定値は、全スパイク時系列データ $N_i(t)$, $i = 1 \dots C$, $t = 1 \dots T$ のもとで、以下の対数尤度関数を最大化することによって得られる。

$$L(R) = \sum_i \sum_t \{N_i(t) \log p_i(t) + (1 - N_i(t))(1 - p_i(t))\} \quad (3)$$

ただし $p_i(t) = \Pr(N_i(t) = 1)$ とした。なお、なめらかな形状の応答関数を得るために以下のシミュレーションでの推定にはガンマ分布密度関数基底 [2] と L2 正則化項を用いた。

比較対照として 2 ニューロン i と c が生じさせるスパイク活動の相互相関を端的に示す指標として以下で定義されるクロスコレログラムを考えよう。

$$CG_{ic}(s) = \sum_{t=1}^T N_i(t) N_c(t-s). \quad (4)$$

ここで s は負の値もとりに得るものとする。

クロスコレログラムは、ニューロン間の機能的結合を反映する指標として古くから使用されてきたが、擬似相関による偽陽性が避けられない。たとえばニューロン i_1 が i_2 と i_3 への影響を与えており、 i_2 と i_3 の間の直接の関係が無い場合に、この二種類のモデルの違いが顕著に現れる。すなわちニューロン i_2 と i_3 の間に直接接続関係が無くとも、クロスコレログラムでは影響が見えてしまう。これを擬似相関と呼ぶ。しかし GLM によるモデリングによればこれら 3 ニューロンのスパイク時系列データを用いた推定値は、 $R_{i_2 i_3}(s)$ および $R_{i_3 i_2}(s)$ の値が全ての $s = 1 \dots M$ についてゼロに近い値となり、擬似相関の影響も、それによる偽陽性の危険も、少い。

GLM クロスコレログラムとは、GLM によって推定された応答関数をクロスコレログラムと同様の形式に並べ直したものである。

$$CG_{ic}^{GLM}(s) = \begin{cases} R_{ic}(s) & \text{if } s > 0 \\ R_{ci}(-s) & \text{if } s < 0 \end{cases} \quad (5)$$

なお、 $s = 0$ における値は未定義とする。GLM クロスコレログラムによれば、通常のクロスコレログラムと同様に $s > 0$ の領域で c から i への、 $s < 0$ の領域で i から c への方向性を持った機能的結合を同時に可視化することができる。

3. 偽陽性要因

3.1 スパイク検出時刻決定誤差に由来するアーティファクト

偽陽性要因の第一は、スパイク検出時刻決定の誤差によるアーティファクトである。スパイク検出時刻決定に誤差が乗るとき、一定確率で本来のスパイク信号の時間的前後関係が逆になってしまう場合が含まれるため、これが逆方向の機能的結合のように見えてしまうのである。

スパイク検出時刻決定誤差が避けられない場合は多い。典型例は、カルシウムイメージングによって高フレームレートかつ高解像度による観測を行う場合が挙げられる。フレームレートが高く解像度が高いとき、各ピクセルにおけるフレームあたりの光子数が少なくなり、蛍光信号に一定のショットノイズが乗る。スパイク時刻推定誤差の分散の大きさは、個々神経細胞に対応するピクセルの平均蛍光量の時系列にノイズが乗るとき、そのノイズ分散の大きさに比例する。

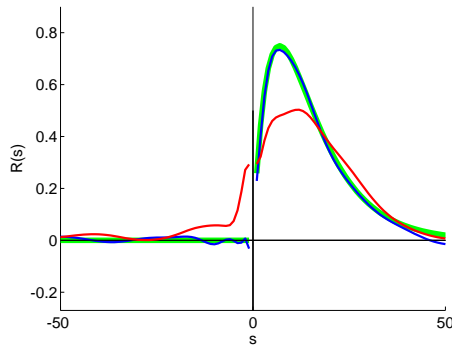


図 1 GLM クロスコレログラムによる応答関数推定結果の可視化例。横軸に時間遅れ s を示し、縦軸に s におけるニューロン i のニューロン c のスパイクへの応答 $R_{ic}(s)$ を示す。ただし、 $s < 0$ においては逆因果に対応する応答関数 $R_{ci}(-s)$ を示している。緑色の太線はシミュレーションに用いた真値、青線・赤線はそれぞれスパイク時刻に含まれる誤差標準偏差が 0.0 の場合と 5.0 の場合を示す。ニューロン c から i への影響は本来存在しないが、誤差のせいで顕著な偽陽性が観測されている。GLM クロスコレログラムによって左右を同時に見れば、これが偽陽性である可能性に気づきやすい。

スパイク時刻推定誤差の影響を示すシミュレーションを行った。線形応答によってスパイクを生成するモデルに基づいて 2 ニューロンのオリジナルのスパイク時系列を生成し、これに標準偏差 5 フレームの誤差を含めた誤差つきスパイク時系列を生成した。両時系列に対してそれぞれ応答関数推定を行い、その結果を GLM クロスコレログラムで表示した。

図 1 に、シミュレーションに用いた真の応答関数、オリジナル時系列に基づく推定応答関数、誤差つき時系列に基づく推定応答関数を比較する図を示した。オリジナル時系列に基づく推定応答関数は真の応答関数をよく近似している。一方で誤差つき時系列に基づく推定応答関数はピーク形状がぼやけており、左半分に浸透している。左半分だけを単独で見たときに、これは因果を遡る方向の応答の存在を示す偽陽性である。

3.2 非線形ダイナミクスに由来するアーティファクト

偽陽性要因の第二は、非線形性の強いダイナミクスを線形なモデルでフィッティングすることによって生じる歪みである。

非線形性の強いダイナミクスを示す神経細胞モデルとして Hodgkin-Huxley (HH) モデル [11] を考える。HH モデルは、神経細胞表面にある電位依存チャネルのうち 3 種 (カリウムイオンチャネル、ナトリウムイオンチャネル、塩化物イオンチャネル) の開閉と、これによるイオン電流の変化を微分方程式でモデル化したものであり、イカの巨大軸索におけるスパイク状の活動電位の生成や活動電位生成直後の不応期をよく説明し、他種の神経細胞の挙動のモデルとしても妥当性があると考えられている。HH モデルに従

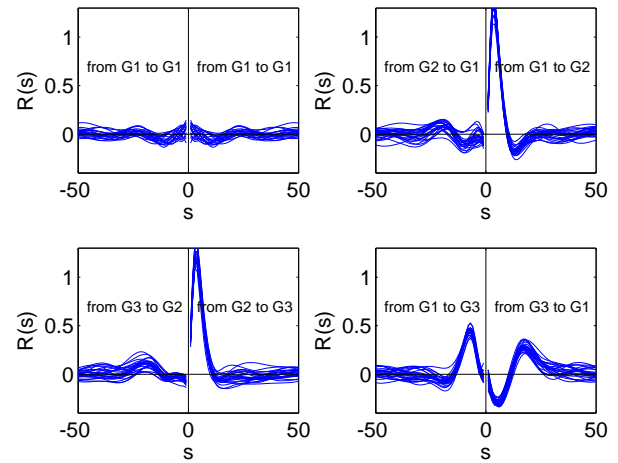


図 2 非線形ダイナミクスを持つ Hodgkin-Huxley モデルニューロンの活動シミュレーションに対する、応答関数推定結果。4 つのパネルはそれぞれ 5 ニューロンからなる 3 グループ G1,G2,G3 の間の機能的結合の推定結果を、GLM クロスコレログラムによって表している。2 群間には 25 組のニューロンペアがあるのでそれぞれに対応する曲線を重ね描いている。G1 に所属するニューロン間の結合は存在せず、G2 から G1、G3 から G2、G1 から G3 の結合も存在しないため、これら組み合わせにおいて見られる非ゼロ応答関数は偽陽性を示している。

う神経細胞のスパイク入力に対する応答は内部状態 (チャネル開閉の度合い) に依存するため、入力の大きさに単純に比例しない。これを線形応答モデルでフィッティングすると、どのような歪みが生じるであろうか。

神経細胞電氣的挙動のシミュレータのひとつである NEST [12], [13] を用いたシミュレーション実験を行った。5 細胞からなる細胞群を 3 つ (G1,G2,G3) を用意し、細胞群 G1 から G2、および、G2 から G3 は、それぞれ興奮性の全結合を、G3 から G1 は抑制性の全結合をしているネットワークを構築し、すべての細胞は外部から独立した平均発火率 50Hz から 150Hz のポアソンスパイク入力を受けているものとした。ただし、細胞群内では全く結合がないものとした。また、各細胞は 1 細胞 1 コンパートメントの標準的 HH モデルにしたがって活動するものとし、ダイナミクスのパラメータはシミュレータのデフォルト値を利用した [13]。以上合計 15 細胞について、5 ミリ秒単位 30 万フレームのシミュレーションを行ったところ、ニューロン 1 つあたりのスパイク数は 1000 から 3000 個程度となった。

図 2 は、こうして得られたスパイク統計データに基づき、応答関数を推定した結果を GLM クロスコレログラムによって表示したものである。左上の (G1-G1) の図は接続が無く、真の応答はゼロであることが期待される、有限データに基づく推定結果であるので一定の統計的ゆらぎを含んでいる。右上 (G1-G2)、左下 (G2-G3) の図は一方方向の興奮性接続に対応した応答を示す。これらのグラフのそれぞれ右半分は興奮性接続による約 3 フレーム (15 ミリ

秒) をピークとした応答を顕著に示しており、左半分は機能的接続が無いことを示しているが、左半分のさらに詳細を見ると 20 フレームをピークとした興奮性接続のような波形が 25 サンプルの全てで系統的に見られる。これらは HH モデルの非線形性によるものと考えられ、ピーク検出の閾値によっては偽陽性を生じさせる可能性がある。右下 (G3-G1) の図は一方方向の抑制性接続に対応した応答を示す。グラフの右半分では約 5 フレームをピークとした抑制性の応答を示すが、約 20 フレームをピークとした興奮性の応答が顕著に見られる。スパイク状の電位抑制入力が入るとき、そのタイミングによっては興奮性の応答を引き起こすという HH モデルの非線形効果を示していると解釈できる。さらに重要なのはグラフの左半分に見られる約 5 フレームをピークとした正方向の応答である。これは因果の方向を逆方向に遡る応答であり、顕著であるため深刻な偽陽性要因となる。

以上から以下の 2 点が言える。

- HH モデルの非線形性に起因して、実際には単方向性の機能的接続であるものが、双方向性の接続のように見えてしまう場合があり、偽陽性の要因となる
- 興奮性・抑制性違いによって両方向それぞれの応答関数において特徴的な形状が見られる

3.3 その他に考える偽陽性要因

以上に述べた 2 つの要因の他に、信号の混線も偽陽性の起因となる。信号の混線が生じる場合として、具体的には以下が想定される。

- 複数ニューロンから発生したスパイクを誤って同一ニューロンとしてまとめて扱ってしまう場合
- 同一ニューロンから発生したスパイクを誤って複数ニューロンとして分けて扱ってしまった場合

どちらも、カルシウムイメージングデータでは関心領域を構成するピクセルのセグメンテーションの失敗や、複数電極の信号ではスパイクソーティングの失敗によって生じ得る状況である。

ここでは詳細な議論を行わないが、両者ともに GLM クロスコレログラム形状として一定の傾向の存在が期待できる。

4. 偽陽性対策

ここまで見てきた偽陽性の可能性の対策として以下の二つが必要であり、また有効である。

4.1 GLM クロスコレログラム使用

第一の対策は GLM クロスコレログラムの積極的使用である。

前節で見たように、GLM ではニューロンペア A,B について、 $A \rightarrow B$ と $B \rightarrow A$ の双方向応答関数をそれぞれ独立

に取り扱うが、これらを独立に解釈する場合に偽陽性の危険が高まる。その一方で、双方向応答関数を同時に可視化する GLM クロスコレログラムでは、各種偽陽性要因のそれぞれに特有の形状が見られるため、これを手がかりとして解釈精度を向上させることができる。

そこで、GLM による応答関数推定を行ったとき、GLM クロスコレログラムによる双方向応答関数パターンをまとめてクラスタリングし、特徴的な形状毎に解釈ラベリングをすることを推奨する。これにより、時刻誤差アーティファクト、非線形性アーティファクト、混線起源アーティファクトそれぞれに特有の形状の検出が期待でき、後処理解釈によって偽陽性を回避する可能性が増す。

4.2 スパース推定の不使用

第二の対策は、GLM の推定においてスパース推定をあえて使用しないことである。GLM による応答関数推定においてスパース解を得るための手法として、これまでに階層的ベイズ推定に基づく手法 [2], [3], [4] や正則化に工夫をした手法 [5] などが提案されている。これらによれば、比較的応答の弱いニューロンペア間の応答関数の値が自動的にゼロ値と推定され、これによって有向リンク有無に関して自動的にクリアカットな判定が得られる。しかし、この判定基準はスパースさを制御するハイパーパラメタ値の設定に強く依存している。ハイパーパラメタ決定のためにクロスバリデーションなどの方法があり、スパイクデータを全般的に説明するうえで最もリーズナブルなモデルを得ることができるが、この基準が個々有向接続有無の判定精度を最高にする根拠は無い。その一方で、弱い信号をスパース推定によって一律に消してしまうことで本来ならば GLM クロスコレログラムの形状から読み取ることのできた、ノイズや非線形性などアーティファクト要因特有の双方向応答形状の手がかり情報を失ってしまう。

L2 正則化による GLM クロスコレログラム推定によれば、真の応答関数がゼロに近い場合にも形状情報を失わない。一方で真の応答関数がゼロから遠いような強い有向リンクにおける応答関数推定精度はスパース法と大きく違わない。

なお、正則化を敢えて行わない場合 [6] にも形状パターンを保持した応答関数が得られるが、推定に用いるスパイク統計データの分量 (フレーム数もしくはスパイク数) が少ない場合のパラメタ推定の精度低下が顕著であり、その結果応答関数推定精度と検出力ともに大幅に低下してしまうことがわかっている [5]。これも L2 正則化が適切と考えられる理由の一つである。

5. まとめと今後の課題

以上、応答関数推定において偽陽性につながり得るアーティファクト要因を 2 点指摘した。またこれらを回避す

るために各細胞ペア毎の双方向接続を同時に可視化する GLM クロスコレログラムの特徴的形状を見ることが有効であることを主張した。

今後の課題として、偽陽性排除プロセスの自動化が挙げられる。そのためには、GLM クロスコレログラムに基づいて偽陽性要因に関する典型的な双方向接続パターンを蓄積してゆく必要がある。本発表で示した第一の要因(時刻推定誤差)に従うパターンは比較的検出しやすい。一方で、第二の要因(HH モデル等非線形効果)に従うパターンには多様性があり、その蓄積の必要性が高い。なお我々が行ったプレリミナリな実験によれば、スパースなネットワーク設定におけるシミュレーションでは非線形効果特有の偽陽性パターンは弱くなる傾向がみられた。図2で示したような状況は極端なものであり、実問題において非線形効果は無視できる場合が多いのかもしれない。引き続き検証のための経験蓄積が必要である。

カルシウムイメージング特有の偽陽性要因の中でもセグメンテーションの失敗に起因するものは、本来ならば前処理段階において取り除かれるべきである。しかし、近年のチャレンジにあるような、1000を超える非常に大きな個数のニューロンを対象としたカルシウムイメージングでは、前処理での完全なセグメンテーションは必ずしも期待できない。GLM による後処理解析結果に基づいて事後的に排除する方法の開発が必要であり、ここで我々が示した知見が生かされることを期待する。

謝辞 本研究の一部は文部科学省科研費(22115012)の助成を受けて行われた。

参考文献

- [1] Stevenson, IH., Rebecco, JM., Miller, LE., and Kording, KP.: Inferring functional connections between neurons, *Current Opinion in Neurobiology*, 18, 6, (2008).
- [2] Gerwinn, S., Macke, JH., and Bethge, M.: Bayesian inference for generalized linear models for spiking neurons *Frontiers in Computational Neuroscience*, 4, 12, (2010).
- [3] Mishchenko, Y., Vogelstein, T., and Paninski, L.: A Bayesian approach for inferring neuronal connectivity from calcium fluorescent imaging data, *The Annals of Applied Statistics*, 5, 2, 1229-1261 (2011).
- [4] 吉本 潤一郎, 銅谷 賢治. 神経スパイク列からのモデルベースシナプス同定法. *IPJS SIG Technical Report*, 2011-BIO-25(4), (2011).
- [5] 安藝 俊介, 大羽 成征, 中江 健, 石井 信. スパイク信号列に基づく神経細胞間ネットワーク構造推定: group LASSO によるアプローチ. *IPJS SIG Technical Report*, 2011-BIO-26(4), (2011).
- [6] Kim, S., Putrino, D., Ghosh, S., and Brown, EN.: A Granger causality measure for point process models of ensemble neural spiking activity. *PLoS Computational Biology*, 7(3):e1001110, (2011).
- [7] Valdes-Sosa, PA., Roebroek, A., Daunizeau, J., and Friston, K.: Effective connectivity: Influence, causality and biophysical modeling. *NeuroImage*, 58(2):339-361, (2011).
- [8] Pillow, JW., Shlens, J., Paninski, L., Sher, A., Litke, AM., Chichilnisky, EJ., and Simoncelli, EP.: Spatio-temporal correlations and visual signaling in a complete neuronal population, *Nature*, 454, 995-1000, (2008).
- [9] Pillow, JW., Ahmadian, Y., and Paninski, L.: Model-based decoding, information estimation, and change-point detection techniques for multi neuron spike trains, *Neural Computation*, 23, 1-45, (2011).
- [10] Stetter, O., Battaglia, D., Soriano, J., and Geisel, T.: Model-Free Reconstruction of Excitatory Neuronal Connectivity from Calcium Imaging Signals, *pLoS Computational Biology*, 8(8): e1002653, (2012).
- [11] Hodgkin, AL. and Huxley, AF.: A quantitative description of ion currents and its applications to conduction and excitation in nerve membranes. *Journal of Physiology*, 117, 500-544, (1952).
- [12] Diesmann M., Gewaltig MO.: NEST: an environment for neural systems simulations. In *Forschung und wissenschaftliches Rechnen*, Beitrage zum Heinz-Billing-Preis 2001, Vol. 58 of GWDG-Bericht, Plesser T., Macho V., editors. , eds (Gottingen, Ges. für Wiss. Datenverarbeitung;), pp. 43?70. (2002)
- [13] Davison, AP., Brüderle, D., Eppler, J., Kremkow, J., Müller, E., Pecevski, D., Perrinet, L., Yger, P.: PyNN: A Common Interface for Neuronal Network Simulators. *Frontiers in Neuroinformatics*, 2008; 2:11, (2008).